

Thesis, FRANZEN Rachelle

Auteur : Bouchri, Sarah

Promoteur(s) : Briganti, Giovanni; GILBERT, Allison

Faculté : Faculté de Médecine

Diplôme : Master en sciences biomédicales, à finalité spécialisée en biomédical data management

Année académique : 2025-2026

URI/URL : <http://hdl.handle.net/2268.2/25983>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

Université de Liège - Faculté de médecine

Développement d'un score de stratification du risque de dégradation clinique chez les patients hospitalisés au CHU de Liège : vers une amélioration de la sécurité des patients

Mémoire

En vue de l'obtention du grade académique de Master en Sciences biomédicales à finalité spécialisée en Data Management

Présenté par :

Bouchri Sarah

Promoteurs

Pr. Briganti Giovanni
Dr. Gilbert Allison

Année académique 2025-2026

REMERCIEMENTS

Je tiens tout d'abord à exprimer ma profonde gratitude au Professeur Briganti pour m'avoir accordé l'opportunité de travailler sur un projet d'une telle envergure. Ce sujet, particulièrement enrichissant, m'a permis d'explorer de nombreux domaines et d'acquérir des compétences précieuses. Je le remercie également pour la confiance qu'il m'a accordée tout au long de ce travail. Ce fut un réel privilège de l'avoir comme promoteur.

Je souhaite ensuite remercier chaleureusement le Dr Gilbert pour son accompagnement constant durant la réalisation de ce mémoire. Sa disponibilité, ses conseils avisés et sa bienveillance ont été essentiels, notamment dans les moments de doute et de stress où elle a su me rassurer. Je lui suis reconnaissante pour son soutien du début à la fin de ce projet. Ce fut un véritable plaisir de l'avoir comme promotrice.

Je remercie également mes amies, du fond du cœur, pour leur présence indéfectible depuis le début de mes études. Leur soutien, leur écoute et leur réconfort, même à des heures improbables, m'ont été d'une aide précieuse tout au long de ce parcours.

Enfin, je souhaite adresser toute ma reconnaissance à ma famille pour son soutien inconditionnel. Merci pour votre présence, vos encouragements et votre confiance en mes choix. Je remercie tout particulièrement mes parents pour les nombreux sacrifices qu'ils ont consentis afin de m'offrir toutes ces opportunités. Leur soutien a été déterminant et a largement contribué à faire de moi l'étudiante et la personne que je suis aujourd'hui.

UTILISATION DE L'INTELLIGENCE ARTIFICIELLE

Dans le cadre de la rédaction de ce mémoire, l'intelligence artificielle a été utilisée comme outil d'aide à la correction linguistique. Son usage s'est limité à la vérification de l'orthographe, de la grammaire et de la syntaxe, à l'amélioration ponctuelle de certaines tournures de phrase, à l'aide à la constitution de la liste des abréviations, ainsi qu'à l'aide à la compréhension de certains articles scientifiques.

ABRÉVIATIONS

Abréviation	Signification
AA	Apprentissage automatique
ACVPU	Alert, Confusion, Verbal, Pain, Unresponsive
Ada	AdaBoost Classifieur
A.P.E.	Analyse, Protection et Exploitation des Données
AVPU	Alert, Verbal, Pain, Unresponsive
CHU	Centre hospitalier universitaire
CRP	Protéine C réactive
CSV	Comma-Separated Values
DPI	Dossier patient informatisé
DT	Decision Tree Classifier
ELISA	Échelle liégeoise de l'indice de sévérité à l'admission
EODA	Éveil, Ordres, Douleur, Aucune
ET	Extra Trees Classifier
EWS	Early Warning Score
FDA	Food and Drug Administration
FN	Faux négatifs
FP	Faux positifs
GBC	Gradient Boosting Classifier
GCS	Glasgow Coma Scale
IA	Intelligence artificielle
IDE	Integrated Development Environment
KNN	K Neighbors Classifier
LASSO	Least Absolute Shrinkage and Selection Operator
LDA	Linear Discriminant Analysis
LightGBM	Light Gradient Boosting Machine
MEWS	Modified Early Warning System
ML	Machine learning

MTS	Manchester Triage System
NA	Not Available / Données manquantes
NB	Naïve Bayes
NEWS	National Early Warning Score
NEWS2	National Early Warning Score 2
OMS	Organisation mondiale de la santé
OOF	Out-of-fold
PAD	Pression artérielle diastolique
PAS	Pression artérielle systolique
PR	Precision-Recall
PR-AUC	Precision-Recall Area Under the Curve
Q1	Premier quartile
Q3	Troisième quartile
RF	Random Forest
RGPD	Règlement général sur la protection des données
RL	Régression logistique
ROC	Receiver Operating Characteristic
ROC-AUC	Receiver Operating Characteristic – Area Under the Curve
SHAP	Shapley Additive exPlanations
SpO ₂	Saturation pulsée en oxygène
USI	Unité de soins intensifs
VN	Vrais négatifs
VP	Vrais positifs
VPN	Valeur prédictive négative
VPP	Valeur prédictive positive

RÉSUMÉ

Garantir une qualité de soins adéquate pour l'ensemble des patients constitue un enjeu majeur à l'hôpital, d'autant plus que la détection précoce de la dégradation clinique est essentielle pour prévenir des issues défavorables telles qu'une admission en USI ou un décès. Cette détection repose habituellement sur le jugement clinique, parfois appuyé par des scores empiriques, mais la charge de travail élevée et la fatigue peuvent diminuer la vigilance et accroître le risque d'évènements indésirables. Dans ce contexte, un outil permettant de prioriser les patients à risque apparaît particulièrement pertinent afin de soutenir une surveillance ciblée et une meilleure allocation de l'attention.

Ce travail vise donc à développer un score de stratification du risque basé sur des méthodes de machine learning afin de soutenir la décision clinique et d'améliorer la sécurité des patients. La dégradation clinique a été définie ici par une admission en USI ou un décès, et le score a été construit sur cinq niveaux à partir de données rétrospectives issues du CHU de Liège. Les données concernaient des patients admis via les urgences, dont l'hospitalisation s'est poursuivie en salle classique, sur la période de janvier 2024 à juin 2025. Les variables utilisées incluaient des signes vitaux, des résultats de laboratoire et des paramètres administratifs.

Le projet s'articule autour de deux approches. La première repose sur une régression logistique avec pénalisation Elastic net, utilisée comme modèle de référence en raison de sa simplicité et de sa facilité d'interprétation. Les prédictions ont ensuite été transformées en score, et les résultats montrent une augmentation du taux d'évènements lorsque le score augmente. Son évaluation a mis en évidence une capacité de discrimination satisfaisante, mais limitée par la nature intrinsèque de cette méthode. La seconde adopte une démarche plus parcimonieuse : un modèle Random Forest a d'abord permis de réduire le nombre de variables prédictives, puis plusieurs modèles ont été comparés, conduisant à la sélection d'un LightGBM. Les résultats obtenus sont particulièrement prometteurs et la transformation des prédictions en score suggère que le modèle stratifie correctement le risque, avec une concentration des évènements dans la catégorie de risque la plus élevée. L'évaluation de ce dernier suggère ainsi un intérêt clinique potentiel pour la stratification du risque et l'aide à la surveillance.

ABSTRACT

Ensuring an adequate level of care for all patients remains a major challenge in hospitals, especially because early detection of clinical deterioration is essential to prevent unfavorable outcomes such as intensive care unit admission or death. In practice, this detection usually relies on clinical judgment, sometimes supported by empirical scores, but high workload and fatigue can reduce vigilance and increase the risk of adverse events. In this context, a tool designed to prioritize at risk patients appears particularly relevant to support targeted monitoring and a better allocation of attention.

This work therefore aims to develop a discrimination score based on machine learning methods to support clinical decision-making and improve patient safety. Clinical deterioration was defined here as intensive care unit admission or death, and the score was built across five levels using retrospective data from the CHU of Liège. The data concerned patients admitted through the emergency department whose stay continued in a conventional ward, over the period from January 2024 to June 2025. The variables used included vital signs, laboratory results, and administrative parameters.

The project is structured around two approaches. The first is a logistic regression model with elastic net penalization, used as a benchmark because it is simple and easy to interpret. The predictions were then converted into a score, and the results showed an increase in the event rate as the score increased. Its evaluation demonstrated satisfactory discriminative ability, although it remained limited by the intrinsic nature of the method. The second approach follows a more parsimonious strategy: a Random Forest was first used to reduce the number of predictive variables, and several models were then compared, leading to the selection of a LightGBM model. The results are quite promising, and the transformation of predictions into a score suggests that the model stratifies risk well, with events concentrated in the highest-risk category. Its evaluation therefore suggests a potential clinical interest in risk stratification and monitoring support.

TABLE DES MATIERES

1. INTRODUCTION	1
1.1. L'HOPITAL ET LA QUALITE DES SOINS	1
1.1.1. <i>Qualité des soins</i>	1
1.1.2. <i>Freins à la qualité des soins</i>	2
1.2. LE PATIENT AIGU AU SEIN DE L'HOPITAL	3
1.2.1. <i>L'arrivée du patient à l'hôpital</i>	3
1.2.2. <i>L'orientation du patient et son hospitalisation</i>	5
1.3. IDENTIFIER LA DETERIORATION CLINIQUE DU PATIENT HOSPITALISE	6
1.3.1. <i>La détérioration clinique du patient hospitalisé</i>	6
1.3.2. <i>Mesures de détection usuelles</i>	7
1.4. L'APPORT DU MACHINE LEARNING	9
1.4.1. <i>Concept théorique du ML</i>	9
1.4.2. <i>Le ML au sein de la médecine aiguë</i>	10
1.4.3. <i>Intérêt du ML par rapport aux scores classiques</i>	12
2. OBJECTIFS.....	16
2.1. OBJECTIFS	16
3. MATÉRIEL ET MÉTHODES	17
3.1. DESIGN DE L'ETUDE	17
3.2. CONTEXTE ET POPULATION ETUDIEE	17
3.3. VARIABLES D'INTERET	17
3.4. COMITE D'ETHIQUE ET REGLEMENTATIONS	18
3.5. CREATION DE LA BASE DE DONNEES ET COLLECTE DES DONNEES	18
3.6. LOGICIELS ET LIBRAIRIES	18
3.7. MODÉLISATION ET ANALYSES STATISTIQUES	19
3.7.1. <i>Préparation des données (RStudio)</i>	19
3.7.2. <i>Analyses statistiques (RStudio)</i>	22
3.7.3. <i>Modélisation de la RL (RStudio)</i>	23
3.7.4. <i>Évaluation sur le jeu de test (RStudio)</i>	24
3.7.5. <i>Modélisation (Python)</i>	24
4. RÉSULTATS	26
4.1. VARIABLES NUMÉRIQUES	26
4.2. VARIABLES CATÉGORIELLES	27
4.3. LE DÉVELOPPEMENT DE LA RÉGRESSION LOGISTIQUE	29
4.4. PERFORMANCE DU MODÈLE SUR LE JEU DE VALIDATION INTERNE ET SUR LE JEU DE TEST	29
4.5. VARIABLE CONTRIBUTIVE AU MODÈLE	30
4.6. TRANSFORMATION DE LA PROBABILITÉ EN SCORE	31
4.7. ANALYSES COMPLÉMENTAIRES DU MODÈLE	33
4.8. RANDOM FOREST	34
4.9. TEST DE PLUSIEURS MODÈLES	36
4.10. L'ANALYSE DU MODÈLE LIGHTGBM	37
5. DISCUSSION	41
6. BIBLIOGRAPHIE	47

Introduction

1. INTRODUCTION

1.1. L'hôpital et la qualité des soins

1.1.1. Qualité des soins

Au sein du milieu hospitalier, garantir une qualité de soins adéquate pour l'ensemble des patients constitue un enjeu central. La qualité des soins peut être définie comme la capacité d'un établissement à assurer la sécurité du patient tout en lui offrant une prise en charge adaptée et efficace¹⁻⁴. L'Organisation mondiale de la santé (OMS) souligne que le patient doit être placé au centre du système de soins, avec une approche centrée sur ses besoins et sa situation clinique⁵.

L'évaluation de cette qualité repose sur différents indicateurs de performance, parmi lesquels figurent notamment les admissions non planifiées ou tardives en unité de soins intensifs (USI), l'échec de sauvetage (*failure to rescue*) ainsi que la durée de séjour^{1,3,6}.

Les admissions tardives en USI sont considérées comme un indicateur pertinent pour mesurer la qualité des soins en raison de leur association directe avec le pronostic vital^{1,5,7}. Les admissions non planifiées traduisent quant à elles fréquemment une inadéquation initiale de la prise en charge, pouvant résulter d'une sous-estimation de la gravité clinique ou d'une erreur diagnostique⁸.

L'échec de sauvetage constitue également un indicateur clé permettant d'évaluer la capacité d'un établissement à détecter et prendre en charge efficacement les complications. Il a été démontré que les hôpitaux présentant des taux de mortalité élevés ne sont pas nécessairement ceux où les complications sont les plus fréquentes, mais plutôt ceux où leur gestion est moins efficiente (Figure 1)^{3,8,9}.

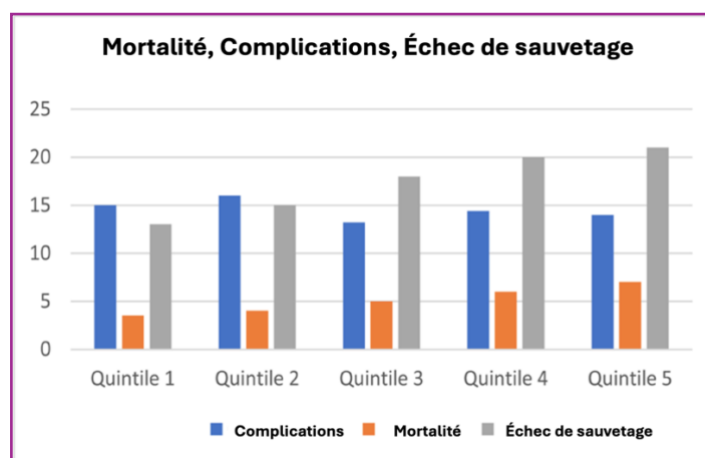


Figure 1 : Comparaison des taux de mortalité entre différents hôpitaux et leur association avec l'échec de sauvetage (figure reprise de Hatchimonji et al. 2019)⁹.

Ces indicateurs sont étroitement liés à la durée de séjour, qui reflète en partie la complexité de la prise en charge et la survenue éventuelle de complications^{1,3,4,10}.

1.1.2. Freins à la qualité des soins

Malgré ces objectifs, assurer une qualité des soins adéquate reste un défi en raison de nombreux facteurs limitants. Le personnel soignant constitue un élément déterminant, étant directement impliqué dans la surveillance et la prise en charge des patients.

En Belgique, une pénurie infirmière est actuellement observée, compromettant la qualité du suivi et de la prise en charge hospitalière^{8,11,12}. Par ailleurs, les professionnels exerçant dans des services à forte charge émotionnelle et exposition à la mortalité, tels que les USI, les urgences ou la gériatrie, présentent un risque accru de *burnout*^{5,13}. Ces conditions favorisent l'absentéisme et les départs, alimentant un cercle vicieux dans lequel la diminution des effectifs augmente la charge de travail des soignants restants⁸.

Ces contraintes ont des répercussions directes sur les patients et sur la qualité des soins prodigués^{5,7,8}. La surcharge de travail limite le temps consacré à chaque patient, rendant plus difficile la détection précoce d'une détérioration clinique. Des délais peuvent apparaître dans la réévaluation, exposant au risque de manquer certains signes d'alerte ou de retarder des interventions nécessaires^{5,14,15}.

De plus, le *burnout* altère la vigilance et la capacité de concentration des soignants, réduisant leur aptitude à identifier des modifications cliniques parfois subtiles^{8,14,16}. Dans ce contexte, il devient difficile d'accorder une attention équivalente à tous les patients rendant nécessaire une priorisation des patients à risque de détérioration.

Lorsque la qualité des soins est altérée, les conséquences peuvent être significatives, avec une augmentation de la morbidité liée aux complications. Il a notamment été montré que chaque heure de retard avant une admission en USI augmente le taux de mortalité d'environ 1,5 %^{4,5,15}.

Ces impacts s'étendent également à l'organisation hospitalière. Un allongement de la durée de séjour peut entraîner une saturation des services, en immobilisant des lits de manière imprévue et en nécessitant une adaptation du parcours de soins. La mise en place d'examen ou de traitements non anticipés, souvent plus lourds et spécialisés, engendre également des coûts supplémentaires pour l'hôpital (Figure 2)^{14,17-19}.

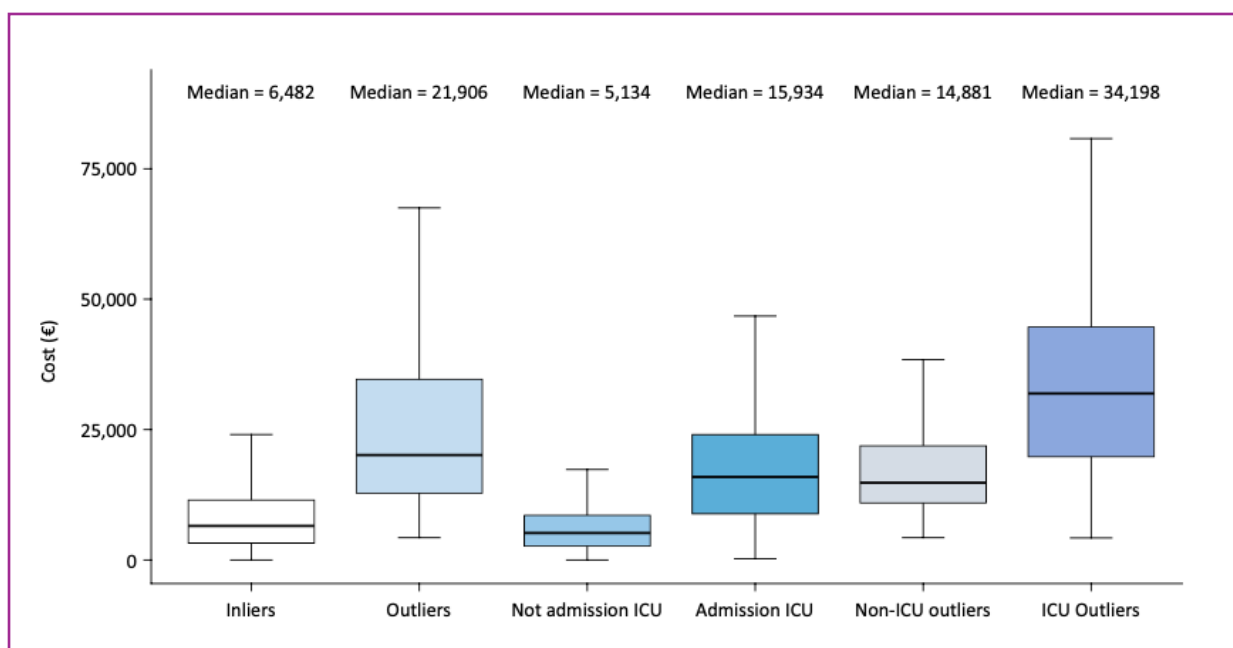


Figure 2 : Comparaison des coûts d'un séjour hospitalier en fonction du type d'admission (USI vs non-USI) et du caractère atypique du séjour (figure reprise de Bruyneel et al. 2025)¹⁹.

Dans ce contexte à la fois organisationnel et humain contraint, marqué par une surcharge du personnel et des ressources limitées, l'identification précoce des patients à risque apparaît essentielle afin d'optimiser leur prise en charge.

1.2. Le patient aigu au sein de l'hôpital

1.2.1. L'arrivée du patient à l'hôpital

Lors de son arrivée à l'hôpital, le patient est orienté vers le secteur le plus adapté à sa situation clinique, que ce soit une unité d'hospitalisation conventionnelle ou, en cas de gravité, une admission directe en USI. Une proportion importante des hospitalisations transite par le service des urgences, qui constitue une porte d'entrée majeure du système hospitalier.

À son admission, le patient bénéficie d'une évaluation clinique réalisée par l'infirmier, comprenant le recueil des paramètres vitaux, des symptômes, des antécédents médicaux et du niveau de douleur^{17,20}. Cette évaluation initiale appelée « triage » représente une priorisation des patients en fonction de la gravité de leur état et de leurs besoins en soins^{20,21}.

Le triage repose sur l'utilisation d'algorithmes standardisés. En Belgique, le *Manchester Triage System* (MTS) est largement utilisé. Il s'appuie sur des diagrammes de flux adaptés au motif de consultation et aux données cliniques recueillies (Figure 3)²².

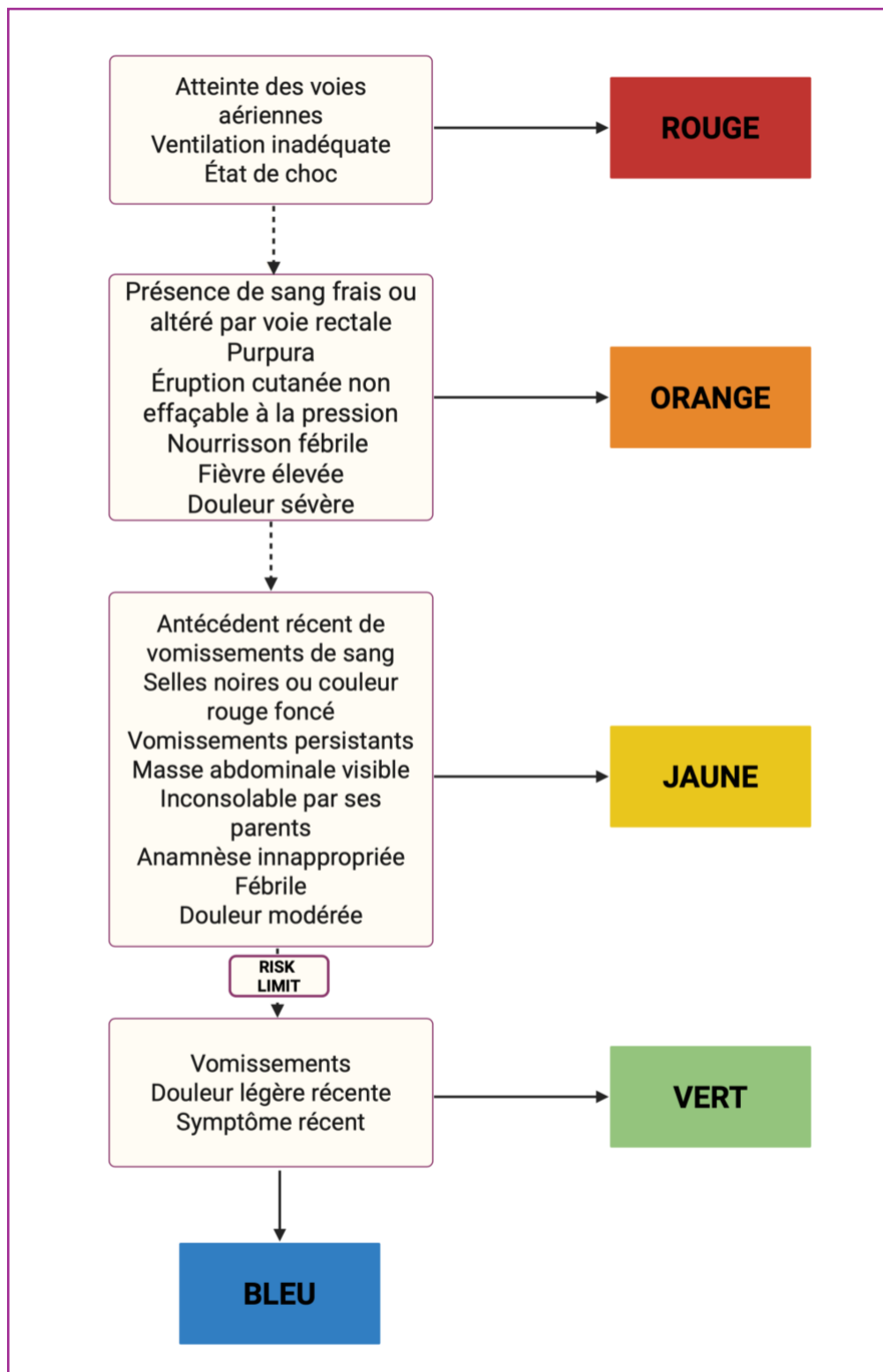


Figure 3 : Diagramme de flux utilisé chez les enfants en bas âge se présentant aux urgences pour des douleurs abdominales. Selon les symptômes rapportés par les parents, le patient est orienté vers la catégorie correspondante (figure traduite et adaptée du Kurs: Emergency Triage Hub | MTS)²².

Ce système permet de classer les patients en cinq catégories de priorité, associées à des codes de couleur, qui déterminent le délai avant la prise en charge.

Les catégories bleue et verte correspondent aux situations les moins urgentes, avec des délais plus longs, tandis que les catégories jaune, orange (10 minutes) et rouge indiquent des niveaux d'urgence croissants, la catégorie rouge nécessitant une prise en charge immédiate¹⁷.

Cependant, le MTS présente certaines limites. Il est notamment moins performant chez les patients âgés, qui présentent fréquemment des symptômes atypiques, alors qu'ils constituent une proportion importante des admissions en USI^{14,23}. De plus, il n'intègre pas systématiquement les paramètres vitaux, pourtant essentiels pour détecter précocement une détérioration clinique. Une étude menée au centre hospitalier universitaire (CHU) de Liège a montré que 69 % des décès inattendus concernaient des patients âgés^{14,17,24}.

À Liège, l'Échelle liégeoise de l'indice de sévérité à l'admission (ELISA) repose sur une stratification similaire en cinq niveaux (U1 à U5), mais se distingue par l'intégration d'une orientation du patient en plus du délai de prise en charge²¹.

Ces outils constituent une aide à la décision, sans toutefois se substituer au jugement clinique. Un triage inadéquat, qu'il s'agisse de sur-triage ou de sous-triage, peut avoir des conséquences importantes. Le sur-triage mobilise inutilement des ressources et peut retarder la prise en charge de patients réellement urgents, tandis que le sous-triage expose le patient à un retard de prise en charge malgré la présence de signes de gravité^{17,21,25}.

1.2.2. L'orientation du patient et son hospitalisation

Au moment de l'orientation du patient, une transmission structurée des informations entre le service des urgences et l'unité d'accueil est essentielle. Cette communication concerne à la fois l'état clinique du patient et les éléments relatifs à sa prise en charge (examens réalisés, paramètres vitaux, évolution), permettant d'assurer une continuité des soins et de disposer d'une vision actualisée de la situation²⁶.

Une fois admis en unité d'hospitalisation, le patient fait l'objet d'une surveillance régulière par le personnel infirmier, dont l'intensité dépend de son orientation initiale et de son état clinique. Cette surveillance comprend la mesure des paramètres vitaux ainsi qu'une évaluation clinique globale, avec une adaptation de la fréquence des observations en fonction de l'évolution du patient^{1,2,27}. Les données collectées sont consignées dans le dossier patient, garantissant un suivi longitudinal et une réévaluation continue de son état.

Ce suivi constitue un élément central dans la détection précoce des signes de détérioration en permettant d'identifier les variations cliniques au cours du séjour^{3,4,10}.

Au cours de l'hospitalisation, l'état du patient peut se dégrader, parfois rapidement, pouvant engager le pronostic vital. Les patients récemment admis présentent une vulnérabilité accrue, augmentant le risque de complications. Ainsi, un patient initialement pris en charge en unité conventionnelle peut nécessiter secondairement une admission en USI^{1,3,6,10}.

Une admission en USI ne traduit pas nécessairement une défaillance dans la prise en charge ou dans l'orientation initiale du patient. En revanche, une admission tardive ou non planifiée constitue un indicateur plus pertinent d'un dysfonctionnement, en particulier au niveau de la détection de la dégradation.

Une identification précoce de cette détérioration permet une adaptation rapide de la prise en charge, limitant ainsi la survenue de complications^{3,28,29}. Dans ce contexte, il est nécessaire de définir précisément ce que recouvre la notion de dégradation clinique chez le patient hospitalisé afin d'en améliorer la détection.

1.3. Identifier la détérioration clinique du patient hospitalisé

1.3.1. La détérioration clinique du patient hospitalisé

La détérioration clinique correspond à une évolution défavorable de l'état de santé d'un patient au cours du temps. Elle dépend de multiples facteurs, notamment du contexte médical, de l'état initial, du service d'hospitalisation, et de la qualité de la surveillance. Elle se manifeste principalement par des altérations des paramètres vitaux, mais peut également inclure des signes cliniques plus discrets, tels qu'une confusion ou une agitation²⁴.

L'évaluation repose en grande partie sur la surveillance régulière des paramètres vitaux incluant la fréquence cardiaque, la saturation pulsée en oxygène (SpO₂), la fréquence respiratoire, la température, la pression artérielle systolique (PAS) ainsi que le niveau de conscience^{27,29,30}.

En complément, l'évaluation clinique globale permet d'identifier des signes dits « discrets » (*soft signs*), non objectivables par les paramètres standard, tels qu'un sentiment de malaise, une diminution de l'alimentation ou des changements comportementaux rapportés par le patient et son entourage^{2,3,14,24}.

En cas de suspicion de détérioration, des examens complémentaires peuvent être réalisés afin d'affiner le diagnostic et d'adapter la prise en charge^{17,25}.

Les épisodes de dégradation surviennent fréquemment durant la nuit, période caractérisée par une surveillance réduite, ce qui peut retarder leur identification et favoriser l'apparition de complications évitables^{3,10,29-32}.

Les conséquences de la dégradation clinique sont multiples. Elle est associée à une augmentation de la morbidité, à un allongement de la durée de séjour et à une complexification de la prise en charge, avec un risque accru de mortalité^{1,8,33}. Elle peut également nécessiter une réorientation vers des unités spécialisées, la réalisation d'examens complémentaires et la mise en place de traitements plus intensifs^{8,34}.

Ces conséquences ont également un impact organisationnel et économique, en raison de l'augmentation des ressources mobilisées et des coûts associés aux soins^{19,35}.

La dégradation clinique constitue ainsi un enjeu majeur pour le patient et pour le système hospitalier^{1,3}. Son identification précoce est essentielle afin d'améliorer la qualité des soins, de limiter les complications et d'optimiser la prise en charge globale^{4,36}. Elle souligne également l'importance d'une orientation initiale adaptée et d'une surveillance rigoureuse tout au long du séjour^{17,18,30}.

Afin d'assurer une surveillance efficace et de faciliter l'identification des patients à risque, des outils de détection ont été développés et intégrés au parcours de soins du patient.

1.3.2. Mesures de détection usuelles

Durant le séjour hospitalier, la surveillance du patient peut être complétée par l'utilisation de scores de gravité. Ces outils ont été développés afin de standardiser la communication entre les différents acteurs hospitaliers (infirmiers, médecins, kinésithérapeutes, etc.), en fournissant un langage commun permettant d'exprimer de manière homogène l'état clinique et le degré de gravité d'un patient^{28,30,36}.

Néanmoins, ils constituent une aide à la décision et ne remplacent pas le jugement clinique : un score faible n'exclut pas une inquiétude clinique, et le soignant peut choisir de ne pas s'y conformer si la situation le justifie^{28,36,37}.

Ces scores permettent ainsi d'objectiver l'état du patient et d'orienter la prise de décision^{30,38,39}. Parmi les plus utilisés figurent le *Modified Early Warning System* (MEWS), le *National Early Warning Score* (NEWS) et sa version actualisée, le *National Early Warning Score 2* (NEWS 2)^{30,39,40}. Ils sont utilisés tout au long de l'hospitalisation afin d'assurer un suivi dynamique du patient et de faciliter la transmission d'informations entre professionnels de santé^{31,36,37}. Leur calcul repose

principalement sur des paramètres vitaux, le niveau de conscience, et, dans certains cas, des variables complémentaires (Tableau 1)^{29,30,37,40,41}. Chaque paramètre se voit attribuer un score en fonction de son écart par rapport aux valeurs normales, et le score global correspond à la somme de ces points^{30,37}.

Tableau 1: Les différents scores utilisés en hospitalisation et leurs paramètres^{30,37,39-41}.

SCORE	Paramètres utilisés pour le calcul du score
NEWS	Fréquence cardiaque, fréquence respiratoire, SpO ₂ , PAS, niveau de conscience (échelle AVPU*), température, apport d'oxygène (réponse binaire)
NEWS2	Fréquence cardiaque, fréquence respiratoire, SpO ₂ (avec une échelle spécifique pour les patients présentant une insuffisance respiratoire hypercapnique), PAS, niveau de conscience (échelle AVPU avec ajout de la nouvelle confusion = ACVPU), température, apport d'oxygène (réponse binaire) ; certaines versions incluent également l'urée ou la protéine C réactive (CRP).
MEWS	Fréquence cardiaque, fréquence respiratoire, PAS, température ; certaines versions incluent également l'âge, les antécédents médicaux et le niveau de conscience (AVPU) ou <i>Glasgow Coma Scale</i> (GCS),

* *Alert Verbal Pain Unresponsive* (AVPU)

Les différents scores reposent sur des compromis entre sensibilité et spécificité, rendant difficile l'obtention d'un équilibre optimal (Tableau 2)⁴². Cette problématique rejoint celle du triage, avec un risque de sur-détection pour certains patients et de sous-détection pour d'autres. Une stratégie privilégiant la sensibilité, dans une optique de sécurité, s'accompagne généralement d'une diminution de la spécificité et d'une augmentation du nombre d'alertes. Cela peut entraîner une surcharge pour les soignants et une perte de confiance dans l'outil^{24,30,31,36}.

Tableau 2 : Performances des scores de gravité et seuils d'utilisation pour la prédiction d'un transfert en USI ou d'un décès⁴².

Scores	SEUIL	ROC-AUC*	Sensibilité	Spécificité
NEWS	≥ 5	82,9 %	51,5 %	94,5 %
NEWS2	≥ 6	83,1 %	50,1 %	94,2 %
MEWS	≥ 3	75,7 %	43,7 %	94,7 %

* *Receiver Operating Characteristic – Area Under the Curve*

Ces scores reposent sur un nombre limité de paramètres, facilitant leur utilisation en pratique clinique, mais réduisant la richesse des informations prises en compte. Cela peut conduire à une évaluation incomplète, en particulier chez les patients âgés, qui présentent fréquemment des manifestations atypiques^{14,28,43}.

Le choix des variables intégrées dans ces scores est donc déterminant, car elles doivent refléter de manière suffisamment globale l'état du patient³⁹. Les paramètres vitaux constituent une base essentielle, mais l'intégration d'échelles neurologiques telles que le GCS ou l'AVPU apporte une dimension supplémentaire à l'évaluation clinique^{21,39,41,44}.

Par ailleurs, ces scores reposent sur des seuils fixes et un système additif qui ne prend pas en compte les interactions entre les variables^{29,40,45}. Cette simplification limite leur capacité à capturer certaines subtilités cliniques, notamment dans les situations atypiques, et peut réduire leur sensibilité^{30,31}. De plus, leur calcul est souvent manuel, exposant à des erreurs pouvant entraîner une sous-estimation du score réel^{10,28,31}.

Une autre limite importante concerne la gestion des données manquantes. En l'absence de certaines variables, le score peut être incomplet ou biaisé, conduisant potentiellement à une sous-estimation de la gravité et à un risque de sous-triage^{10,37,39,40}.

Il existe également ce que l'on appelle le « paradoxe de l'alerte »³⁰. Les performances des scores sont généralement évaluées en fonction de leur capacité à prédire la mortalité. Toutefois, lorsqu'un score permet une intervention précoce et évite une dégradation ou un décès, le patient est considéré, dans les analyses statistiques, comme un faux positif (FP), alors même que l'outil a rempli son objectif préventif³⁰.

Ainsi, bien que ces scores constituent des outils utiles pour soutenir la décision clinique et standardiser l'évaluation des patients, ils présentent plusieurs limites, notamment en termes de performance, de prise en compte des données, de rigidité des seuils et d'absence de modélisation des interactions entre variables. Dans ce contexte, le recours à des approches plus avancées, telles que le *machine learning* (ML), apparaît comme une piste pertinente pour dépasser certaines de ces limites^{29,31,40}.

1.4. L'apport du machine learning

1.4.1. Concept théorique du ML

Le ML, ou apprentissage automatique (AA), constitue une branche de l'intelligence artificielle (IA) qui repose sur l'apprentissage de relations à partir des données disponibles. Il permet notamment d'analyser de grandes quantités de données cliniques et d'en extraire des informations pertinentes avec un niveau de précision difficilement atteignable par une analyse humaine seule.

Il existe principalement deux types d'apprentissage : l'apprentissage supervisé, dans lequel la variable à prédire est connue (*classification ou régression*), et l'apprentissage non supervisé, qui vise à identifier des structures ou des regroupements dans les données sans variable cible prédéfinie (Figure 4)⁴⁶.

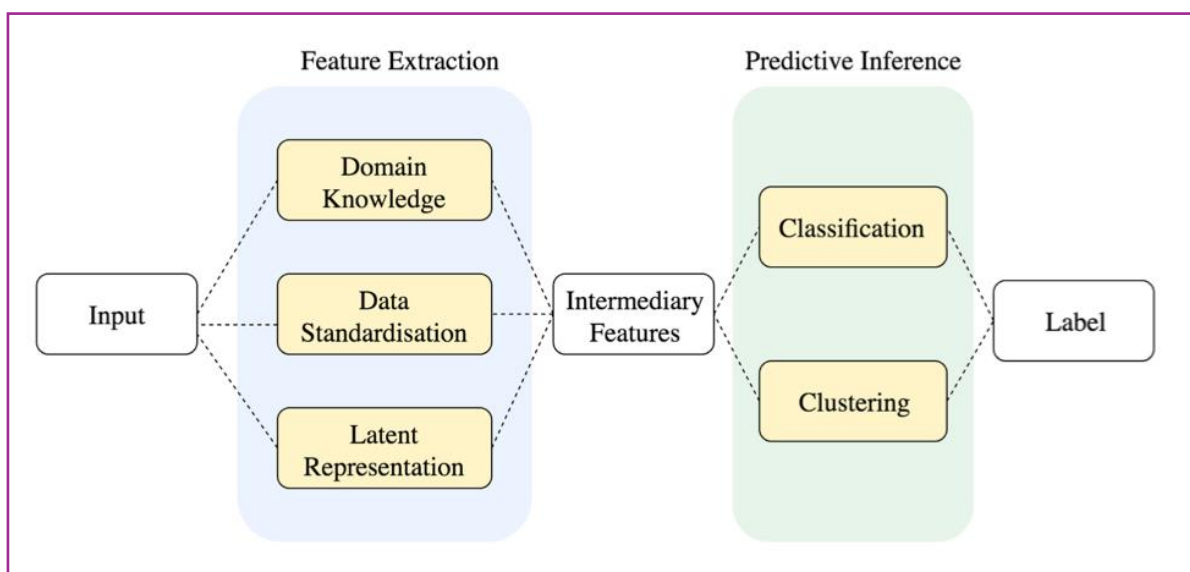


Figure 4 : Flux de travail standard en ML appliqué à la médecine d'urgence. Les données en entrée (*input*), constituées de variables explicatives, sont d'abord prétraitées et transformées (*feature extraction*) selon différentes approches (expertise métier, standardisation, représentations latentes). Ces transformations permettent d'obtenir des caractéristiques intermédiaires utilisées par les modèles pour produire des prédictions (*classification ou clustering*), aboutissant à une sortie (*label*). Figure reprise de Shamout et al. (2021)⁴⁶.

1.4.2. Le ML au sein de la médecine aiguë

Dans le domaine médical, le ML est utilisé comme un outil visant à soutenir et assister les praticiens dans leur prise de décision.

La médecine aiguë représente un contexte particulièrement propice à son utilisation, en raison de la rapidité des prises en charge, de la saturation fréquente des services et de l'impact immédiat des décisions sur le pronostic des patients. Dans cet environnement, des outils capables d'améliorer l'efficacité décisionnelle constituent un apport significatif. Le ML s'inscrit

ainsi comme un support à la pratique clinique sans se substituer au praticien, et contribue notamment à alléger la charge de travail.

En apprentissage supervisé, les modèles peuvent être utilisés pour prédire différents événements cliniques, tels qu'un risque de détérioration, un décès, une admission en soins intensifs ou encore le diagnostic de certaines pathologies. À l'inverse, l'apprentissage non supervisé permet d'identifier des profils de patients homogènes ou atypiques, facilitant ainsi une meilleure compréhension des populations prises en charge. Ces approches contribuent également à optimiser le flux des patients en accélérant la prise de décision.

Une application classique de l'apprentissage supervisé consiste à développer un score à partir des prédictions du modèle. Ces prédictions peuvent être exprimées sous forme de probabilité, de classification binaire ou transformées en score de risque. Dans une optique d'utilisation clinique, un score de stratification reste généralement le plus pertinent, car il est plus facilement interprétable et directement exploitable par les soignants⁴⁷.

Dans le cadre de la construction de scores cliniques, l'utilisation du ML est particulièrement pertinente, car il permet d'intégrer un grand nombre de variables issues de sources variées (paramètres vitaux, données cliniques, biologiques ou administratives), souvent disponibles sous forme de données tabulaires structurées. Ces caractéristiques correspondent précisément au type de données fréquemment rencontrées aux urgences, où plusieurs algorithmes peuvent être mobilisés, enrichissant ainsi la capacité de prédiction.

Des exemples concrets illustrent l'intérêt de ces approches, tels que le modèle eCARTv5, basé sur des algorithmes de ML et autorisé par la *Food and Drug Administration* (FDA). Comparé au NEWS, il présente de meilleures performances en termes de discrimination (ROC-AUC : 0,895 vs 0,829), ainsi qu'une meilleure valeur prédictive positive (VPP) pour un niveau de sensibilité similaire, traduisant une réduction du nombre de FP⁴².

Cependant en santé comme dans d'autres domaines, les performances du ML dépendent de nombreux facteurs, notamment la qualité des données, leur prétraitement, la définition de la variable cible, la population étudiée, les méthodes de validation et l'intégration du modèle dans le flux clinique réel. Un modèle performant en phase de développement ne garantit pas nécessairement une efficacité en conditions réelles. Ainsi, bien que le ML constitue un outil prometteur, sa mise en œuvre nécessite une rigueur méthodologique importante afin d'assurer sa pertinence et son applicabilité clinique.

1.4.3. Intérêt du ML par rapport aux scores classiques

Les scores classiques tels que le NEWS, le NEWS2 ou le MEWS permettent de standardiser et d'objectiver la prise de décision clinique, tout en facilitant la communication entre professionnels de santé. Ils offrent également un gain de temps et permettent d'anticiper une éventuelle dégradation clinique, en laissant au soignant la possibilité d'adapter précocement la prise en charge.

Cependant, ces scores présentent plusieurs limites. Ils reposent sur des seuils fixes et sur une évaluation indépendante de chaque paramètre, ce qui simplifie fortement la réalité clinique. Cette approche peut être insuffisante pour détecter certains profils atypiques, notamment des patients apparemment stables mais présentant un risque élevé de détérioration. Dans ces situations, les signaux faibles sont souvent peu détectés par les scores classiques, alors que certains modèles de ML sont capables de capter certaines relations plus complexes entre les variables.

Un autre avantage du ML réside dans sa capacité à s'adapter à la population étudiée, puisqu'il est entraîné sur des données spécifiques et peut apprendre des relations propres à cette population^{31,40,48,49}. Cela permet une meilleure prise en compte de la variabilité clinique et une détection plus fine des profils à risque

Contrairement aux scores classiques, les modèles ML peuvent également intégrer des relations complexes et non linéaires entre les variables. Dans le contexte hospitalier, les interactions entre paramètres sont fréquentes et porteuses d'informations cliniques importantes. Leur prise en compte constitue donc un avantage majeur⁴⁰.

Bien que le ML soit également confronté au compromis entre sensibilité et spécificité, certains modèles issus de ces approches peuvent présenter de meilleures performances globales, avec une amélioration de la détection des patients à risque et une réduction des fausses alertes. Toutefois, l'évaluation des performances ne doit pas se limiter à un seul indicateur, mais prendre en compte l'ensemble des métriques pertinentes.

Plusieurs algorithmes peuvent être considérés pour la détection de la dégradation, notamment la Régression Logistique (RL), le *Random Forest (RF)* et le *Light Gradient Boosting Machine (LightGBM)*, en raison de leur capacité à exploiter efficacement des données tabulaires et à extraire des informations pertinentes pour la prise de décision clinique^{31,40,42}.

La RL constitue un modèle de référence (*benchmark*) largement utilisé, notamment pour sa simplicité et son interprétabilité. Elle repose sur une formule mathématique dans laquelle

chaque variable est associée à un coefficient estimé à partir des données. Ces coefficients déterminent l'importance de chaque variable dans la probabilité prédite de l'évènement.

La RL permet d'intégrer un grand nombre de variables et de modéliser des relations linéaires (sur l'échelle du *logit*), tout en restant facilement explicable aux praticiens, ce qui favorise son acceptabilité^{42,49}. Elle peut également être associée à des méthodes de régularisation, telles que la pénalisation L1 (*LASSO*), L2 (*Ridge*) ou *Elastic Net*, permettant respectivement la sélection de variables, la gestion de la colinéarité ou une combinaison des deux approches (Tableau 3).

Tableau 3 : Méthodes de régularisation en RL et leurs caractéristiques.

Pénalisation	Effet sur les coefficients	Intérêt
LASSO* (L1)	Possibilité de mettre certains coefficients à zéro.	Sélection de variables.
Ridge (L2)	Réduction des coefficients.	Gestion de la colinéarité.
Elastic Net	Combinaison de réduction des coefficients et de mise à zéro de certains coefficients.	Sélection de variables et gestion de la colinéarité.

*Least Absolute Shrinkage and Selection Operator

Le RF est un algorithme de ML reposant sur un ensemble d'arbres de décision (Figure 5)^{50,51}. Il est capable de modéliser des relations complexes et non linéaires entre les variables, ce qui constitue un avantage par rapport aux approches plus simples.

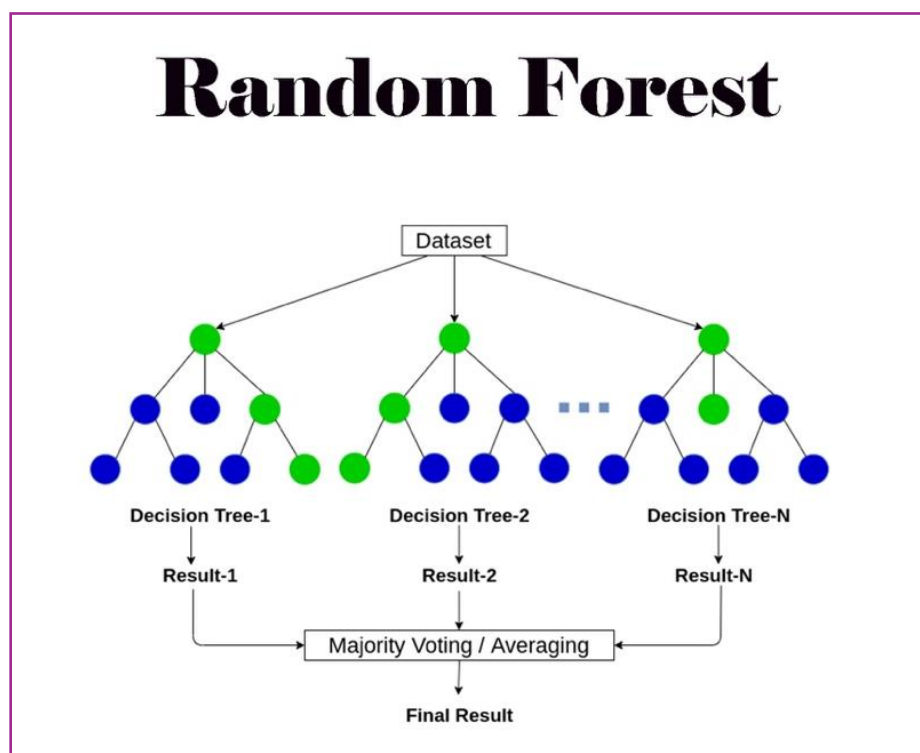


Figure 5 : Principe général du RF. À partir du jeu de données initial (*dataset*), celui-ci est subdivisé en plusieurs sous-ensembles, chacun étant utilisé pour entraîner un arbre de décision. Chaque arbre génère une prédiction, et le résultat final est obtenu par agrégation de ces prédictions, soit par vote majoritaire, soit par moyenne. Figure reprise de Consuledge^{50,51}.

Il peut être utilisé directement pour la construction d'un score, de manière similaire à la RL, mais également comme outil intermédiaire dans le processus de modélisation. En effet, l'identification des variables les plus pertinentes est une étape essentielle, car le nombre et la nature des variables influencent directement l'applicabilité du modèle en milieu hospitalier.

Dans ce cadre, le RF permet d'estimer l'importance des variables et de sélectionner celles qui contribuent le plus à la prédiction. Cette approche permet de réduire la dimension du modèle tout en conservant des performances satisfaisantes, ce qui facilite son utilisation en pratique^{52,53}.

Le LightGBM est particulièrement adapté aux données tabulaires structurées et présente l'avantage de gérer les données manquantes sans recourir systématiquement à une imputation. Il permet également de modéliser des relations non linéaires complexes, ce qui constitue un avantage important dans ce contexte^{31,45}.

La gestion des données manquantes représente un enjeu majeur en pratique clinique. Certains algorithmes, comme le LightGBM, peuvent les prendre en compte directement, tandis que d'autres, comme la RL, nécessitent des étapes préalables de traitement, telles que l'imputation par la moyenne ou la médiane⁴².

Les modèles de ML nécessitent également des étapes rigoureuses de validation. Leur développement repose classiquement sur quatre étapes principales : la préparation des données, l'entraînement du modèle, son évaluation et sa validation (Figure 6)^{50,54}.

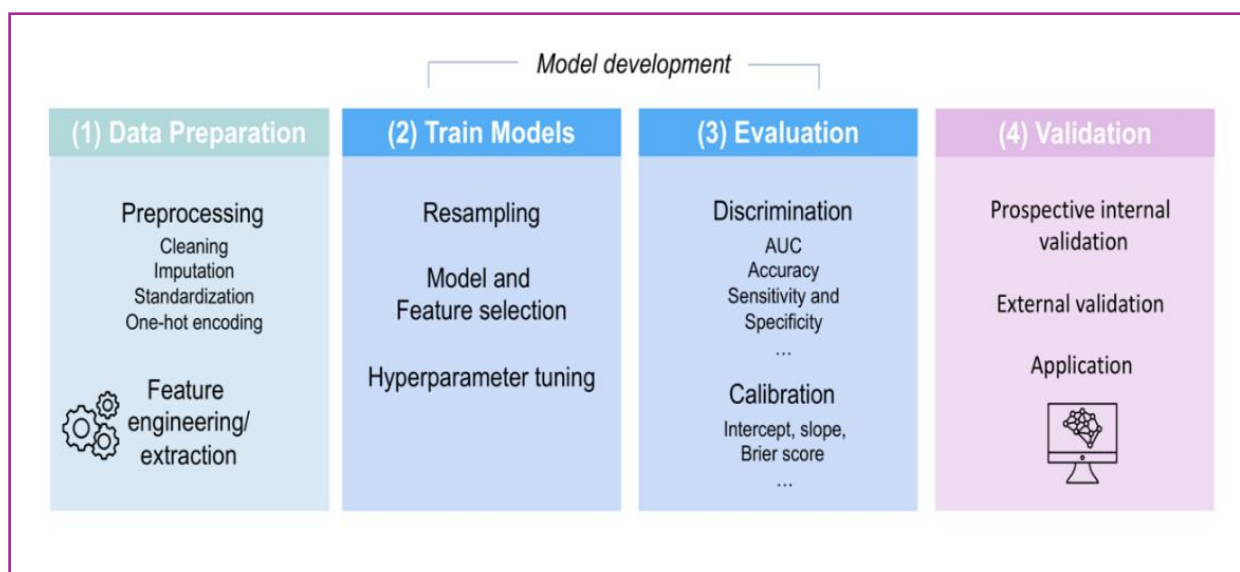


Figure 6 : Flux de développement d'un modèle de ML. (1) Préparation des données ; (2) entraînement des modèles ; (3) évaluation des performances ; (4) validation interne et externe du modèle. Figure reprise de Kernbach et staartjes (2020)⁵⁴.

Dans ce cadre, le point de départ repose sur un jeu de données qui est préalablement divisé en deux sous-ensembles non équivalents : un ensemble utilisé pour le développement et l'entraînement du modèle, et un second, conservé indépendamment, destiné à l'évaluation finale⁵⁵.

Plus précisément, les étapes de préparation des données et d'entraînement du modèle reposent exclusivement sur le jeu de données d'entraînement, permettant au modèle d'apprendre les relations entre les variables.

À l'inverse, les étapes d'évaluation et de validation s'appuient sur le jeu de données conservé, qui n'a pas été exposé au modèle au cours de son apprentissage. Cette séparation permet de simuler des conditions proches de la pratique réelle, en testant les performances du modèle sur des données nouvelles, et constitue un élément essentiel pour évaluer sa capacité de généralisation, limiter le risque de surapprentissage et garantir la robustesse du modèle^{56,57}.

Par ailleurs, les modèles les plus complexes peuvent souffrir d'un manque d'interprétabilité. Des méthodes telles que *Shapley Additive exPlanations (SHAP)* permettent d'expliquer les prédictions en identifiant la contribution de chaque variable, réduisant ainsi l'effet de « boîte noire » et améliorant la confiance des utilisateurs⁴⁵.

Ainsi, les approches basées sur le ML permettent de dépasser plusieurs limites des scores classiques, en améliorant la détection de la détérioration clinique, notamment celle conduisant à une admission en USI ou un décès. Elles offrent une meilleure prise en compte des interactions entre variables, une adaptation à la population étudiée et une réduction des fausses alertes^{31,40,48,49}.

Cette réflexion met en évidence l'importance d'une identification précoce des patients à risque au sein de l'hôpital, non seulement dans une optique de sécurité individuelle, mais également en raison de l'impact que cela peut avoir sur l'organisation des soins et le parcours global des patients.

Dans ce contexte, cette étude a été réalisée afin d'évaluer si le ML peut être mobilisé comme une solution permettant de détecter les patients à risque de dégradation clinique, dans le but de favoriser la mise en place de mesures précoces et, ainsi, améliorer la qualité des soins au sein de l'hôpital.

Objectifs

2. OBJECTIFS

2.1. Objectifs

Dans le milieu hospitalier, la dégradation clinique d'un patient peut survenir à tout moment, aussi bien lors de son admission aux urgences que durant une hospitalisation en salle classique. Plus cette dégradation est détectée tardivement, plus le risque de complications augmente, pouvant conduire à une admission en USI, voire au décès du patient.

Dans un contexte hospitalier marqué par une surcharge des services, la priorisation des patients constitue un enjeu majeur. Identifier précocement les patients nécessitant une attention accrue permet une meilleure allocation des ressources, notamment du personnel infirmier, et contribue à améliorer la qualité des soins. Une surveillance ciblée permet ainsi d'adapter la prise en charge en fonction du niveau de risque de chaque patient.

L'objectif principal de ce travail est de développer un outil d'aide à la décision permettant d'identifier les patients à prioriser en matière de surveillance, afin de prévenir une éventuelle dégradation clinique. Cet outil prend la forme d'un score de stratification du risque, allant de 1 à 5, permettant de stratifier les patients selon leur risque d'admission en USI ou de décès.

Notre hypothèse de recherche principale était qu'il était possible, sur base de variables issues du DPI, de construire un score de stratification du risque de dégradation clinique.

Matériel et Méthodes

3. MATÉRIEL ET MÉTHODES

3.1. Design de l'étude

Le projet de création du score de stratification du risque a été mené sur la base d'analyses rétrospectives des admissions au sein du Centre Hospitalier Universitaire (CHU) de Liège pour la période de janvier 2024 à juin 2025.

3.2. Contexte et population étudiée

La population sélectionnée pour l'étude a été celle du CHU de Liège, sur deux de ses sites, le site Sart Tilman (ST) et le site Notre Dame des Bruyères (NDB).

Pour être sélectionné, un patient devait respecter des critères d'inclusion : (1) être admis via les urgences ; (2) être majeur (≥ 18 ans) ; (3) avoir une admission suivie d'une hospitalisation classique (hors hospitalisation provisoire, hors service de psychiatrie, hors USI et hors maternité).

3.3. Variables d'intérêt

Les variables d'intérêt à étudier ont fait l'objet d'une réflexion préalable au processus de collecte des données. Cela a nécessité une recherche bibliographique approfondie portant sur les scores existants, avec et sans IA, ainsi que sur la définition de la dégradation clinique. L'objectif était d'identifier les paramètres influençant cette dégradation et ceux associés au risque d'admission en USI ou de décès.

L'objectif de cette première étape était donc la réalisation d'un *codebook* (voir annexe 1) en collaboration avec des cliniciens afin de structurer la collecte des données issues du CHU de Liège et identifier les variables les plus pertinentes.

À l'issue de ces réflexions, plusieurs types de variables ont été sélectionnés :

- Les signes vitaux (température, fréquence cardiaque, SpO_2 , PAS, pression artérielle diastolique (PAD), échelle Éveil, Ordres, Douleur, Aucune (EODA), échelle de *Glasgow*, NEWS (appelé *Early Warning Score* (EWS) dans la suite de ce travail), administration d'oxygène) ;
- Les résultats de laboratoire (CRP, créatinine, glucose, potassium, troponine I, urée) ;
- Autres (service d'admission, sexe, âge et antécédents de passage en USI).

L'ensemble des variables est exprimé et détaillé dans le *codebook*.

3.4. Comité d'éthique et réglementations

L'accès aux informations nécessaires de la population ciblée a nécessité la soumission d'un dossier au comité d'éthique, précisant les objectifs de l'étude ainsi que les modalités d'utilisation des données. L'étude a donc été soumise au comité hospitalo-universitaire de Liège sous la référence 2025-460. Celui-ci a constaté qu'elle n'entrait pas dans le cadre de la loi du 7 mai 2004 relative aux expérimentations sur la personne humaine et aucune objection n'a donc été émise quant à sa réalisation. Le règlement sur la protection des données a également été respecté. Les données étant issues du CHU de Liège et traitées par l'équipe de recherche de l'ULiège dans le cadre du mémoire, un *data transfer and processing agreement* a été signé entre les deux institutions.

3.5. Création de la base de données et collecte des données

Après identification de la population cible et des variables d'intérêt, le processus de création de la base de données a été réalisé. Les données utilisées pour le développement des modèles de ML proviennent des dossiers patients informatisés (DPI) du CHU de Liège. Une demande d'extraction auprès du service Analyse, Protection et Exploitation des Données (A.P.E) a été réalisée, le *codebook* a été fourni et discuté afin d'obtenir les variables sélectionnées.

Les variables ont été collectées à 7 points de mesure différents. Lorsque le séjour ne comportait pas de passage en USI, les variables des 6 jours en hospitalisation classique ont été recueillies, incluant celles du jour de l'admission aux urgences. Aucun séjour ne présentait un décès survenu avant une admission en USI. Ainsi, pour les séjours avec évènement, les variables ont été relevées à l'admission aux urgences, puis durant les 3 jours suivants en hospitalisation classique, avec l'addition des variables des 3 jours avant le passage en USI (jour de l'admission en USI inclus) (Figure 7).

Lors d'un passage aux urgences, l'état d'un patient peut se dégrader et mener à une prise en charge en USI ou un décès. L'évènement d'intérêt correspond à la dégradation clinique qui se reflète ici par cette admission en USI ou par un décès, cela a permis de définir deux groupes de patients (évènement vs non-évènement). Chaque groupe contribue à la création d'un profil type de patient le reflétant.

3.6. Logiciels et librairies

Pour le projet, le logiciel R version 4.4.1 (2024-06-14) a été utilisé, dans l'environnement RStudio version 2024.09.0+375 (2024.09.0+375). Les librairies principalement utilisées sont tidyverse⁵⁸ pour la manipulation de données et glmnet⁵⁹ pour le modèle de régression logistique.

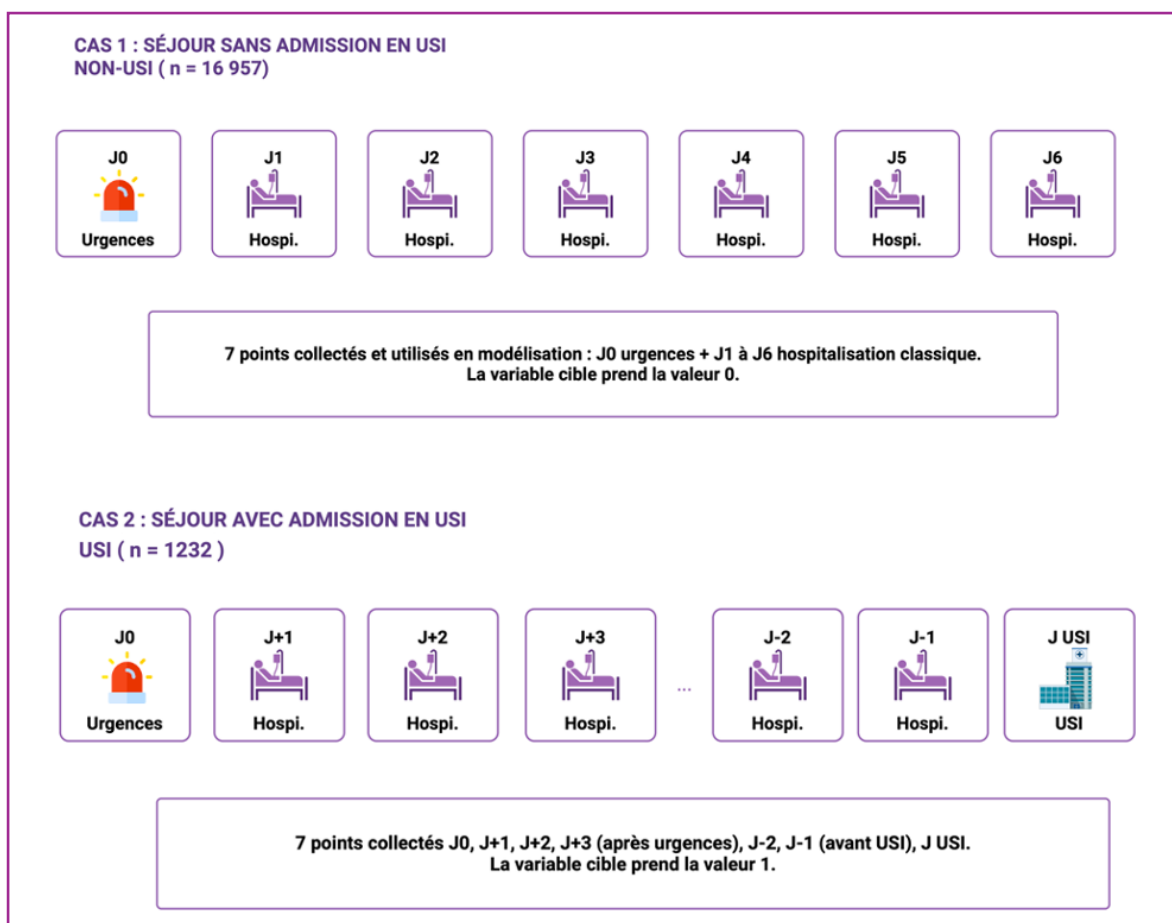


Figure 7 : Schéma de construction de la variable cible selon la survenue d'un évènement clinique (admission en USI ou décès). Dans la cohorte étudiée, aucun décès ne précédait une admission en USI ; le schéma illustre donc le cas d'une admission en USI. Figure réalisée à l'aide de BioRender.

Afin d'explorer et de tester plus facilement de nombreux types de modèles de ML, le logiciel Python version 3.11.14 a également été utilisé par le biais de l'Integrated Development Environment (IDE) PyCharm. La base de données préalablement nettoyée et préparée a été importée depuis RStudio et plusieurs modèles ont été entraînés grâce à la librairie PyCaret⁶⁰.

La majeure partie du code, comprenant le *data management* et le développement des modèles a été réalisée depuis une machine personnelle. Pour l'entraînement des modèles, l'hyperparamétrage et le réentraînement, un ordinateur possédant des ressources plus puissantes a été utilisé. L'ensemble du code est disponible sous forme de document généré avec Quarto, et hébergé via GitHub Pages, accessible via le lien suivant :

https://prvtsarah.github.io/Code_memoire/

3.7. Modélisation et analyses statistiques

3.7.1. Préparation des données (RStudio)

Les principales étapes méthodologiques du projet, depuis la définition de la population étudiée jusqu'à l'évaluation des modèles, sont résumées dans la figure 8.

La base de données a été obtenue en format *comma-separated values* (CSV) et importée sur RStudio. Les données étaient déjà anonymisées, de sorte que seuls les identifiants de patients et de séjours étaient visibles. Une étape de pseudonymisation a ensuite été réalisée via un script autonome, indépendant du *pipeline* principal, qui génère des chiffres aléatoires à partir des identifiants originaux afin de respecter les exigences du règlement général sur la protection des données (RGPD). Le jeu de données pseudonymisé obtenu a été utilisé pour l'ensemble de la recherche.

Des contrôles de structure ont été effectués. Durant ces contrôles, certaines valeurs chiffrées accompagnées d'un signe d'infériorité (« < ») ont été identifiées. Afin de traiter ces données sans retirer cette information, un indicateur binaire a été ajouté pour signaler quand une valeur est accompagnée de ce signe.

Des bornes cliniques de plausibilité ont également été appliquées aux variables numériques afin d'identifier d'éventuelles valeurs aberrantes ou incompatibles avec des valeurs cliniques cohérentes. Les valeurs situées en dehors de ces intervalles ont été considérées comme potentiellement liées à des erreurs d'encodage ou de mesure et ont été remplacées par des valeurs manquantes (Tableau 4).

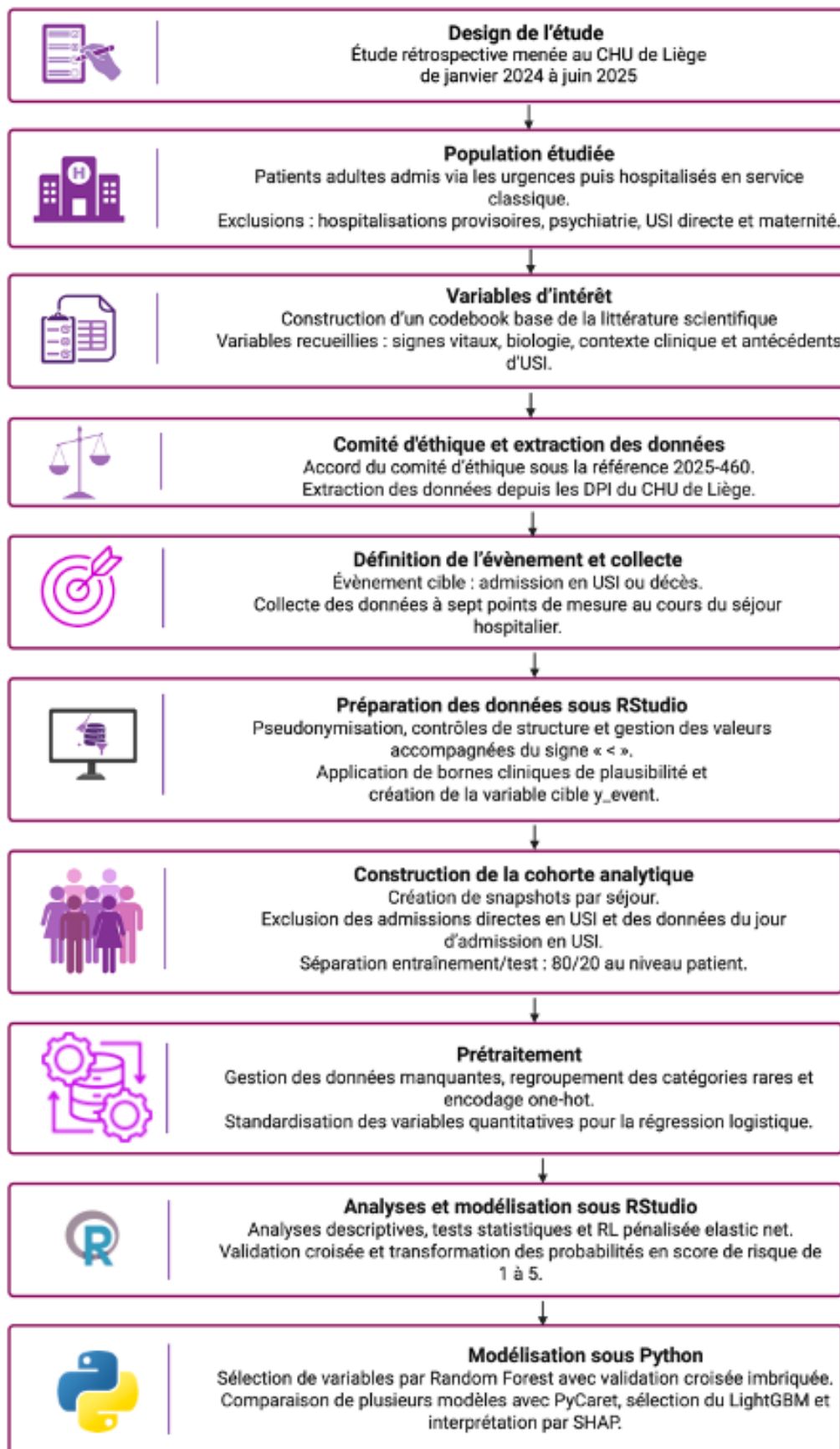


Figure 8 : Schéma reprenant les principales étapes du développement du score de stratification du risque de dégradation clinique. Figure réalisée à l'aide de BioRender.

Tableau 4 : Bornes cliniques appliquées aux variables quantitatives.

Variable	Borne inférieure	Borne supérieure	Unité
Âge	18	120	ans
Créatinine enzymatique	0,1	25	mg/dL
CRP	0	1000	mg/L
EWS	0	20	score
Fréquence cardiaque	20	250	bpm
Score de Glasgow	3	15	score
Glucose	20	1000	mg/dL
PAD	20	200	mmHg
PAS	40	300	mmHg
Potassium	1,0	10	mmol/L
SpO ₂	30	100	%
Température	25	44	°C
Troponine I	0	100 000	ng/L
Urée	0	500	mg/dL

Une variable cible binaire a ensuite été créée, nommée *y_event*. Lorsque le séjour d'un patient contient un passage en USI ou un décès, la variable cible prend la valeur 1, et 0 dans le cas contraire.

Des fenêtres temporelles (*snapshots*) ont été créées afin d'obtenir une observation par patient et par jour. Pour les séjours contenant un passage en USI, les observations des variables prédictives juste avant le passage ont été retenues. Pour les autres séjours, une moyenne des observations aux différents points de mesure a été réalisée.

Les séjours comportant une admission directe en USI ont été retirés, l'objectif étant de prédire une dégradation clinique et non une situation déjà constituée. De la même manière, les snapshots du jour de l'admission en USI ont également été exclus, car leur proximité temporelle avec l'évènement cible risquait d'influencer fortement les modèles (Tableau 5).

La cohorte a été divisée en deux parties : un jeu d'entraînement composé de 80 % des données, utilisé pour la création du *pipeline* de nettoyage, de structuration et pour l'entraînement des modèles, et un jeu de test de 20 %, exploité pour tester et analyser les modèles réalisés. Une graine aléatoire a été utilisée via la fonction `set.seed()` pour assurer la reproductibilité du découpage.

Tableau 5. Description de la cohorte finale selon le jeu de données et la présence de l'évènement cible.

Jeu de données	Nombre de séjours	Nombre d'évènements	Prévalence de l'évènement (%)
Cohorte finale	18 189	1 232	6,8
Jeu d'entraînement	14 583	990	6,8
Jeu de test	3 606	242	6,7

La cohorte finale comprenait donc 18 189 séjours hospitaliers, dont 1 232 présentaient l'évènement cible, soit une prévalence de 6,8 %. Les données ont ensuite été séparées en un jeu d'entraînement et un jeu de test au niveau du patient, afin d'éviter qu'un même patient ne se retrouve dans les deux ensembles. La prévalence observée dans les deux sous-groupes était comparable, avec 6,8 % dans le jeu d'entraînement et 6,7 % dans le jeu de test.

Les variables ont été séparées en deux catégories :

- Variables de routine : relevées de manière fréquente lors des admissions hospitalières ;
- Variables sur indication : collectées uniquement dans des cas spécifiques (ex. troponine I en cas de suspicion de problème cardiaque).

Pour éviter le bruit créé par un trop grand nombre de données manquantes, des variables ont été éliminées selon leur taux d'absence⁶¹ :

- Variables de routine : élimination si > 5 % de données manquantes ;
- Variables sur indication : élimination si >10 % de données manquantes.

Afin de pallier les algorithmes ne supportant pas les données manquantes, celles-ci ont été imputées par la médiane pour les variables quantitatives de routine et d'indication, car la médiane est plus robuste que la moyenne. Pour les variables qualitatives, les valeurs manquantes ont été imputées par le mode, c'est-à-dire par la modalité la plus fréquemment observée dans le jeu d'entraînement.

Les modalités sous-représentées (fréquence < 0,01) ont été agrégées dans une nouvelle catégorie « catégories rares » afin d'éviter l'instabilité des modèles.

Les variables catégoriques à deux modalités ont été transformées en variables binaires. Pour les variables à plus de deux modalités, chaque modalité a été convertie en une colonne binaire (*one-hot encoding*).

La régression logistique nécessitant des données standardisées, une standardisation z-score a été appliquée sur les variables quantitatives continues via la fonction *step_normalize*, amenant la moyenne μ à 0 et l'écart-type σ à 1.

Les variables seront donc écrites telles que :

$$X_{standard} = \frac{X - \mu}{\sigma}$$

Où :

- X est la valeur initiale,
- μ est la moyenne de la variable,
- σ est l'écart-type de la variable,
- $X_{standard}$ est la valeur standardisée.

3.7.2. Analyses statistiques (RStudio)

Une description des variables numériques a été effectuée, avec un tableau reprenant le nombre d'observations, le pourcentage de données manquantes, la moyenne, l'écart-type, la médiane, le premier quartile (q1), le troisième quartile (q3), le minimum (min) et le maximum (max) pour l'ensemble des variables retenues. Pour les variables qualitatives, un tableau recensant le nombre d'observations, la modalité, le pourcentage de données manquantes et observées a été réalisé.

Des représentations graphiques ont été réalisées pour une meilleure compréhension des données : histogrammes et diagrammes en boîtes pour les variables numériques (valeurs extrêmes identifiées en rouge), et diagrammes en barres de proportions pour les variables qualitatives.

Des analyses sur les variables quantitatives par rapport à la variable cible ont été menées, avec le nombre d'observations, la médiane, le q1 et le q3 pour les deux valeurs de la variable cible (0 et 1). Le test de Wilcoxon-Mann-Whitney a été utilisé afin de comparer deux groupes indépendants sans supposer une distribution normale des variables.

Pour les variables qualitatives, le tableau reprend la variable, sa modalité, le nombre d'observations et le pourcentage d'observations pour chacune des valeurs de la variable cible. Le test du khi-carré d'indépendance a été utilisé pour comparer la répartition des modalités entre les deux groupes.

Des diagrammes en boîtes ont été utilisés pour les données quantitatives et des diagrammes en barres de proportions pour les variables qualitatives.

3.7.3. Modélisation de la RL (RStudio)

Un modèle de régression logistique a été construit avec une pénalisation de type *Elastic net*, mélangeant les pénalisations *LASSO* et *ridge*. La probabilité d'observer l'évènement cible s'écrit :

$$P(Y = 1 | X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}}$$

Où :

Y est la variable cible,

$P(Y = 1 | X)$ est la probabilité que l'évènement cible soit observé,

β_0 est l'intercept,

$\beta_1, \dots, \beta_p$ mesurent l'effet de chaque variable X_1 à X_p sur le logit de la probabilité.

La validation croisée à 5 sous-ensembles a été utilisée pour l'entraînement, avec un regroupement par patient et une stratification par évènement. Une graine aléatoire a été appliquée pour assurer la reproductibilité des plis.

Différentes combinaisons des deux paramètres de l'*Elastic net* ont été testées : la mixture (proportion entre *LASSO* et *Ridge*) et la plage de pénalisation (intervalle de -4 à 0 sur l'échelle $\log_{10}$). Une grille combinant plusieurs valeurs de pénalisations et de mixture a été utilisée. La sélection du meilleur modèle a été orientée par la métrique *Precision-Recall Area Under the Curve* (PR-AUC), privilégiant la performance sur la classe positive, nécessaire dans le cas de classes déséquilibrées. Le modèle final a été réentraîné sur l'entièreté du jeu d'entraînement sans validation croisée.

Les prédictions du modèle ont été transformées en un score allant de 1 à 5, sur la base de quintiles des probabilités prédites. Les probabilités *out-of-fold* (OOF) du jeu d'entraînement ont

été classées de la plus faible à la plus forte, puis divisées en 5 groupes de taille égale : le score 1 regroupant les 20 % les plus faibles, et le score 5 les 20 % les plus élevés. Les seuils obtenus sur le jeu d'entraînement ont ensuite été appliqués au jeu de test.

Une classification binaire a été adoptée afin de permettre le calcul d'indicateurs cliniquement interprétables (sensibilité, spécificité, etc.). Lorsque le score est égal à 5, le profil du séjour indique une ressemblance élevée avec un profil avec évènement : en dessous de 5, le profil correspond davantage à un séjour sans évènement.

3.7.4. Évaluation sur le jeu de test (RStudio)

Le jeu de test a été préparé sur la base du jeu d'entraînement et a permis d'évaluer la performance du modèle sur des données jamais rencontrées.

Le score de Brier a été calculé pour évaluer la précision des prédictions de probabilité et comparé à la prévalence de l'évènement, afin de vérifier que le modèle démontre une performance supérieure à celle d'un modèle naïf. Un tableau et un graphique de calibration ont été réalisés.

Une matrice de confusion a été construite pour obtenir la sensibilité, la spécificité, la VPP et la Valeur prédictive négative (VPN). Une analyse d'erreurs a été réalisée : vérification si les probabilités prédites pour les FP et les faux négatifs (FN) étaient proches du seuil, puis sélection et exploration de variables cliniques (âge, sexe, etc.) pour identifier les profils associés aux erreurs.

Une analyse de la tendance du score a été réalisée pour vérifier que lorsque le score augmente, la proportion d'évènements augmente également.

Une courbe décisionnelle a été produite pour évaluer le bénéfice net du modèle. Différents seuils (autour de la prévalence de l'évènement) ont été testés, la classification d'un séjour comme similaire à un profil avec évènement étant comparée aux stratégies extrêmes : (1) classer tous les profils comme à risque élevé de présenter l'évènement, (2) classer tous les profils comme à bas risque de présenter l'évènement.

3.7.5. Modélisation (Python)

Le *pipeline* se structure en deux temps complémentaires. Dans un premier temps, un modèle de RF a été utilisé comme étape intermédiaire pour la sélection de variables. Ce modèle établit un classement des variables les plus importantes pour la prédiction, sur la base duquel des sous-ensembles contenant les 5, 7 et 10 variables les plus contributives ont été construits.

Les données préalablement préparées sous R, c'est-à-dire imputées, normalisées, encodées par *one-hot encoding* et divisées en jeux d'entraînement et de test, ont été reprises. Des vérifications ont été ajoutées afin de s'assurer qu'aucun patient ne soit présent simultanément dans les deux jeux et que les variables exclues lors du prétraitement sous RStudio ne soient pas réintroduites. Une table de correspondance a également été construite pour relier les colonnes issues du *one-hot encoding* à leur variable d'origine. Une fonction a ensuite été développée afin d'agréger les importances fournies par le RF et d'obtenir une importance globale par variable.

Comme le RF n'a pas mobilisé le jeu de test à ce stade, une validation croisée imbriquée a été mise en place. La validation croisée externe, composée de 5 sous-ensembles, a permis d'évaluer l'ensemble du processus sans recourir au jeu de test. La validation croisée interne a été utilisée pour comparer les tailles des sous-ensembles (5, 7 et 10 variables) et sélectionner le nombre optimal de variables, en maximisant la PR-AUC.

L'erreur standard a été calculée afin de retenir le sous-ensemble le plus parcimonieux présentant une performance proche de la meilleure observée. Le sous-ensemble le plus fréquemment choisi au cours des *folds* externes a été conservé. Le RF a ensuite été réentraîné sur l'ensemble du jeu d'entraînement afin d'identifier définitivement les variables retenues, puis des jeux de données réduits, pour l'entraînement et le test, ont été reconstruits.

Dans un second temps, plusieurs modèles de ML ont été comparés à l'aide de la librairie *PyCaret*, en utilisant les données issues de la sélection réalisée par RF. Les modèles ont été entraînés sur le jeu d'entraînement avec une validation croisée à 5 sous-ensembles, puis comparés selon la métrique PR-AUC. Le modèle le plus performant a été sélectionné, optimisé par ajustement des hyperparamètres, et conservé uniquement s'il présentait de meilleures performances que le modèle initial. Il a ensuite été réentraîné sur l'ensemble du jeu d'entraînement.

Les probabilités prédites ont été transformées en un score de risque allant de 1 à 5, selon une méthodologie similaire à celle utilisée pour la RL.

Le modèle final sélectionné a été évalué sur le jeu de test avec PR-AUC, ROC-AUC, score de Brier, une courbe de calibration, une matrice de confusion, la VPP et la VPN.

Les valeurs SHAP, qui montrent la contribution de chaque variable à la prédiction, ont été calculées sur le jeu de test. Les analyses globales (SHAP) ont permis de visualiser l'influence des variables, leur importance relative et leur variabilité entre patients. Les analyses individuelles (SHAP) ont consisté à regarder des cas de FP et de FN pour identifier les variables qui ont conduit à des erreurs.

Résultats

4. RÉSULTATS

4.1. Variables numériques

Une première description des variables numériques a été réalisée afin d'évaluer, pour chacune d'entre elles, la proportion de données manquantes, ainsi que les mesures de tendance centrale et de dispersion. Les bornes cliniques ont également été appliquées afin de conserver uniquement des valeurs cohérentes et d'exclure d'éventuelles erreurs d'encodage (Tableau 6).

L'âge médian de la cohorte était de 70 ans, indiquant une population majoritairement âgée. La plupart des paramètres vitaux présentaient des valeurs cohérentes avec une population hospitalisée. Les variables biologiques affichaient en revanche davantage de valeurs manquantes, suggérant qu'elles n'étaient mesurées que dans certains contextes cliniques. Lors de la comparaison entre les séjours avec et sans évènement, les valeurs étaient globalement moins favorables dans le groupe associé à l'évènement cible, notamment l'âge, la fréquence cardiaque, la PAD, la CRP et la SpO₂.

Tableau 6 : Description des variables numériques du jeu d'entraînement.

Variable	NA (%)	Total	Sans évènement	Avec évènement	p-valeur
Âge (ans)	0,0	70 [55 - 80]	69 [54 - 79]	76 [67 - 85]	<0,001
Créatinine (mg/dL)	5,7	0,91 [0,72 - 1,21]	0,9 [0,72 - 1,19]	1,08 [0,75 - 1,83]	<0,001
CRP (mg/L)	5,6	32,05 [6,15 - 93,17]	30,02 [5,75 - 89,12]	85,1 [30,95 - 176,85]	<0,001
EWS (score)	7,5	1 [0 - 3]	1 [0 - 3]	2 [0 - 4]	0,045
Fréquence cardiaque (bpm)	0,9	79 [70,67 - 87,71]	78,57 [70,5 - 87]	85 [74 - 100]	<0,001
Glucose (mg/dL)	9,1	110,5 [97,33 - 132]	110,2 [97,45 - 131,33]	117 [96 - 148]	0,010
PAD (mmHg)	1,1	72,86 [66,5 - 80]	73,14 [67 - 80]	65 [58 - 78,25]	<0,001
PAS (mmHg)	1,0	128,33 [117 - 140,31]	128,57 [117,57 - 140,57]	120 [106 - 137]	<0,001
Potassium (mmol/L)	7,2	4,11 [3,85 - 4,38]	4,12 [3,85 - 4,38]	4,08 [3,69 - 4,46]	0,093

Variable	NA (%)	Total	Sans évènement	Avec évènement	p-valeur
SpO ₂ (%)	1,4	96,14 [94,5 - 97,43]	96,2 [94,67 - 97,5]	95 [92 - 97]	<0,001
Température (°C)	1,3	36,21 [36,06 - 36,41]	36,21 [36,07 - 36,4]	36,2 [36 - 36,6]	0,428

4.2. Variables catégorielles

Les variables catégorielles ont également été explorées afin d'évaluer leur répartition selon la survenue de l'évènement cible (Tableau 7). La répartition des séjours selon les campus d'hospitalisation et d'urgence était globalement comparable entre les séjours avec et sans évènement. Les différences observées entre les principaux sites n'étaient pas statistiquement significatives, que ce soit pour le campus d'hospitalisation ($p = 0,374$) ou pour celui d'urgence ($p = 0,538$).

En revanche, l'administration d'oxygène à l'admission était nettement plus fréquente chez les patients ayant présenté l'évènement cible ($p < 0,001$). De même, la répartition des disciplines post-urgence différait significativement selon la survenue de l'évènement cible ($p < 0,001$). Certaines disciplines post-urgence, notamment la gériatrie, la cardiologie, la médecine interne générale et la pneumologie, affichaient une proportion plus élevée d'évènements. La proportion de patients de sexe masculin était également plus élevée dans le groupe avec évènement que dans le groupe sans évènement ($p = 0,001$).

Ces résultats suggèrent des tendances intéressantes sans permettre de conclure à ce stade, puisqu'il s'agit uniquement d'analyses descriptives.

Tableau 7. Description des variables catégorielles du jeu d'entraînement.

Variable	Modalité	Total	Sans évènement	Avec évènement	p-valeur
Campus d'hospitalisation	ST	9 016 (61,8 %)	8 420 (61,9 %)	596 (60,2 %)	0,374
Campus d'hospitalisation	NDB	5 565 (38,2 %)	5 171 (38 %)	394 (39,8 %)	0,374
Campus d'hospitalisation	ES	1 (0 %)	1 (0 %)	0 (0 %)	0,374

Variable	Modalité	Total	Sans évènement	Avec évènement	p-valeur
Campus d'hospitalisation	FR	1 (0 %)	1 (0 %)	0 (0 %)	0,374
Campus d'urgence	ST	8 642 (59,3 %)	8 065 (59,3 %)	577 (58,3 %)	0,538
Campus d'urgence	NDB	5 941 (40,7 %)	5 528 (40,7 %)	413 (41,7 %)	0,538
Discipline post-urgence	Médecine interne générale	2 666 (18,3 %)	2 492 (18,3 %)	174 (17,6 %)	<0,001
Discipline post-urgence	Cardiologie	1 771 (12,1 %)	1 625 (12 %)	146 (14,7 %)	<0,001
Discipline post-urgence	Pneumologie	1 545 (10,6 %)	1 400 (10,3 %)	145 (14,6 %)	<0,001
Discipline post-urgence	Gériatrie	1 277 (8,8 %)	1 137 (8,4 %)	140 (14,1 %)	<0,001
Discipline post-urgence	Chir Abdo-Endo Transplantations	1 138 (7,8 %)	1091 (8 %)	47 (4,7 %)	<0,001
Discipline post-urgence	Chir Ortho	935 (6,4 %)	905 (6,7 %)	30 (3 %)	<0,001
Oxygène à l'admission	non	12 491 (85,7 %)	12 015 (88,4 %)	476 (48,1 %)	<0,001
Oxygène à l'admission	oui	1863 (12,8 %)	1411 (10,4 %)	452 (45,7 %)	<0,001
Oxygène à l'admission	Manquant	229 (1,6 %)	167 (1,2 %)	62 (6,3 %)	<0,001
Sexe	F	7389 (50,7 %)	6937 (51 %)	452 (45,7 %)	0,001
Sexe	M	7194 (49,3 %)	6656 (49 %)	538 (54,3 %)	0,001

4.3. Le développement de la régression logistique

Le modèle final de RL présentait un paramètre de pénalisation de 0,000339 et un paramètre de mélange de 0,789. Une pénalisation *Elastic net* avait été retenue, la valeur du paramètre de mélange indique que la pénalisation est proche du *LASSO*, ce qui permet de réduire l'influence des variables peu informatives tout en conservant une bonne stabilité du modèle (Tableau 8).

Tableau 8 : Hyperparamètres du modèle de RL.

Paramètre de pénalisation	Paramètre de mélange	PR-AUC OOF
0,000339	0,789	0,489

4.4. Performance du modèle sur le jeu de validation interne et sur le jeu de test

Le jeu de test a été utilisé pour évaluer les performances du modèle. La ROC-AUC était de 0,835 [0,807 ; 0,863] (Figure 9), indiquant une bonne capacité globale du modèle à distinguer les séjours avec évènement de ceux sans évènement. La courbe ROC se situait nettement au-dessus de la diagonale de référence, correspondant à une discrimination aléatoire.

La PR-AUC était de 0,386 [0,323 ; 0,455]. La courbe PR montre que la précision diminue progressivement lorsque le rappel augmente, traduisant une augmentation de FP lorsque le modèle cherche à détecter davantage d'évènements. La PR-AUC restait toutefois supérieure à la prévalence de l'évènement (ligne horizontale) (Figure 10).

Le jeu de test contenait un nombre plus limité d'évènements, ce qui rend la PR-AUC particulièrement sensible aux variations d'échantillonnage, comme le montre l'intervalle de confiance.

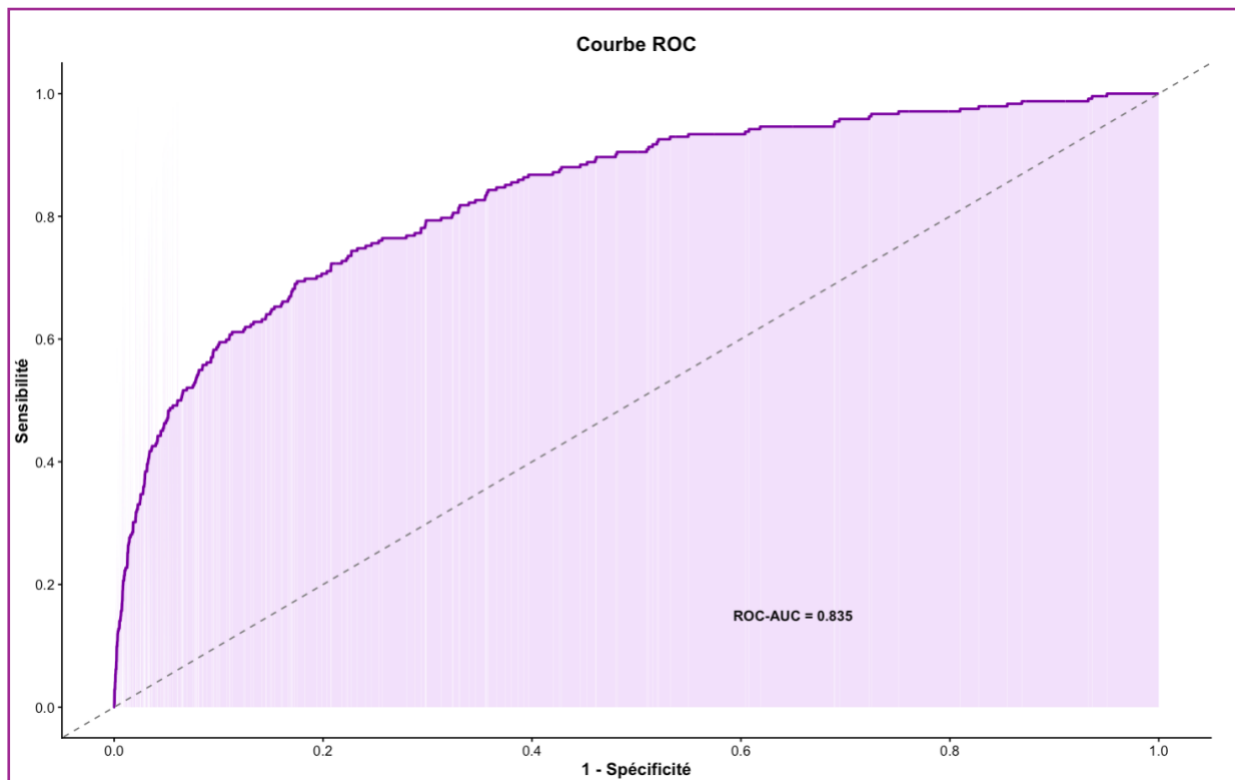


Figure 9 : Courbe ROC du modèle de RL sur le jeu de test. La ligne bleue représente la performance du modèle selon les différents seuils de décision, tandis que la diagonale pointillée correspond à une classification aléatoire. L'aire bleue représente l'aire sous la courbe ROC.

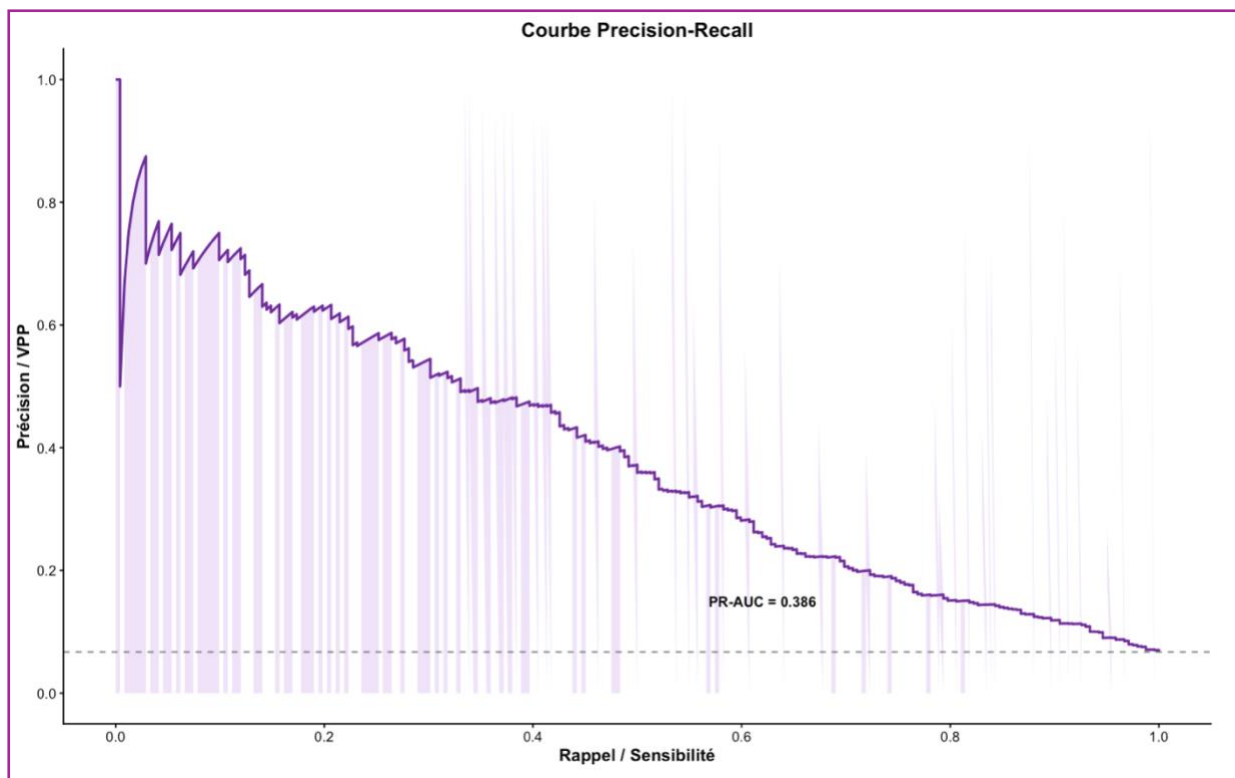


Figure 10 : Courbe PR du modèle de RL sur le jeu de test. La ligne bleue représente la performance du modèle, l'aire bleue correspond à la PR-AUC, et la ligne horizontale pointillée indique la prévalence de l'évènement.

Le score de Brier du modèle était de 0,05 contre 0,063 pour un modèle naïf basé uniquement sur la prévalence. Cela suggère que les probabilités prédites apportaient une amélioration par rapport à une stratégie consistant à prédire systématiquement la probabilité moyenne de l'évènement (Tableau 9).

Tableau 9. Performance du modèle sur le jeu de validation interne et sur le jeu de test.

Métrique	Résultat [IC 95 %]
PR-AUC out-of-fold	0,489 [0,453 ; 0,522]
PR-AUC test	0,386 [0,323 ; 0,455]
ROC-AUC test	0,835 [0,807 ; 0,863]
Score de Brier test	0,05 [0,045 ; 0,056]
Score de Brier naïf	0,063 [0,056 ; 0,069]

4.5. Variable contributive au modèle

Le modèle fondait ses prédictions sur l'ensemble des variables retenues, mais certaines contribuaient davantage que d'autres. Les coefficients absolus les plus élevés concernaient l'EWS, la fréquence cardiaque, l'âge, la PAD, la CRP, l'administration d'oxygène à l'admission, la SpO₂ ainsi que certaines disciplines post-urgence. Ces résultats montrent que les prédictions reposaient à la fois sur les paramètres vitaux, biologiques et contextuels liés à l'hospitalisation.

L'EWS présentait l'importance en valeur absolue la plus élevée, ce qui suggère un rôle majeur dans la prédiction de l'évènement cible. Les variables de fréquence cardiaque, d'âge, de PAD et de CRP apparaissent également cohérentes avec la littérature clinique, car elles sont souvent associées à une dégradation de l'état de santé. Les coefficients ne permettent pas ici d'interpréter directement l'effet de chaque variable sur le risque, car les données numériques ont été standardisées et les coefficients ont été modifiés par la pénalisation de type *Elastic net*.

Cette analyse sert donc surtout à identifier les variables qui contribuent le plus à la prédiction (Tableau 10).

Tableau 10. Variables les plus contributives au modèle de RL.

Variable	Coefficient	Importance en valeur absolue
EWS	1,342	1,342
Fréquence cardiaque	-0,660	0,660
Âge	-0,540	0,540
PAD	0,514	0,514
Indicateur de signe (<) CRP	0,461	0,461
Discipline post-urgence gynécologie	0,415	0,415
Oxygène à l'admission (non)	0,346	0,346
Oxygène à l'admission (oui)	-0,307	0,307
Discipline post-urgence urologie	0,290	0,290
SpO ₂	0,248	0,248
Sexe	-0,212	0,212
Discipline post-urgence chirurgie ortho.	0,193	0,193
Discipline post-urgence ORL	0,185	0,185
PAS	0,166	0,166
Discipline post-urgence cardiologie	-0,143	0,143

4.6. Transformation de la probabilité en score

Après obtention des probabilités prédites par le modèle, celles-ci ont été transformées en score de risque, à partir des quintiles des prédictions OOF du jeu d'entraînement, afin de faciliter l'utilisation en milieu hospitalier. Les mêmes seuils ont ensuite été appliqués au jeu de test. Cette transformation a également permis de calculer des métriques à partir d'une décision binaire. Le score variait de 1 à 5, avec un seuil de classification fixé à 5.

Les métriques du modèle montraient une sensibilité de 68,2 %, indiquant qu'un score de 5 permettait d'identifier une proportion importante d'évènements, tout en laissant persister une part non négligeable de FN. La spécificité atteignait 82,9 %, montrant que la plupart des séjours sans évènement étaient correctement classés. La VPP restait faible, ce qui est attendu compte tenu de la faible prévalence de l'évènement. En revanche, la VPN était élevée, à 0,973, suggérant qu'un score inférieur à 5 était associé à une faible probabilité de survenue de l'évènement (Tableau 11).

Tableau 11 : Performance du score binaire issu du modèle.

Métrique	Valeur
Sensibilité	0,682
Spécificité	0,829
Valeur prédictive positive	0,223
Valeur prédictive négative	0,973
Accuracy	0,819

L'évolution du score montrait une augmentation progressive du taux d'évènements observés, suggérant une cohérence globale entre les différents niveaux du score et le risque clinique associé (Figure 11).

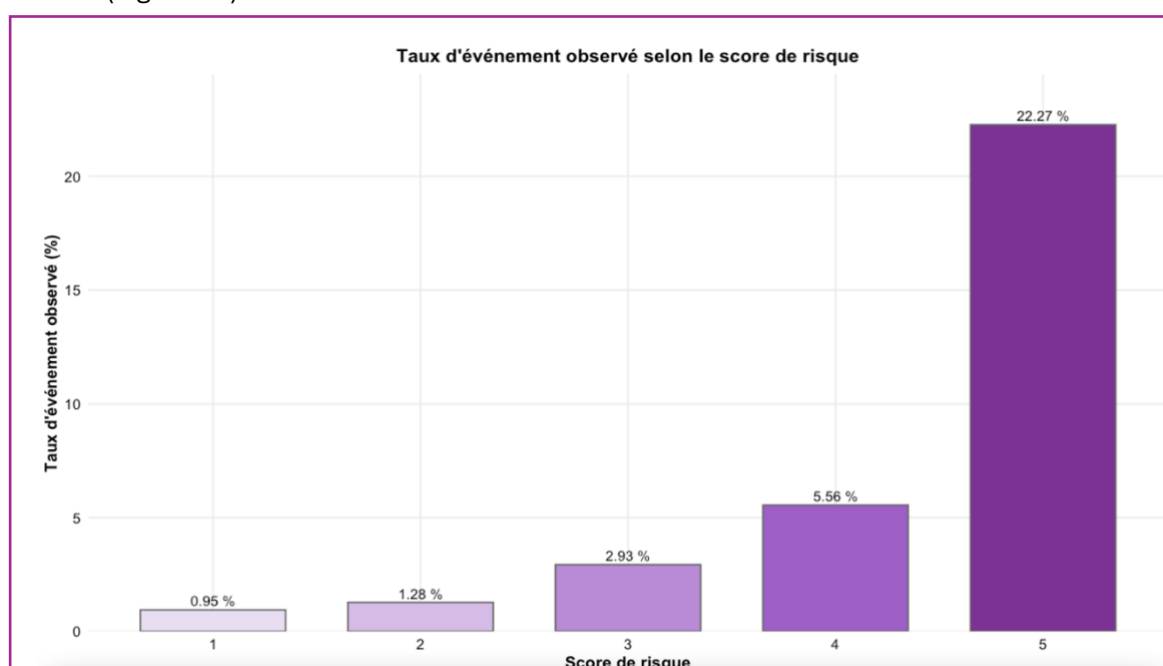


Figure 11: Variation du taux d'évènements observés selon le niveau du score.

La règle binaire a permis de classer les scores de 5 comme à haut risque et les scores inférieurs comme à bas risque. Sur le jeu de test, 165 évènements et 2788 non-évènements ont été

correctement identifiés, tandis que 576 séjours sans évènement ont été classés à tort comme à haut risque et 77 séjours avec évènement ont été classés à tort comme à bas risque (Tableau 12).

Tableau 12 : Matrice de classification du score binaire sur le jeu de test.

Indicateur	Abréviation	Nombre
Vrais positifs	VP	165
Faux positifs	FP	576
Faux négatifs	FN	77
Vrais négatifs	VN	2 788

4.7. Analyses complémentaires du modèle

La calibration du modèle a ensuite été évaluée à l'aide d'une courbe de calibration, qui permet de comparer une calibration parfaite, servant de référence, à celle obtenue pour notre modèle. Cette courbe confronte le taux d'évènements observé à la probabilité moyenne prédite par le modèle. Pour la RL, une bonne concordance avec la tendance attendue est observée, même si, pour les probabilités les plus élevées, le modèle passe légèrement sous la courbe idéale, ce qui traduit une légère surestimation du risque (Figure 12).

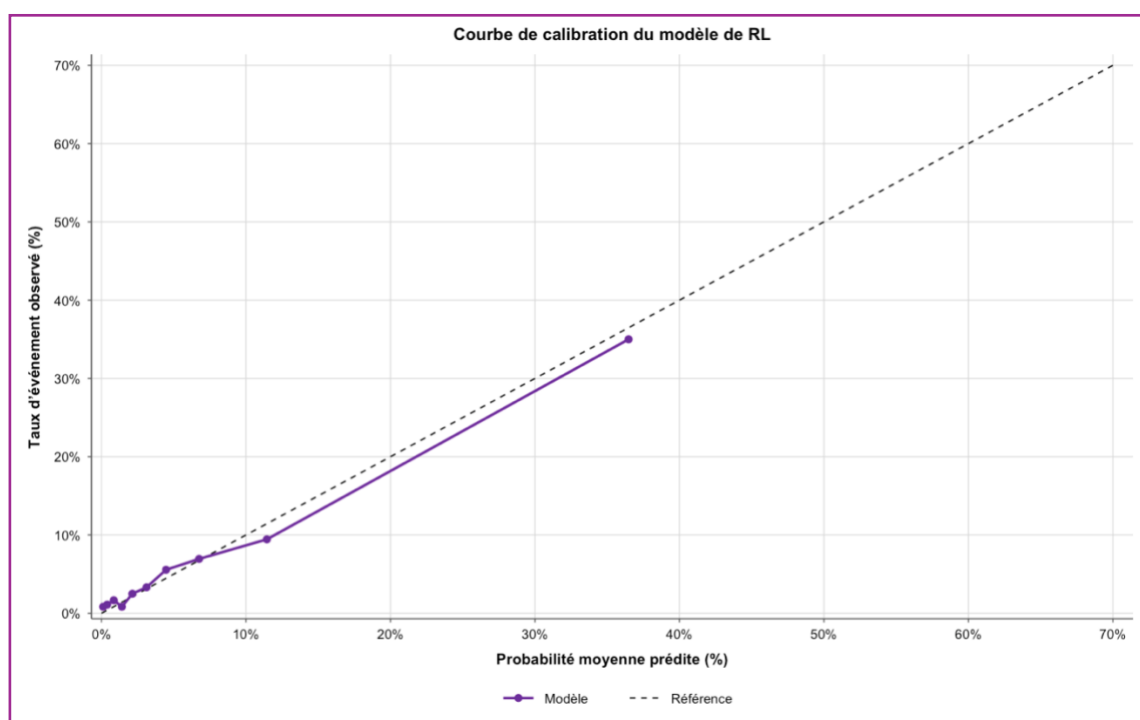


Figure 12 : Courbe de calibration du modèle de RL. La courbe de référence est représentée en pointillés, tandis que la courbe du modèle est représentée en mauve.

Une analyse complémentaire a été réalisée afin d'examiner les erreurs de classification produites par le modèle (Tableau 13).

Tableau 13 : Analyse des erreurs de classification selon la probabilité prédite et le score médian.

Type de classification	Nombre de séjours	Probabilité médiane prédite (%)	Q1 (%)	Q3 (%)	Score médian
VP	165	31,39	17,00	52,81	5
FP	576	13,91	10,36	22,26	5
FN	77	3,82	2,16	5,63	4
VN	2 788	1,65	0,58	3,58	2

L'analyse des erreurs montrait que les VP présentaient la probabilité prédite la plus élevée, tandis que les VN affichaient la plus faible. Les FP présentaient une probabilité médiane intermédiaire, à 13,9 %, alors que les FN avaient une probabilité médiane de 3,82 % et un score médian de 4. Ces erreurs concernaient donc principalement des profils intermédiaires ou situés juste en dessous du seuil de décision.

Par la suite, une courbe de décision a été réalisée afin d'évaluer si le modèle apportait un bénéfice net par rapport à deux stratégies extrêmes : classer tous les séjours comme à haut risque ou, au contraire, classer tous les séjours comme à bas risque (Figure 13).

La courbe montrait que le modèle présentait un bénéfice net supérieur à ces deux stratégies de référence. Elle restait constamment au-dessus de la stratégie consistant à classer tous les patients comme à risque nul, ce qui indique un bénéfice net positif. Elle demeurait également supérieure à la stratégie consistant à classer tous les patients comme à haut risque, en particulier lorsque le seuil de décision augmentait.

4.8. Random Forest

Après le modèle de RL développé sous RStudio, une seconde approche a été menée sous Python afin de comparer plusieurs modèles simultanément et d'identifier celui présentant les meilleures performances. Cette étape visait également à explorer le nombre minimal de variables permettant de conserver une performance satisfaisante. Trois configurations ont ainsi été testées, avec une prédiction fondée sur 5, 7 et 10 variables.

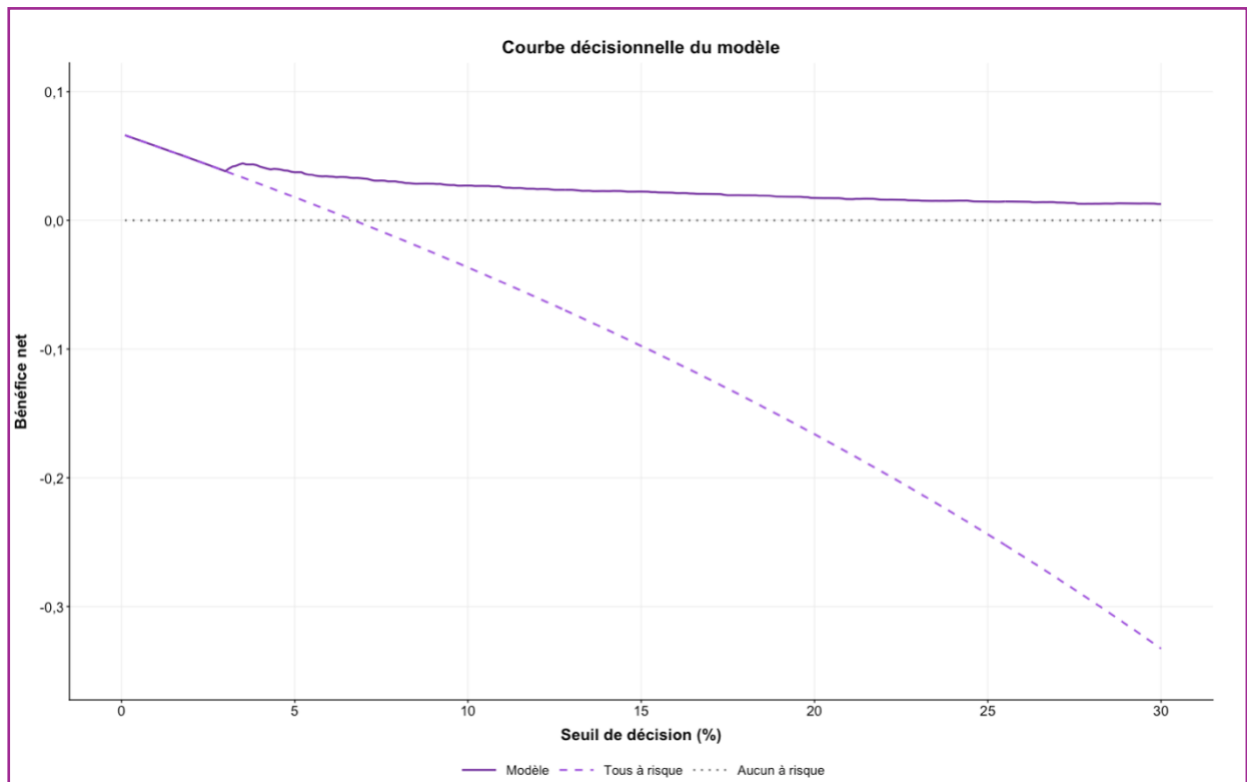


Figure 13 : Courbe d'analyse décisionnelle comparant le modèle aux stratégies extrêmes. La ligne violette pleine représente le bénéfice net du modèle. La ligne violette en grands pointillés correspond à la stratégie « tous à risque ». La ligne grise en petits pointillés correspond à la stratégie « aucun à risque ».

Pour cela, un modèle de RF a été entraîné à l'aide d'une validation croisée imbriquée. La validation croisée interne a servi à déterminer le nombre optimal de variables à conserver. Afin de privilégier un modèle parcimonieux, la règle du *one standard error* a été appliquée. Cette approche permet de retenir un modèle comportant moins de variables, tout en conservant des performances comparables à celles d'un modèle plus complexe.

La validation croisée externe a ensuite servi d'évaluation finale, sur des données jamais observées par le modèle, en appliquant le nombre de variables sélectionné par la validation interne.

Le top 7 a été retenu dans 3 *folds* sur 5, ce qui conduit à conserver 7 variables pour la suite de l'analyse. Les PR-AUC externes variaient entre 0,775 et 0,862, indiquant des performances globalement satisfaisantes (Tableau 14).

Tableau 14 : Sélection du nombre optimal de variables par validation croisée imbriquée.

Fold externe	Top k sélectionné	PR-AUC interne Top 5	PR-AUC interne Top 7	PR-AUC interne Top 10	PR-AUC interne moyenne	Écart-type interne	PR-AUC externe
1	10	0,777	0,807	0,837	0,837	0,028	0,862
2	7	0,801	0,825	0,833	0,825	0,024	0,775
3	10	0,798	0,819	0,852	0,852	0,017	0,853
4	7	0,788	0,828	0,815	0,828	0,021	0,836
5	7	0,781	0,806	0,815	0,806	0,018	0,826

Les variables retenues pour la suite des modèles produits sous PyCaret étaient les suivantes : EWS, administration d'oxygène à l'admission, CRP, PAD, glucose, SpO₂ et créatinine enzymatique (Figure 14). L'EWS apparaissait comme la variable la plus contributive. Les autres variables retenues étaient également liées à la sévérité clinique du patient, ce qui paraît cohérent. Elles couvrent en effet différents aspects de l'état physiologique, notamment le statut respiratoire, l'équilibre hémodynamique, ainsi que les fonctions métaboliques et rénales.

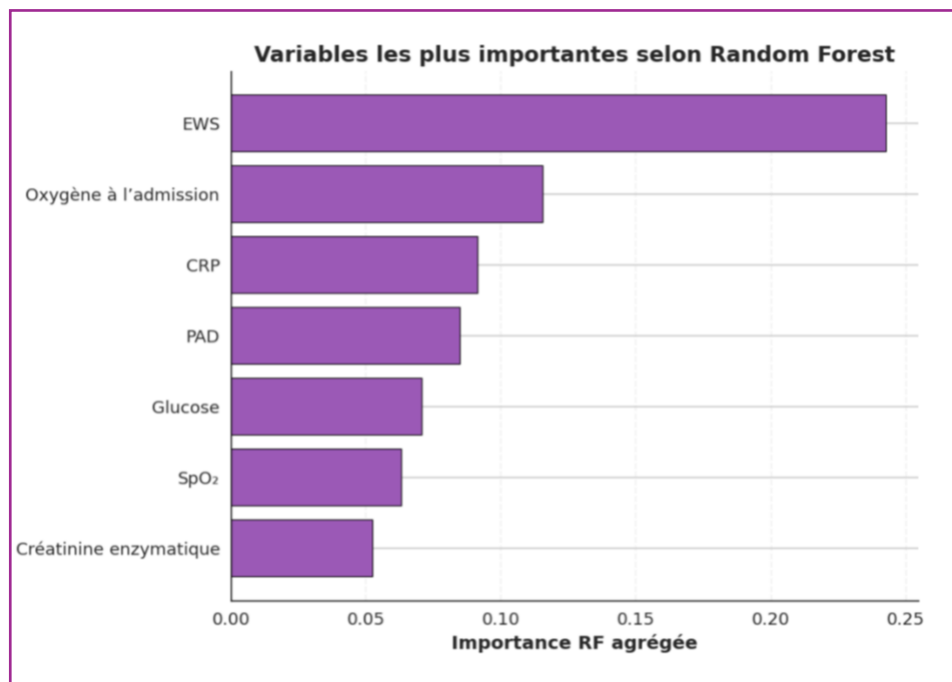


Figure 14: Variables sélectionnées à l'issue de la validation croisée imbriquée.

4.9. Test de plusieurs modèles

PyCaret a ensuite été utilisé pour tester dix modèles de classification en maximisant la PR-AUC. Les modèles testés incluaient notamment le *LightGBM*, le *Gradient Boosting Classifier* (GBC), le *Random Forest Classifier* (RF), l'*AdaBoost Classifier* (Ada), l'*Extra Trees Classifier* (ET), le *Decision Tree Classifier* (DT), le *K Neighbors Classifier* (KNN), le *Naive Bayes* (NB), la *Linear Discriminant Analysis* (LDA) et la RL. Le modèle présentant la meilleure performance a été retenu, à savoir le *LightGBM*.

Ce modèle présentait une performance (*accuracy*) de 0,974, indiquant une bonne capacité de discrimination en validation croisée. La ROC-AUC atteignait 0,973, ce qui montre une très bonne capacité à distinguer les séjours avec évènement de ceux sans évènement. Compte tenu de la faible prévalence de l'évènement cible, la PR-AUC a également été examinée, avec une valeur de 0,654, ce qui reste très satisfaisant dans ce contexte (Tableau 15).

Tableau 15. Performances du modèle *LightGBM*.

Model	Accuracy	ROC-AUC	PR-AUC
LightGBM	0,974*	0,973	0,654

*calculée au seuil de classification par défaut *PyCaret* (0,5)

Un ajustement (*tuning*) des hyperparamètres du *LightGBM* a ensuite été réalisé, avant une évaluation sur le jeu de test. Ce dernier était identique à celui utilisé pour la RL. La prévalence ainsi que le score de *Brier* naïf y étaient donc les mêmes.

Les performances observées sur le jeu de test étaient très élevées. La courbe ROC se situait nettement au-dessus de la diagonale de référence, avec une ROC-AUC de 0,986, traduisant une excellente capacité globale de discrimination (Figure 16). La courbe PR était également largement supérieure à la ligne de prévalence, avec une PR-AUC de 0,894, montrant que le modèle parvenait à concentrer les séjours avec évènement parmi les prédictions les plus à risque (Figure 17). Enfin, le score de Brier du modèle, égal à 0,018, était nettement inférieur à celui du modèle naïf, suggérant que les probabilités prédites apportaient une information bien plus pertinente qu'une simple estimation fondée sur la prévalence (Tableau 16).

Tableau 16. Performances du modèle LightGBM sur le jeu de test.

Métrique	Valeur
Prévalence test	0,067
PR-AUC test	0,894 [0,862 ; 0,924]
ROC-AUC test	0,986 [0,981 ; 0,991]
Score de Brier test	0,018 [0,015 ; 0,022]
Score de Brier naïf	0,063 [0,056 ; 0,070]

4.10. L'analyse du modèle LightGBM

L'analyse du taux d'évènement selon les tranches de score montrait une progression nette du risque. Aucun évènement n'était observé dans les trois premiers niveaux du score. Le score 5 présentait un taux d'évènements nettement plus élevé, atteignant 30,22 % (Figure 15).

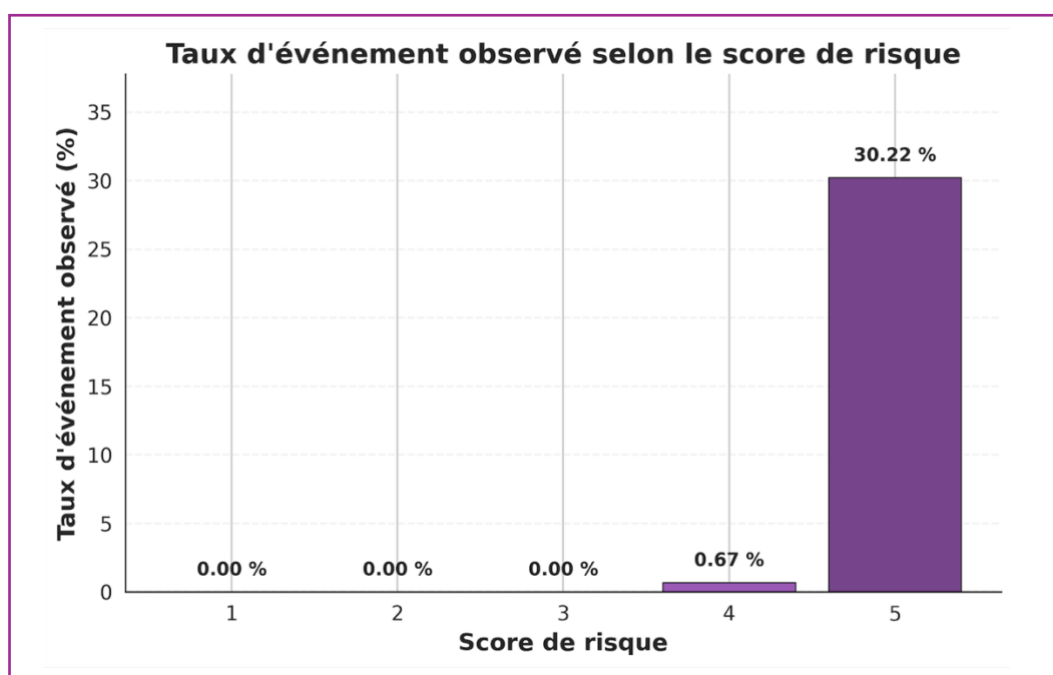


Figure 15 : Taux d'évènement observé selon le niveau du score de risque du LightGBM.

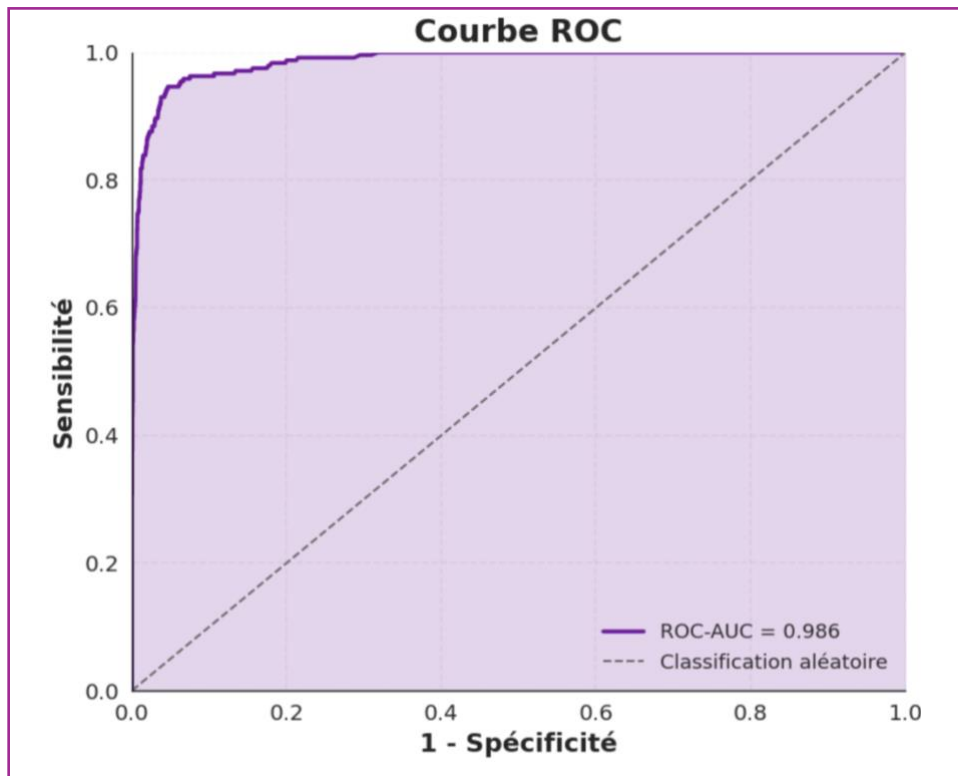


Figure 16 : Courbe ROC du modèle LightGBM sur le jeu de test. La ligne violette représente la performance du modèle selon les différents seuils de décision, tandis que la diagonale pointillée correspond à une classification aléatoire. L'aire violette représente l'aire sous la courbe ROC.

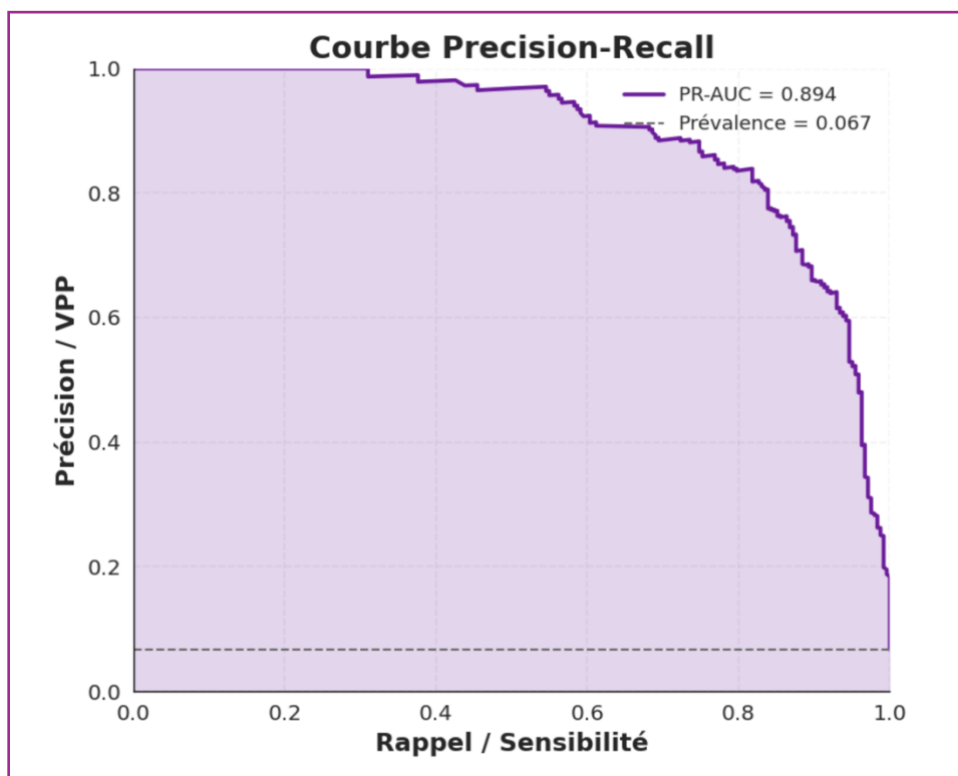


Figure 17 : Courbe PR du modèle LightGBM sur le jeu de test. La ligne violette représente la performance du modèle, l'aire violette correspond à la PR-AUC, et la ligne horizontale pointillée indique la prévalence de l'évènement.

Après l'évaluation globale des performances, une analyse d'interprétabilité a été réalisée afin de mieux comprendre les variables influençant les prédictions du modèle, une analyse *SHAP* a été réalisée sur le jeu de test (Figure 18). Les valeurs *SHAP* positives indiquent une contribution vers une probabilité prédite plus élevée de l'évènement, tandis que les valeurs négatives indiquent une contribution inverse. Toutefois, certains effets n'étaient pas strictement linéaires : par exemple, certaines valeurs élevées d'EWS étaient associées à des valeurs *SHAP* négatives, suggérant que le modèle capte des interactions complexes entre variables plutôt qu'un effet isolé de chaque paramètre.

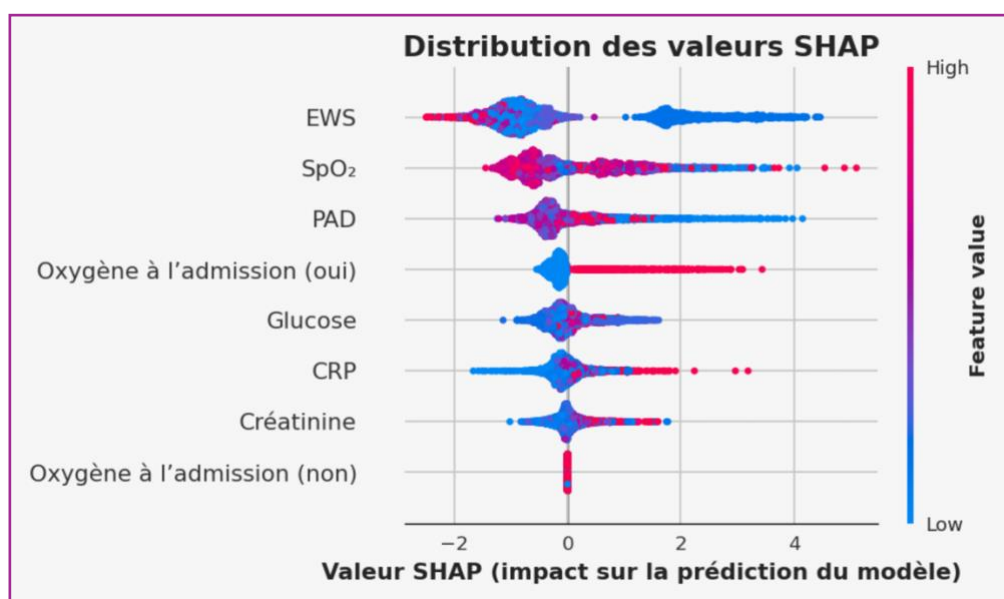


Figure 18 : Distribution des valeurs *SHAP* du modèle LightGBM sur le jeu de test. Chaque point représente un séjour hospitalier. L'axe horizontal indique l'impact de chaque variable sur la prédiction du modèle, tandis que la couleur représente la valeur de la variable, des valeurs faibles en bleu aux valeurs élevées en rouge.

Le score a ensuite été transformé en décision binaire, avec la règle suivante ; un score égal à 5 correspond à un patient classé à haut risque, tandis qu'un score inférieur à 5 correspond à un patient classé à bas risque. La performance du modèle fondée sur cette décision binaire atteignait 0,847.

Le modèle présentait une sensibilité très élevée, à 0,975, indiquant un nombre faible de FN. Seuls 6 séjours avec évènement n'étaient pas classés comme à haut risque, contre 236 VP correctement identifiés. La VPN était également très élevée, à 0,998, ce qui suggère qu'un séjour non classé comme à haut risque avait très peu de chances d'évoluer vers une admission en USI ou un décès. La spécificité atteignait 0,838, avec 2 819 VN, traduisant une bonne capacité du modèle à reconnaître les séjours sans évènement.

En revanche, la VPP restait plus limitée à 0,302. Cela signifie qu'une part importante des séjours classés à haut risque ne présentaient finalement pas l'évènement cible, avec 545 FP. Cette limite

s'explique en grande partie par la faible prévalence de l'évènement dans le jeu de test (Tableau 17).

Tableau 17. Performances de classification binaire du score LightGBM.

Métrique	Valeur
Sensibilité	0,975
Spécificité	0,838
Valeur prédictive positive	0,302
Valeur prédictive négative	0,998
Accuracy	0,847

Une évaluation de la calibration du modèle a ensuite été menée afin de vérifier si les probabilités prédites étaient cohérentes avec les évènements effectivement observés. Les probabilités moyennes prédites ont été comparées au taux d'évènement observé par groupe de risque. La ligne pointillée représentait la calibration parfaite. La courbe du modèle restait globalement proche de cette référence, même si un léger écart apparaissait en fin de courbe, suggérant une légère sous-estimation des probabilités prédites (Figure 19).

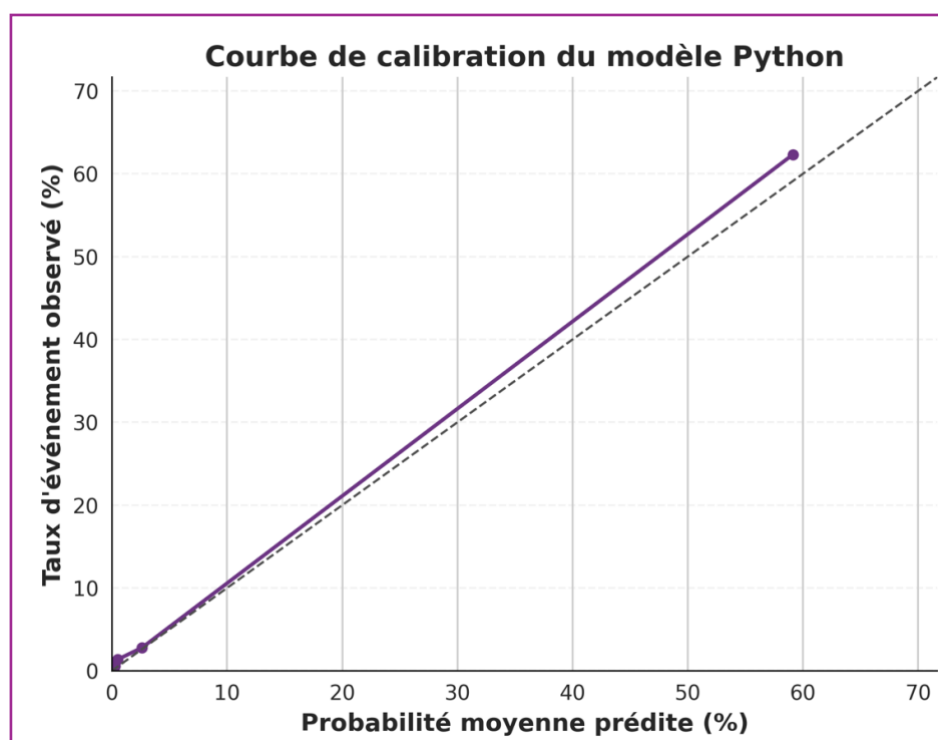


Figure 19 : Courbe de calibration du modèle LightGBM. La courbe de référence est représentée en pointillés, tandis que la courbe du modèle est représentée en mauve.

Une analyse des erreurs de classification a également été réalisée. Pour les séjours réellement associés à un évènement, les VP présentaient une probabilité médiane très élevée, autour de 86,6 %, tandis que les FP affichaient une probabilité médiane beaucoup plus faible, de 2,41 %. Pour les VN, la probabilité médiane était très basse, à 0,08 %, avec un score médian de 3. Les FN présentaient quant à eux une probabilité médiane de 0,44 % et un score médian de 4, ce qui indique qu'un état clinique défavorable avait tout de même été partiellement détecté (Tableau 18).

Tableau 18. Analyse des erreurs de classification du modèle LightGBM.

Type de classification	Nombre de séjours	Probabilité médiane prédite (%)	Q1 (%)	Q3 (%)	Score médian
VP	236	86,620	60,040	97,670	5
FP	545	2,410	1,160	7,390	5
FN	6	0,440	0,300	0,550	4
VN	2 819	0,080	0,040	0,180	3

Enfin, une courbe décisionnelle a été produite afin d'évaluer l'intérêt clinique du modèle (Figure 20). La courbe du modèle se situait au-dessus des deux stratégies extrêmes, consistant soit à considérer tous les patients comme à risque, soit à ne considérer aucun patient comme à risque. Cela indique que le modèle apporte un bénéfice net supérieur à ces approches de référence. La stabilité de la courbe suggère également que le modèle conserve un intérêt pour plusieurs seuils de décision.

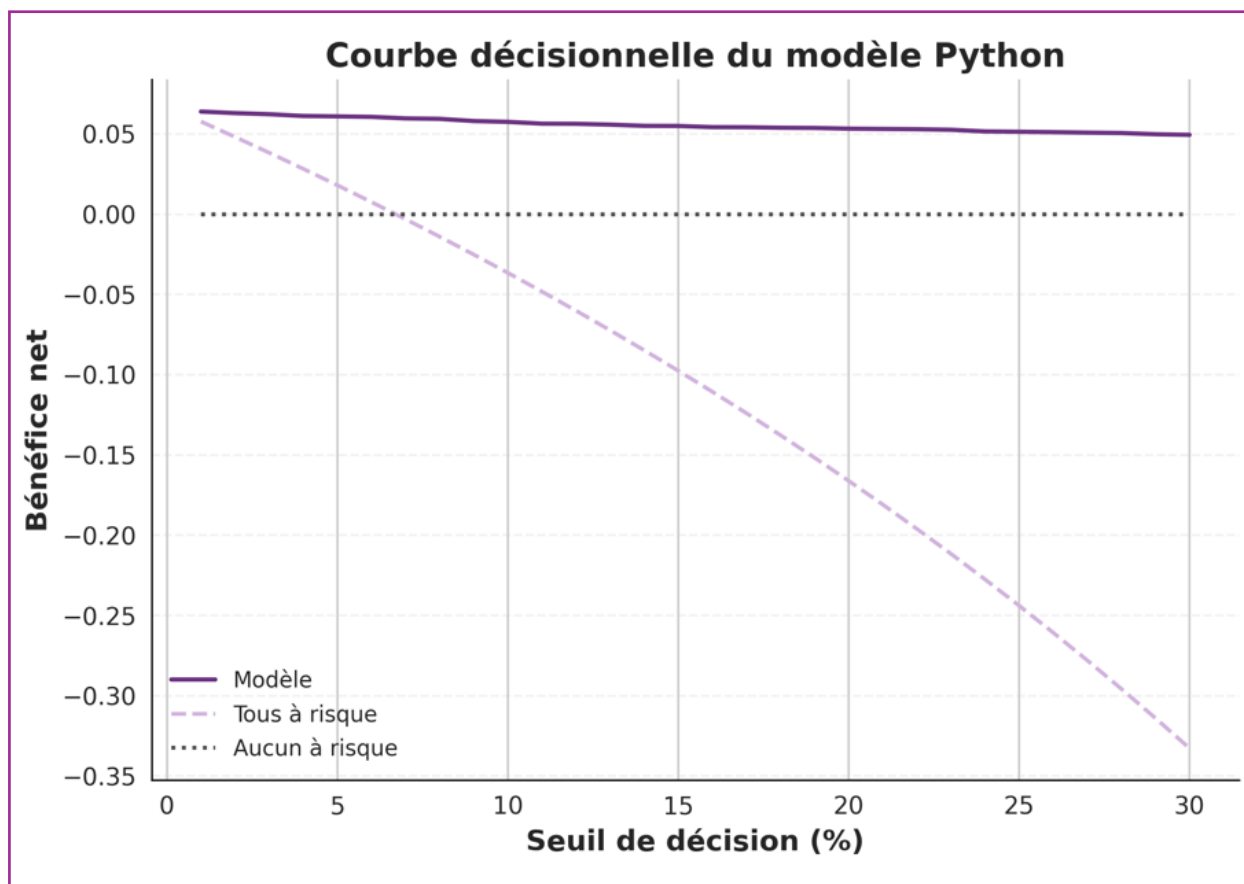


Figure 20 : Courbe décisionnelle du modèle LightGBM. La ligne violette pleine représente le bénéfice net du modèle. La ligne violette en grands pointillés correspond à la stratégie « tous à risque ». La ligne grise en petits pointillés correspond à la stratégie « aucun à risque ».

Discussion

5. DISCUSSION

L'environnement hospitalier est actuellement confronté à des problématiques majeures pour assurer la qualité et la sécurité des soins. En particulier, la pénurie de personnel soignant, l'augmentation de la fatigue et des *burnouts* parmi les professionnels de la santé, crée une tension sur la dispense de soins^{13,62}. En corollaire, la charge de travail est grandissante avec une population vieillissante et une majoration des comorbidités rencontrées⁶³. Ainsi, assurer une surveillance adéquate en milieu hospitalier chez des patients parfois complexes est devenu une tâche ardue pour un personnel en nombre limité^{32,64}. En conséquence, la dégradation clinique des patients au sein de l'hôpital constitue une problématique complexe qui s'avère difficile à anticiper et à prendre en charge de manière adaptée. Cette dégradation a des répercussions importantes pour la sécurité des soins et le pronostic de ces patients^{3,7,53}. Elle influence directement le déroulement du séjour hospitalier, en augmentant le risque de morbidité, mais également en impactant l'organisation des services et en générant des coûts supplémentaires pour l'hôpital⁶⁵⁻⁶⁷. Dans ce contexte, la stratification des patients selon leur niveau de risque apparaît comme une approche pertinente, permettant de prioriser certains profils et de mettre en place des mesures précoces, afin d'améliorer la prise en charge et de limiter les conséquences associées à cette dégradation⁶⁸. Ce projet avait donc pour objectif de développer un outil permettant la détection de cette dégradation clinique, définie ici par la survenue d'une admission en USI ou d'un décès. L'outil développé repose sur un score de stratification en cinq niveaux, construit à partir de modèles de ML, et visant à identifier les séjours à risque. Ce score, stratifiant ainsi les séjours selon leur niveau de risque, pourrait faciliter la priorisation des patients nécessitant une surveillance renforcée.

La régression logistique et ses perspectives

Plusieurs travaux ont montré que les scores cliniques développés à partir de méthodes de ML présentent des performances supérieures à celles des approches traditionnelles^{42,69}. Ce projet a donc permis d'explorer diverses opportunités dans ce domaine. Parmi celles-ci, deux ont retenu l'attention par leur potentiel. Dans un premier temps, une RL a été utilisée comme modèle de référence, en raison de sa simplicité et de son interprétabilité. La prévalence de l'évènement étudié étant de 6,8 %, le déséquilibre entre classes a été pris en compte dans la construction des modèles et l'évaluation des performances, notamment par l'analyse conjointe de la ROC-AUC et de la PR-AUC. La RL a présenté des performances globalement satisfaisantes, avec une ROC-AUC de 0,835 et une PR-AUC de 0,386 sur l'échantillon test, supérieures à la prévalence de l'évènement.

La deuxième étape s'est donc intéressée à la possibilité de transformer les probabilités issues du modèle en score de risque. En effet, ce type de modèle est développé dans une perspective d'utilisation en milieu hospitalier, où des probabilités brutes sont généralement moins pertinentes

qu'une échelle de stratification plus lisible et plus directement exploitable. La littérature montre d'ailleurs que les modèles convertis en scores sont plus interprétables, plus simples à comprendre et davantage adaptés aux professionnels de santé, ce qui facilite leur utilisation en pratique clinique⁷⁰. Ce score a montré une capacité globale de stratification du risque, avec une augmentation progressive du taux d'évènements en fonction des niveaux de score, traduisant une bonne cohérence du modèle dans la hiérarchisation des patients en captant bien le risque et son évolution.

En analysant de plus près les évènements qui n'ont pas été détectés, on observe que la médiane du score prédit pour ces évènements est de 4, ce qui signifie que, malgré le fait que le modèle ne les ait pas détectés comme à haut risque dans sa transformation binaire, il a tout de même attribué un score élevé et qu'une dégradation a bien été captée par le modèle. La spécificité élevée (0,829) indique une bonne capacité à identifier les patients ne présentant pas d'évènement, limitant ainsi le nombre de fausses alertes.

Les courbes de calibration et décisionnelle ont apporté une information complémentaire importante, en mettant en évidence un intérêt clinique potentiel pour l'utilisation de ce modèle. À l'aide de la courbe de calibration, on observe une bonne concordance entre les probabilités prédites et les taux observés, malgré une légère tendance à la surestimation pour les niveaux de risque les plus élevés. La courbe décisionnelle montre que le modèle suggère une plus-value potentielle dans un cadre clinique. L'ensemble de ces éléments permet de suggérer que, de manière globale, le modèle parvient à stratifier le risque chez les patients.

En tant que modèle de référence, il est intéressant à utiliser et son interprétation est aisée. Ce constat est cohérent avec l'étude Chen et al. (2025), qui vise à développer un outil d'aide à la décision clinique de même nature que celui de notre travail. La RL est également utilisée comme modèle de référence en raison de sa simplicité et de sa clarté d'interprétation. Elle sert ainsi de *benchmark* afin de comparer les performances des autres modèles⁷¹.

Réduction de dimensionnalité

Dans un second temps, une approche orientée vers une application plus proche des conditions réelles a été adoptée. Un modèle de RF a été utilisé afin de sélectionner un nombre réduit de variables, dans le but de faciliter l'utilisation du modèle.

L'étude de Speiser et al. (2021) montre que la réduction du nombre de variables utilisées pour construire un modèle de prédiction constitue une stratégie pertinente, car elle rend le modèle plus simple, plus pratique à utiliser et plus facilement intégrable en pratique clinique⁵². Elle permet également de diminuer la charge liée à la collecte des données et, par conséquent, de réduire les coûts associés, sans nécessairement altérer les performances du modèle.

En appliquant la règle du « one standard error », il a été possible de construire un modèle parcimonieux, en passant de 18 à 7 variables. Parmi les variables retenues certaines sont facilement accessibles (EWS, SpO₂, PAD, oxygène à l'admission) tandis que d'autres nécessitent des analyses biologiques. Les variables retenues sont cohérentes avec le contexte clinique de la dégradation, suggérant que le modèle s'appuie sur des éléments pertinents pour ses prédictions.

Le LightGBM, une opportunité intéressante

Plusieurs modèles ont ensuite été testés, avec une sélection basée sur la maximisation de la PR-AUC, à l'aide de la bibliothèque *PyCaret*. Cette approche permettait de tester de nombreux modèles différents sans nécessiter un développement manuel complet de chaque modèle, afin d'avoir un premier aperçu puis d'approfondir uniquement le modèle le plus prometteur.

Le modèle retenu, basé sur LightGBM, a présenté des performances élevées avec une PR-AUC de 0,894 et une ROC-AUC de 0,986 sur l'échantillon test. Cela permet de suggérer que le modèle a une bonne capacité à discriminer les patients avec évènement de ceux sans évènement ; sa PR-AUC appuie également le fait que même si l'évènement est rare, le modèle parvient bien à placer les séjours avec évènement parmi les prédictions les plus à risque.

Ensuite, la progression du taux d'évènements observés selon les tranches de score suggère qu'en plus de bonnes performances, le modèle semble stratifier efficacement le risque dans cet échantillon test. Les trois premiers niveaux ne contiennent aucun évènement, ce qui est particulièrement intéressant car, dans un cadre de priorisation, il est essentiel de réussir à identifier les patients qui nécessitent moins de ressources afin de pouvoir en privilégier d'autres pour la surveillance.

Le score 5 contient, en revanche, la plus forte concentration d'évènements, ce qui indique que le modèle parvient à isoler les séjours présentant un risque supérieur au reste de la cohorte. Cela confirme encore une fois l'intérêt du modèle en tant qu'outil de stratification, car il permet d'identifier directement les patients à cibler et à surveiller de manière plus rapprochée.

Les performances élevées de ce modèle s'expliquent notamment par sa capacité à modéliser des relations complexes et non linéaires entre les variables.

Ces résultats montrent que le modèle pourrait être intéressant en tant qu'outil de surveillance, car il présente un faible nombre de FN. La courbe de calibration confirme la bonne adéquation entre les probabilités prédites et les observations, renforçant l'intérêt du modèle pour la stratification du risque.

La courbe décisionnelle montre également un bénéfice clinique, le modèle étant supérieur aux stratégies consistant à surveiller tous les patients ou aucun, sur une plage de seuils cliniquement pertinents. Notre courbe suggère que, même en modifiant les différents seuils du modèle, il semble conserver un bénéfice décisionnel et supérieur aux deux stratégies extrêmes (surveiller tout le monde

ou ne surveiller personne), ce qui soutient le fait que le modèle présente un intérêt clinique à être utilisé.

La transformation des prédictions en score a permis de rendre l'outil plus facilement utilisable en pratique clinique. L'analyse conjointe des courbes de calibration et décisionnelle suggère que ce score présente un intérêt potentiel pour la stratification et la priorisation des patients.

Ces résultats sont cohérents avec ceux de Logothetis et al. (2023), qui montrent que les modèles LightGBM appliqués à des données tabulaires structurées présentent des performances supérieures aux autres modèles de classification, notamment en raison de leur capacité à capturer des interactions complexes⁴⁵.

Limitations des modèles

Certaines limitations doivent néanmoins être prises en compte. D'une part, il existe des limitations liées aux modèles eux-mêmes. Pour la RL, la nature même du modèle limite sa capacité à capturer les interactions complexes entre variables et à identifier certains profils atypiques. Une sensibilité de 0,682 indique qu'une proportion non négligeable d'évènements n'est pas détectée, ce qui constitue une limitation importante pour un outil d'alerte. Les performances de notre RL restent donc limitées à la fois par la nature intrinsèque du modèle et par la faible prévalence de l'évènement étudié.

La littérature souligne d'ailleurs que, bien que les modèles basés sur le ML soient globalement plus performants que les scores classiques, les modèles plus complexes peuvent offrir des performances supérieures à celles de la RL.

À l'échelle globale, la RL conserve un intérêt en tant que modèle de référence, car elle permet une stratification générale du risque. Toutefois, sa performance reste insuffisante pour une prédiction fiable à l'échelle d'un patient individuel, ce qui constitue une limite importante.

Pour le modèle LightGBM, la VPP reste toutefois modérée (0,302), traduisant la présence de FP. Étant donné que l'évènement est rare, une VPP faible est attendue ; cela ne signifie pas que le modèle n'est pas intéressant, mais plutôt qu'il doit être envisagé comme un outil d'accompagnement à la prise de décision des soignants. Dans un contexte de surveillance clinique, il peut être préférable de privilégier la détection des patients à risque, même au prix d'un certain nombre de fausses alertes.

Cependant, les performances élevées du LightGBM reposent sur sa capacité à modéliser des relations complexes entre les variables. Cette complexité s'accompagne d'un enjeu d'interprétabilité, nécessitant l'utilisation d'outils explicatifs afin de limiter l'effet de « boîte noire ».

D'autre part, il existe également des limitations liées aux choix méthodologiques du projet. L'étude repose sur des données rétrospectives, nécessitant une validation externe pour confirmer la

généralisation du modèle. Bien que l'utilisation de données issues du CHU de Liège permette une bonne adaptation à la population locale, elle peut limiter l'extrapolation à d'autres contextes hospitaliers.

Par ailleurs, la taille importante de la cohorte, associée à une forte variabilité des données, peut entraîner une sous-représentation de certains profils atypiques.

Une autre limitation concerne la stratégie retenue pour représenter un séjour. Bien que des informations aient été disponibles sur plusieurs jours, celles-ci ont été résumées en une seule ligne par séjour, ce qui a entraîné une perte d'informations. Pour les séjours avec évènement, le choix d'utiliser la donnée la plus proche de l'évènement cible permettait une meilleure stratification, mais a pu influencer les performances du modèle.

L'utilisation des scores est largement répandue en pratique clinique, mais leur construction repose souvent sur des approches statistiques, ce qui peut limiter leur pertinence clinique.

Implications pratiques et perspectives pour le futur

En conclusion, les résultats obtenus montrent un intérêt réel pour l'utilisation de modèles de ML dans la stratification du risque et la priorisation des patients. Toutefois, leur intégration en pratique clinique nécessite des étapes supplémentaires d'optimisation et de validation prospective. Le choix du modèle ne doit pas uniquement reposer sur ses performances, mais également sur son interprétabilité et sa facilité d'utilisation, éléments essentiels pour une adoption en milieu hospitalier.

Pour une utilisation clinique réelle, trois étapes clés seraient nécessaires : la validation prospective, une phase pilote et une étude d'impact. La validation prospective consisterait à tester le modèle de manière silencieuse, c'est-à-dire à calculer le score en temps réel à partir des données disponibles à l'hôpital, sans l'afficher aux soignants ni modifier la prise en charge. Cette étape permettrait de vérifier la qualité des données en conditions réelles, d'évaluer la stabilité des performances du modèle et de s'assurer de la cohérence du score.

Ensuite, une phase pilote pourrait être mise en place en collaboration avec une équipe de cliniciens afin d'intégrer le score dans le flux de soins. Cette étape permettrait d'ajuster légèrement le seuil si nécessaire, de définir une conduite à tenir selon le niveau de score et d'intégrer l'outil directement dans le DPI. Cette phase serait testée dans quelques services avant une éventuelle généralisation.

Enfin, une étude d'impact prospective permettrait d'évaluer le score dans des conditions réelles. Il serait alors possible de comparer des périodes avec et sans utilisation du score, en s'appuyant sur des indicateurs de qualité des soins afin d'estimer le bénéfice réel de son intégration clinique^{56,57}.

Les scores proposés dans ce projet permettent de stratifier les patients selon le niveau de risque estimé de survenue d'un évènement. Cependant, une perspective d'évolution intéressante serait le développement d'un score dynamique, capable de mieux capturer l'évolution de la dégradation clinique au cours du temps.

D'un point de vue pratique, ce type de score nécessiterait d'intégrer des valeurs mesurées à plusieurs moments distincts, par exemple à l'admission aux urgences, au cours des jours suivants et juste avant l'admission en USI, afin de conserver la dimension temporelle de l'information sans la résumer en une seule valeur^{40,72,73}.

Une telle approche supposerait donc le recours à des modèles adaptés aux données longitudinales, capables de prendre en compte plusieurs mesures d'un même patient sur plusieurs jours. Elle permettrait potentiellement une meilleure détection de la dégradation progressive, mais nécessiterait une structuration plus complexe des données ainsi qu'une validation prospective spécifique.

Bibliographie

6. BIBLIOGRAPHIE

1. Li, L., He, H., Xiang, L. & Wang, Y. A warning model for predicting patient admissions to the intensive care unit (ICU) following surgery. *Perioper. Med.* **14**, 60 (2025).
2. England, N. H. S. NHS England » Managing acute physical deterioration through the ‘prevention, identification, escalation, response’ (PIER) approach. <https://www.england.nhs.uk/patient-safety/managing-acute-physical-deterioration-through-the-prevention-identification-escalation-response-pier-approach/>.
3. Prediction of Patient Deterioration on the General Care Floor: Making the Connection with ICU Admissions & Resource Utilization – A Special Series on Perioperative Patient Safety. <https://www.socca.org/vol31-iss1-prediction-of-patient-deterioration-on-the-general-care-floor>.
4. Anagi, S. *et al.* Delayed Admission to the Intensive Care Unit Is Associated With Increased Mortality Risk in Critically Ill Patients: A Prospective Cohort Study. *J. Adv. Nurs.* <https://doi.org/10.1111/jan.70493> (2026) doi:10.1111/jan.70493.
5. Bruyneel, A., Lucchini, A. & Dauvergne, J. E. The relationship between nursing workload and quality of care in intensive care units. *Crit. Care Sci.* **37**, e20250381 (2025).
6. Safiri, S. & Ayubi, E. Risk factors for unplanned transfer to the intensive care unit after emergency department admission: Methodological issues. *Am. J. Emerg. Med.* **35**, 1573 (2017).
7. Aiken, L. H. *et al.* Physician and nurse well-being, patient safety and recommendations for interventions: cross-sectional survey in hospitals in six European countries. *BMJ Open* **14**, e079931 (2024).
8. Bruyneel, A. *et al.* Association of burnout and intention-to-leave the profession with work environment: A nationwide cross-sectional study among Belgian intensive care nurses after two years of pandemic. *Int. J. Nurs. Stud.* **137**, 104385 (2023).
9. Hatchimonji, J. S. *et al.* FAILURE TO RESCUE IN SURGICAL PATIENTS: A REVIEW FOR ACUTE CARE SURGEONS. *J. Trauma Acute Care Surg.* **87**, 699–706 (2019).
10. Pimentel, M. A. F. *et al.* Detecting Deteriorating Patients in the Hospital: Development and Validation of a Novel Scoring System. *Am. J. Respir. Crit. Care Med.* **204**, 44–52.
11. Densité infirmière par secteur d'activité. *Iweeps* <https://www.iweeps.be/indicateur-statistique/densite-infirmiere-par-secteur-dactivite/>.
12. La force de travail des infirmiers en 2046 : projections alternatives à partir de l'influx réel jusqu'en 2022.
13. Khan, Y., Bruyneel, A. & Smith, P. Determinants of the risk of burnout among nurses during the first wave of the COVID-19 pandemic in Belgium: A cross-sectional study. *J. Nurs. Manag.* **30**, 1125–1135 (2022).
14. Gontier, R. *et al.* Predicting the unpredictable: unveiling hidden patterns of emergency department unexpected deaths - a retrospective study. *Acta Clin. Belg.* **81**, 18–27 (2026).
15. Tack, J. *et al.* Intensive care patients eligible for intermediate care unit-level care: a single-centre prospective cohort study in Belgium. *Crit. Care Sci.* **38**, (2025).
16. Kohnen, D. *et al.* Engaging leadership and nurse well-being: the role of the work environment and work motivation—a cross-sectional study. *Hum. Resour. Health* **22**, 8 (2024).
17. Jesus, A. P. S. D., Okuno, M. F. P., Campanharo, C. R. V., Lopes, M. C. B. T. & Batista, R. E. A. Manchester Triage System: assessment in an emergency hospital service. *Rev. Bras. Enferm.* **74**, e20201361 (2021).
18. Magali, P. *et al.* Retrospective study on the health and economic burden of hospitalized patients due to pneumonia and invasive pneumococcal infections in Belgium settings. *Vaccine* **42**, 3018–3023 (2024).
19. Bruyneel, A., den Bulcke, J. V., Leclercq, P. & Pirson, M. Frequency, financial impact, and factors associated with cost outliers in intensive care units: a cohort study in Belgium. *Crit. Care Sci.* **37**, e20250207 (2025).
20. Morreel, S. *et al.* Triage and referring in adjacent general and emergency departments (the TRIAGE trial): A cluster randomised controlled trial. *PLoS ONE* **16**, e0258561 (2021).
21. Jobe, J., Ghuysen, A. & D'orio, V. ELISA : ECHELLE LIÉGEOISE DE L'INDICE DE SÉVÉRITÉ À L'ADMISSION. *Rev Med Liège* (2012).
22. Kurs: Emergency Triage Hub | MTS. <https://www.triagenet.net/classroom/course/view.php?id=1173&lang=pl#module-25056>.
23. Profil USIs - MICA Program. <https://www.micaprogram.be/fr/icu-profile/>.
24. Byrne, A. *et al.* When Nurses Worry: A Concept Analysis of Intuition in Clinical Deterioration. *J. Adv. Nurs.* **81**, 4566–4583 (2025).
25. Revue Médicale de Liège - Le triage infirmier avancé aux Urgences. <https://rmlg.uliege.be/article/3004?lang=en>.

26. Cross, R., Considine, J. & Currey, J. Nursing handover of vital signs at the transition of care from the emergency department to the inpatient ward: An integrative review. *J. Clin. Nurs.* **28**, 1010–1021 (2019).
27. Recommendations | Acutely ill adults in hospital: recognising and responding to deterioration | Guidance | NICE. <https://www.nice.org.uk/guidance/cg50/chapter/Recommendations#identifying-patients-whose-clinical-condition-is-deteriorating-or-is-at-risk-of-deterioration> (2007).
28. Downey, C. L., Tahir, W., Randell, R., Brown, J. M. & Jayne, D. G. Strengths and limitations of early warning scores: A systematic review and narrative synthesis. *Int. J. Nurs. Stud.* **76**, 106–119 (2017).
29. Evaluation of machine learning-based models for prediction of clinical deterioration: A systematic literature review - Recherche Google. <https://pubmed.ncbi.nlm.nih.gov/37156168/>.
30. Williams, B. The National Early Warning Score: from concept to NHS implementation. *Clin. Med.* **22**, 499–505 (2022).
31. Watson, M. et al. Performance of machine learning versus the national early warning score for predicting patient deterioration risk: a single-site study of emergency admissions. *BMJ Health Care Inform.* **31**, e101088 (2024).
32. Musy, S. N. et al. The association between nurse staffing and inpatient mortality: A shift-level retrospective longitudinal study. *Int. J. Nurs. Stud.* **120**, 103950 (2021).
33. Wang, C., Xu, R., Zeng, Y., Zhao, Y. & Hu, X. A comparison of qSOFA, SIRS and NEWS in predicting the accuracy of mortality in patients with suspected sepsis: A meta-analysis. *PLoS ONE* **17**, e0266755 (2022).
34. Furqan, A., Naseer, M. & Tabassum, R. Comparison of Sensitivity, Specificity and Accuracy of APACHE II, SAPS II and SOFA Scoring Systems as Predictors. *Pak. J. Med. Health Sci.* **15**, 3364–3366 (2021).
35. Bruyneel, A. et al. Cost comparisons and factors related to cost per stay in intensive care units in Belgium. *BMC Health Serv. Res.* **23**, 986 (2023).
36. Berg, A. M. N., Werner, A., Knutsen, I. R. & Johannessen, A.-K. Hospital nurses and physicians' experiences practicing patient safety work to recognize deteriorating patients: a qualitative study. *BMC Health Serv. Res.* **24**, 1429 (2024).
37. Zhang, K. et al. The Prognostic Accuracy of National Early Warning Score 2 on Predicting Clinical Deterioration for Patients With COVID-19: A Systematic Review and Meta-Analysis. *Front. Med.* **8**, 699880 (2021).
38. Díaz-González, S., Martín-Conty, J. L., López-Izquierdo, R., Martín-Rodríguez, F. & Sanz-García, A. Impact of point-of-care biomarkers on the improvement of the predictive capacity of early warning scores in prehospital care: a systematic review and meta-analysis.
39. Durr, D. et al. National Early Warning Score (NEWS) Outperforms Quick Sepsis-Related Organ Failure (qSOFA) Score for Early Detection of Sepsis in the Emergency Department. *Antibiotics* **11**, 1518 (2022).
40. Muralitharan, S. et al. Machine Learning-Based Early Warning Systems for Clinical Deterioration: Systematic Scoping Review. *J. Med. Internet Res.* **23**, e25187 (2021).
41. Carr, E. et al. Evaluation and improvement of the National Early Warning Score (NEWS2) for COVID-19: a multi-hospital study. *BMC Med.* **19**, 23 (2021).
42. Edelson, D. P. et al. Early Warning Scores With and Without Artificial Intelligence. *JAMA Netw. Open* **7**, e2438986 (2024).
43. Zachariasse, J. M. et al. Improving the prioritization of children at the emergency department: Updating the Manchester Triage System using vital signs. *PLOS ONE* **16**, e0246324 (2021).
44. Covino, M. et al. Predicting ICU admission and death in the Emergency Department: A comparison of six early warning scores. *Resuscitation* **190**, 109876 (2023).
45. Boulitsakis Logothetis, S., Green, D., Holland, M. & Al Moubayed, N. Predicting acute clinical deterioration with interpretable machine learning to support emergency care decision making. *Sci. Rep.* **13**, 13563 (2023).
46. Shamout, F., Zhu, T. & Clifton, D. A. Machine Learning for Clinical Outcome Prediction. *IEEE Rev. Biomed. Eng.* **14**, 116–126 (2021).
47. Perlis, R. & Abbasi, J. Researchers Compared Hospital Early Warning Scores for Clinical Deterioration- Here's What They Learned. *JAMA* **333**, 363–365 (2025).
48. The early warning paradox - PubMed. <https://pubmed.ncbi.nlm.nih.gov/39900787/>.
49. Caraballo, P. J. et al. Trustworthiness of a machine learning early warning model in medical and surgical inpatients. *JAMIA Open* **8**, ooae156 (2025).
50. Kar, A. A Very Short Introduction of Random Forests. *Consuledge* <https://consuledge.com.au/blog/mastering-random-forests-a-beginners-guide-to-smarter-machine-learning/> (2024).
51. An improved random forest algorithm for tracing the origin of metastatic renal cancer tissues - PMC. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12305792/>.

52. Speiser, J. L. A random forest method with feature selection for developing medical prediction models with clustered and longitudinal data. *J. Biomed. Inform.* **117**, 103763 (2021).
53. Sung, S., Kim, Y., Kim, S. H. & Jung, H. Identification of Predictors for Clinical Deterioration in Patients With COVID-19 via Electronic Nursing Records: Retrospective Observational Study. *J. Med. Internet Res.* **26**, e53343 (2024).
54. Kernbach, J. M. & Staartjes, V. E. Machine learning-based clinical prediction modeling -- A practical guide for clinicians. *arXiv.org* <https://arxiv.org/abs/2006.15069v1> (2020).
55. Efthimiou, O. *et al.* Developing clinical prediction models: a step-by-step guide. *BMJ* **386**, e078276 (2024).
56. Riley, R. D. *et al.* Evaluation of clinical prediction models (part 2): how to undertake an external validation study. *BMJ* **384**, e074820 (2024).
57. Collins, G. S. *et al.* Evaluation of clinical prediction models (part 1): from development to external validation. *The BMJ* **384**, e074819 (2024).
58. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
59. Friedman, J. H., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **33**, 1–22 (2010).
60. pycaret/pycaret. PyCaret (2026).
61. Heymans, M. W. & Twisk, J. W. R. Handling missing data in clinical research. *J. Clin. Epidemiol.* **151**, 185–188 (2022).
62. Garcia, C. de L. *et al.* Influence of Burnout on Patient Safety: Systematic Review and Meta-Analysis. *Medicina (Mex.)* **55**, 553 (2019).
63. Ageing and health. <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health>.
64. Winters, B. D., Rosen, M., Sharma, R., Zhang, A. & Bass, E. B. Patient Monitoring Systems To Prevent Failure To Rescue: Rapid Review. in *Making Healthcare Safer IV: A Continuous Updating of Patient Safety Harms and Practices* (Agency for Healthcare Research and Quality (US), Rockville (MD), 2023).
65. Galen, L. S. van *et al.* Delayed Recognition of Deterioration of Patients in General Wards Is Mostly Caused by Human Related Monitoring Failures: A Root Cause Analysis of Unplanned ICU Admissions. *PLOS ONE* **11**, e0161393 (2016).
66. Belk, K. W., Beals, J. & McInnis, S. J. Mortality and Length of Stay Implications of Deterioration-Associated Transfer to the Intensive Care Unit over Different Time Frames. *Health Serv. Insights* **18**, 11786329241312877 (2025).
67. Khanna, A. K., Moucharite, M. A., Benefield, P. J. & Kaw, R. Patient Characteristics and Clinical and Economic Outcomes Associated with Unplanned Medical and Surgical Intensive Care Unit Admissions: A Retrospective Analysis. *Clin. Outcomes Res. CEOR* **15**, 703–719 (2023).
68. Muralitharan, S. *et al.* Machine Learning-Based Early Warning Systems for Clinical Deterioration: Systematic Scoping Review. *J. Med. Internet Res.* **23**, e25187 (2021).
69. Al-Nafjan, A., Aljuhani, A., Alshebel, A., Alharbi, A. & Alshehri, A. Artificial Intelligence in Predictive Healthcare: A Systematic Review. *J. Clin. Med.* **14**, 6752 (2025).
70. Zhao, Z. *et al.* Prediction model and risk scores of ICU admission and mortality in COVID-19. *PLoS ONE* **15**, e0236618 (2020).
71. Frontiers | A machine learning model based on emergency clinical data predicting 3-day in-hospital mortality for stroke and trauma patients. <https://www.frontiersin.org/journals/neurology/articles/10.3389/fneur.2025.1512297/full>.
72. Zhu, Y. *et al.* Dynamic individual vital sign trajectory early warning score (DyniEWS) versus snapshot national early warning score (NEWS) for predicting postoperative deterioration. *Resuscitation* **157**, 176–184 (2020).
73. Gonem, S. *et al.* Dynamic early warning scores for predicting clinical deterioration in patients with respiratory disease. *Respir. Res.* **23**, 203 (2022).

ANNEXES

Annexe 1 : Codebook constituée pour la demande d'extraction des données.

ID_Patient	Description
Température_J0	Température à l'admission aux urgences
Température_J1	Température à J1 hospitalisation
Température_J2	Température à J2 hospitalisation
Température_J3	Température à J3 hospitalisation
Température_USI0	Température le jour de l'admission en réa
Température_USI-1	Température à J-1 avant l'admission en réa
Température_USI-2	Température à J-2 avant l'admission en réa
Température_USI-3	Température à J-3 avant l'admission en réa
Fréquence cardiaque_J0	Fréquence cardiaque à l'admission aux urgences
Fréquence cardiaque_J1	Fréquence cardiaque à J1 hospitalisation
Fréquence cardiaque_J2	Fréquence cardiaque J2 hospitalisation
Fréquence cardiaque_J3	Fréquence cardiaque à J3 hospitalisation
Fréquence cardiaque_USI0	Fréquence cardiaque le jour de l'admission en réa
Fréquence cardiaque_USI-1	Fréquence cardiaque à J-1 avant l'admission en réa
Fréquence cardiaque_USI-2	Fréquence cardiaque à J-2 avant l'admission en réa
Fréquence cardiaque_USI-3	Fréquence cardiaque à J-3 avant l'admission en réa
SpO ₂ _J0	Saturation en oxygène à l'admission aux urgences
SpO ₂ _J1	Saturation en oxygène à J1 hospitalisation
SpO ₂ _J2	Saturation en oxygène J2 hospitalisation
SpO ₂ _J3	Saturation en oxygène à J3 hospitalisation
SpO ₂ _USI0	Saturation en oxygène le jour de l'admission en réa
SpO ₂ _USI-1	Saturation en oxygène à J-1 avant l'admission en réa
SpO ₂ _USI-2	Saturation en oxygène à J-2 avant l'admission en réa
SpO ₂ _USI-3	Saturation en oxygène à J-3 avant l'admission en réa
Pression artérielle systolique_J0	Pression artérielle systolique à l'admission aux urgences
Pression artérielle systolique_J1	Pression artérielle systolique à J1 hospitalisation
Pression artérielle systolique_J2	Pression artérielle systolique J2 hospitalisation
Pression artérielle systolique_J3	Pression artérielle systolique à J3 hospitalisation
Pression artérielle systolique_USI0	Pression artérielle systolique le jour de l'admission en réa
Pression artérielle systolique_USI-1	Pression artérielle systolique à J-1 avant l'admission en réa
Pression artérielle systolique_USI-2	Pression artérielle systolique à J-2 avant l'admission en réa
Pression artérielle systolique_USI-3	Pression artérielle systolique à J-3 avant l'admission en réa
Pression artérielle diastolique_J0	Pression artérielle diastolique à l'admission aux urgences

Pression artérielle diastolique_J1	Pression artérielle diastolique à J1 hospitalisation
Pression artérielle diastolique_J2	Pression artérielle diastolique J2 hospitalisation
Pression artérielle diastolique_J3	Pression artérielle diastolique à J3 hospitalisation
Pression artérielle diastolique_USI0	Pression artérielle diastolique le jour de l'admission en réa
Pression artérielle diastolique_USI-1	Pression artérielle diastolique à J-1 avant l'admission en réa
Pression artérielle diastolique_USI-2	Pression artérielle diastolique à J-2 avant l'admission en réa
Pression artérielle diastolique_USI-3	Pression artérielle diastolique à J-3 avant l'admission en réa
AVPU/GCS_J0	AVPU/GCS à l'admission aux urgences
AVPU/GCS_J1	AVPU/GCS à J1 hospitalisation
AVPU/GCS_J2	AVPU/GCS J2 hospitalisation
AVPU/GCS_J3	AVPU/GCS à J3 hospitalisation
AVPU/GCS_USI0	AVPU/GCS le jour de l'admission en réa
AVPU/GCS_USI-1	AVPU/GCS à J-1 avant l'admission en réa
AVPU/GCS_USI-2	AVPU/GCS à J-2 avant l'admission en réa
AVPU/GCS_USI-3	AVPU/GCS à J-3 avant l'admission en réa
Besoin O2_J0	Besoin O2 à l'admission aux urgences
Besoin O2_J1	Besoin O2 à J1 hospitalisation
Besoin O2_J2	Besoin O2 J2 hospitalisation
Besoin O2_J3	Besoin O2 à J3 hospitalisation
Besoin O2_USI0	Besoin O2 le jour de l'admission en réa
Besoin O2_USI-1	Besoin O2 à J-1 avant l'admission en réa
Besoin O2_USI-2	Besoin O2 à J-2 avant l'admission en réa
Besoin O2_USI-3	Besoin O2 à J-3 avant l'admission en réa
CRP_J0	CRP à l'admission aux urgences
CRP_J1	CRP à J1 hospitalisation
CRP_J2	CRP J2 hospitalisation
CRP_J3	CRP à J3 hospitalisation
CRP_USI0	CRP le jour de l'admission en réa
CRP_USI-1	CRP à J-1 avant l'admission en réa
CRP_USI-2	CRP à J-2 avant l'admission en réa
CRP_USI-3	CRP à J-3 avant l'admission en réa
Urée_J0	Urée à l'admission aux urgences
Urée_J1	Urée à J1 hospitalisation
Urée_J2	Urée J2 hospitalisation
Urée_J3	Urée à J3 hospitalisation
Urée_USI0	Urée le jour de l'admission en réa
Urée_USI-1	Urée à J-1 avant l'admission en réa

Urée_USI-2	Urée à J-2 avant l'admission en réa
Urée_USI-3	Urée à J-3 avant l'admission en réa
Créatinine_J0	Créatinine à l'admission aux urgences
Créatinine_J1	Créatinine à J1 hospitalisation
Créatinine_J2	Créatinine J2 hospitalisation
Créatinine_J3	Créatinine à J3 hospitalisation
Créatinine_USI0	Créatinine le jour de l'admission en réa
Créatinine_USI-1	Créatinine à J-1 avant l'admission en réa
Créatinine_USI-2	Créatinine à J-2 avant l'admission en réa
Créatinine_USI-3	Créatinine à J-3 avant l'admission en réa
Potassium_J0	Potassium à l'admission aux urgences
Potassium_J1	Potassium à J1 hospitalisation
Potassium_J2	Potassium J2 hospitalisation
Potassium_J3	Potassium à J3 hospitalisation
Potassium_USI0	Potassium le jour de l'admission en réa
Potassium_USI-1	Potassium à J-1 avant l'admission en réa
Potassium_USI-2	Potassium à J-2 avant l'admission en réa
Potassium_USI-3	Potassium à J-3 avant l'admission en réa
NEWS_J0	Score NEWS dans dossier infirmier à l'admission aux urgences
NEWS_J1	Score NEWS dans dossier infirmier à J1 hospitalisation
NEWS_J2	Score NEWS dans dossier infirmier J2 hospitalisation
NEWS_J3	Score NEWS dans dossier infirmier à J3 hospitalisation
NEWS_USI0	Score NEWS dans dossier infirmier le jour de l'admission en réa
NEWS_USI-1	Score NEWS dans dossier infirmier à J-1 avant l'admission en réa
NEWS_USI_2	Score NEWS dans dossier infirmier à J-2 avant l'admission en réa
NEWS_USI-3	Score NEWS dans dossier infirmier à J-3 avant l'admission en réa
Service_Admission_posturg	Service d'hospitalisation post urgences
Type_sortieurg_monito	Type de sortie urgences
Site_hospi	Site d'hospitalisation
age	Age du patient
Sexe	Sexe du patient
Atcds_USI	Hospitalisation en USI lors de l'hospitalisation précédente
Deces	Décès durant l'hospitalisation
Admission_USI	Admission en USI durant l'hospitalisation
Jour_Deces	Jour du décès depuis l'admission en salle
Jour_USI	Jour d'admission en USI depuis l'admission en salle

