



HAL
open science

Prédiction du Parcours Patient aux Urgences : 3P-U

Émilien Arnaud

► **To cite this version:**

Émilien Arnaud. Prédiction du Parcours Patient aux Urgences : 3P-U. Médecine humaine et pathologie. Université de Picardie Jules Verne, 2024. Français. <NNT : 2024AMIE0005>. <tel-05268716>

HAL Id: tel-05268716

<https://theses.hal.science/tel-05268716v1>

Submitted on 19 Sep 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization



Thèse de Doctorat

Mention Santé

Spécialité Médecine d'Urgence

présentée à l'École Doctorale en Sciences, Technologies, Santé (ED 585)

de L'Université de Picardie Jules Verne

par

Émilien ARNAUD

pour obtenir le grade de Docteur de l'Université de Picardie Jules Verne

***Prédiction du Parcours Patient aux Urgences
3P-U***

Soutenue le 11 avril 2024, après avis des rapporteurs, devant le jury d'examen :

M. P. Preux, Professeur, Université de Lille
M. E. Casalino, Professeur, Université de Paris

Rapporteur
Rapporteur

Mme C. Lucet, Professeure, Université Picardie Jules Verne
Mme C. Ammirati, Professeure, Université Picardie Jules Verne
Mme S. Charpentier, Professeure, Université Paul Sabatier
M. T. Desmettre, Professeur, Université de Genève

Examinatrice
Examinatrice
Examinatrice
Examinateur

M. G. Dequen, Professeur, Université Picardie Jules Verne
M. D.A. Ghazali, Professeur, Université Picardie Jules Verne

Directeur de thèse
Co-directeur de thèse

Remerciements

Avant de plonger dans le cœur de cette aventure scientifique, je tiens à exprimer ma profonde gratitude envers toutes les personnes qui ont contribué à la réalisation de ce travail. Leur expertise, leur soutien et leurs encouragements ont été essentiels à chaque étape de ce parcours. Je souhaite tout d'abord remercier chaleureusement les membres du jury qui m'ont fait l'honneur de juger ce travail.

Professeur Philippe PREUX, je vous remercie pour votre lecture attentive du manuscrit, votre analyse scientifique informatique, vos questions pertinentes en sollicitant un éclaircissement nécessaire. Me laisser l'opportunité d'apporter des précisions a permis d'enrichir ce travail.

Professeur Enrique CASALINO, je vous remercie pour votre lecture attentive du manuscrit et votre analyse médicale de ce travail. Votre expertise a été un guide précieux dans la finalisation de cette thèse.

Professeure Corinne LUCET-VASSEUR, je vous remercie pour pour l'examen attentif de mon travail et l'apport de perspectives enrichissantes. Je vous remercie d'avoir participé à la qualité de mon accueil au laboratoire.

Professeure Christine AMMIRATI, je vous remercie pour l'examen attentif de mon manuscrit et de votre analyse médicale de ce travail. Votre expertise dans le domaine a été une source d'inspiration profonde.

Je tiens également à exprimer ma sincère gratitude envers le Professeur Gilles DEQUEN et Professeur Daniel Aiham GHAZALI. En tant que directeur et co-directeur de cette thèse, vous avez non seulement partagé votre vaste connaissance, mais aussi votre passion pour la recherche. Votre soutien indéfectible, vos encouragements constants et votre capacité à me guider à travers les défis scientifiques ont été les piliers de ce projet. Votre confiance en mes capacités m'a permis de me surpasser et de mener à bien ce travail.

La réalisation de cette thèse a été une période de croissance intellectuelle et personnelle, marquée par des rencontres enrichissantes et des échanges stimulants. Chaque membre de ce jury a contribué, à sa manière, à faire de ce parcours une expérience inoubliable et formatrice.

Avec toute ma gratitude,

Émilien ARNAUD

Remerciements

En premier lieu, je voudrais remercier Bénédicte, mon épouse. Sans son soutien inconditionnel, son amour profond et sa patience incroyable, ce travail n'aurait jamais pu aboutir. Son expertise de médecin biologiste et sa vision de la médecine et des patients enrichissent à mon quotidien de praticien et de chercheur. Mes enfants, Philippine et Louis, sont de loin ce que j'ai réussi de plus beau. Je les remercie sincèrement car leur joie, leur insouciance et leur amour sont des moteurs pour moi au quotidien. Je suis particulièrement reconnaissant envers mes parents. Ils ont fait de moi ce que je suis et c'est grâce à eux que j'en suis là aujourd'hui. Je n'oublie pas mes frères et sœurs, beaux-frères et belles-sœurs, Pauline et Aymeric, Joséphine et Antoine, Victor et Jeanne, et Louise et Andrieu qui m'ont soutenu toutes ces années.

Je remercie chaleureusement les Professeurs Gilles DEQUEN, Maxime GIGNON et Daniel Aïham GHAZALI. Ils m'ont accompagné jusqu'à l'aboutissement de ce travail. Ils m'ont appris ce qu'était la recherche, tant en sciences informatiques qu'en sciences médicales. Merci à l'ensemble des membres du laboratoire Modélisation, Information et Systèmes : Juliette et Isabelle, le point de convergence de toutes les solutions, Christophe pour son appui technique, à tous les doctorants, les post-doc (et tout particulièrement au Docteur Mahmoud ELBATTAH pour nos travaux communs), et plus globalement à tous les membres permanents pour leur accueil.

Je remercie tout particulièrement le Professeure Christine AMMIRATI. La confiance qu'elle m'accorde en m'associant à ses projets, son accompagnement dans mon développement médical et son soutien dans mes projets ont été structurants pour ma personnalité médico-chercheur. Les quelques situations exceptionnelles, opérationnelles ou sanitaires, passées à vos côtés, sont gravées à jamais dans ma mémoire.

Je remercie également le Docteur Christophe BOYER qui m'accorde sa confiance tous les jours au SAMU, dans l'opérationnel comme dans la construction de la régulation de demain. Merci pour ton accompagnement dans les projets novateurs en pré-hospitalier permettant de concrétiser des idées en des dispositifs réels et utilisables. Merci pour ta personnalité et tes compétences humaines et managériales, justes, droites et constantes qui nous ont permis de passer cette crise COVID-19, en première ligne et très précocement aussi bien qu'il le fût possible.

Je remercie le Docteur Paulo HENRIQUES pour m'avoir associé aux travaux numériques des urgences. Nos discussions en garde, et hors garde, la réflexion sur la pertinence des indicateurs statistiques de l'activité des urgences, la co-conception d'outils d'amélioration du quotidien, l'imagination du DPI urgences de demain, et les réflexions sur l'éthique et la réglementation des données (cher DPO) ont contribué à mon développement de médecin-chercheur en IA.

Un immense merci aux équipes soignantes et administratives des Urgences Adultes et du SAMU 80! Travailler avec vous et à vos côtés est un moteur et une source d'inspiration. Merci!

Je voudrais aussi remercier les équipes de la Direction des Services Numériques du CHU Amiens-Picardie et plus particulièrement Monsieur Sébastien FLOREK et l'équipe InfoCentre pour leur appui sur ce travail.

Je finirais par remercier les amis qui ont toujours été présents depuis le début des études et dans leur continuité : Chub et Marine, Philippe et Laure, Pierre et Gaëlle, Marion et Franek. Je n'oublie pas les amis de BG : Julien et Audrey, Aude et Jean-Charles, Joëlle et Guillaume pour leur amitié. Merci.

Table des matières

Table des matières	3
Acronymes	6
Table des figures	7
Liste des tableaux	10
Note de synthèse	13
Généralités	20
Introduction	20
1 Parcours de soin dans les Structures d’Urgence	21
1.1 Le parcours de soin à l’hôpital	22
1.2 Le parcours de soin intra-hospitalier dans les SU	22
1.3 Les indicateurs du parcours de soin	24
1.3.1 Nombre de passages	24
1.3.2 Durée de passage	26
1.3.3 Évaluation du besoin en lits	26
1.4 Les urgences du CHU Amiens-Picardie	26
1.5 Problématiques rencontrées	28
1.5.1 Surcharge aux urgences	28
1.5.2 Synthèse manuelle	29
1.6 L’aide à la décision par l’IA, une solution potentielle	29
2 Intelligence Artificielle (IA)	31
2.1 Jeu de données	33
2.1.1 Notations	34
2.1.2 Jeu d’exemple	35
2.1.3 Donnée numérique	35
2.1.4 Donnée catégorielle	37
2.1.5 Données non structurées (<i>e.g.</i> verbatim)	38
2.1.6 Qualité des données	38
2.2 Apprentissage machine	42
2.2.1 Mesure de performance	42
2.2.2 Apprentissage supervisé	45
2.2.3 Apprentissage non supervisé	50
2.2.4 Fonctions de coût	55
2.3 Évaluation du modèle	58

2.3.1	Performance de l'apprentissage	58
2.3.2	Évaluation par jeu de test	58
2.3.3	Évaluation par validation croisée	58
2.3.4	Validation externe	59
2.4	Explicabilité de l'intelligence artificielle	60
2.4.1	Enjeux	60
2.4.2	Techniques d'XAI <i>post-hoc</i>	60
2.5	Utilisation des données en santé pour entraîner de les modèles	64
2.5.1	Données structurées	64
2.5.2	Données non structurées	64
2.6	Garantie humaine	66
Rationnel des contributions		67
Contributions		69
3 Définition d'un modèle prédictif du parcours de soins		69
3.1	Méthodologie	70
3.1.1	Description des données	70
3.1.2	Approche d'apprentissage	70
3.1.3	Pré-traitement des données structurées	70
3.1.4	Pré-traitement des données non structurées	71
3.2	Résultats	71
3.2.1	Données expérimentales	71
3.2.2	Structure du modèle	73
3.2.3	Performances du modèle	73
4 Gestion des données manquantes en médecine d'urgence		76
4.1	Introduction	77
4.1.1	Données manquantes en médecine d'urgence	77
4.1.2	Techniques de complétion	77
4.2	Matériels et méthodes	81
4.3	Résultats	83
4.3.1	Caractéristique des études	83
4.3.2	Modèles et stratégies d'imputation	83
4.4	Discussion	88
4.4.1	Qualité de la description des stratégies	88
4.4.2	Fréquence d'utilisation des stratégies	88
4.4.3	Évolution au cours du temps	90
4.4.4	Comparaison de la performance des stratégies	90
4.4.5	Améliorer le rapport des données manquantes et des stratégies appliquées	91
4.4.6	Limites de l'étude	92
4.4.7	Force de l'étude	92
5 Comment expliquer la prédiction de 3P-U ?		93
5.1	Matériel et méthodes	94
5.1.1	Modèle 3P-U	94
5.1.2	Techniques d'explicabilité	95
5.1.3	Interprétation médicale	95

5.1.4	Autorisation réglementaire	95
5.2	Résultats	95
5.2.1	Modèle 3P-U	95
5.2.2	Techniques d'explicabilité	97
5.2.3	Interprétation médicale	109
5.3	Discussion	109
5.3.1	Comparaison avec les travaux existants	109
5.3.2	Méthodes d'explicabilité et alternatives	110
5.3.3	Implications	110
5.3.4	Limites de l'étude	111
5.4	Conclusions	111
6	Application concrète du modèle de prédiction du parcours patient aux urgences dans le contexte de la COVID-19	112
6.1	Méthodologie	113
6.1.1	Construction de l'étude	113
6.1.2	Participants	113
6.1.3	Intervention	113
6.1.4	Analyses statistiques	115
6.2	Résultats	116
6.2.1	Caractéristiques des patients de l'étude	116
6.2.2	Performances du modèle	116
6.2.3	Prédiction individuelle	116
6.3	Discussion	121
6.3.1	Seuil décisionnel de 3P-U	121
6.3.2	Comparaison aux études similaires	121
6.3.3	Implications cliniques	121
6.3.4	Forces et limites	122
7	Garantir la place centrale de l'humain dans la conception et l'utilisation des modèles d'IA en médecine d'urgence	123
7.1	La réglementation	124
7.2	Risques et limites	125
7.2.1	Liés au modèle	125
7.2.2	Liés à l'utilisation du modèle	126
7.3	La formation	126
7.4	La garantie humaine	126
7.4.1	Définition	126
7.4.2	Cadre d'exercice de cette garantie humaine	128
7.4.3	Collège de garantie humaine	128
7.4.4	Projet de déploiement	128
7.4.5	Déploiement	128
7.4.6	Utilisation	129
	Conclusion	131
	Conclusions	131
	Perspectives	133

Table des figures

1.1	Parcours de soin à l'hôpital	22
1.2	Parcours de soins complexe avec avis d'autre spécialité.	24
1.3	Principes d'évaluation du parcours de soin par les experts visiteurs tels que recommandé par la haute autorité de santé[1]	25
1.4	Nombre de passages total sur l'ensemble des centres hospitaliers d'après la Direction de la Recherche, des Études, de l'Évaluation et des Statistiques (DREES)[30]	26
1.5	Représentation spatiale et répartition des Infirmier Diplômé d'État (IDE) (👤) et des médecins (👨) et des effectifs du service des urgences adultes du CHU Amiens-Picardie (Amiens, Picardie, France) (CHU-AP). Les aides-soin, les brancardiers, et les cadres de santé ne sont pas représentés dans cette représentation spatiale.	28
2.1	Représentations des différents types d'Intelligence Artificielle (IA), des tâches qu'elle peut effectuer et des algorithmes sur lesquels elle peut se baser.	32
2.2	Pré-traitement d'un texte médical (<i>i.e.</i> , données non structurées). Les valeurs des vecteurs après le plongement lexical ne représentent pas réellement la phrase en question et sont données uniquement à titre de représentation conceptuelle[23].	39
2.3	Courbe "Receiver operating characteristic" (ROC)	44
2.4	Fonction logistic	45
2.5	Réseau de neurones[69]	46
2.6	Neurone fonctionnel[69]	46
2.7	Rétropropagation de l'erreur lors de l'apprentissage d'un réseau de neurones[69].	47
2.8	Réseau de neurones convolutifs[69]	48
2.9	Représentation logique d'un arbre décisionnel. Les cercles représentant les embranchements et les rectangles les feuilles.	49
2.10	Architecture d'un auto-encodeur[81]	52
2.11	Représentation d'un auto-encodeur entraîné pour une tâche de segmentation d'image[82].	53
2.12	Différence entre l'encodeur et le mécanisme d'attention[84]	54
2.13	Jeu de test et jeu d'apprentissage	58
2.14	Exemple d'une validation croisée à $n = 3$ segmentations, d'un même jeu de données ordonné de manière identique dans toutes les itérations.	58
2.15	Exemple de courbe d'apprentissage.	59
3.1	Structuration des réseaux de neurones en vue d'une tâche de classification (<i>i.e.</i> , hospitalisation vs. sortie des urgences) à partir des données structurées et non structurées[23]. RNP : réseau de neurones profond, RNC : réseau de neurones convolutionnel.	71
3.2	Structure détaillée du modèle de prédiction du parcours patient aux urgences[23].	74
3.3	Fonction de coût d'entraînement et de validation du modèle[23].	75
3.4	Courbe ROC du modèle[23].	75

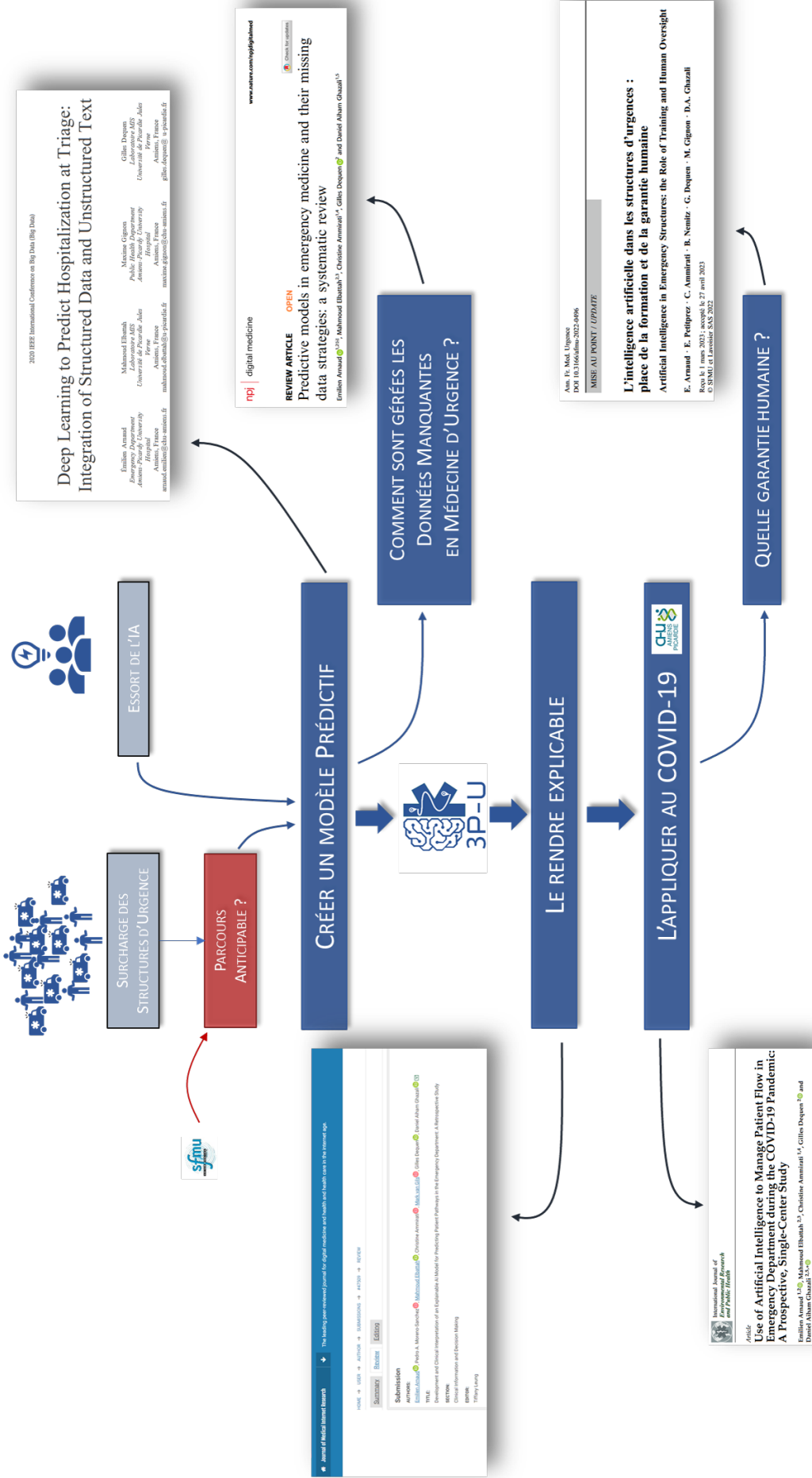
4.1	Stratégies appliquées aux données manquantes avant le modèle prédictif final[24]. A Stratégies des données manquantes en utilisant un pré-traitement autonome lié à l’algorithme final. B Stratégies de données manquantes utilisant un pré-traitement non autonome. C Stratégies additionnelles de données manquantes.	78
4.2	Méthodologie d’inclusion des études	82
4.3	Identification, analyse et inclusion des études via les bases de données et les registres[24].	84
4.4	Types d’algorithmes utilisés par année[24].	85
4.5	Évolution de la fréquence d’utilisation des stratégies de données manquantes[24]. Certaines publications utilisant plus d’une stratégie, le total peut excéder 100%.	87
4.6	Un schéma méthodologique explicite des données manquantes[24], adaptées de Martinez <i>et al.</i> [119]	89
5.1	Courbe ROC des modèles en utilisant le jeu de test	98
5.2	Courbe ROC du meilleur modèle de Réseau de Neurones Profonds (RNP), et le seuil décisionnel pour le meilleur F1-score.	98
5.3	Le graphique à point résumant l’explication globale de Explications additives de Shapley (SHAP) pour le modèle d’hospitalisation : plus la valeur SHAP est élevée (pour les variables quantitatives), plus la probabilité d’hospitalisation est élevée, et <i>vice versa</i> . Pour les variables binaires ou les variables catégorielles binarisées, la couleur rouge et bleue correspondent respectivement aux valeurs 1 et 0. L’échelle de tri French Emergency Nurses Classification in-Hospital triage (FRENCH) est détaillée dans le tableau 1.1.	99
5.4	Explication individuelle d’un cas réellement positif (a) et d’un cas réellement négatif (b) en utilisant SHAP.	100
5.5	Graphique Graphe des dépendances partielles, “ <i>Partial Dependence Plot</i> ” (PDP) pour la variable quantitative de l’âge.	100
5.6	Graphique PDP pour la variable binaire du motif de consultation “traumatisé sévère multiple”.	101
5.7	Graphique PDP pour la variable binaire du niveau 2 de l’échelle French.	101
5.8	Graphique PDP pour la variable binaire du motif de consultation “altération de l’état général”.	102
5.9	Graphique PDP pour la variable quantitative de la fréquence cardiaque.	102
5.10	Graphique PDP pour la variable quantitative de la cétonémie capillaire.	103
5.11	Graphique PDP pour la variable quantitative de la glycémie capillaire.	103
5.12	Graphique PDP pour la variable binaire du motif de consultation “glycémie supérieure à 13 mmol/L et cétonémie capillaire supérieure à 1.5mmol/L”.	104
5.13	Graphique PDP pour la variable binaire du motif de consultation “ascite”.	104
5.14	Graphique PDP pour la variable binaire du motif de consultation “ictère”.	105
5.15	Graphique PDP pour la variable binaire du motif de consultation “anémie”.	105
5.16	Graphique PDP pour la variable binaire du motif de consultation “score de Glasgow inférieur à 8”.	106
5.17	Graphique PDP pour la variable binaire du motif de consultation “score de Glasgow entre 9 et 12”.	106
5.18	Graphique PDP pour la variable binaire du constat d’une dyspnée majeure.	107
5.19	Graphique PDP pour la variable binaire du constat d’une hypoxémie inférieure à 85%.	107
5.20	Graphique PDP pour la variable du taux d’alcool expiré.	108
5.21	Graphique PDP pour la variable du débit d’oxygène administré.	108
6.1	Parcours du patient (et de ses données) au cours d’une utilisation réelle de 3P-U.	114
6.2	Tableau de bord disponible à la cellule de gestion des lits.	115
6.3	Évolution des résultats Réaction en Chaîne Polymérase (“ <i>polymerase chain reaction</i> ”) (PCR) du département de la Somme. Les données ont été extraites depuis data.gouv.fr, le 1 ^{er} juillet 2022[150]	118
6.4	Performances du modèle 3P-U et interprétation par la cellule de gestion des lits[26].	119
6.5	Lits nécessaires pour les patients présumés non-COVID-19 et les patients suspects COVID-19 par semaine, tel qu’utilisé par la cellule de gestion des lits pour optimiser les unités d’hospitalisation[26].	119

6.6	Prédiction individuelle de triage aux urgences du CHU-AP pendant la période COVID-19[26]. . .	120
7.1	Exemple de recours à ChatGPT, utilisant l'intelligence artificielle, pour résoudre un problème médical[27].	127
7.2	Composition et rôle du Collège de Garantie Humaine (CGH) dans un projet d'IA en santé au sein d'une institution[27].	129

Liste des tableaux

1.1	<i>FRench Emergency Nurses Classification in-Hospital triage</i> (FRENCH)[11]	23
1.2	Nombre de médecins seniors nécessaires au fonctionnement d'un service d'urgences en fonction du nombre de passages. Équivalent Temps Plein (ETP).[32]	25
1.3	Nombre d'IDE nécessaires au fonctionnement d'un service d'urgence (hors Unité d'Hospitalisation de Courte Durée (UHCD), hors Salle d'Accueil des Urgences Vitales (SAUV))[32]	25
2.1	5 premières lignes d'un jeu de données d'exemple	35
2.2	Tableau de contingence.	42
3.1	Jeu de données entraînant le modèle de prédiction du patient aux urgences[23]. Σ représente la somme et \bar{x} la moyenne sur l'ensemble du jeu de données.	70
3.2	Variables utilisées pour l'entraînement du classifieur visant à prédire l'hospitalisation du patient à la fin de la prise en charge[23].	72
4.1	Exemples de situations où des données peuvent manquer chez un patient se présentant aux urgences[24].	77
4.2	Nombre de publications pour chaque type de variable cible (primaire ou secondaire)[24].	83
4.3	Stratégies appliquées aux données manquantes utilisées dans les publications relues[24].	86
4.4	Description des données manquantes dans les publications.	86
5.1	Principales caractéristiques de la population d'étude. L'échelle de tri FRENCH[11] est détaillée dans la tableau 1.1. Les données sont exprimées en moyenne (écart-type) ou nombre (pourcentage).	97
5.2	Résultat des différentes itérations d'apprentissage et comparaison des différentes moyennes à l'aide du test de Friedman.	97
5.3	Performances des meilleurs modèles sur le jeu de test	97
6.1	Caractéristiques de la population d'étude[26]. Les variables continues sont représentés par leur moyenne (\pm écart-type) et les variables catégorielle par le nombre de patients (pourcentage).	117

Note de synthèse



Généralités

Le parcours de soin à l'hôpital est l'ensemble des étapes par lesquelles le patient passe pour sa prise en charge diagnostique (*i.e.*, caractériser sa pathologie) et thérapeutique (*i.e.*, soigner sa pathologie). Le parcours de soin peut se voir de manière macroscopique (*e.g.*, consultation aux urgences, admission dans le service de cardiologie, sortie) ou plus microscopique (*e.g.*, aux urgences : accueil et priorisation, biologie, imagerie, administration de traitement, avis spécialisé; en cardiologie : admission en soins intensif, coronarographie, surveillance, transfert en cardiologie conventionnelle, scanner, sortie). Appliqué au contexte de l'urgence, il faut distinguer le parcours pré-hospitalier du parcours intra-hospitalier (figure 1.1). Chacune de ces étapes est un parcours en lui-même. Le parcours de soin à l'intérieur d'un service d'urgence est particulier à chaque patient. Le parcours dépend du motif de venue aux urgences et des différents besoins pour le patient. La Haute Autorité de Santé (HAS) reconnaît l'importance du concept de parcours de soin et de son évaluation au sein d'un établissement de santé (figure 1.3)[1]. **Les mesures de l'activité d'un Structures d'Urgence (SU) (*e.g.*, temps de passage moyen, nombre de passages quotidiens moyen, taux d'hospitalisation) sont institutionnelles et non spécifiques du patient : elles ne permettent pas une adaptation fine des besoins en fonction des patients réellement présents dans les SU.**

Régulièrement, les SU sont en état de surcharge (*i.e.*, dépassement de la capacité d'un SU par le nombre de patients en attente d'être vus, traités ou hospitalisés[2]) : les indicateurs (*e.g.*, nombre de passages, nombre de patients présents, lits disponibles) sont très élevés et les ressources (*i.e.*, humaines et matérielles) ne permettent pas une prise en charge adaptée. Plusieurs études internationales arrivent à la même conclusion : les SU sont régulièrement surchargées, conduisant à une diminution de la sécurité et de la qualité des soins[2-4]. La combinaison d'une surcharge dans les SU et un nombre insuffisant de lits conduisent souvent à l'admission d'un patient dans une unité inappropriée. Cette situation a comme conséquences : une Durée Moyenne de Séjour (DMS) plus élevée (et des coûts plus importants)[5], une mortalité plus élevée[6], une morbidité plus élevée[7], une augmentation des patients partis sans attendre[8], *etc.* De nombreux groupes de recherche travaillent sur des moyens de contrer ces problèmes et de limiter les effets de cette surcharge : une régulation en amont du passage aux urgences (et permettre l'utilisation de l'offre de soin ambulatoire), une meilleure gestion des lits, un tri par une équipe d'urgentistes, évaluation de l'afflux, systèmes de tris automatiques, optimisation des transferts dans les services (même si le lit n'est pas prêt), et une augmentation du nombre de lits disponibles[2, 3, 9, 10].

Dès son arrivée, et une fois l'inscription administrative réalisée, le patient est vu par l'Infirmier Organisateur de l'Accueil (IOA). Lors de cette étape du parcours de soin aux urgences, l'IOA recueille les éléments de contexte (*e.g.*, motif de consultation, quel parcours l'ayant conduit aux urgences, *etc.*) et cliniques (*e.g.*, fréquence cardiaque, pression artérielle, saturation en oxygène, *etc.*). L'ensemble de ces éléments permet d'effectuer un *triage*, c'est-à-dire prioriser le patient en fonction de son état dans une catégorie indiquant le délai acceptable avant sa prise en charge. En France, il est recommandé que le triage soit réalisé par l'IOA suivant l'échelle French Emergency Nurses Classification in-Hospital triage (FRENCH) incluant cinq catégories[11] (tableau 1.1).

La Société Française de Médecine d'Urgence (SFMU) a dédié les journées thématiques interactives 2022 à l'utilisation concrète des outils d'Intelligence Artificielle (IA) en médecine d'urgence : pour l'organisation, pour le soin, pour la supervision et pour la formation[12]. Par exemple, en intrahospitalier, l'IA peut être utilisée pour détecter les patients à hospitaliser, pour cibler les patients à risque de se dégrader à la fin de leur prise en charge aux urgences[13], pour prédire les flux entrants aux urgences avec sept jours d'avance (*e.g.*, projet du centre hospitalier de Valenciennes[14]), en préhospitalier pour identifier les motifs d'appels (*e.g.*, projet du CHU de Bordeaux[15]), ou encore comme outil d'aide au diagnostic visant à éclairer le médecin dans son analyse des données brutes[16]. Son usage est déjà largement acquis dans le domaine de l'imagerie médicale[17]. Plus récemment, un outil de reconnaissance automatique des fractures fondée sur l'IA a été déployé au centre

hospitalier Paris Saint-Joseph avec un accompagnement privé permettant la mise en place d'un collège de garantie humaine à l'échelle de l'établissement de santé[12]. **Existe-t-il un modèle permettant de prédire l'hospitalisation, dans le contexte du système de santé français, à partir des données d'accueil et comment pourrait-il être utilisé ?**

Les modèles d'IA, pouvant être basés sur des algorithmes très différents (*e.g.*, [Regression Logistique \(RL\)](#), [Réseau de Neurons Profonds \(RNP\)](#), [Forêts Aléatoires \(FA\)](#), *etc.*), sont tous basés sur des données ([figure 2.1](#)). La manière dont les données sont collectées (*i.e.*, manuellement, automatiquement) et le processus menant à cette collecte (*i.e.*, prospectif, rétrospectif, systématique ou à l'indication) font varier la *qualité* du jeu de données. La quantité de données manquantes est un élément important influant sur la qualité du jeu de données. Or par définition, dans les [SU](#), toutes les données collectables ne le sont pas systématiquement (*e.g.*, la mesure de la glycémie capillaire est non systématique lorsqu'un patient consulte pour une entorse de la cheville). **Est-il licite de considérer ces données *manquantes* dans le contexte des [SU](#) de la même manière que dans le contexte général ?**

La crédibilité et la confiance en ces modèles prédictifs restent des préoccupations pour les cliniciens[18, 19]. Une meilleure acceptabilité de ces outils requiert des résultats générés plus accessibles et interprétables[20]. L'[Intelligence Artificielle eXplicable \(XAI\)](#) a été développée pour permettre de répondre à ces objectifs[21], et ainsi de permettre aux experts médicaux, en confiance, de prendre des décisions basées sur des données, et de fournir des diagnostics et des traitements personnalisés[22]. Aux urgences, l'[XAI](#) pourrait permettre (i) aux urgentistes de comprendre comment les données sont prises en compte dans le modèle, (ii) aux administrations comment fonctionne la logique du modèle et justifier son déploiement dans leur institution, et (iii) aux patients de comprendre comment leurs données sont utilisées.

Ces modèles proposent de prédire la variable cible de manière rapide, reproductible, systématique et infatigable quelle que soit l'heure du jour ou de la nuit. Le modèle se fonde sur des données collectées lors d'un processus humain (*e.g.*, interrogatoire, prise de constante, *etc.*). En revanche, la prise en compte, l'analyse et l'intégration de ces informations ainsi que la proposition de décision finale relèvent de processus informatiques. Ainsi, un modèle est utilisable fidèlement à toute heure du jour et de la nuit, sans variation de performance contrairement aux performances humaines (*e.g.*, baisse de vigilance nocturne, interruptions de tâches, émotion, *etc.*). De plus, même s'il traite les patients séquentiellement, le modèle proposera une réponse pour plus de patients qu'un médecin pour une même unité de temps. En limitant les facteurs humains et les biais cognitifs, il peut être pertinent que des outils d'aide à la décision viennent en appui à un médecin aux performances diminuées par plusieurs heures de garde. Ils permettent de gagner du temps et d'éprouver parfois le raisonnement médical lorsque les résultats proposés sont différents de l'idée médicale initiale. **Comment ces modèles de prédiction du parcours patient pourront être expliqués au clinicien et quelle garantie humaine pourrions-nous assurer quant à son utilisation ?**

Contributions

Nous avons vu précédemment qu'une des solutions pour lutter contre le problème de la surcharge dans les SU consiste à prédire précocement l'orientation du patient à la fin de la prise en charge : sortie au domicile ou admission dans un service d'hospitalisation, mais qu'il n'existait pas de modèle français permettant de faire cette prédiction.

📌 Question de recherche #1

Comment construire un modèle prédictif du parcours de soins d'un patient dans une SU ?

Cette contribution a été publiée sous la référence: Emilien ARNAUD *et al.* "Deep Learning to Predict Hospitalization at Triage : Integration of Structured Data and Unstructured Text". In : *2020 IEEE International Conference on Big Data (Big Data)*. 2020 IEEE International Conference on Big Data (Big Data). Atlanta, GA, USA : IEEE, 10 déc. 2020, p. 4836-4841. ISBN : 978-1-72816-251-5. DOI : [10.1109/BigData50022.2020.9378073](https://doi.org/10.1109/BigData50022.2020.9378073). URL : <https://ieeexplore.ieee.org/document/9378073/> (visité le 18/05/2021).

Nous avons utilisé l'apprentissage profond à cette fin. À partir d'un jeu de données massives d'environ 260k passages aux urgences du [CHU Amiens-Picardie \(Amiens, Picardie, France\) \(CHU-AP\)](#), les données structurées et non structurées (*i.e.*, notes textuelles) disponibles à l'étape du triage ont été utilisées. L'idée principale consiste à appliquer en entrée deux processus d'apprentissage : d'une part un [RNP](#) pour les données structurées, constitué de trois couches de neurones comprenant respectivement 128, 64 et 32 neurones et d'autre part un [Réseau de Neurones Convolutifs \(RNC\)](#) pour les données non structurées (*i.e.*, notes de l'IOA), constitué de trois couches convolutionnelles à une dimension ayant respectivement une taille de filtres définie à 16, 32 et 64, puis une couche de fusion de type *global-max* et enfin une couche dense. Ces deux processus d'apprentissages sont fusionnés dans un [RNP](#) pour entraîner au final un modèle de classification de l'hospitalisation : [Prédiction du Parcours Patient aux Urgences \(3P-U\)](#).

Les résultats démontrent que le modèle de classification permet de faire une prédiction de bonne précision ([Aire sous la courbe ROC, "Area Under the ROC curve" \(AUROC\)](#) à 83%). Cette étude montre qu'il est pertinent d'utiliser à la fois les données structurées et non structurées en utilisant des techniques d'analyse du langage naturel dans le domaine de la médecine d'urgence et particulièrement pour la prédiction de l'hospitalisation des patients.

Lors de l'entraînement de [3P-U](#), nous avons été confrontés à des données manquantes. En effet, dans certains cas les données sont volontairement omises et constituent ainsi les "[Données volontairement non collectées](#)" ("*data not purposely collected*") ([DNPC](#)). Ce biais d'information assumé peut être géré de différentes manières : exclure les patients avec des données manquantes, imputer par la moyenne ou utiliser des techniques automatiques (*e.g.*, d'apprentissage machine) pour gérer ou imputer les données manquantes.

📌 Question de recherche #2

Comment sont gérées les données manquantes dans les autres études impliquant l'IA en médecine d'urgence ?

Cette contribution a été publiée sous la référence: Emilien ARNAUD *et al.* "Predictive Models in Emergency Medicine and Their Missing Data Strategies : A Systematic Review". In : *npj Digital Medicine* 6.1 (1 23 fév. 2023), p. 1-9. ISSN : 2398-6352. DOI : [10.1038/s41746-023-00770-6](https://doi.org/10.1038/s41746-023-00770-6). URL : <https://www.nature.com/articles/s41746-023-00770-6> (visité le 24/02/2023).

Nous avons effectué une revue systématique de la littérature des méthodes utilisées pour gérer les données manquantes dans la recherche sur l'IA en médecine d'urgence. Nous avons évalué toutes les publications en langue anglaise qui traitaient de la construction ou de l'application d'un modèle prédictif basé sur des données structurées de patients adultes dans les services d'urgences. Cette revue systématique a été effectuée en questionnant PubMed avec la requête "(emergency medicine OR emergency service) AND (artificial intelligence OR machine learning)".

Après une première analyse des titres, nous avons sélectionné 213 publications pour l'analyse des articles. Au total, 141 publications ont été exclues sur la base des résumés (102) ou des textes complets (41). Ainsi, 72 études ont été incluses dans cette revue de la littérature. Les modèles entraînés visaient à prédire un diagnostic dans 25 (35%) des publications, la mortalité dans 21 (29%) des publications et la probabilité d'admission dans 21 (29%) des publications. Huit publications (11%) prédisaient deux variables. Seulement 15 (21%) publications décrivaient leurs données manquantes. Quarante-huit (67%) publications utilisent une seule stratégie (tableau 4.3), 14 (19%) publications ont utilisé plus d'une stratégie, et 10 (14%) publications n'ont pas spécifié la stratégie utilisée. La stratégie la plus fréquemment utilisée était l'exclusion de patients contenant des données manquantes (30 publications ; 41%), suivie par l'imputation par la moyenne (17 publications ; 23%), la prise en compte native (no-op ; 13 publications ; 18%), l'imputation algorithmique (6 publications ; 8%) et enfin l'utilisation de constantes physiologiques/de référence (3 publications ; 4%). Sept (10%) publications ont utilisé l'indicateur d'absence de valeur. Même si les DNPC ont été décrites de manière plus rigoureuse depuis 2020, les descriptions dans la littérature ne sont ni exhaustives, ni systématiques, ni homogènes. L'imputation multiple apparaît comme la meilleure stratégie mais requiert plus de temps et plus de ressources de calcul.

Pour augmenter la qualité, et la comparabilité des études, nous recommandons l'inclusion de la liste de vérification "*The Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis*" (TRIPOD) pour toute nouvelle publication, d'une synthèse du processus de machine learning dans un diagramme méthodologique explicite et de systématiquement publier les aires sous la courbe "*Receiver operating characteristic*" (ROC) même quand ce n'est pas le critère de jugement principal.

La mesure de la performance d'un outil d'IA *in silico* ne suffit pas à garantir son utilisation par l'ensemble des personnels soignants. La crédibilité du modèle est importante pour les cliniciens, et augmenter le taux d'utilisation de celui-ci nécessite une réponse lisible et interprétable. L'XAI a été développée spécifiquement pour répondre à ces problématiques et pour permettre aux experts de santé de prendre des décisions basées sur les données pour une médecine personnalisée et une organisation du parcours améliorée.

📌 Question de recherche #4

Comment expliquer la prédiction du modèle de parcours de soins d'un patient dans une SU ?

Cette contribution a été publiée sous la référence: Émilien ARNAUD *et al.* "Development and Clinical Interpretation of an Explainable AI Model for Predicting Patient Pathways in the Emergency Department : A Retrospective Study". In : *Journal of Medical Internet Research* UNDER REVIEW (21 sept. 2023).

Pour développer un modèle explicable et expliqué médicalement, nous avons mené une étude monocentrique, rétrospective sur les données structurées uniquement des patients ayant consulté aux urgences adultes du CHU-AP entre 2015 et 2019. Nous avons entraîné plusieurs modèles (incluant une validation croisée à cinq itérations d'un modèle de RNP, d'un modèle des Plus Proches Voisins (PPV), d'un modèle de RL et d'un modèle de FA) sur un jeu d'entraînement (80%). Nous avons évalué la performance de ces modèles en comparant leur AUROC. La validation interne des meilleurs modèles a utilisé le jeu de test (20%) pour calculer le rappel, la précision et l'exactitude. Le seuil décisionnel utilisé pour ces mesures était celui du plus grand F1-score. Les techniques d'XAI, comme les Explications additives de Shapley (SHAP) et le Graphe des dépendences partielles,

“*Partial Dependence Plot*” (PDP) ont été appliquées au modèle. Deux médecins urgentistes ont apporté des explications critiques cliniques aux résultats donnés par l’XAI. Confirmant nos études précédentes, le RNP donne une meilleure performance (AUROC : 83.1%, SD : 0.2%) que le modèle FA (AUROC : 71.6%, SD : 0.1%), le modèle PPV (AUROC : 67.2%, SD : 0.2%) et le modèle RL (AUROC : 71.5%, SD : 0.2%). La validation interne, en utilisant le jeu de test, l’AUROC du RNP est de 83.2%. Le meilleur F1-score (0.67) a permis de déterminer que 0.35 était le seuil décisionnel optimal. Le rappel, la précision et l’exactitude correspondants étaient respectivement de 75.7%, 59.7% et 75.3%. Les techniques d’XAI de type SHAP et PDP (évaluées par les médecins urgentistes) mettent en évidence que l’âge, la fréquence cardiaque et la présentation pour de multiples traumatismes ont le plus influencé la réponse du modèle sur l’admission patient. Ainsi, le modèle explicable 3P-U confirme les mécanismes cognitifs de réflexion des urgentistes en utilisant, entre autre, l’âge, les paramètres vitaux et le motif de consultation dans le processus décisionnel de l’admission ou de la sortie du patient. La transparence du modèle devrait améliorer la facilité d’utilisation de tels modèles dans la pratique clinique.

Le développement d’un modèle prédictif n’a de sens que s’il a une application clinique. Pendant la pandémie du coronavirus de 2019 (COVID-19), le calcul du nombre de lits nécessaires aux patients avec et sans suspicion de COVID-19 présents dans les SU était un vrai problème de santé publique. Nous avons mis à profit le modèle 3P-U pour optimiser la gestion des lits au CHU-AP.

📌 Question de recherche #5

Comment peut être appliqué le modèle de prédiction du patient dans un contexte clinique ?

Cette contribution a été publiée sous la référence: Emilien ARNAUD *et al.* “Use of Artificial Intelligence to Manage Patient Flow in Emergency Department during the COVID-19 Pandemic : A Prospective, Single-Center Study”. In : *International Journal of Environmental Research and Public Health* 19.15 (5 août 2022), p. 9667. ISSN : 1660-4601. DOI : [10.3390/ijerph19159667](https://doi.org/10.3390/ijerph19159667).

Nous avons mené une étude prospective, monocentrique des patients consultant aux urgences adultes du CHU-AP en 2020 et 2021. L’objectif était de déterminer le nombre minimum et maximum de lits nécessaires, en temps réel, d’après le modèle 3P-U. À leur arrivée aux urgences, une IOA priorisait les patients en suivant l’échelle de tri FRENCH et triait les patients selon s’ils étaient suspects d’être atteints de la COVID-19 ou non. La cellule de gestion des lits avait accès à un tableau de bord (figure 6.2) avec le statut réel des patients : “orientation inconnue”, “à hospitaliser” (*i.e.*, décision médicale définitive), “probablement à hospitaliser” (*i.e.*, recommandation de 3P-U), “probable sortie” (*i.e.*, recommandation 3P-U), “retour au domicile” (*i.e.*, décision médicale définitive). Nous avons défini un patient comme “évident” lorsque la probabilité prédite d’admission était inférieure à 0.2 (*i.e.*, interprétée comme “probable sortie”) ou supérieure à 0.6 (*i.e.*, interprétée comme “probablement à hospitaliser”). Entre 0.2 et 0.6, nous avons considéré que le patient était dans une “zone grise” et l’avons indiqué par le statut : “orientation inconnue”. Au total, 105 457 patients ont été inclus. L’AUROC était de 82% pour tous les patients et de 90% pour les patients évidents ($P_{admission} \in [0; 0.2] \cup [0.6; 1]$). Plus précisément, 38 353 (36.4%) patients étaient catégorisés “probablement sortants”, 18 815 (17.8%) étaient catégorisés “probablement admis”, et 48 297 (45.8%) patients n’ont pu être catégorisés. En se basant sur le nombre minimum de lits nécessaires (pour les cas évidents) et le nombre maximal de lits nécessaires (pour tous les patients), la cellule de crise hospitalière a coordonné la conversion des services en unités dédiés aux patients COVID-19. L’AUROC du modèle 3P-U se situe au milieu de l’intervalle des AUROC rapportés par d’autres études avec des classifieurs similaires. En considérant l’intervalle des lits nécessaires, la perte de ressources (*i.e.*, de temps et de lits) a pu être limitée. Cette étude confirme que l’application d’un modèle prédictif peut aider considérablement la management organisationnel des ressources hospitalières pendant les événements mondiaux, tel que celui de la COVID-19.

L'IA doit être considérée comme un outil d'aide dans la démarche diagnostique et thérapeutique des patients et d'amélioration de l'organisation des SU, notamment par la prise en compte de contraintes "métiers", contextuelles, relatives aux patients et plus généralement structurelles. L'IA comporte des avantages (reproductibilité, rapidité) mais aussi des risques (erreur, perte d'esprit critique).

📌 Question de recherche #6

Comment garantir la place de l'humain dans la conception et l'utilisation d'un modèle d'IA dans une SU ?

Cette contribution a été publiée sous la référence: E. ARNAUD *et al.* "L'intelligence artificielle dans les structures d'urgences : place de la formation et de la garantie humaine". In : *Annales françaises de médecine d'urgence* 13.3 (3^{1^{er}} mai 2023), p. 169-178. ISSN : 2108-6524, 2108-6591. DOI : [10.3166/afmu-2022-0496](https://doi.org/10.3166/afmu-2022-0496). URL : https://afmu.revuesonline.com/articles/lvafmu/abs/2023/03/lvafmu_2023_sprurge001324/lvafmu_2023_sprurge001324.html (visité le 20/02/2024).

La Commission Européenne (CE) a publié une proposition de règlement visant à établir des règles harmonisées concernant l'IA au sein de l'Union Européenne (UE), baptisée "Artificial Intelligence Act", ou "IA Act"[28]. La proposition de règlement retient une approche fondée sur les risques[29]. Les utilisations de l'IA sont classées selon leur niveau de risque et les règles juridiques varient en fonction de ce niveau. Les pratiques présentant un risque inacceptable sont ainsi interdites (Article 5). Les Dispositifs Médicaux (DMs) sont notamment visés par l'annexe II de la proposition de règlement. Une IA qui serait utilisée dans une structure de soins notamment de soins d'urgences serait donc potentiellement concernée. En tant qu'IA à haut risque, en plus de l'exigence de certification (norme CE et DM), son utilisation serait conditionnée par différentes obligations. L'article 14 en particulier est relatif au contrôle humain, qui vise à prévenir ou à réduire au minimum les risques pour la santé, la sécurité ou les droits fondamentaux, notamment au moyen d'interfaces hommes-machines appropriées.

Dans la mesure où l'apprentissage supervisé d'une IA est le reflet d'une expertise humaine, unique ou basée sur un consensus et s'appuyant sur une masse de données, la qualité de la réponse peut être impactée lorsque les données d'entraînement sont biaisées par leur sélection, leur mesure, leur analyse ou leur interprétation. Intégrer un outil complexe, basé sur l'IA, dans un processus de soins multidimensionnel, au service de la santé d'une personne humaine, implique nécessairement des compétences spécifiques et transdisciplinaires. La garantie humaine doit s'appuyer sur plusieurs profils complémentaires pour appréhender l'implémentation et le déploiement d'un modèle dans un contexte de soins. La mise en place d'un collège de garantie humaine pluriprofessionnel visant à garantir la supervision des outils d'IA de la conception au développement, au déploiement et à l'utilisation quotidienne permettra ainsi d'assurer durablement la sécurité des patients.

Généralités

Introduction

Le nombre de passages dans les [SU](#) ne cesse d'augmenter[[30](#)] d'années en années (18.8 millions en 2013 contre 20.4 millions en 2021, [figure 1.4](#)). Il en résulte un dépassement régulier des capacités des [SU](#) : un état de surcharge[[2](#)]. Plusieurs solutions existent pour lutter contre cette surcharge, dont l'utilisation de l'[IA](#).

L'utilisation de l'[IA](#) dans l'optimisation des parcours de soins des patients pour diminuer cet effet de surcharge est une question qui mérite d'être posée. Nous définirons dans cette partie notion de parcours de soin ([chapitre 1](#)) et la notion d'[IA](#) (dans le [chapitre 2](#)), puis nous étudierons la partie contributions le développement d'un modèle de prédiction du parcours de soin ([chapitre 3](#)), puis le problème des données manquantes ([chapitre 4](#)), l'explicabilité du modèle ([chapitre 5](#)), l'application du modèle en situation clinique ([chapitre 6](#)), et enfin la question de la garantie humaine ([chapitre 7](#)).

Chapitre 1

Parcours de soin dans les Structures d'Urgence

Nous décrirons dans ce chapitre le parcours de soin d'un patient dans les [Structures d'Urgence \(SU\)](#).

1.1 Le parcours de soin à l'hôpital

Le parcours de soin à l'hôpital est l'ensemble des étapes par lesquelles le patient passe pour sa prise en charge diagnostique (caractériser sa pathologie) et thérapeutique (soigner sa pathologie).

Le parcours de soin peut se voir de manière macroscopique (*e.g.*, consultation aux urgences, admission dans le service de cardiologie, sortie) ou plus microscopique (*e.g.*, aux urgences : accueil et priorisation, biologie, scanner, administration de traitement, avis spécialisé; en cardiologie : admission en soins intensif, coronarographie, surveillance, transfert en cardiologie conventionnelle, scanner, sortie). Appliqué au contexte de l'urgence, il faut distinguer le parcours pré-hospitalier du parcours intra-hospitalier (figure 1.1). Chacune de ces étapes est un parcours en lui-même.

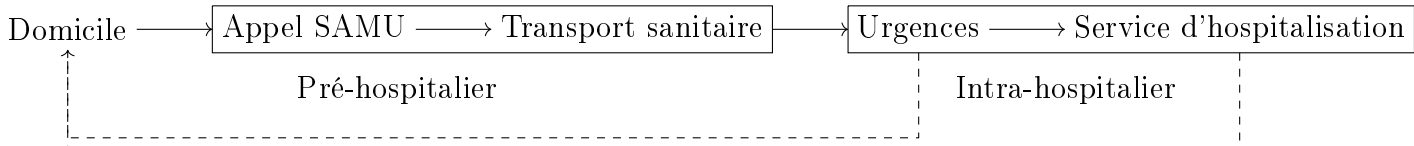


FIGURE 1.1 – Parcours de soin à l'hôpital

1.2 Le parcours de soin intra-hospitalier dans les SU

Le rôle d'un service d'accueil des urgences consiste à prodiguer les traitements urgents le cas échéant à un patient puis l'orienter vers un service d'hospitalisation ou organiser sa prise en charge ambulatoire. À cette fin, l'équipe en charge du patient peut prescrire des examens complémentaires urgents : bilan biologique, examens d'imagerie ou examens spécialisés (*e.g.*, électro-encéphalogramme). Enfin, d'autres spécialistes peuvent intervenir dans la prise en charge (*i.e.*, donner un avis spécialisé, réaliser un geste technique urgent).

Le parcours de soin à l'intérieur d'un service d'urgence est propre à chaque patient. Le parcours dépend du motif de venue aux urgences et des différents besoins pour la patient. Nous proposons ici différents types de parcours d'un patient aux urgences.

1.2.0.1 Évaluation initiale et systématique par l'Infirmier Organisateur de l'Accueil (IOA)

Dès son arrivée, et une fois l'inscription administrative réalisée, le patient est vu par l'IOA. Lors de cette étape du parcours de soin aux urgences, l'IOA recueille les éléments de contexte (*e.g.*, motif de consultation, quel parcours l'ayant conduit aux urgences, *etc.*) et cliniques (*e.g.*, fréquence cardiaque, pression artérielle, saturation en oxygène, *etc.*). L'ensemble de ces éléments permettent d'effectuer un triage, c'est-à-dire prioriser le patient en fonction de son état dans une catégorie indiquant le délai acceptable avant sa prise en charge. En France, il est recommandé que le triage soit réalisé par l'IOA suivant l'échelle [FRench Emergency Nurses Classification in-Hospital triage \(FRENCH\)](#)[11] (tableau 1.1)

Les tris 1 et 2 sont orientés dans la "Salle d'Accueil des Urgences Vitales (SAUV)", les tris 3 sont orientés dans le "Circuit Long (CL)", et les tris 4 et 5 sont pris en charge dans le "Circuit Court (CC)" (figure 1.5).

1.2.0.2 Parcours simple : consultation médicale

Le parcours de soin le plus simple consiste en un examen médical suffisant à orienter le patient vers une hospitalisation ou la sortie des urgences (*i.e.*, poursuite de la prise en charge en ambulatoire). Ces patients sont souvent regroupés et pris en charge dans un secteur nommé CC (figure 1.5).

Niveau de triage	Description	Action
1 - Risque vital immédiat	Patients présentant une détresse vitale patente ou une lésion traumatique qui engage le pronostic vital dans les minutes (<i>e.g.</i> , coma profond, amputation de membre, <i>etc.</i>)	Intervention médicale immédiate
2 - Risque vital imminent	Patients présentant une atteinte patente d'un organe ou une lésion traumatique sévère (<i>e.g.</i> , dyspnée sévère, déficit neurologique, <i>etc.</i>)	Intervention médicale dans les 20 min
3A - Urgence latente	Patients présentant une atteinte potentielle d'un organe ou d'une lésion traumatique instable, présentant une comorbidité en rapport avec la lésion ou adressé par un médecin	Intervention médicale dans les 60 min
3B - Urgence latente	Idem tri 3A sans comorbidité ou se présentant d'eux-même	Intervention médicale dans les 60 min
4 - Stable	Patients dont le motif de recours ne menace par le pronostic vital ou fonctionnel mais nécessitant la réalisation d'acte hospitalier (<i>e.g.</i> , examen biologique, examen radiologique, <i>etc.</i>)	Intervention médicale dans les 120min
5 - Sans urgence patente	Symptômes bénins ou chroniques stables	Intervention médicale dans les 240min

TABLE 1.1 – *FRENch Emergency Nurses Classification in-Hospital triage* (FRENCH)[11]

1.2.0.3 Parcours simple avec réalisation d'examen(s) complémentaire(s)

Lorsque l'examen médical ne suffit pas à orienter le patient, le médecin prescrit des examens complémentaires. Ces examens peuvent se réaliser dans le box des urgences (*e.g.*, bilan biologique) ou dans un service annexe (*e.g.*, radiographie, échographie, électro-encéphalogramme). Chacune de ces étapes (figure 1.2) prend un certain temps : prescription, acheminement du patient ou du prélèvement vers la zone d'analyse, réalisation de l'examen, attente de l'analyse médicale ou biologique, retour du patient. Si l'état du patient nécessite seulement une radiographie alors le patient est pris en charge dans le secteur des CC sinon, il est pris en charge dans le secteur des CL (figure 1.5).

Sur la base de ces résultats, le médecin urgentiste peut prescrire d'autres examens complémentaires ou prendre la décision de la poursuite des soins en milieu hospitalier ou en ambulatoire (*e.g.*, malaise vagal).

1.2.0.4 Parcours complexe avec avis d'autre(s) spécialité(s)

Certaines situations cliniques nécessitent le concours d'autres spécialités cliniques, souvent après les examens complémentaires, afin de participer à la prise en charge du patient dans sa spécialité (figure 1.2). Dans certains cas, ils peuvent être amenés à requérir de nouveaux examens complémentaires.

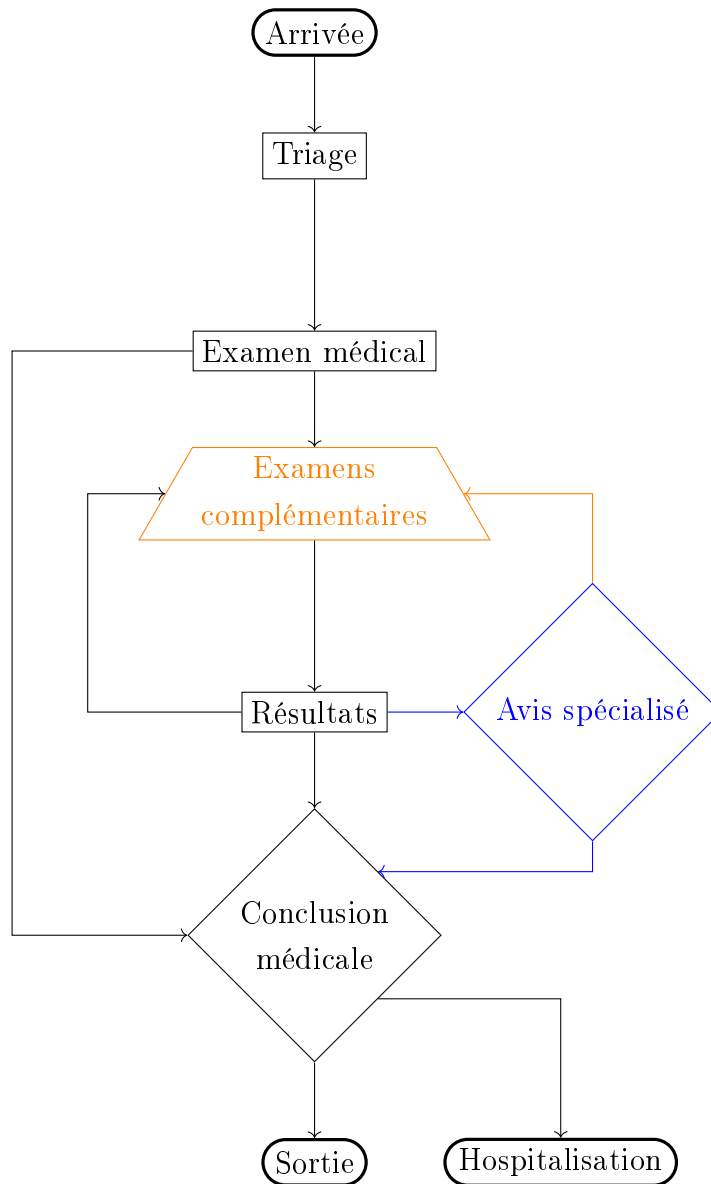


FIGURE 1.2 – Parcours de soins complexe avec avis d’autre spécialité.

1.3 Les indicateurs du parcours de soin

La [Haute Autorité de Santé \(HAS\)](#) reconnaît l’importance du concept de parcours de soin et de son évaluation au sein d’un établissement de santé ([figure 1.3](#))[\[1\]](#).

1.3.1 Nombre de passages

Le nombre de passages correspond au nombre de patients consultant aux urgences dans une journée. Cet indicateur est utilisé par la [Direction de la Recherche, des Études, de l’Évaluation et des Statistiques \(DREES\)](#) du ministère de la santé pour mesurer l’activité aux urgences en France. En 2022, 20480137 passages ont été enregistrés dans les urgences françaises[\[31\]](#) ([figure 1.4](#)).

La [Société Française de Médecine d’Urgence \(SFMU\)](#) et [SAMU Urgences de France \(SUdF\)](#) ont établi une graduation des effectifs médicaux ([tableau 1.2](#)) et non médicaux (dont [Infirmier Diplômé d’État \(IDE\)](#), [tableau 1.3](#)) en fonction du nombre de passages annuels[\[32\]](#)

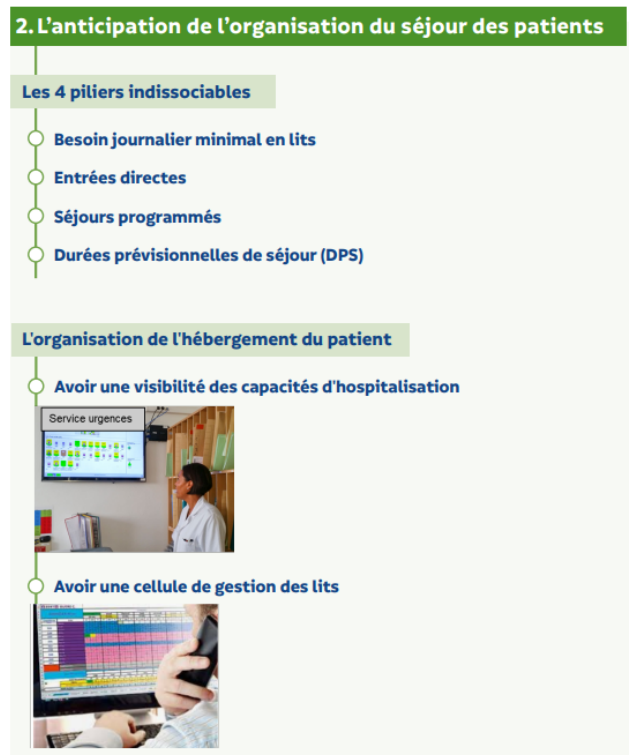


FIGURE 1.3 – Principes d'évaluation du parcours de soin par les experts visiteurs tels que recommandé par la haute autorité de santé[1]

Passages annuels	Médecins H24	Renfort diurne	Renfort nocturne	Total au pic d'activité	ETP
10k – 20k	1	1	0	2	8.4
20k – 30k	1	2	0	3	10.8
30k – 40k	2	1.5	0	4	15.6
40k – 50k	2	2	1	4	19.2
50k – 60k	2	2.5	1	5	20.4
60k – 70k	3	2.5	0	5	24.0
70k – 80k	3	2.5	1	6	26.4
80k – 90k	3	3	1	6	27.6
90k – 100k	4	3	0	7	31.2

TABLE 1.2 – Nombre de médecins seniors nécessaires au fonctionnement d'un service d'urgences en fonction du nombre de passages. ETP.[32]

Passages annuels	IDE H24	IOA H24	IDE Gypso
< 10k	1	0	0
10k – 17.5k	2	0.5	0
17.5k – 25k	3	0.5	0.5
25k – 35k	4	1	0.5
35k – 45k	5	1	1
45k – 55k	6	1.5	1
55k – 65k	7	1.5	1
65k – 75k	8	2	1
75k – 85k	9	2	1.5
85k – 95k	10	2.5	1.5

TABLE 1.3 – Nombre d'IDE nécessaires au fonctionnement d'un service d'urgence (hors UHCD, hors SAUV)[32]

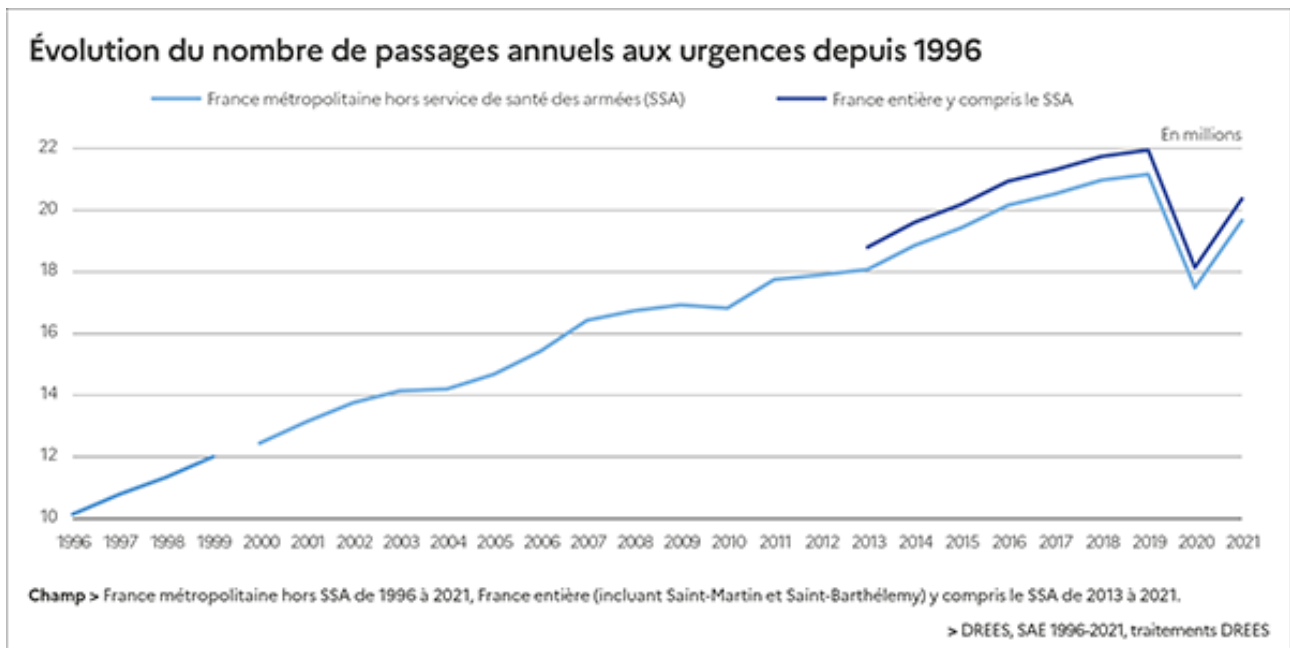


FIGURE 1.4 – Nombre de passages total sur l’ensemble des centres hospitaliers d’après la DREES[30]

1.3.2 Durée de passage

La durée de passage est définie par le temps passé par le patient entre son enregistrement administratif (*i.e.*, lors de son arrivée) et la clôture de son dossier médical (*i.e.*, lors de son départ du service des urgences à son domicile ou dans le service d’hospitalisation). Cet indicateur est scrupuleusement suivi par les chefs de service car plus cette durée augmente, plus la morbi-mortalité des patients augmente[5, 7]. En France en 2022, 61% des patients ont eu une durée de passage inférieure à 4 heures[33].

1.3.3 Évaluation du besoin en lits

Le **Besoin Journalier Minimal en Lits (BJML)** est un indicateur construit à partir de l’activité antérieure des urgences. Il s’agit du 25^{ème} percentile du nombre d’hospitalisations des jours de la même semaine sur les trois dernières années. Ce nombre correspond au nombre minimal de lits que le centre hospitalier doit avoir libérés. Trois fois sur 4, ce nombre est dépassé par le besoin réel.

Cet indicateur reste un indicateur institutionnel, qui ne tient pas compte de l’état clinique individuel de chaque patient, ni du besoin en temps réel du nombre de lits, ni de leur disponibilité.

Points clés

Les institutions et les sociétés savantes recommandent l’utilisation d’indicateurs (*e.g.*, nombre de passages, temps de passage) pour mesurer l’évolution de l’activité, anticiper les besoins en ressources humaines et le capacitaire. Néanmoins, ces indicateurs sont institutionnels et non spécifiques du patient : ils ne permettent pas une adaptation fine des besoins en fonction des patients réellement présents aux urgences.

1.4 Les urgences du CHU Amiens-Picardie

L’organisation d’un service d’urgence fait l’objet de recommandations de la SFMU[34]. Les urgences actuelles du **CHU Amiens-Picardie (Amiens, Picardie, France) (CHU-AP)** ont été ouvertes en novembre 2014. Elles ont

été construites en suivant au mieux les contraintes et les recommandations de la **SFMU**. Ainsi, il existe une entrée séparée pour les patients arrivant par leurs propres moyens et les patients arrivant en ambulance. Il existe également une zone d'accueil, une zone pour le **CC**, une zone pour le **CL**, et deux zones d'attente pour les patients en attente d'être vus par le médecin et ceux déjà vus ([figure 1.5](#)).

La **SAUV** est indépendante et possède sa propre entrée. Elle reste néanmoins facilement accessible de l'**IOA** pour y orienter les patients les plus graves. La **SAUV** est composée de six box de "déchocage" permettant de prendre en charge un patient en détresse vitale. L'équipe en charge d'accueillir ces patients est constituée d'un médecin urgentiste, de deux infirmières et d'une aide-soin. Ils sont tous expérimentés. Les box de **SAUV** sont équipés avec un moniteur multi-paramétrique avec report dans la salle de soin, permettant de surveiller les constantes vitales du patient : fréquence cardiaque, pression artérielle (invasive et non invasive), saturation en oxygène, mesure de la capnie expirée.

Le secteur **CL** est le secteur prenant en charge la grande majorité des patients sur 18 box de consultation. L'équipe de soin est composée d'un binôme aide-soin/**IDE**, le plus souvent d'un médecin junior (*i.e.*, interne ou docteur junior), et systématiquement d'un médecin sénior. Le nombre et la répartition des **IDE** et des médecins sont présentés dans la figure [figure 1.5](#). A cet effectif, il faut rajouter trois aide-soins, trois brancardiers et un cadre de santé.

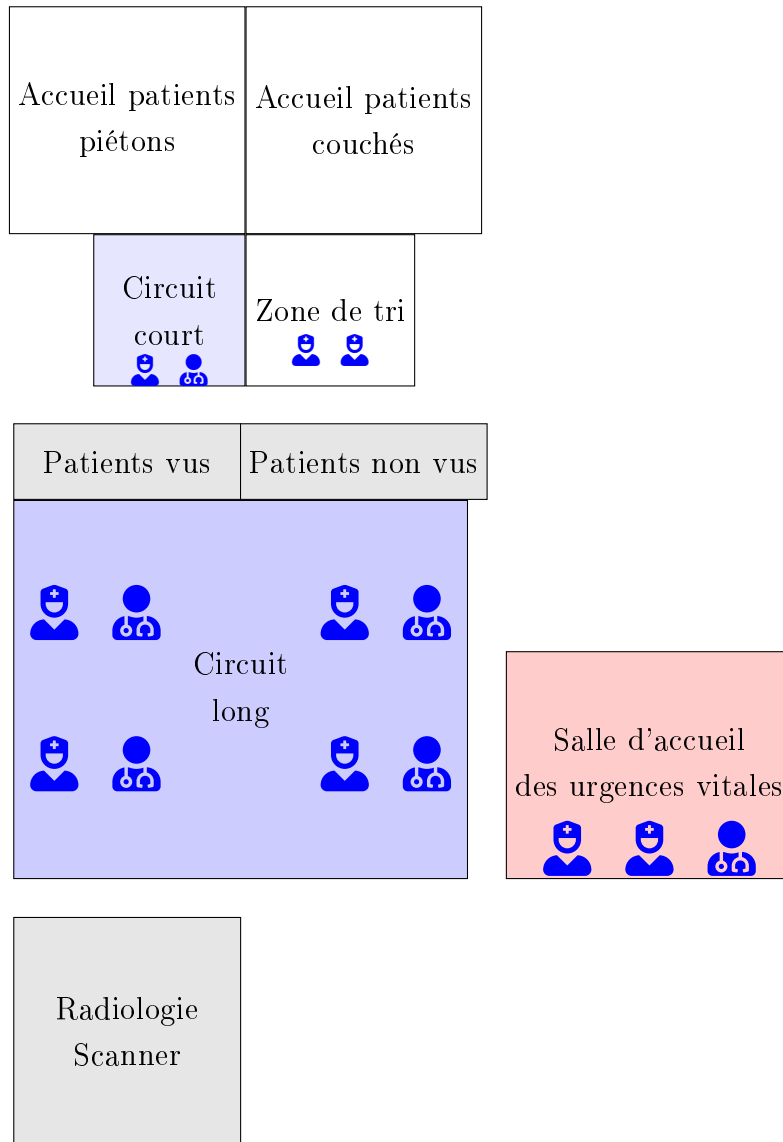


FIGURE 1.5 – Représentation spatiale et répartition des IDE (👤) et des médecins (👨‍⚕️) et des effectifs du service des urgences adultes du CHU-AP. Les aides-soin, les brancardiers, et les cadres de santé ne sont pas représentés dans cette représentation spatiale.

1.5 Problématiques rencontrées

1.5.1 Surcharge aux urgences

Les services d'urgence sont régulièrement en état de surcharge. La capacité d'un service d'urgence est définie par la quantité de patient qu'un service peut prendre en charge en simultanément. La surcharge est définie par le dépassement de la capacité d'un service d'urgence, par le nombre de patients en attente d'être vus, d'être traités ou d'être hospitalisés[2]. Claret *et al.* en 2014 proposaient déjà des mesures pour lutter contre cette surcharge : (i) **diminuer le nombre d'entrées aux urgences** en impliquant les SAMU Centre 15, en renforçant la permanence des soins, en facilitant l'accès direct aux spécialistes d'organe et aux plateaux techniques depuis la ville, en aidant le maintien à domicile, en réinsérant les patients précarisés et en éduquant la population, (ii) **améliorer l'efficacité du service** en augmentant le personnel médical et paramédical, en renforçant la collaboration avec les spécialistes d'organe, en créant un circuit court, en harmonisant les protocoles de soin institutionnels, en limitant les investigations superflues, et en suivant un tableau de bord des indicateurs, et (iii) **augmenter la sortie des patients des urgences** en augmentant les capacités hospitalières, en ayant une politique claire

d'attribution des lits, en favorisant une régulation médico-administrative, en accélérant les sorties des services conventionnels, en favorisant les patients urgents aux programmés, en disposant de lits pour les explorations rapides, et en hébergeant les patients médicaux dans un service de chirurgie.

Sur cette base, dans la suite des développements des tableaux de bord d'activité, développer un outil numérique qui permettrait d'aider au triage et à repérer les patients à haute probabilité d'être hospitalisés pour les admettre rapidement dans le service adapté. Le transfert rapide de ces patients permettrait de diminuer le nombre de patients présents aux urgences et ainsi participer à la lutte contre la surcharge[35].

L'augmentation de la morbi-mortalité liée à la surcharge a été démontrée en 2006 par RICHARDSON[6], confirmée en 2011 par GUTTMANN *et al.*[5] et plus récemment pour les personnes âgées par une étude française menée en 2023 par ROUSSEL *et al.*[36].

1.5.2 Synthèse manuelle

Afin de pouvoir organiser le flux de patients à hospitaliser, et pour anticiper les problématiques de recherche de lits, la direction des établissements de santé demande au service des urgences de pouvoir remonter rapidement le nombre de lits nécessaires. Or, faire une synthèse en temps réel du besoin en lits est consommateur de temps : il faut en permanence prendre en compte les nouveaux patients, retirer les patients sortis, reclasser les spécialités en fonction des nouveaux résultats, *etc.* Cette tâche est dévolue à un médecin et l'occupe une grande partie de son temps, et l'information produite est rapidement obsolète. Ce processus est donc répété régulièrement. Développer un outil numérique permettrait de faire cette synthèse rapidement et automatiquement en intégrant les éléments de triage et de gestion de flux : listing des patients occupant les box, entrées, attentes de prise en charge médicale, attentes d'examen complémentaires, attentes des résultats, attentes d'avis, attentes de sortie ou d'hospitalisation.

1.6 L'aide à la décision par l'Intelligence Artificielle (IA), une solution potentielle

Plusieurs approches ont été étudiées pour diminuer la surcharge dans les SU en utilisant l'apprentissage machine (*e.g.*, améliorer le triage[37]) ou prédire un diagnostic (*e.g.*, un accident vasculaire cérébral[38] ou une insuffisance cardiaque[39]).

La SFMU a dédié les journées thématiques interactives 2022 à l'utilisation concrète de l'IA en médecine d'urgence : pour l'organisation, pour le soin, pour la supervision et pour la formation[12]. Par exemple en intra-hospitalier, l'IA peut être utilisée pour détecter les patients à hospitaliser, pour cibler les patients à risque de se dégrader à la fin de leur prise en charge aux urgences[13], ou pour prédire les flux entrants aux urgences avec 7 jours d'avance. L'IA peut-être également utilisable en pré-hospitalier pour identifier les motifs d'appels (*e.g.*, projet du CHU Bordeaux[15]).

L'IA est aussi utilisée dans les SU comme outil d'aide au diagnostic visant à éclairer le médecin dans son analyse des données brutes[16]. Son usage est déjà largement acquis dans le domaine de l'imagerie médicale[17]. Plus récemment, un outil de reconnaissance automatique des fractures basée sur l'IA a été déployé au CH Paris Saint-Joseph avec un accompagnement privé permettant la mise en place d'un collègue de garantie humaine à l'échelle de l'Établissement de Santé (ES)[12].

 Points clés

Le parcours de soin d'un patient dans une structure d'urgence est rapidement complexe, le nombre de patients multiplie cette complexité et aboutit à la surcharge des urgences. La création et l'application d'un modèle de prédiction du parcours patients aux urgences d'un établissement de santé français en situation clinique réelle n'a jamais été réalisée ni étudiée.

Chapitre 2

Intelligence Artificielle (IA)

Dans ce chapitre, nous présenterons une définition de l'IA et quelles techniques peuvent être utilisées pour développer un modèle d'IA : analyse de données, pré-traitement et apprentissage machine. Nous resterons néanmoins synthétiques et insisterons plus particulièrement sur les réseaux de neurones utilisés par 3P-U et les fonctions de coûts. Nous aborderons enfin la problématique de la garantie humaine.

L'**IA** est définie comme “l’ensemble des théories et des techniques développant des programmes informatiques complexes capables de simuler certains traits de l’**Intelligence Humaine (IH)** (raisonnement, apprentissage)” [40]. D’après RUSSELL *et al.*, pour qu’une machine soit considérée comme “intelligente”, il faudrait non seulement qu’elle passe le test de Turing (*i.e.*, un observateur humain, après avoir posé des questions par écrit n’est pas capable de distinguer la réponse venant d’un humain ou d’un la machine)[41, p. 20], mais qu’elle soit également capable d’apprécier avec sa propre subjectivité l’expérience subjective d’un alter-ego[41, p. 1037]. Une telle machine devrait être capable de maîtriser (i) le traitement du langage naturel, (ii) la représentation des connaissances, (iii) le raisonnement automatisé, (iv) l’apprentissage automatique, (v) l’introspection, et (vi) les expériences psychologiques.[41, p. 20]

L'**IA** peut être classée en deux catégories (figure 2.1) : l'**IA faible** et l'**IA forte**[41, p. 1032]. Une machine douée d'**IA faible** se définit comme une machine qui agit comme si elle était intelligente. Elle a pour objectif de résoudre des problèmes simples (classification, prédiction de valeur, *etc.*) ou capable de modéliser des idées abstraites. Une machine douée d'**IA forte** (ou générale) se définit comme une machine capable d’apprendre et d’effectuer n’importe quelle tâche réalisable par un humain. Bien qu’à cette heure, la question de savoir si “les machines pensent réellement” reste ouverte[41, p. 1035], PITRAT envisage que les **Chercheurs Artificiels en IA (CAIAs)** puissent un jour développer la recherche en **IA** sans l’aide humaine[42, 43].





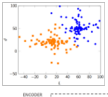

TYPE D'IA	TÂCHE RÉALISÉE	EXEMPLE D'ALGORITHME
IA FORTE Réalisation de n'importe quelle tâche	Une IA générale est une IA capable d'apprendre et de réaliser n'importe quelle tâche complexe réalisée par un humain	
IA FAIBLE Réalisation de tâches simples	Génération de texte	TRANSFORMERS  EX: CHAT GPT
	Génération d'image	
	Apprentissage supervisé : Apprentissage sur des données antérieures	RÉGRESSION LOGISTIQUE 
	Classification Classer des individus dans des groupes préalablement définis (ex: classer les patients dans le groupe « à hospitaliser » ou « sortant des urgences ») Régression Prédire une valeur numérique (ex: prédire le nombre probable d'admissions aux urgences pour chaque semaine)	ARBRES DÉCISIONNELS  RÉSEAUX DE NEURONES 
	Apprentissage non supervisé : Apprentissage sur des données observées	K-MOYENNE 
	Regroupement Identifier des groupes dans une population naïve (ex: identifier des clusters en période épidémique) Encodage Représentation abstraite des données (ex: repérer les différents types de structures sur une coupe de scanner)	AUTO-ENCODEUR 

FIGURE 2.1 – Représentations des différents types d'**IA**, des tâches qu'elle peut effectuer et des algorithmes sur lesquels elle peut se baser.

Un algorithme est défini comme étant un “ensemble de règles opératoires dont l’application permet de résoudre un problème énoncé au moyen d’un nombre fini d’opérations” [44]. Un algorithme d'**IA** est donc un ensemble de règles opératoires permettant de simuler certains traits de l'**IH** et plus spécifiquement de résoudre une suite de problèmes simples. Ces algorithmes appartiennent par conséquent à la catégorie des **IA faibles**. La conjonction et la coordination d'**IA faibles** permet, trivialement, de renforcer la capacité de simulation du raisonnement humain sans toutefois permettre la construction d’une **IA forte**.

Les algorithmes d’apprentissage machine et d’optimisation constituent les briques techniques de l'**IA faible**. Ils apprennent des modèles de manière supervisée, semi-supervisée ou non supervisée. Un modèle à apprentissage

supervisé consiste à apprendre les réponses d’une variable cible (*e.g.*, l’issue d’une consultation aux urgences : admission ou sortie) en fonction des autres variables explicatives (*e.g.* le motif de consultation, l’âge, le sexe, les paramètres vitaux) sur un jeu de données rétrospectif. La variable cible à prédire peut-être une catégorie, on parle alors de “tâche de classification”, ou une variable continue, on parle alors de “tâche de régression”. Un modèle à apprentissage non supervisé consiste à regrouper les individus par groupes de ressemblance (“tâche de regroupement”) ou à représenter de manière résumée les données (“tâche d’encodage”). Le terme “algorithme” représente alors le moteur technique, et le terme de “modèle” représente une occurrence paramétrée du moteur technique. L’apprentissage semi-supervisé satisfait les mêmes tâches que l’apprentissage supervisé sur la base d’un jeu de données rétrospectif où seule une partie des valeurs des variables cibles est connue (*e.g.* l’unique renseignement du patient “sain” dans le cadre de l’aide au diagnostic d’une ou plusieurs pathologies à faible prévalence).

La transparence des approches algorithmiques est variable. L’explicabilité (*i.e.*, la compréhension de la manière dont un modèle répond à la question) peut-être naturelle d’une part lorsque le modèle mathématique sous-jacent peut correspondre à une réalité appréhendable par l’humain. Nous pouvons citer en exemple la [Regression Logitique \(RL\)](#) qui correspond aux coefficients d’une équation mathématique dont le résultat est la réponse attendue, ou les arbres décisionnels[45] qui correspondent à une itération de subdivisions du jeu de données d’entraînement (*i.e.*, branches) aboutissant à une réponse commune (*i.e.*, feuille). L’explicabilité peut être artificielle d’autre part lorsque le modèle mathématique sous-jacent ne correspond à aucune grandeur appréhendable par l’humain (*e.g.*, les réseaux de neurones artificiels[46] : constitués de plusieurs couches de neurones reliés entre eux par des poids à l’image de l’organisation du cerveau humain). Dans ce dernier cas, le modèle produit ne correspond pas à une représentation compréhensible, que ce soit sur le plan physiologique ou clinique, du processus décisionnel. On parle alors de “boîte noire” : la décision du modèle peu transparente et faiblement explicable.

La majorité des modèles actuels sont fondés sur un apprentissage supervisé ou semi-supervisé et visent à prédire une variable d’intérêt (*e.g.*, hospitalisation, mortalité). Dans un tel apprentissage, le modèle est généralement entraîné sur un sous-ensemble aléatoire de données tirées aléatoirement et représentatives du jeu de données complet, le reste étant destiné à la validation et au test. La phase d’entraînement consiste à donner les variables explicatives et la variable cible réelle au modèle pour qu’il apprenne à répondre. La phase de test du modèle consiste à prédire la variable cible à partir des variables explicatives, et de comparer cette prédiction à la valeur réelle connue.

Dans ce cadre, l’entraînement d’un modèle suppose une quantité suffisante de données disposant par ailleurs d’une meilleure qualité et d’une meilleure complétude de renseignement possibles (*e.g.*, données de santé, produites lors de la prise en charge des patients ; [sous-section 2.1.6](#)).

Dans ce manuscrit, quelque soit la forme du modèle d’[IA](#) développé, il se base sur des données. Nous verrons donc dans un premier temps comment on constitue un jeu de données, puis dans la suite de notre propos, nous détaillerons le mode de fonctionnement d’un modèle d’[IA](#). Nous prendrons l’exemple du modèle [Prédiction du Parcours Patient aux Urgences \(3P-U\)](#) et une vue partielle du jeu de données (quelques colonnes) tout au long du propos comme illustration des différentes démonstrations.

2.1 Jeu de données

Un modèle d’[IA](#) utilise des données pour répondre au problème posé. Nous verrons les types de données[47], leur qualité et les problématiques soulevées en cas de défaut de qualité. Une donnée peut être structurée ou non structurée.

Une donnée structurée est une donnée qui peut être rangée dans une colonne de tableau : une donnée numérique continue, une donnée numérique discrète, une catégorie (numérique ou textuelle). Une donnée non

structurée est une donnée qui ne peut pas être rangée dans un tableau : un texte ou un média (image, son, vidéo).

Les données structurées représentent une mesure ou une catégorisation à un instant défini. Ces mesures peuvent être répétées dans le temps et forment ainsi une *série temporelle*. Ainsi, une donnée non structurée ayant une dimension temporelle (son, vidéo), peut être décomposée en une série temporelle de données structurées. De la même manière, une image ou un texte peuvent être décomposés ou résumés en données structurées.

2.1.1 Notations

Pour la suite de la présentation, nous considérerons les notations suivantes :

- N le nombre d'individus i utilisés pour l'apprentissage
- K le nombre de sorties possibles dans le cas d'une prédiction à multiples sorties
- X les vecteurs des données d'entrée
- \widehat{X} les vecteurs approximant les données d'entrées
- X_i le vecteur de données d'entrée pour l'individu i
- x_{ij} la $j^{\text{ème}}$ entrée de X_i
- y_i la sortie réelle attendue pour l'individu i
- Y les vecteurs des sorties réelles
- $Y_i = \{y_{i1}, \dots, y_{iK}\}$ le vecteur des K sorties réelles pour l'individu i
- y_{ij} la $j^{\text{ème}}$ sortie réelle de l'individu i
- \widehat{y}_i la sortie prédite par le modèle pour l'individu i
- \widehat{Y} le vecteur des sorties prédites
- $\widehat{Y}_i = \{\widehat{y}_{i1}, \dots, \widehat{y}_{iK}\}$ le vecteur des K sorties prédites pour l'individu i
- \widehat{y}_{ij} la $j^{\text{ème}}$ sortie prédite de l'individu i
- R tel que $y_i \in R$ l'ensemble de tous les y_i possibles (dans le cas d'une classification), ou le domaine d'existence (dans le cas d'une régression)
- b biais d'un neurone
- B vecteur des biais d'un réseau de neurones
- C vecteur contenant une information condensée
- ω_j poids de l'entrée j
- W vecteurs des poids d'un neurone
- l couche d'un réseau de neurones
- $m^{(l)}$ nombre de neurones de la couche l
- $y_i^{(l)}$ sortie du neurone i de la couche l
- $\delta_i^{(l+1)}$ l'erreur rétropropagée du neurone i de la couche $l+1$
- $loss_i$ la valeur du coût pour la prédiction de l'individu i
- $Loss$ la fonction de coût globale du modèle
- r le pas d'apprentissage
- V_i le volume d'entrée d'un réseau de neurones convolutifs
- V_o le volume de sortie d'un réseau de neurones convolutifs
- S la surface des récepteurs d'une couche de convolution
- P le pas de la couche de convolution

- M la marge de la couche de convolution
- f la fonction représentant le modèle
- θ le vecteur des paramètres de l'algorithme
- ψ le vecteur des paramètres à l'origine des données manquantes

2.1.2 Jeu d'exemple

Pour illustrer les propos des sections suivantes, nous nous baserons sur ce jeu d'exemple :

	SEXE	AGE	GLYCEMIE	MOTIF_RECOURS	MOIS	JOUR_SEMAINE	HOSPITALISATION
0	1	25	–	–	6	7	Faux
1	1	53	–	2103	4	3	Vrai
2	1	58	–	–	9	4	Vrai
3	1	20	–	–	10	5	Faux
4	0	39	–	3112	6	4	Faux

TABLE 2.1 – 5 premières lignes d'un jeu de données d'exemple

2.1.3 Donnée numérique

Une donnée numérique peut être **continue**. Une donnée X est continue lorsque son domaine d'existence peut prendre une infinité de valeurs. Son domaine d'existence peut être bornée (*i.e.*, ne prendre que des valeurs positives par exemple, $X \in [0; +\infty]$) ou non (*i.e.*, prendre n'importe quelle valeur réelle, $X \in \mathbb{R}$). L'exemple le plus commun de donnée continue en santé est l'âge. Il faut considérer l'âge comme une donnée continue pouvant prendre une infinité de valeurs entre 0 et un âge maximal physiologique (*e.g.*, $X_{age} \in [0; 150]$). Une donnée numérique peut être **discrète**. Une donnée X est discrète lorsque son domaine d'existence contient un nombre n fini de valeur x_i ($X \in \{x_1, x_2, x_3, \dots, x_n\}; n \in \mathbb{N}^+$)

Plusieurs métriques permettent de résumer une série de données numériques.

- La **moyenne arithmétique** \bar{x} permet de résumer une série de n nombres x_i en une valeur ne dépendant pas de l'ordre de ces nombres, tel que

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.1)$$

- La **médiane** permet de séparer la série numérique de taille n ordonnée par ordre croissant en deux groupes $\frac{n}{2}$ nombres. Dans le cas où n est paire, la médiane est la moyenne arithmétique des deux valeurs centrales. Dans le cas où n est impaire, la médiane est la valeur centrale.
- Le **mode** est la valeur la plus fréquente.
- La **variance** σ_x^2 mesure la dispersion de n valeurs x_i autour de la moyenne arithmétique \bar{x} , en calculant la moyenne des carrés des écarts à la moyenne, tel que

$$\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.2)$$

- L'**écart-type** σ_x représente l'écart *moyen* des n valeurs x_i et la moyenne arithmétique \bar{x} et correspond à la racine carrée de la variance, tel que :

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.3)$$

- Le **minimum** est la valeur la plus basse, le **maximum** est la valeur la plus élevée.

- L'**intervalle de confiance** $IC_{1-\alpha}$ est l'intervalle dans lequel chaque valeur x_i d'une donnée numérique suivant une loi normale a une probabilité $P(x_i \in IC_{1-\alpha}) = (1 - \alpha)$ d'appartenir, tel que :

$$IC_{1-\alpha} = [\bar{x} \pm t_\alpha \frac{\sigma_x}{\sqrt{n}}] \quad (2.4)$$

En médecine, il est admis d'utiliser l'intervalle de confiance à 95% ($IC_{95} \Rightarrow t_5 = 1.96$)[48], ainsi l'équation [équation 2.4](#) prend la forme suivante ([équation 2.5](#)).

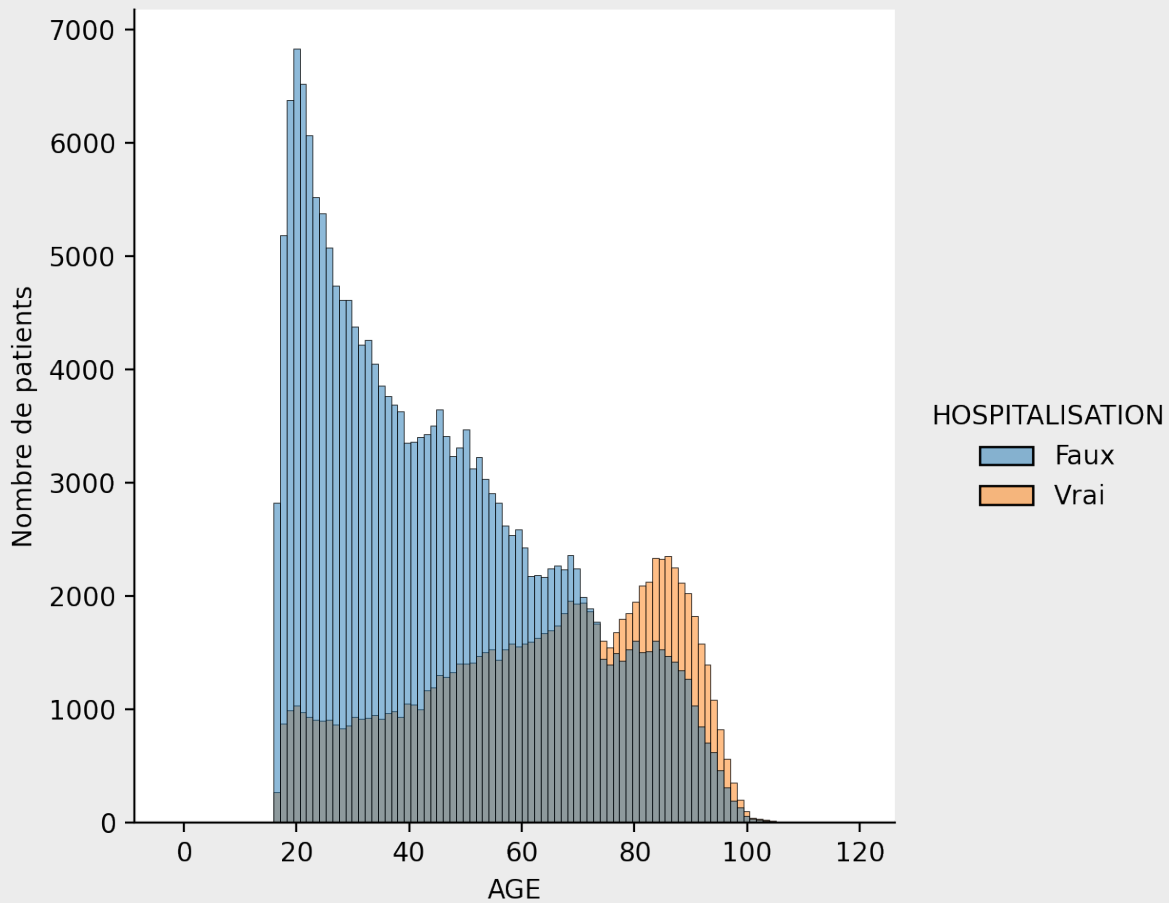
$$IC_{95\%} = [\bar{x} \pm 1.96 \times \frac{\sigma_x}{\sqrt{n}}] \quad (2.5)$$

Exemple – Variable numérique *AGE*

Statistiques de la variable *AGE*

HOSPITALISATION	Faux	Vrai
n	203 626	99 340
moyenne	45.0	62.40
écart-type	21.10	21.71
médiane	41.19	65.30
IC_{95}	[44.90 ; 45.09]	[62.26 ; 62.53]

Distribution de la variable *AGE*



2.1.4 Donnée catégorielle

Une donnée **catégorielle** X est une donnée discrète pouvant prendre un nombre n fini de valeurs x_i (*i.e.*, le domaine $\{x_1, x_2, \dots, x_n\}$) tel que $X \in \{x_1, x_2, \dots, x_n\}; n \in \mathbb{N}^+$. La valeur de chaque x_i peut être une lettre, un mot, ou un nombre. Par exemple, le mode d'attente d'un patient aux urgences $X_{attente}$ peut être debout, assis, ou sur un brancard tel que, $X_{attente} \in \{debout, assis, brancard\}$.

Plusieurs métriques permettent de résumer une série de données catégorielles.

- Le **mode** est la ou les valeurs les plus fréquentes.
- Dans le cas d'une variable catégorielle représentée par des nombres, il est possible d'utiliser les métriques des données numériques (*c.f.*, [sous-section 2.1.3](#))

Exemple – Variable catégorielle *SEXE*

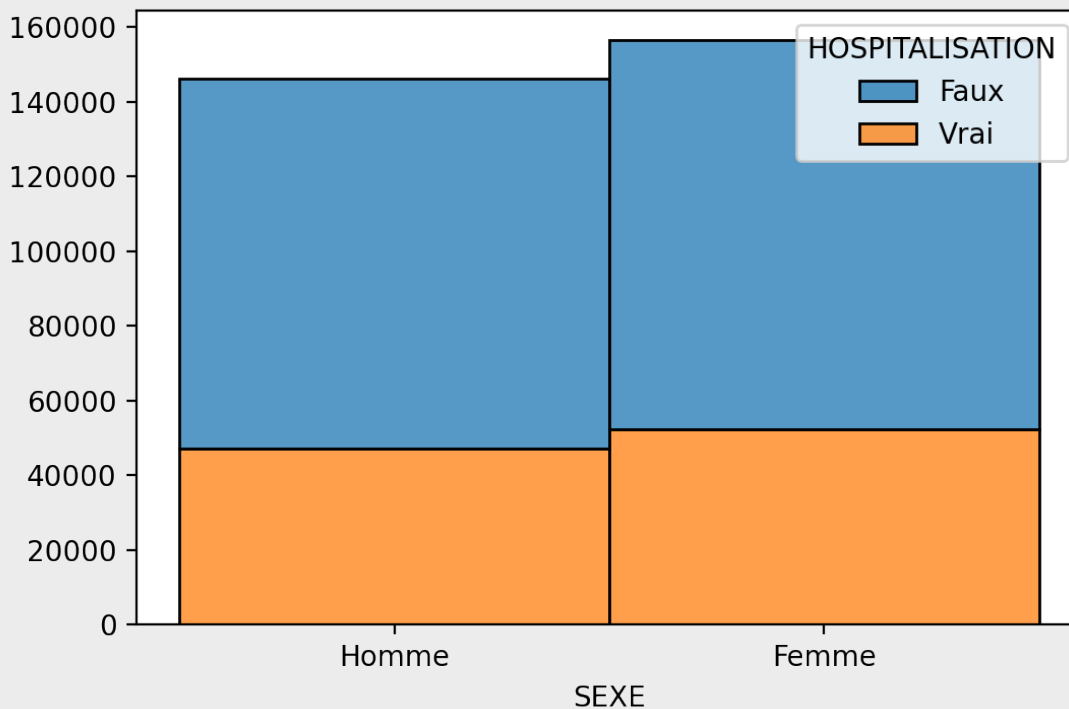
Tableau de contingence de la variable *SEXE*

SEXE	HOSPITALISATION	
	Faux	Vrai
Homme	104 408	53 213
Femme	99 218	47 127

Statistiques de la variable *SEXE*

	HOSPITALISATION	Faux	Vrai
n		203 626	99 346
mode		Homme	Homme

Distribution de la variable *SEXE*



2.1.5 Données non structurées (*e.g.* verbatim)

Les données non structurées représentent les données qui ne peuvent être représentées sous forme de tableau ou de graphe. Les techniques d'analyses utilisées dans le cadre des données structurées (*i.e.*, statistiques descriptives) ne peuvent s'appliquer directement sur du texte brut, un signal audio ou vidéo. L'une des approches consiste à transformer la données non structurée en donnée structurée[49].

A cet effet, le document textuel va subir une série de traitements pour le rendre structuré (*i.e.*, rendre les mots comparables entre eux et compatibles avec l'utilisation de modèles de type [Réseau de Neurones Profonds \(RNP\)](#))

2.1.5.1 Normalisation du texte

Le document textuel est ensuite converti en **minuscules** et les **mots-stop sont retirés**. Les mots sont ensuite extraits en ne considérant que leur racine (*e.g.*, abdominal, abdominale, abdominaux, et abdomen sont considérés comme étant le même mot). La **racinisation** permet de réduire le nombre de degrés de liberté pour chaque mot et représenter les concepts communs. L'algorithme *SnowballStemmer* dans sa version française peut être utilisé pour cette transformation. La normalisation du texte peut être réalisée en utilisant la librairie Python NLTK[50].

2.1.5.2 Individualisation – Tokenization[51]

L'individualisation des mots revient à **couper le texte en plus petite parties indépendantes** appelées token. Les tokens sont alors représentés par un vecteur. Il existe donc un vecteur de token par document. Après la normalisation du texte, le document est prêt à son individualisation en utilisant la librairie Python Keras[52].

2.1.5.3 Encodage binaire – One hot encoding[53]

Par la suite, nous pouvons procéder à un encodage binaire de ces vecteurs de token. Cette méthode consiste à produire un vecteur de variables binaires de la taille du corpus (*i.e.*, ensemble de mots racinisés uniques) où seule la variable binaire correspondant au mot vaut 1.

2.1.5.4 Plongement lexical – Word embedding[54]

L'étape finale consiste à appliquer une technique de plongement lexical. Cette étape est indispensable car le nombre de dimensions résultant de l'encodage binaire est très élevé rendant l'apprentissage très compliqué. L'un des stratégies de plongement lexical consiste à représenter le vecteur de données binaire par un vecteur de nombre réel de dimension inférieure. La bibliothèque Python Keras[52] propose une couche spécialement conçue pour cette opération. Cette couche peut être utilisée comme couche d'entrée d'un [Réseau de Neurones Convolutifs \(RNC\)](#). La [figure 2.2](#) synthétise l'ensemble des opérations de pré-traitement.

2.1.6 Qualité des données

La qualité d'un jeu de données est directement liée à sa représentativité et sa complétude.

2.1.6.1 Représentativité

La représentativité d'un jeu de données dépend de l'échantillonnage. Dans le cas où le jeu de données contient l'exhaustivité des données de la population étudiée, il est possible de porter des conclusions précises sur cette population à partir des données. Dans le cas inverse, lorsque le jeu de données ne contient qu'un sous-ensemble des données disponibles, il est nécessaire de vérifier que ce jeu de données est représentatif de la population.

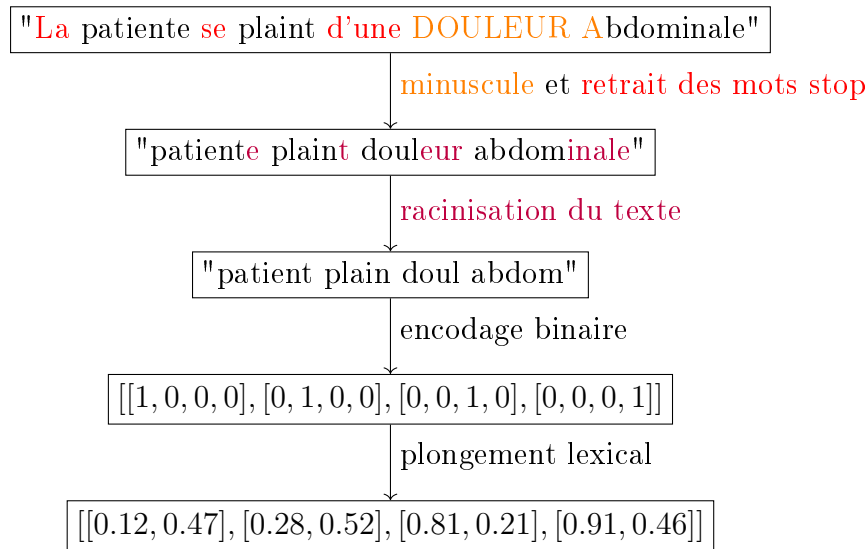


FIGURE 2.2 – Pré-traitement d’un texte médical (*i.e.*, données non structurées). Les valeurs des vecteurs après le plongement lexical ne représentent pas réellement la phrase en question et sont données uniquement à titre de représentation conceptuelle[23].

Pour vérifier si un échantillon est représentatif d’une population plus large, il convient de réaliser des tests statistiques. L’hypothèse principale, dite “hypothèse nulle” (H_0) posée est “il n’existe pas de différence significative entre l’échantillon et la population cible”. L’hypothèse alternative (H_1) est “il existe une différence significative entre l’échantillon et la population cible”. Le chercheur consent à deux risques :

- **Le risque de premier espèce (risque α)**, défini comme la probabilité de se tromper en affirmant qu’il existe une différence alors qu’il n’y en a pas. Il est classiquement fixé à $\alpha = 0.05$ en médecine.
- **Le risque de second espèce (risque β)**, défini comme la probabilité de se tromper en affirmant qu’il n’existe pas de différence alors qu’elle existe. Il est classiquement fixé à $\beta = 0.20$ en médecine.

Pour chaque variable du jeu de données, il conviendra de tester cette hypothèse :

- Dans le cas d’une variable numérique : par un test de Student[55] (si les données suivent une loi normale), un test de Mann-Whitney[56] ou une étude des corrélations.

Exemple – Variable AGE

Les hypothèses testées quand à la distribution de l’âge dans les deux séries de patients :

Hypothèse nulle H_0 : la moyenne de l’âge des patients hospitalisés est identique à la moyenne de l’âge des patients sortants

Hypothèse alternative H_1 : la moyenne de l’âge des patients hospitalisés est significativement différente de la moyenne de l’âge des patients sortants

Test de Student : $\tau = 211$ soit une $p < 0.001$ d’après la table[55]

Interprétation : Nous rejetons l’hypothèse nulle

Conclusion : La moyenne d’âge des patients hospitalisés ($\mu_{hospitalise} = 62.40$) est significativement différente de la moyenne d’âge des patients sortants ($\mu_{sortant} = 45.0$)

- Dans le cas d’une variable catégorielle : par un test du χ^2 [57] (si effectifs attendus sont supérieurs à 5), un test exact de Fisher[58] ou un test de Mac Nemar[59].

Exemple – Variable *SEXE*

Les hypothèses testées quand à la distribution du genre dans les deux séries de patients :

Hypothèse nulle H_0 : la distribution de la variable sexe des patients hospitalisés est identique à la distribution de la variable sexe des patients sortants

Hypothèse alternative H_1 : la distribution de la variable sexe des patients hospitalisés est significativement différente à la distribution de la variable sexe des patients sortants

Test de χ^2 : $\chi^2 = 44$ soit une $p < 0.001$ d'après la table[57]

Interprétation : Nous rejetons l'hypothèse nulle

Conclusion : La distribution de la variable sexe diffère parmi les patients hospitalisés et sortants

2.1.6.2 Complétude

Pour la question de la complétude, nous renvoyons le lecteur à des travaux de référence de la littérature[60-62]. Nous tenterons de synthétiser en disant que la probabilité d'observer des données manquantes dépend de la partie observée du jeu de données (Y_{obs}), des données manquantes du jeu de données (Y_{miss}), du mécanisme à l'origine des données manquantes (formellement décrit par le paramètre ψ), et le fait que les données soient observées ($R = 1$) ou non ($R = 0$) :

$$Pr(R = 0 | Y_{obs}, Y_{miss}, \psi) \quad (2.6)$$

En utilisant le paramètre θ pour représenter le modèle formel aboutissant au jeu de données, la probabilité que les données soient observées peut être représentée par la fonction de densité :

$$f(Y_{obs}, R | \theta, \psi) \quad (2.7)$$

Les données manquantes peuvent être classées en trois catégories : **Complètement aléatoirement manquantes**, "*Missing Completely At Random*" (MCAR), **Aléatoirement manquantes**, "*Missing At Random*" (MAR), et **Manquantes non aléatoirement**, "*Missing Not At Random*" (MNAR)[62]. Dans le cas des données MCAR, la probabilité qu'une donnée soit manquante est indépendante du caractère observé de la donnée mais uniquement du mécanisme aboutissant à l'absence de la donnée. Ainsi dans ce cas précis, la probabilité qu'une donnée soit manquante peut être simplifiée comme ci-dessous :

$$Pr(R = 0 | Y_{obs}, Y_{miss}, \psi) = Pr(R = 0 | \psi) \quad (2.8)$$

Dans le cas des données MAR, la probabilité qu'une donnée soit manquante dépend de son caractère observable et du mécanisme aboutissant à l'absence de la donnée. Ainsi dans ce cas précis, la probabilité qu'une donnée soit manquante peut-être simplifiée comme ci-dessous :

$$Pr(R = 0 | Y_{obs}, Y_{miss}, \psi) = Pr(R = 0 | Y_{obs}, \psi) \quad (2.9)$$

Enfin dans le cas de données MNAR, la probabilité qu'une donnée soit manquante dépend de toutes les

données ainsi que du mécanisme aboutissant à l'absence de la données, et ne peut donc être simplifiée.

Ces trois types de données manquantes peuvent être trouvés dans les jeux de données de médecine d'urgence. Les données **MCAR** peuvent être collectées mais dont la saisie peut être aléatoirement omise (*e.g.*, lorsque l'infirmière d'accueil oublie de saisir une donnée). **Les données MAR peuvent ne pas être collectées car elles ne sont pas nécessaires ou la collecte n'est pas éthique au regard du motif de consultation aux urgences** (*e.g.*, la mesure de la glycémie capillaire n'est pas pertinente pour la prise en charge d'un jeune patient sans antécédents se présentant aux urgences pour la prise en charge d'un traumatisme mineur de la cheville). Enfin, les données **MNAR** peuvent ne pas être collectées car leur mesure est consommatrice de temps (*e.g.*, la mesure de la fréquence respiratoire).



Points clés

Les données peuvent être volontairement manquantes en médecine d'urgence. Cette spécificité est à prendre en considération dès la conception d'un outil qui pourrait être mis en œuvre dans une SU.

2.2 Apprentissage machine

L'apprentissage machine consiste en l'entraînement d'un modèle en utilisant des données. Les apprentissages peuvent être supervisés ou non supervisés, pour réaliser une classification ou une régression. Pour guider l'apprentissage, l'algorithme va chercher à minimiser une fonction de coût.

Régression. La régression consiste à déterminer une valeur numérique.

Exemple – Régression

Déterminer la durée probable de séjour d'un patient

Classification. La classification consiste à déterminer l'appartenance d'un individu à une catégorie.

Exemple – Classification

Déterminer l'appartenance probable d'un patient au groupe *Hospitalisation = Vrai* ou au groupe *Hospitalisation = Faux*

Seuil décisionnel. Dans le cas d'une tâche de classification, la sortie du modèle est un ensemble de probabilité (*i.e.* une par classe) où 0 correspond à la probabilité nulle qu'a l'individu d'appartenir à la classe et 1 correspond à la probabilité certaine que l'individu a d'appartenir à la classe. Or le classifieur doit décider à quelle classe appartient l'individu. Il est donc nécessaire de définir un seuil décisionnel au-delà duquel une probabilité permettra de placer l'individu dans la classe (*i.e.*, prédiction positive; \widehat{Y}_+) et en-deça duquel une probabilité permettra de ne pas placer l'individu dans la classe (*i.e.*, prédiction négative; \widehat{Y}_-).

2.2.1 Mesure de performance

La mesure de la performance des modèles s'appuie sur quatre catégories dont la présentation peut se faire dans un tableau de contingence (tableau 2.2) :

- Les **Vrais Positifs (VPs)** : classés à raison comme positif par le modèle,
- Les **Vrais Négatifs (VNs)** : classés à raison comme négatif,
- Les **Faux Positifs (FPs)** : classés à tort comme positif,
- Les **Faux Négatifs (FNs)** : classés à tort comme négatif.

Prédiction \ Réalité	Réellement positif (Y_+)	Réellement négatif (Y_-)
Prédit positif (\widehat{Y}_+)	VP	FP
Prédit négatif (\widehat{Y}_-)	FN	VN

TABLE 2.2 – Tableau de contingence.

2.2.1.1 Exactitude

L'exactitude correspond à la probabilité d'avoir une bonne prédiction par le modèle en général (équation 2.10).

$$Exactitude = \frac{VP + VN}{VP + VN + FP + FN} \quad (2.10)$$

2.2.1.2 Mesure de la validité intrinsèque de la performance

La sensibilité et la spécificité sont deux mesures de la performance d'un test qui sont indépendantes de la répartition des individus réellement vrais et réellement faux (*e.g.*, ne dépend pas de la prévalence d'une maladie).

La sensibilité est la probabilité de bien détecter les individus réellement positifs (*i.e.*, la probabilité conditionnelle de classer un individu positif sachant qu'il est réellement positif; [équation 2.11](#)). Un modèle est dit sensible s'il est capable de ne pas rater des cas qu'il doit correctement classer. Si on veut augmenter la sensibilité, il faut baisser le seuil décisionnel pour augmenter le nombre d'individus classés positifs (parmi lesquels il existe des VP). Cependant, plus le seuil décisionnel baisse, plus le nombre de FN augmente. La sensibilité est souvent notée *Se*. Elle peut être aussi appelée *Rappel* notamment en anglais ("*Recall*").

$$\text{Sensibilit} = Se = P(\widehat{Y}_+ | Y_+) = \frac{VP}{VP + FN} \quad (2.11)$$

La spécificité est la probabilité de bien détecter les individus réellement négatifs (*i.e.*, la probabilité conditionnelle de classer un individu négatif sachant qu'il est réellement négatif; [équation 2.12](#)). Un modèle est dit spécifique s'il est capable de correctement discriminer les éléments négatifs. Si on veut augmenter la spécificité, il faut augmenter le seuil décisionnel pour augmenter le nombre d'individus classés négatifs (parmi lesquels il existe des VN). Cependant, plus le seuil décisionnel augmente, plus le nombre de FN augmente. La spécificité est souvent notée *Sp*.

$$\text{Specificit} = Sp = P(\widehat{Y}_- | Y_-) = \frac{VN}{VN + FP} \quad (2.12)$$



Exemple – Seuil décisionnel et impact sur la performance

Seuil décision = 0. Les probabilités prédites sont toutes supérieures ou égales à 0. Tous les individus sont classés positifs dès que la probabilité est différente de 0. Tous les vrais positifs sont détectés, mais aucun négatif n'est détecté; on dit que le **classifieur est très sensible**.

Seuil décision = 1. Les probabilités prédites sont toutes inférieures ou égales à 1. Tous les individus sont classés négatifs dès que la probabilité est différente de 1. Tous les vrais négatifs sont détectés, mais aucun positif n'est détecté; on dit que le **classifieur est très spécifique**.

2.2.1.3 Mesure de la validité extrinsèque de la performance

Les valeurs prédictives positives et négatives correspondent à la probabilité que le modèle ait raison quand il prend une décision. Cette mesure de la performance dépend de la distribution des cas dans la population d'individus.

La Valeur Prédictive Positive (VPP) est la probabilité que l'individu soit réellement positif lorsqu'il est classé positif (*i.e.*, la probabilité conditionnelle d'être réellement positif sachant que l'individu a été classé positif; [équation 2.13](#)).

$$VPP = P(Y_+ | \widehat{Y}_+) = \frac{VP}{VP + FP} \quad (2.13)$$

La Valeur Prédictive Négative (VPN) est la probabilité que l'individu soit réellement négatif lorsqu'il est classé négatif (*i.e.*, la probabilité conditionnelle d'être réellement négatif sachant que l'individu a été classé négatif; [équation 2.14](#)).

$$VPN = P(Y_- | \widehat{Y}_-) = \frac{VN}{VN + FN} \quad (2.14)$$

2.2.1.4 “Receiver operating characteristic” (ROC)

Nous avons vu précédemment que dans le cas d’un classifieur, un seuil décisionnel devait être choisi pour classer les individus à partir de la probabilité calculée. En réalité, **il existe une infinité de seuils** entre 0 (*i.e.*, tous les cas classés positifs, sensibilité à 1, spécificité médiocre) et 1 (*i.e.*, tous les cas classés négatifs, spécificité à 1, sensibilité médiocre). Il est possible de représenter cette infinité de seuils en reportant le taux de VPs en ordonné et le taux de VNs en abscisse (figure 2.3). La diagonale représente le classifieur aléatoire où quel que soit le seuil de décision, la probabilité de classer un patient positif est aléatoire. Le point le plus en haut à gauche correspond au meilleur compromis de sensibilité et de spécificité (*Meilleur point* du graphique de la figure 2.3). Cette courbe porte le nom de courbe ROC. L’Aire sous la courbe ROC, “*Area Under the ROC curve*” (AUROC) permet de **mesurer la performance du modèle indépendamment du choix du seuil décisionnel** et permet de comparer les modèles entre eux.

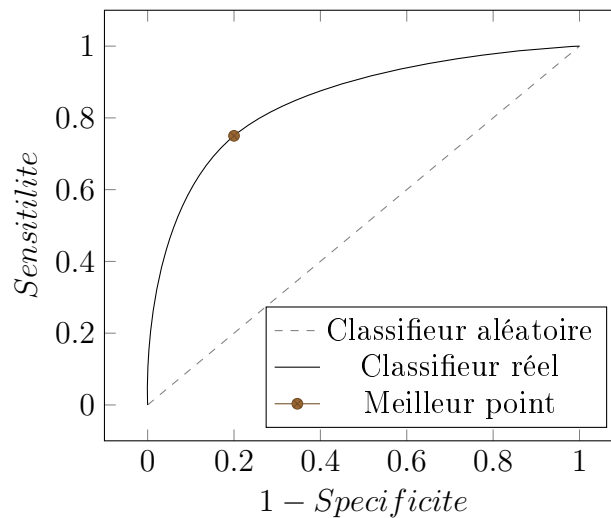


FIGURE 2.3 – Courbe ROC

2.2.1.5 Choix du meilleur seuil

Le choix du meilleur seuil doit se faire en lien avec le contexte et l’objectif dans lesquels le modèle est entraîné. Ainsi, certains modèles chercheront à être très sensibles (*e.g.*, lorsque l’important est de ne pas rater les positifs, au prix de plus de FPs) et d’autres très spécifiques (*e.g.*, lorsque l’important est d’être discriminant et d’être certain des positifs détectés au prix de ne pas tous les détecter).

A titre d’exemple, nous voudrions utiliser un modèle sensible pour détecter les fractures. En effet, nous souhaiterions que toutes les suspicions de fractures, mêmes lorsque le modèle a un doute, soient identifiées. Le médecin sera alors responsable de déterminer la véracité de la fracture, éventuellement en prenant avis auprès du radiologue. Le médecin doit pouvoir s’appuyer sur ce modèle pour être certain qu’un patient non détecté a un risque quasi nul de d’avoir une fracture.

A l’inverse, nous voudrions utiliser un modèle spécifique pour l’anticipation des hospitalisations depuis l’accueil. Dans ce cas, nous souhaiterions n’hospitaliser rapidement que les patients dont la probabilité d’hospitalisation est très élevée. Paramétré de la sorte, ce modèle ne pourra être valable que pour aider à l’hospitalisation des patients évidents et non pour aider à la sortie des patients.

Dans le cas *standard* où l’objectif est d’être aussi performant dans la détection des positifs que dans la discrimination des positifs, le choix du meilleur seuil correspondant au meilleur rapport sensibilité / spécificité peut être identifié en utilisant plusieurs approches : (i) maximiser le score de YOUDEN [63] (équation 2.15), et (ii) maximiser le F1-score[64] (équation 2.16).

$$J = Se + Sp - 1 \quad \text{Score de Youden} \quad (2.15)$$

$$F_1 = 2 \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad \text{F1 score} \quad (2.16)$$

2.2.2 Apprentissage supervisé

2.2.2.1 Principe

Dans le cas des apprentissages supervisés, le modèle est entraîné à répondre à une question. Plus exactement, il est entraîné à prédire une variable de réponse en fonction de variables d'entrée. Dans ce type d'apprentissage, il est nécessaire de posséder de nombreux exemples de variables d'entrées (X_i) avec la variable de réponse réellement constatée (y_i). Le modèle va apprendre à répondre sur la base de ces exemples en produisant une prédiction (\hat{y}_i). Initialement, l'apprentissage est souvent aléatoire, et la distance entre la prédiction du modèle (initialement aléatoire) et la variable de réponse attendue (celle constatée dans l'exemple) est déterminée par une fonction de coût (L ; *c.f.*, [sous-section 2.2.4](#)). L'apprentissage consistera à minimiser cette fonction de coût. L'exposition du jeu de données d'apprentissage au modèle peut être répétée plus d'une fois en bruitant les données d'entrées, on dira alors qu'elle est répétée plusieurs *epochs*

2.2.2.2 Regression Logitique (RL)

En théorie d'après KLEINBAUM et KLEIN [65], la **RL** est un modèle mathématique qui peut être utilisé pour **décrire la corrélation entre plusieurs variables X_n et une variable dépendante et dichotomique y** . La fonction logistique ([équation 2.18](#)) est une fonction comprise dans l'intervalle $[0; 1]$ pour des valeurs de $z \in [-\infty; +\infty]$ ([figure 2.4](#)). Le modèle mathématique z peut se formaliser de la forme de [l'équation 2.17](#). La **RL** cherche à déterminer la probabilité d'obtenir $Y = 1$ sachant les données d'entrées X_1, \dots, X_n en trouvant les $\alpha, \beta_1, \dots, \beta_n$ correspondant ([équation 2.19](#)).

$$z = \alpha + \beta_1 X_1 + \dots + \beta_n X_n \quad \text{Modèle } z \text{ réel} \quad (2.17)$$

$$f(z) = \frac{1}{1 + e^{-z}} \quad \text{Fonction logistique} \quad (2.18)$$

$$P(Y = 1 | X_1, \dots, X_n) = \frac{1}{1 + e^{-(\alpha + \sum_{i=1}^n \beta_i X_i)}} \quad \text{Probabilité d'avoir } Y = 1 \quad (2.19)$$

$$\text{avec } \alpha, \beta_1, \dots, \beta_n \text{ à déterminer} \quad (2.20)$$

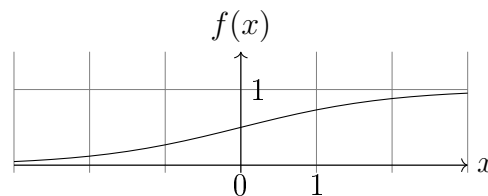


FIGURE 2.4 – Fonction logistique représentant la probabilité $P(Y = 1 | X_1, \dots, X_n)$ ([équation 2.19](#)).

En médecine d'urgence, la **RL** est utilisée pour des problèmes de prédiction de mortalité ou d'évènement clinique.

FAISAL *et al.* [66] compare la **RL** aux autres techniques d'apprentissage machine sur une cohorte de 38 173 patients. L'**AUROC** est de 85.7%, similaire aux autres méthodes pour une rapidité et une simplicité d'entraînement.

RAITA *et al.* [67] cherche à prédire l'hospitalisation d'un patient en secteur conventionnel ou en soins critiques à partir des données d'accueil aux urgences à partir d'une cohorte rétrospective de 135 470 patients. Les performances de leur modèle sont très acceptables : l'**AUROC** est de 84.0% pour la prédiction d'admission en soins critiques et de 81% pour l'admission en secteur conventionnel.

2.2.2.3 Réseau de Neurons Profonds (RNP) ou perceptron multicouches

En théorie d'après BISHOP [46] et WIKISTAT [68], un **RNP** est une suite de neurones organisés en couches (figure 2.5). Chaque neurone est composé par les entrées x_i , un biais propre au neurone b et une fonction d'activation f (figure 2.6). Ainsi, le modèle mathématique du neurone formel est une fonction qui retourne une sortie y comme étant la somme pondérée des p entrées $x_i w_i$ ajoutée au biais b du neurone passée ensuite par une fonction d'activation act (équation 2.21).

$$y = act\left(\sum_{i=0}^p (w_i x_i)\right) + b \quad (2.21)$$

Les fonction d'activation souvent retrouvées sont :

- Fonction *identité* : $act(x) = x$
- Fonction *ReLU* : $act(x) = \max\{0, x\}$ (équation 2.41)
- Fonction *sigmoïde* (équation 2.46)
- Fonction *softmax* (équation 2.50)

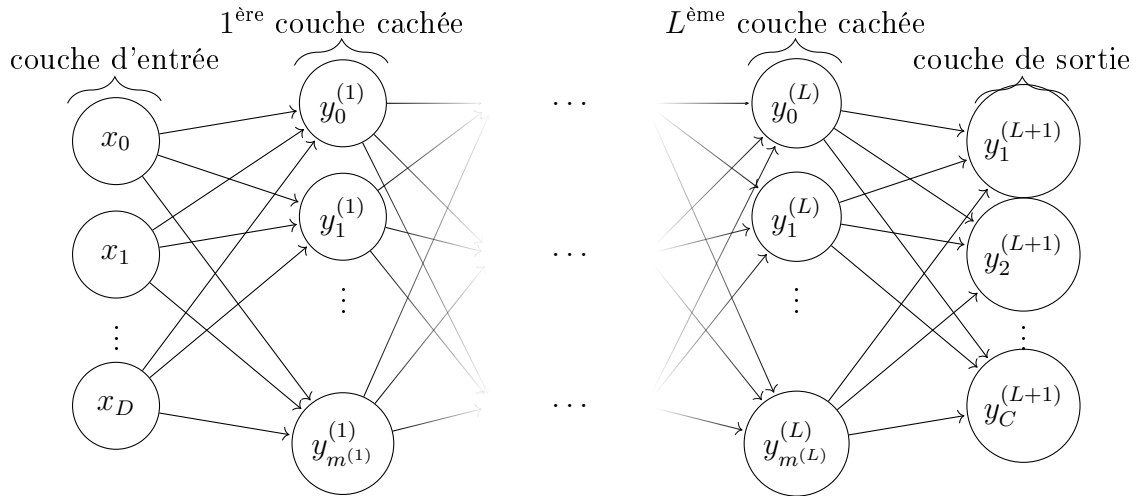


FIGURE 2.5 – Réseau de neurones[69]

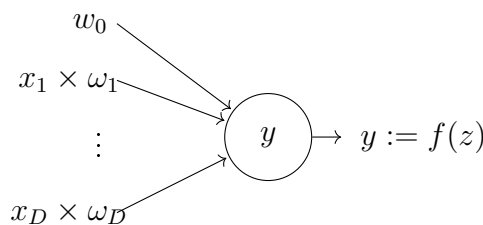


FIGURE 2.6 – Neurone fonctionnel[69]

Un **RNP** peut donc être modélisé par la fonction f (équation 2.26) de la manière suivante pour p variables explicables (X_1, \dots, X_p) d'une variable à expliquer Y en utilisant les poids W et les biais B . W est le vecteur

des poids w_{jkl} pour la $j^{\text{ième}}$ entrée du $k^{\text{ième}}$ neurone de la $l^{\text{ième}}$ couche, et B est le vecteur des biais b_{kl} pour le $k^{\text{ième}}$ neurone de la $l^{\text{ième}}$ couche.

$$\hat{Y}_{l_1} = \text{act}_{l_1}(\sum (XW_{l_1}) + B_{l_1}) \tag{2.22}$$

$$\hat{Y}_{l_2} = \text{act}_{l_2}(\sum (Y_{l_1} W_{l_2} + B_{l_2})) \quad \text{donc} \tag{2.23}$$

$$\hat{Y}_l = \text{act}_l(\sum (Y_{l-1} W_l + B_l)) \quad \text{Modèle de la couche } l \tag{2.24}$$

$$\hat{Y}_{l,k} = \text{act}_l(\sum_{i=1}^{K_{l-1}} (w_{l,i} y_{l-1,i}) + b_{l,k}) \quad \text{Modèle pour un neurone } k \text{ de la couche } l \tag{2.25}$$

$$\hat{Y} = f(X_1, \dots, X_p; W, B) \quad \text{Modèle global} \tag{2.26}$$

Les poids et les biais sont initialisés aléatoirement puis mis à jour en rétro-propageant l'erreur entre la sortie prédite et la sortie réelle attendue afin de minimiser cette erreur. Prenons par exemple la fonction de coût *Loss* de l'erreur au carré (équation 2.27, plus de détail à section 2.2.4.2) :

$$\text{Loss}(W, B) = \sum (Y - \hat{Y})^2 \tag{2.27}$$

$$= \sum (Y_i - f(X_i; W_i, B_i))^2 \tag{2.28}$$

$$\tag{2.29}$$

Pour réaliser cette opération, l'algorithme d'optimisation utilise la descente de gradient à partir de la fonction de coût (*Loss*), appliquée aux N exemples servant à l'entraînement. Cette fonction doit donc être dérivable : partiellement par rapport aux poids (équation 2.30) puis partiellement par rapport aux biais (équation 2.31).

$$\frac{\partial L}{\partial W} = 2 \sum (Y - f(X; W, B)) \frac{\partial f(X; W, B)}{\partial W} \tag{2.30}$$

$$\frac{\partial L}{\partial B} = 2 \sum (Y - f(X; W, B)) \frac{\partial f(X; W, B)}{\partial B} \tag{2.31}$$

La rétro-propagation de l'erreur (δ_i) s'effectue en calculant la descente de gradient pour les poids d'une part, puis pour les biais d'autres part. Pour un poids w_i , la descente de gradient par la fonction $Loss_i$ avec un pas r est exprimée par l'équation 2.32. Puisqu'un réseau de neurones est caractérisé par un vecteur de poids W et non un poids unique, la descente de gradient s'exprime par l'équation 2.33.

$$w_{i,t+1} = w_{i,t} - r \frac{d\text{Loss}}{dw_i}(w_{i,t}) \quad t \text{ le temps, } t + 1 \text{ le pas suivant} \tag{2.32}$$

$$W_{t+1} = W_t - r \nabla \text{Loss}(W_t) \tag{2.33}$$

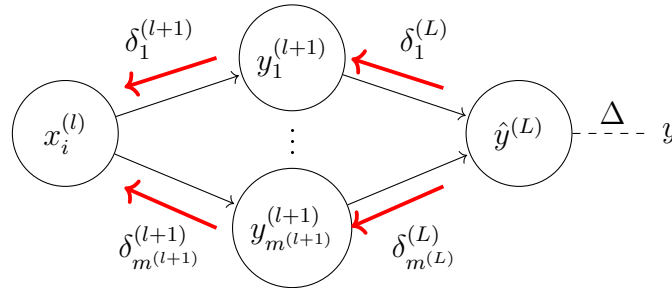


FIGURE 2.7 – Rétropropagation de l'erreur lors de l'apprentissage d'un réseau de neurones[69].

Ce mécanisme boucle jusqu'à l'épuisement des données d'entraînement et/ou jusqu'à une limite posée définie dans l'expérimentation. Plus le nombre de couches cachées est élevé, plus l'apprentissage devient *profond* ("deep

learning”) et permet de réaliser des tâches complexes. En revanche, plus le nombre de couches augmente, plus le nombre de paramètres à calculer et prendre en compte augmente, et donc plus l’entraînement est consommateur de ressource.

En médecine d’urgence, les **RNP** ont été testés entre autre dans la prédiction du parcours de soin[70] : FAISAL *et al.* [66] utilise les **RNP** (une couche cachée de trois neurones) pour prédire la mortalité hospitalière des patients admis aux urgences (**AUROC** 84%), ou encore RAITA *et al.* [67] utilise les **RNP** (six couches cachées) pour prédire l’hospitalisation dans plusieurs hôpitaux américains (**AUROC** 82%).

2.2.2.4 Réseau de Neurones Convolutifs (RNC)

Limite des perceptrons multicouches. Bien qu’efficace sur les entrées de tailles importantes, le **RNP** n’est pas sans poser quelques problèmes de temps de calcul lors de l’utilisation de texte ou d’image en entrée. En effet, une image 32×32 pixels avec trois niveaux de couleurs et un niveau de saturation correspond à une couche à 4 096 entrées, et donc une première couche cachée où chaque neurone serait connecté à ces 4 096 entrées.

Principe du RNC. Pour avoir une vision détaillée et complète des **RNCs**, nous renvoyons le lecteur à des travaux de références : LECUN et BENGIO [71] et LI *et al.* [72]. Pour résumer simplement, un **RNC** est un empilement de bloc de :

1. **couche de convolution** qui fonctionne comme un récepteur d’un sous-ensemble des données de la couche supérieure qui va se déplacer suivant un certain pas (*i.e.*, chevauchement des champs de récepteurs),
2. **couche de pooling** qui synthétisera les données de la couche de convolution en une information compressée suivant une fonction mathématique (*e.g.*, maximum, moyenne, ...),
3. **couche d’activation**, qui joue le même rôle que la fonction d’activation d’un **RNP**,

Le volume de sortie V_o d’un bloc de ces trois couches présentant un volume d’entrée V_i , une surface de traitement S (*i.e.*, nombre de champs récepteurs), un pas P et une marge M correspond à :

$$V_o = \frac{V_i - S + 2M}{P} + 1 \quad (2.34)$$

Après avoir permis la **réduction du volume d’entrées tout en gardant une bonne résolution spatiale** (*i.e.*, synthèse des neurones de proximité), les blocs de neurones convolutifs servent d’entrée à un **RNP** classique.

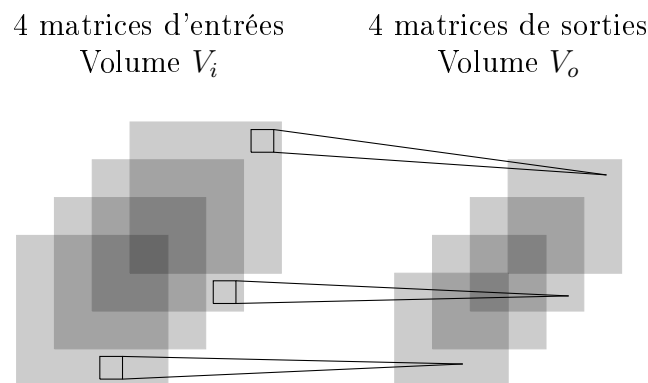
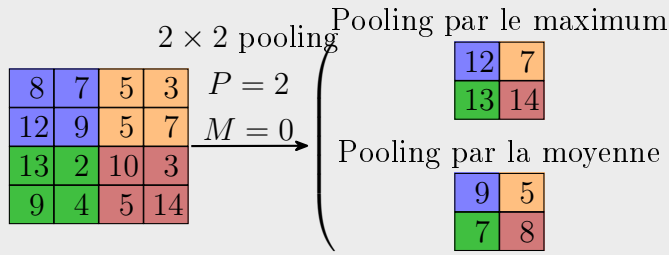


FIGURE 2.8 – Réseau de neurones convolutifs[69]

Exemple – Couche de neurones convolutifs



2.2.2.5 Arbres décisionnels

En théorie, les algorithmes d'apprentissage machine basés sur les arbres décisionnels consistent en la **segmentation répétée du jeu de données** sur une variable d'entrée en utilisant un test logique (embranchement) jusqu'à aboutir à une feuille (réponse du modèle) (figure 2.9).

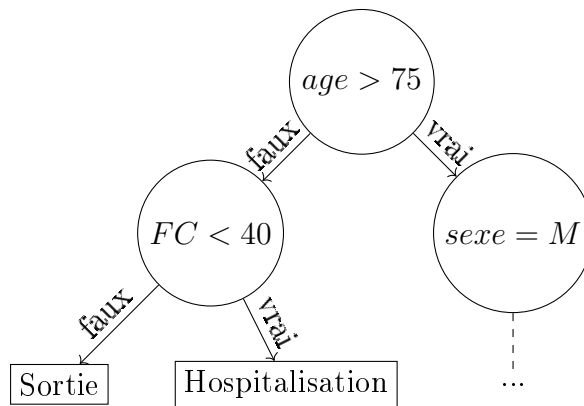


FIGURE 2.9 – Représentation logique d'un arbre décisionnel. Les cercles représentant les embranchements et les rectangles les feuilles.

Il existe plusieurs approches pour choisir comment diviser le jeu de données et quand s'arrêter. Nous prendrons l'exemple de la méthode "*Classification And Regression Trees*" (CART) décrite par BREIMAN *et al.* [45] : chaque embranchement ne doit considérer qu'une seule variable et doit séparer au mieux les classes de réponse. Pour séparer au mieux les classes de réponse, l'indice Gini I est utilisé (équation 2.35) pour le nœud h , la variable v testée, la proportion $P(v = i)$ de la classe i pour les N classes possibles. Plus l'indice Gini I est petit, plus l'une des classes prédomine : meilleur est le nœud.

$$I_h = 1 - \left(\sum_{i=0}^N P(v = i)^2 \right) \quad (2.35)$$

Le risque de la méthode CART est de mal choisir les variables de segmentation, conduisant à la création d'un mauvais classifieur. Afin d'améliorer ces techniques, BREIMAN [73] a décrit la méthode des Forêts Aléatoires (FA) ("*random forest*"). Elle consiste à générer plusieurs arbres décisionnels en faisant varier les données d'entrées en appliquant la technique du :

- tirage avec remise ("*bootstrap*") de la population : sur une population n , on supprime m enregistrements et on répète m enregistrements conservés.
- tirage aléatoire des variables à considérer : seul un sous-ensemble $\text{entier}(\sqrt{P})$ des P variables d'entrées sera considéré par chaque arbre
- vote à la majorité : le classifieur final répond ce que la majorité des arbres de la forêts répond

Des méthodes plus modernes ont fait leur apparition dans le même principe d'application du gradient de descente aux modèles : XGBoost[74] ou encore LightGBM[75].

L'algorithme des FA, proposée par BREIMAN [73] repose sur des arbres décisionnels en appliquant l'algorithme suivant :

```
for  $n$  ensemble apprentissage do  
  Utiliser un échantillon des données d'entrées en appliquant la technique du bootstrap (i.e.,  
  probabilité de rajout de duplicatas de données d'origine);  
  Utiliser un échantillon des prédictors ( $< \sqrt{n_{\text{predicteurs}}}$ );  
  Entraîner un arbre décisionnel;  
end  
Procéder à un vote majoritaire des décisions de tous les arbres de la forêt
```

Algorithme 1: Algorithme des FA

En médecine d'urgence, HONG, HAIMOVICH et TAYLOR [76] prédisent l'admission dans un service et GOTO *et al.* [77] prédit l'admission dans les services de soins critiques, en utilisant XGBoost et les forêts aléatoires à partir de l'historique médical du patient et les données cliniques présentes à l'accueil des urgences.

2.2.3 Apprentissage non supervisé

L'apprentissage non supervisé[78] est à la limite du périmètre de ce travail, ainsi nous nous satisferons d'une introduction simple. L'apprentissage non supervisé consiste à entraîner un modèle sans lui donner de réponse à apprendre : il doit fournir une sortie *intelligente* sans qu'on lui spécifie ce que l'on cherche. Il existe deux grandes tâches réalisables par un algorithme d'IA non supervisé : la détection d'anomalies et le regroupement.

2.2.3.1 La détection d'anomalie

La détection d'anomalie[78, p. 1-32] consiste à modéliser numériquement un jeu de données. Le modèle sera capable d'identifier l'appartenance au jeu de données initial d'une nouvelle donnée présentée.

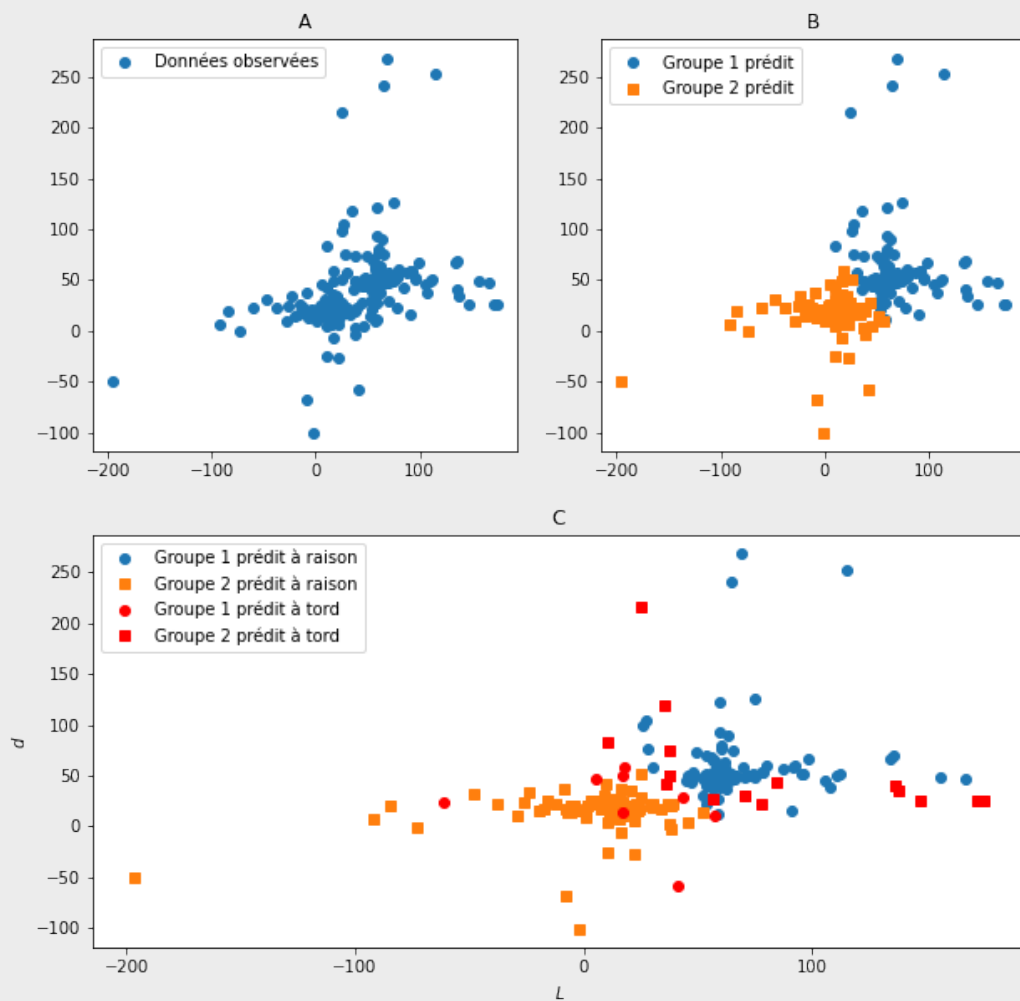
2.2.3.2 Le regroupement

Le regroupement[78, p. 253-302] consiste à regrouper l'ensemble des individus du jeu d'entraînement en fonction de leurs caractéristiques de telle sorte que chaque individu ressemble (*i.e.*, ont des prédictors de valeurs similaires) plus à ceux de son groupe qu'à ceux des autres groupes.

 Exemple – K-Regroupement

Nous prendrons l'exemple du K-regroupement qui consiste à séparer les données en K groupes en minimisant la distance entre chaque point et la moyenne de son groupe.

Dans cet exemple, nous avons généré un jeu de données fictif représentant le diamètre d'inhibition d de colonies bactériennes dans une boîte de Pétri en fonction de la distance L du disque antibiotique. Nous l'avons généré en partant de deux moyennes centrées sur deux points fixes avec un bruit simulé suivant une loi Normale. Le bruit simule la variabilité du vivant. Le K-regroupement permet de repérer la plupart des individus. Il persiste quelques erreurs que nous pouvons observer car nous avons généré manuellement le jeu de données. L'identification d'erreur dans la réalité ne pourrait être effectuée qu'à posteriori.



Exemple de K-regroupement, d représentant le diamètre de la colonie et L la distance du disque antibiotique d'une boîte de Pétri. (A) Les données initialement observées. (B) Les données regroupées par le modèle. (C) Les erreurs de classement après vérification.

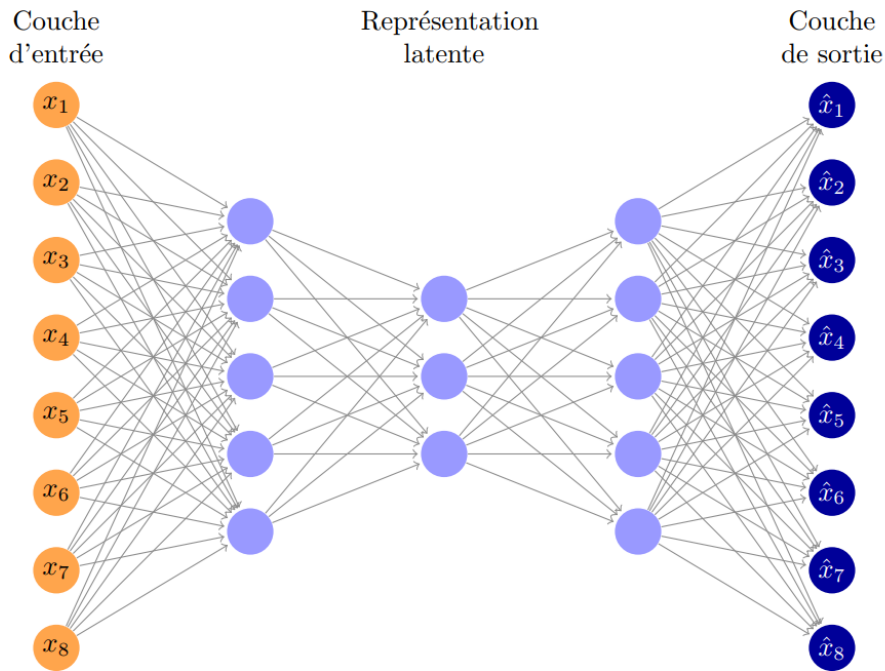


FIGURE 2.10 – Architecture d'un auto-encodeur[81]

2.2.3.3 Les autoencodeurs

Les autoencodeurs sont des **RNP** (ou des **RNC**) qui stockent (*i.e.*, encodent ; équation 2.36) **une représentation abstraite C , compressée et latente d'un jeu de données** (figure 2.10). Cette information peut être décodée (équation 2.37) et retourner une approximation de l'information initiale (\hat{X}). D'après ZHANG *et al.* [79], l'auto-encodeur peut être représenté par deux fonctions :

$$f_e = X \rightarrow C \quad \text{Fonction d'encodage} \quad (2.36)$$

$$f_d = C \rightarrow \hat{X} \quad \text{Fonction de décodage} \quad (2.37)$$

Cette représentation latente peut ensuite être décodée et entraînée pour différentes tâches[80] :

- Stockage d'information à dimension réduite (*i.e.*, résumer, compresser une information ; *e.g.*, segmentation d'image : figure 2.11)
- Reconstruction d'un signal bruité
- Détection d'anomalies
- Génération de nouveaux exemples

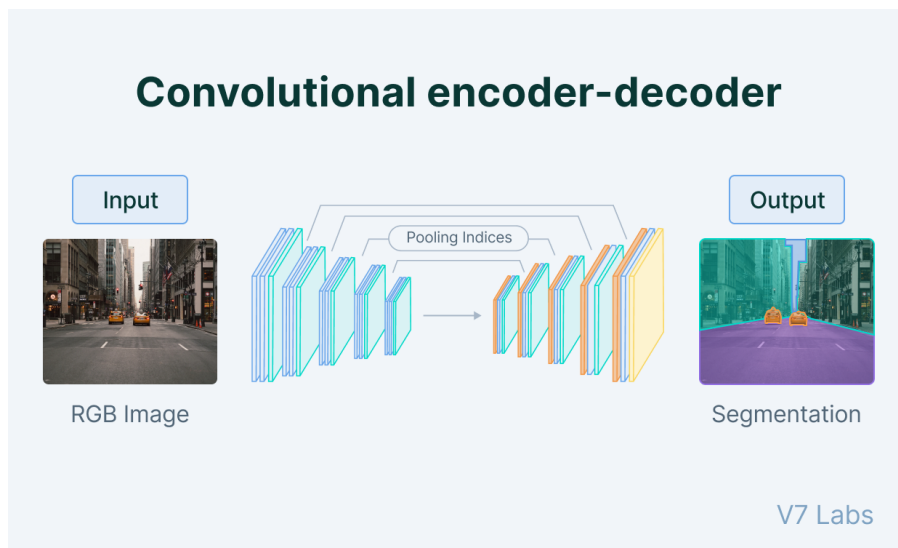


FIGURE 2.11 – Représentation d'un autoencodeur entraîné pour une tâche de segmentation d'image[82].

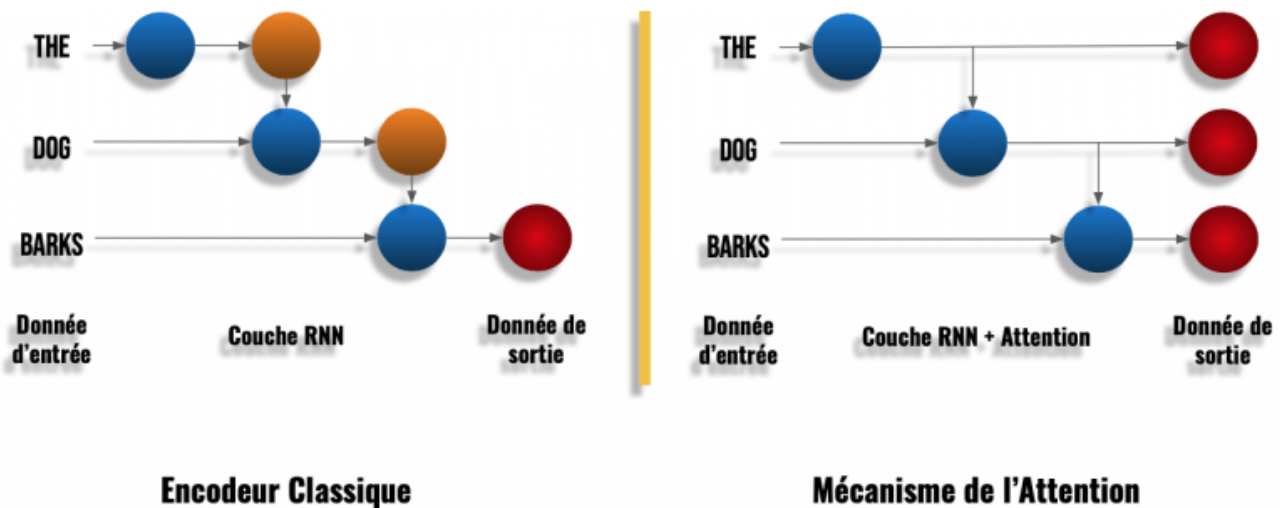


FIGURE 2.12 – Différence entre l'encodeur et le mécanisme d'attention[84]

2.2.3.4 Les transformers

Les transformers sont des modèles d'architecture à base de réseaux de neurones qui évite la récurrence et se basent sur le mécanisme d'attention pour établir une dépendance globale entre l'entrée et la sortie[83]. Le mécanisme de l'attention est une évolution de l'architecture des auto-encodeurs (où l'information A est combinée à l'information suivante B, résultant in fine à une information résumée sur une sortie unique) dans laquelle l'information A est passée comme entrée de l'information B suivante, mais toutes les sorties (A et B) sont utilisées (figure 2.12)[84]. Pour plus de détails, nous renvoyons le lecteur à un article de référence sur le sujet[85].

2.2.4 Fonctions de coût

Une fonction de coût peut s'interpréter comme une *pénalité* imputée au modèle à chaque erreur (ou comme une *réussite* imputée au modèle à chaque bonne prédiction). L'objectif de l'algorithme est de diminuer au maximum la pénalité et donc de minimiser $Loss$. Nous parcourons rapidement les fonctions de coût d'après une analyse de l'article WANG *et al.* [86].

2.2.4.1 Fonctions de coût des problèmes de classification

Dans les problèmes de classification, R peut prendre les valeurs suivantes :

- $R = \{-1, 1\}$ dans le cas d'une classification binaire
- $R = \{1, 2, 3, \dots, K\}$ dans le cas d'une classification à K possibilités de réponse

Fonction de coût binaire. Cette fonction de coût prend la valeur 1 si les valeurs sont différentes (*i.e.*, signes différents) et 0 si les valeurs sont égales (*i.e.*, mêmes signes). Cette fonction ne prend pas en compte la distance entre la prédiction et la valeur réelle, mais simplement l'information binaire de la véracité de la prédiction. Il s'agit d'une fonction non continue.

$$loss_i(y_i, \hat{y}_i) = \begin{cases} 1, & \text{si } y_i \hat{y}_i < 0 \\ 0, & \text{si } y_i \hat{y}_i \geq 0 \end{cases} \quad \text{Fonction de coût de } i \quad (2.38)$$

$$Loss(Y, \hat{Y}) = \sum_{i=1}^N loss_i(y_i, \hat{y}_i) \quad \text{Fonction de coût global} \quad (2.39)$$

Fonction de coût perceptron. Cette fonction de coût ressemble à la fonction 0-1. Si les valeurs sont identiques, le coût est de zéro. En revanche, si les valeurs sont différentes, l'erreur admise est la valeur absolue de la valeur prédite. Dans ce cas, la pénalité infligée au modèle est proportionnelle à la distance entre la valeur prédite et la valeur réelle. Cette fonction est continue et dérivable.

$$loss_i(y_i, \hat{y}_i) = \max\{0, -\hat{y}_i\} \quad \text{Fonction de coût de } i \quad (2.40)$$

$$\max\{a, b\} = \begin{cases} a, & \text{si } a \geq b \\ b, & \text{si } a < b \end{cases} \quad \text{Fonction maximum} \quad (2.41)$$

Fonction de coût logarithmique. Cette fonction de coût prend en compte la distribution des réponses en utilisant la distribution de la probabilité conditionnelle de chacune des réponses. Soit p_i , la probabilité conditionnelle que y_i vaille 1 sachant les entrées X_i pour l'individu i (équation 2.42). Soit $1 - p_i$ la probabilité conditionnelle que y_i ne vaille pas 1 sachant les entrées X_i pour l'individu i (équation 2.43). Nous noterons \tilde{p} la probabilité conditionnelle correspondant à la sortie réelle y_i . La fonction de coût logarithmique l_i d'un individu i correspond à son logarithme, et la fonction de coût logarithmique globale L correspond à la moyenne des logarithmes.

L'interprétation est la suivante : plus la probabilité de i est d'être prédit sur le bon label, plus la probabilité conditionnelle est élevée, plus la fonction de coût diminue.

$$p_i = P(y = 1 | X_i) = \frac{1}{1 + e^{-\hat{y}_i}} \quad \text{Probabilité conditionnelle } p \quad (2.42)$$

$$1 - p_i = P(y = -1 | X_i) = \frac{1}{1 + e^{\hat{y}_i}} \quad \text{Probabilité conditionnelle } 1 - p \quad (2.43)$$

$$loss_i(y_i, \tilde{p}_i) = -\log(\tilde{p}_i) \text{ avec } \tilde{p}_i = \begin{cases} p_i, & \text{si } y = 1 \\ 1 - p_i, & \text{si } y \neq 1 \end{cases} \quad \text{Fonction de coût pour } i \quad (2.44)$$

$$Loss(Y, \hat{Y}) = -\frac{1}{N} \sum_{i=1}^N \log(\tilde{p}) \quad \text{Fonction de coût globale} \quad (2.45)$$

Fonction de coût sigmoïde à entropie croisée. Cette fonction de coût est très similaire à la fonction de coût logarithmique (équation 2.44). Au lieu d'utiliser la probabilité conditionnelle, nous utiliserons la fonction d'activation sigmoïde σ (équation 2.46). L'entropie croisée correspond à la distance entre la valeur prédite activée (équation 2.47) par la fonction sigmoïde et la valeur réelle. L'interprétation est la suivante : plus l'entropie croisée est basse, plus la distribution des valeurs prédites et des valeurs réelles sont proches. Cette fonction de coût permet un apprentissage plus rapide avec les réseaux de neurones[87].

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad \text{Fonction sigmoïde} \quad (2.46)$$

$$p_i = \sigma(\hat{y}_i) \quad \text{Passage dans la fonction } \sigma \quad (2.47)$$

$$loss_i(y_i, \tilde{p}_i) = \text{c.f. équation équation 2.44} \quad \text{Fonction de coût de } i \quad (2.48)$$

$$Loss(Y, \hat{Y}) = \text{c.f. équation équation 2.45} \quad \text{Fonction de coût globale} \quad (2.49)$$

Fonction de coût softmax à entropie croisée. Cette fonction est identique à la fonction de coût sigmoïde à entropie croisée à l'exception de la fonction d'activation : la sigmoïde est substituée par la fonction *softmax*. La fonction softmax permet de transformer un vecteur de nombres réels ($Z = z_1, \dots, z_K$) en un autre vecteur de nombres réels ($\sum_{i=1}^K \text{softmax}(Z)_i = 1$) dont la somme vaut 1. Cette fonction représente la loi des probabilités sur Z .

$$\text{softmax}(Z)_i = \frac{e^{z_i}}{\sum_{k=1}^K e^{z_k}} \quad \text{Fonction softmax} \quad (2.50)$$

$$p_{ij} = \text{softmax}(\hat{Y}_i)_j \quad \text{pour le label } j \text{ de l'individu } i \quad (2.51)$$

$$loss_i(y_i, \tilde{p}_i) = \text{c.f. équation équation 2.44} \quad \text{Fonction de coût de } i \quad (2.52)$$

$$Loss(Y, \hat{Y}) = \text{c.f. équation équation 2.45} \quad \text{Fonction de coût globale} \quad (2.53)$$

2.2.4.2 Fonctions de coût des problèmes de régression

Dans les problèmes de régression, R peut prendre n'importe quelle valeur réelle dans \mathbb{R} .

Fonction de coût de l'erreur au carré. Cette fonction de coût correspond à la somme des différences entre la valeur prédite et la valeur réelle au carré. Cette fonction est la plus utilisée. Elle est efficace mais très sensible aux extrêmes : plus une erreur est grande, plus son poids dans l'erreur globale est importante. Ainsi, le modèle aura tendance à porter plus d'attention aux erreurs des extrêmes (équation 2.54) car la correction sera plus importante sur la fonction de coût globale (équation 2.55), au prix potentiellement d'une diminution de la précision globale.

$$loss_i(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2 \quad \text{Erreur individuelle de } i \quad (2.54)$$

$$Loss(Y, \hat{Y}) = \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad \text{Erreur globale} \quad (2.55)$$

Fonction de coût de l'erreur absolue. Cette fonction correspond à la simple somme des différences entre la valeur prédite et la valeur réelle. Le fait que cette différence ne soit pas élevée au carré rend cette fonction moins sensible aux extrêmes. Elle devrait être préférée à la fonction de coût d'erreur au carré lorsqu'il existe des extrêmes.

$$loss_i(y_i, \hat{y}_i) = |y_i - \hat{y}_i| \quad \text{Erreur individuelle de } i \quad (2.56)$$

$$Loss(Y, \hat{Y}) = \sum_{i=1}^N |y_i - \hat{y}_i| \quad \text{Erreur globale} \quad (2.57)$$

Fonction de coût de Huber. Cette fonction, proposée par HUBER [88], tente d'exploiter les avantages des deux fonctions précédentes en définissant un paramètre δ strictement positif comme frontière. Les individus i à l'intérieur de la frontière δ utiliseront la fonction de coût du carré des erreurs alors que les individus i au-delà de celle-ci utiliseront la fonction de coût de l'erreur absolue. L'inconvénient de cette fonction de coût est de correctement déterminer δ , et est donc surtout utilisée en apprentissage profond.

$$loss_i(y_i, \hat{y}_i)_\delta = \begin{cases} \frac{1}{2}(y_i - \hat{y}_i)^2 & \text{si } |y_i - \hat{y}_i| \leq \delta \\ \delta|y_i - \hat{y}_i| - \frac{1}{2}\delta^2 & \text{sinon} \end{cases} \quad \text{Fonction de coût individuelle pour } \delta \text{ donné} \quad (2.58)$$

$$Loss(Y_i, \hat{Y}_i)_\delta = \sum_{i=1}^N loss_i(y_i, \hat{y}_i)_\delta \quad \text{Fonction de coût globale pour } \delta \text{ donné} \quad (2.59)$$

Fonction de coût du logarithme du cosinus hyperbolique. Cette fonction utilise le logarithme du cosinus hyperbolique (équation 2.60). Elle a des propriétés similaires à la fonction de coût de Huber mais n'est pas conditionnée à un quelconque paramètre et est directement dérivable au second ordre.

$$\cosh(x) = \frac{e^x + e^{-x}}{2} \quad \text{Fonction cosinus hyperbolique} \quad (2.60)$$

$$loss_i(y_i, \hat{y}_i) = \log(\cosh(\hat{y}_i - y_i)) \quad \text{Fonction de coût individuelle} \quad (2.61)$$

$$Loss(Y, \hat{Y}) = \sum_{i=1}^N loss_i(y_i, \hat{y}_i) \quad \text{Fonction de coût globale} \quad (2.62)$$

2.3 Évaluation du modèle

Lorsqu'un modèle est entraîné, il est indispensable de mesurer sa performance, pendant l'apprentissage (pour guider l'évolution des paramètres au fil des itérations) et en fin d'apprentissage (pour informer l'utilisateur sur le *niveau de confiance* qu'il peut accorder à ce modèle).

2.3.1 Performance de l'apprentissage

Les mesures de performances sont utilisées pendant l'entraînement à l'aide des données d'entraînement. Elles peuvent néanmoins conduire à du sur-apprentissage[89]. Le sur-apprentissage est la situation dans laquelle se trouve un modèle lorsqu'il devient trop spécifique de la population apprise. Il ne sera alors plus performant sur des populations similaires. C'est pourquoi il est indispensable de mesurer les performances du modèle avec des données n'ayant pas servi à l'entraînement (*i.e.*, jeu de test) afin d'éviter le sur-apprentissage.

2.3.2 Évaluation par jeu de test

Le jeu de test est souvent une fraction du jeu de données initial tirée aléatoirement (figure 2.13). Seuls les 80% des données sont utilisées pour l'entraînement. Les 20% restants sont utilisés pour mesurer la performance du modèle. Le ratio 80%/20% est un ratio communément utilisé dans la littérature[46].

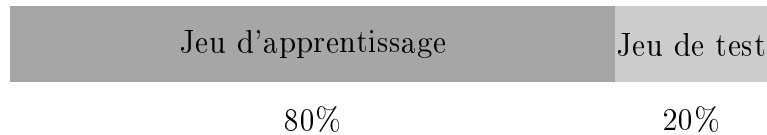


FIGURE 2.13 – Jeu de test et jeu d'apprentissage

2.3.3 Évaluation par validation croisée

Cette méthode d'évaluation consiste à segmenter le jeu de données en n parties égales, d'utiliser $n - 1$ parties pour l'entraînement et une partie pour le test (figure 2.14). L'entraînement est répété n fois et la mesure de performances est réalisée n fois. Cette segmentation est réalisée de manière randomisée, souvent stratifiée sur la variable de réponse. Cette validation croisée permet de vérifier la reproductibilité des résultats en prenant à chaque fois une autre partie du jeu de données pour évaluer le modèle.

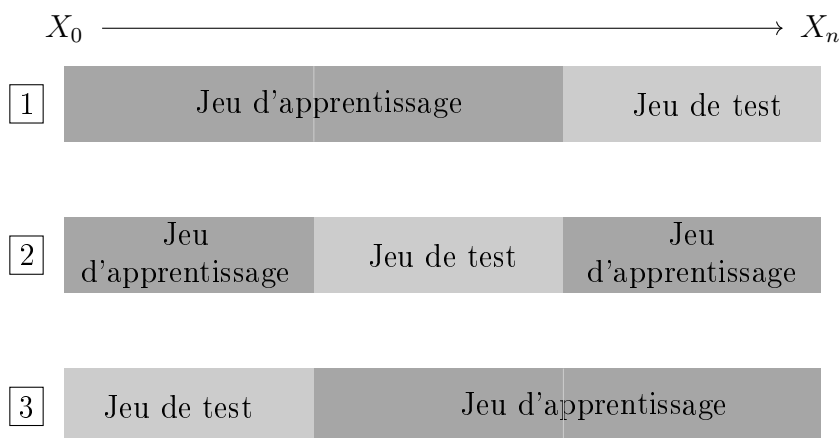


FIGURE 2.14 – Exemple d'une validation croisée à $n = 3$ segmentations, d'un même jeu de données ordonné de manière identique dans toutes les itérations.

2.3.4 Validation externe

La validation externe consiste à prendre un jeu de données complètement indépendant du jeu d'entraînement et est nécessaire dans la recherche[90]. Il peut être indépendant géographiquement (*e.g.*, un autre centre hospitalier[91]) et/ou temporellement (*e.g.*, sur une période ultérieure). Dans cette méthode de validation, il convient de comparer les deux populations. Il n'est pas obligatoire que les populations soient strictement comparables, mais cette analyse est importante pour la discussion des résultats de validation :

- Si la population d'entraînement et de validation sont différentes mais que
 - les performances sont identiques : le modèle semble fonctionner de manière stable et reproductible en dehors de sa population d'entraînement
 - les performances sont moindres : alors le modèle n'est pas stable. Le plus souvent il a été sur-entraîné[89], c'est-à-dire que le modèle est devenu trop spécifique de la population entraînée et n'est plus valable dans d'autres populations similaires.
- Si les populations d'entraînement et de validation sont comparables et que les performances des itérations sont différentes, alors le modèle est sur-entraîné de manière outrancière.

Le sur-apprentissage peut être objectivé lorsqu'une courbe trace la fonction de coût pour le jeu d'apprentissage et le jeu de test. Lorsque la courbe de la fonction de coût sur le jeu de validation stagne alors que la courbe de la fonction de coût sur le jeu d'entraînement diminue, le sur-apprentissage commence (figure 2.15). Lorsque la courbe de la fonction de coût sur le jeu de validation augmente, le sur-apprentissage a un effet délétère sur la généralisation du modèle.

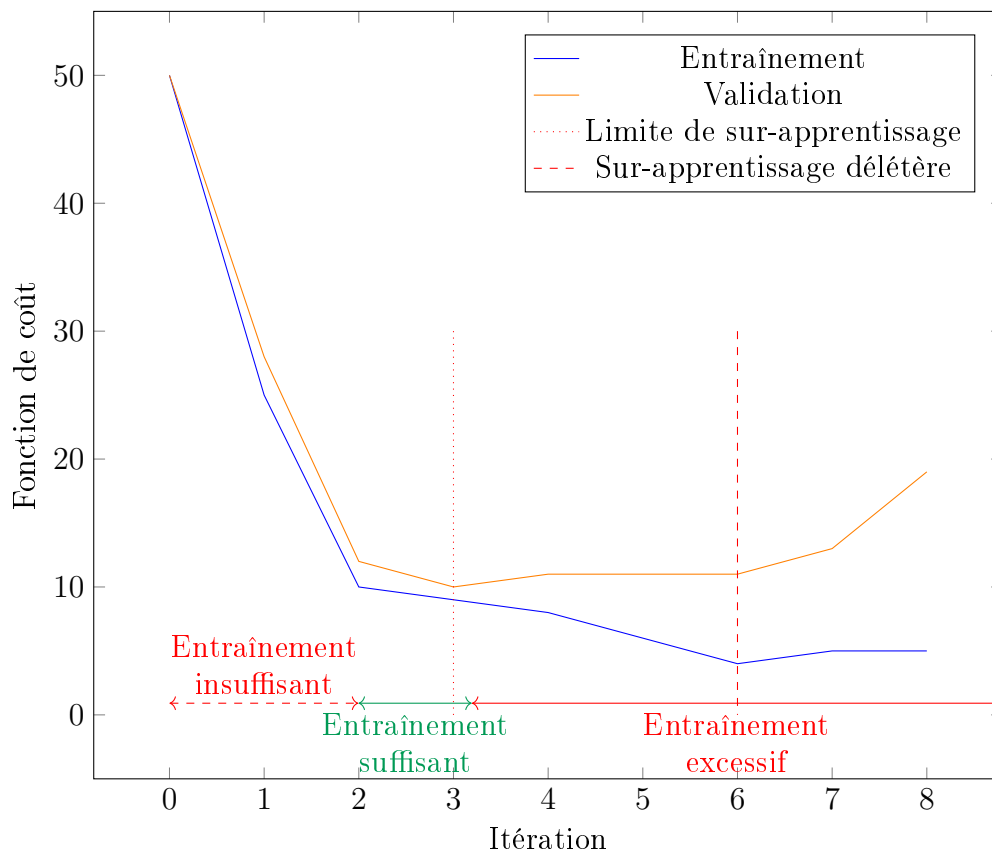


FIGURE 2.15 – Exemple de courbe d'apprentissage.

2.4 Explicabilité de l’intelligence artificielle

2.4.1 Enjeux

La crédibilité des modèles prédictifs est un enjeu pour les cliniciens[18, 19]; une plus grande utilisation clinique nécessitera des résultats plus facilement accessibles et interprétables[20]. L’**Intelligence Artificielle eXpliable (XAI)** a été développée pour répondre à ces contraintes[21] et **pour permettre aux experts médicaux de prendre, en toute confiance, des décisions guidées par les données et de fournir une prise en charge personnalisée**[22]. Aux urgences, un modèle **XAI** pourrait aider : (i) les médecins urgentistes à comprendre comment les données sont considérées en interne par la logique du modèle, et (ii) les administratifs à mieux comprendre la logique du modèle et justifier leurs choix de déploiement dans leur hôpital. De plus, l’**XAI** pourrait aider le patient à comprendre comment leurs données sont utilisées, et ce conformément à la réglementation de l’Union Européenne : le **Règlement Européen de Protection des Données (RGPD)** et le “droit à l’explication” pour les patients pour lesquels les données sont utilisées par un modèle d’**IA**[92].

L’**XAI** au sens général (*i.e.*, expliquer l’**IA**) est une étape qui peut se réaliser

- L’**XAI pre-hoc** : avant l’entraînement du modèle, on cherchera à expliquer les données (*e.g.*, résumer par la moyenne, tests de corrélation, *etc.*),
- L’**XAI post-hoc** : après l’entraînement du modèle, on cherchera à expliquer le fonctionnement du modèle (i) à travers les paramètres directement accessibles dans le cas des modèles transparents (*i.e.*, boîtes blanches ; *e.g.*, arbres décisionnels, **RL**), et (ii) à travers des techniques spécifiques d’**XAI post-hoc** pour les modèles non transparents (*i.e.*, boîtes noires ; *e.g.*, **RNP** ; les modèles transparents peuvent aussi recourir aux techniques d’**XAI post-hoc**).

2.4.2 Techniques d’XAI *post-hoc*

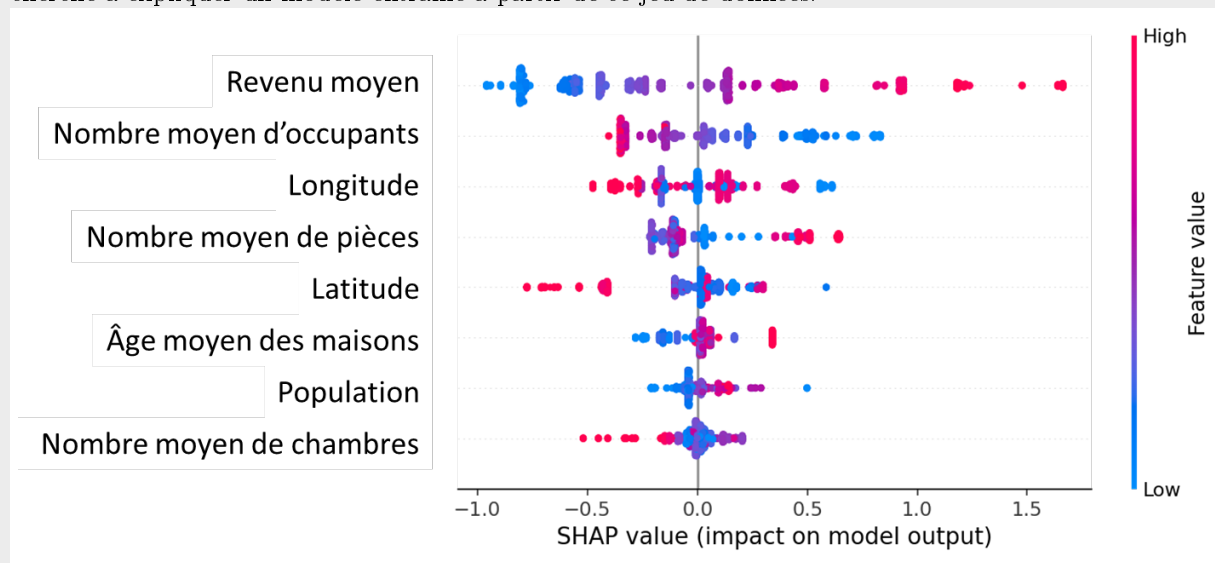
En général, les techniques d’**XAI** produisent des explications globales sur la manière dont le modèle a été entraîné à faire des prédictions. Ces techniques aident à comprendre la distribution des probabilités de sortie comme une fonction transformant les données d’entrée[21]. Certaines techniques *post-hoc* peuvent fournir des explications locales sur la manière dont le modèle arrive à la prédiction d’un cas précis. Les explications globales et locales peuvent être fournies par deux techniques *post-hoc* agnostiques du modèle : **Explications additives de Shapley (SHAP)** et le **Grphe des dépendances partielles, “Partial Dependence Plot” (PDP)**.

2.4.2.1 Explications additives de Shapley (SHAP)

SHAP est une technique analysant **l’importance des variables explicatives en fournissant des explications locales**. Celles-ci sont calculées à partir de la mesure additive (*i.e.*, la valeur de Shapley) avec l’exactitude locale et la consistance de chaque variable explicative de chaque instance[93]. **SHAP** fournit également une explication globale en agrégeant toutes les valeurs Shapley de chaque instance du jeu de données. **SHAP** permet de générer (i) un graphique résumant **SHAP**[93] permettant de quantifier la contribution de la valeur des variables explicatives à la prédiction du modèle, et (ii) des diagrammes en cascade pour décrire l’influence spécifique des valeurs des variables explicatives sur la prédiction locale d’une instance.

Exemple – SHAP – Explication générale

Un cas de test classique en IA est la prédiction de la valeur moyenne des maisons des différentes parties de l'état de Californie à partir des caractéristiques moyennes des maisons (*c.f.*, [scikit-learn](#)). Cet exemple cherche à expliquer un modèle entraîné à partir de ce jeu de données.

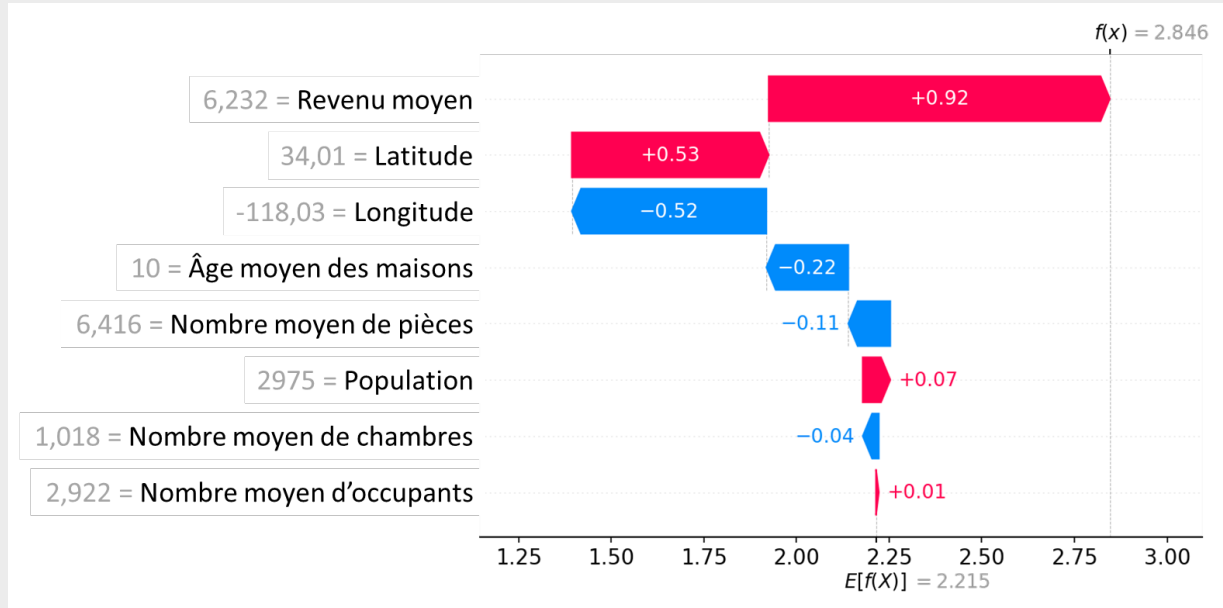


Sur ce graphique, nous pourrions tirer, entre autre, les conclusions suivantes :

- Plus le revenu moyen est élevé (vers le rouge), plus il contribue positivement à la valeur moyenne de la maison (vers la droite du graphique)
- Plus le nombre d'occupant est faible (vers le bleu), plus il contribue positivement à la valeur moyenne de la maison (vers la droite du graphique)
- Plus le nombre de chambres est important (vers le rouge), plus il contribue négativement à la valeur de la maison (vers la gauche du graphique)

Exemple – SHAP – Explication individuelle

Dans la suite de l'exemple précédent, nous pouvons utiliser SHAP pour expliquer individuellement la contribution de chaque variable explicatrice dans la prédiction du modèle.



Sur ce graphique, nous pourrions tirer, entre autre, les conclusions suivantes :

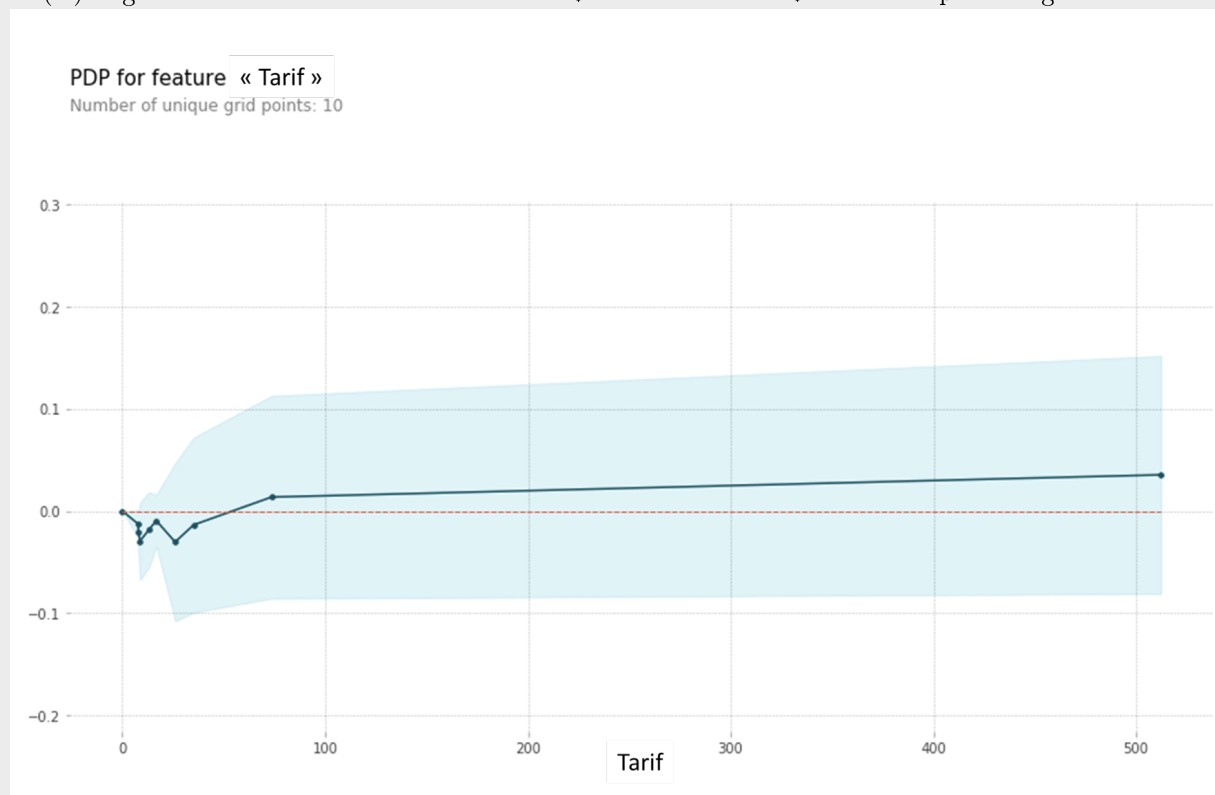
- Sur cette instance précise, la valeur moyenne des maisons du secteur prédit est de $\$2.846 \times 10^5$
- La contribution la plus importante (+0.92) à augmenter la valeur moyenne des maisons est le revenu moyen (de 6.232 dans ce secteur)
- L'âge moyen des maisons de ce secteur (10) contribue négativement (-0.22) à la valeur moyenne des maisons du secteur

2.4.2.2 Graphe des dépendances partielles, “*Partial Dependence Plot*” (PDP)

PDP produit une explication visuelle sur l'effet marginal de la prédiction de sortie pour une variable explicative dans l'éventail des valeurs observées[94]. De cette manière, les graphiques PDP permettent de visualiser la tendance de l'influence d'une variable explicative particulière sur la prédiction globale.

Exemple – Exemple sur la survie

Dans cet exemple ^a basé sur les données des passagers du Titanic ^b, le graphique ci-dessous montre l'influence du tarif du billet sur la prédiction de la survie : (i) plus le prix est élevé, plus la contribution sur la survie est élevée, (ii) les billets premiers prix avait une contribution très négative sur la survie, et (iii) le gain de contribution entre les billets à \$90 et les billets à \$550 semble plus marginal.



- a. https://github.com/SauceCat/PDPbox/blob/v0.2.1/tutorials/pdpbox_binary_classification.ipynb
b. <https://www.kaggle.com/competitions/titanic/data>

2.5 Utilisation des données en santé pour entraîner de les modèles

En médecine d’urgence, les données sont utilisées pour entraîner des modèles très variés pour prédire la mortalité, la survenue d’un arrêt cardiaque, la survenue d’un syndrome infectieux, la survenue d’une insuffisance rénale, *etc.*[95]

2.5.1 Données structurées

L’étude de RAITA *et al.* [67] menée sur les données 2007-2015 du National Hospital and Ambulatory Medical Care Survey (NHAMCS) contenait 200 000 passages aux urgences. Les données utilisées proviennent de 12 variables et sont structurées : âge, sexe, mode d’arrivée, paramètres vitaux à l’arrivée (*e.g.*, température, fréquence respiratoire, saturation en oxygène, *etc.*). Différentes approches algorithmiques ont été testées pour prédire l’hospitalisation à partir des données de triage, parmi lesquelles la **RL**, **FA**, les **Classifieurs améliorés par le gradient**, “*Gradient Boosting Classifier*” (**GBC**), et apprentissage profond. Les modèles basés sur **GBC** et apprentissage profond proposent les meilleures performances en présentant **AUROC** respectives de 85% et 86%.

L’étude de PARKER *et al.* [96] était basée sur un très large jeu de données de plus de 1.2 millions de passages aux urgences sur plus de 10 ans d’un hôpital de Singapour. Le modèle prédictif n’utilisait que huit variables : le groupe d’âge, la race, le code postal, le jour de la semaine, l’heure du jour, la priorisation à l’entrée, le mode d’arrivée et la présence d’une hyperthermie. Le modèle de **RL** qu’ils ont entraîné affiche une **AUROC** de 82.5%.

L’étude transversale de RAHIMIAN *et al.* [97] combinait les données informatisées de 4.5 millions de patients à travers l’Angleterre de 1985 à 2015. Les prédicteurs étaient constitués de 43 variables comprenant les données démographiques, déterminants de styles de vie, tests biologiques, morbidités et passages antérieurs aux urgences. Des informations externes aux données hospitalières (*e.g.*, statut marital, consultations en médecine générale, *etc.*) et des données temporelles (*e.g.*, délais depuis le premier diagnostic) ont été intégrées dans le jeu de données. En réalisant une validation croisée à 5 itérations, les modèles basés sur **GBC** et les **FA** ont permis d’afficher une **AUROC** respectives de 85% et 83%.

L’étude de BARAK-CORREN, ISRAELIT et REIS [98] a développé un modèle de **RL** prédisant l’hospitalisation suivant une approche progressive. Ce modèle a été entraîné pour prédire l’hospitalisation à 3 moments consécutifs de la prise en charge aux urgences : à 10 minutes, à une heure et à deux heures.

2.5.2 Données non structurées

L’étude STERLING *et al.* [99] a utilisé l’approche “sac de mots” (*i.e.*, représentation du texte qui décrit la présence de mots dans un document) pour représenter les notes des infirmières d’accueil. Basé sur un jeu de données de plus de 250 000 passages, le **RNP** entraîné à prédire les hospitalisations affiche une **AUROC** prometteuse de 74%.

L’étude ZHANG *et al.* [100] a comparé la performance d’un modèle sans/avec l’inclusion des données textuelles. L’analyse comparative était basée sur une **RL** et un **RNP** pour prédire l’hospitalisation ou le transfert juste après le triage. Leur résultat principal montrait une amélioration de la précision de la prédiction avec l’ajout du texte. L’étude CHEN *et al.* [101] s’appuie sur la technique du plongement lexical et affiche des performances bien supérieures à toute autre méthode.

L’étude CHANG, HONG et TAYLOR [102] a utilisé la technologie basée sur la **Représentation à encodage bi-directionnel des transformers**, “*Bidirectional Encoder Representation from Transformers*” (**BERT**). Il s’agit d’une méthode de référence dans le traitement du langage naturel. En utilisant un jeu de données de plus 1.8 millions de notes, le modèle **BERT** a été utilisé pour dériver le plongement lexical contextualisé. Ils ont pu

démontrer que le plongement lexical contextualisé permettait de représenter les données textuelles de manière similaire à des points dans un espace vectoriel.

 Points clés

Il n'existe pas à ce jour en France, de modèle explicable permettant de prédire le parcours de soin d'un patient aux urgences à partir des données du dossier médical.

2.6 Garantie humaine

Les modèles d'IA proposent de prédire la variable à expliquer de manière rapide, reproductible, systématique et infatigable quelle que soit l'heure du jour ou de la nuit. Le modèle se base sur des données collectées lors d'un processus humain (*e.g.*, interrogatoire, prise de constantes, *etc.*). En revanche la prise en compte, l'analyse, et l'intégration de ces informations, ainsi que la proposition de décision finale relèvent de processus informatiques. Ainsi, un modèle est utilisable fidèlement à toute heure du jour et de la nuit, sans variation de performance contrairement aux performances humaines (*e.g.*, baisse de vigilance nocturne, interruptions de tâches, émotion, *etc.*). De plus, il peut effectuer la tâche de prédiction ou de classification pour plusieurs patients en même temps alors qu'un médecin ne peut prendre une décision que patient par patient. En limitant les facteurs humains et les biais cognitifs, il peut être pertinent que des outils d'aide à décision viennent en appui à un médecin aux performances diminuées par plusieurs heures de garde. Ils permettent de gagner du temps et d'éprouver parfois le raisonnement médical lorsque les résultats proposés sont différents de l'idée médicale initiale.

Dans une démarche de gestion des risques, les modèles d'IA peuvent être entraînés à limiter les erreurs et demander une confirmation lorsqu'un risque est identifié. En effet, par exemple, si le seuil de positivité d'hospitalisation d'un patient est réglé pour obtenir une VPP maximale (c'est-à-dire la probabilité maximale que le modèle ait raison lorsqu'il répond qu'il faut hospitaliser le patient) alors qu'un médecin organise dans le même temps une prise en charge ambulatoire du patient, cette divergence pousserait à la recherche d'un deuxième avis médical. Dans ce cas spécifique, la réciproque est non nécessairement applicable.

Toute prise en charge médicale est associée à un risque d'erreur et de préjudice. L'utilisation de l'IA dans le domaine de la santé pose la question suivante : qui est responsable en cas de préjudice médical au cours d'une prise en charge impliquant l'utilisation de l'IA ? Est-ce le médecin ? L'institution ? Le distributeur de l'outil ? Le créateur de l'outil ? En droit français, des règles sont susceptibles d'encadrer la responsabilité en cas de préjudice subi en lien avec l'utilisation de l'IA en pratique médicale. Elles varient selon la cause du préjudice. La responsabilité du fait personnel pourrait s'appliquer en cas de faute dans l'utilisation de l'IA [103]. La responsabilité du fait des choses pourrait également s'appliquer, en cas de fait de l'IA indépendamment de tout comportement fautif [104]. Auquel cas, le responsable sera le gardien de l'IA, celui qui en a l'usage, la direction et le contrôle [105], la désignation du gardien relevant de l'appréciation souveraine des juges. On peut également songer à la responsabilité du fait des produits défectueux ou à la responsabilité du fait des produits de santé. La détermination du régime de responsabilité n'est toutefois pas aisée, notamment cela dépendra du cadre dans lequel l'IA est utilisée.

Points clés

La recherche sur l'IA appliquée à la médecine d'urgence et son utilisation au quotidien dans les SU a augmenté significativement ces dernières années. L'IA doit être considérée comme un outil d'aide à la prise en charge diagnostique et thérapeutique des patients et d'amélioration de l'organisation des SU, notamment par la prise en compte de contraintes "métiers", contextuelles, relatives aux patients et plus généralement structurelles.

Rationnel des contributions

Une des solutions pour lutter le problème de la surcharge dans les **SU** consiste à prédire précocement l'orientation du patient à la fin de la prise en charge : sortie au domicile ou admission dans un service d'hospitalisation. **A ce jour, il n'existe aucun modèle visant à prédire le parcours de soins d'un patient aux urgences.** Nous construirons un tel modèle dans le [chapitre 3 : 3P-U](#).

En médecine d'urgence, l'utilisation des outils d'aide à la décision basés sur l'**IA** est en augmentation significative ces dernières années. Dans certains cas, les données sont volontairement omises et constituent ainsi les **“Données volontairement non collectées”** (*“data not purposely collected”*) (**DNPC**). Ce biais d'information assumé peut être géré de différents manières : exclure les patients avec des données manquantes, imputer par la moyenne ou utiliser des techniques automatiques (*e.g.*, d'apprentissage machine) pour gérer ou imputer les données manquantes. **A ce jour, aucune étude ne s'est posé la question de la gestion des données manquantes dans les études construisant des modèles d'IA dans les SU.** Nous étudierons les stratégies utilisées pour gérer les manquantes dans les études construisant un modèle d'**IA** aux urgences dans le [chapitre 4](#).

La mise en place d'un outil d'**IA** en pratique clinique ne suffit pas à garantir son utilisation par l'ensemble des personnels soignants. La crédibilité du modèle est importante pour les cliniciens, et augmenter le taux d'utilisation de celui-ci nécessite une réponse lisible et interprétable. L'explicabilité de l'intelligence artificielle (*“explainable artificial intelligence”*; **XAI**) a été développée spécifiquement pour répondre à ces problématiques et pour permettre aux experts de la santé de prendre des décisions basées sur les données pour une médecine personnalisée et une organisation du parcours améliorée. Dans les **SU**, un modèle **XAI** peut aider (i) les urgentistes à mieux comprendre comment les données sont prises en compte à l'intérieur du modèle, (ii) les directions à mieux comprendre la logique du modèle et justifier le déploiement dans leur structure hospitalière, et (iii) les patients à mieux comprendre comment leurs données sont utilisées. **Le modèle 3P-U généré dans le cadre du travail du chapitre 3 n'a pas été expliqué.** Nous rendrons **3P-U** explicable dans le [chapitre 5](#).

L'application du modèle de prédiction du parcours patients aux urgences a pour objectif d'être utilisé en pratique clinique. Nous étudierons une application concrète de **3P-U** dans le [chapitre 6](#).

L'**IA** doit être considérée comme un outil d'aide à la prise en charge diagnostique et thérapeutique des patients et d'amélioration de l'organisation des **SU**, notamment par la prise en compte de contraintes *“métiers”*, contextuelles, relatives aux patients et plus généralement structurelles. L'**IA** comporte des avantages (reproductibilité, rapidité) mais aussi des risques (erreur, perte d'esprit critique). **La place de l'expertise humaine médicale dans la conception et l'utilisation d'un modèle d'IA dans les SU n'a pas été étudiée.** Nous étudierons la problématique de la formation des soignants et la place de la supervision humaine dans le cadre de l'utilisation de l'**IA** dans le [chapitre 7](#).

Contribution à la prédiction du parcours de soins aux urgences

Chapitre 3

Définition d'un modèle prédictif du parcours de soins

Points clés

La question de la création d'un modèle prédictif du parcours de soins a déjà été abordées sous plusieurs angles dans de récentes publications, concomitantes du début des travaux de cette thèse. Aucune étude française n'étudie l'applicabilité et la validité de ces approches sur le modèle de santé français. Ainsi, l'objectif principal de cette contribution est de construire un modèle prédictif du parcours patient aux urgences en se basant sur les données structurées et non structurées.

Cette contribution a été publiée sous la référence: Emilien ARNAUD *et al.* "Deep Learning to Predict Hospitalization at Triage : Integration of Structured Data and Unstructured Text". In : *2020 IEEE International Conference on Big Data (Big Data)*. 2020 IEEE International Conference on Big Data (Big Data). Atlanta, GA, USA : IEEE, 10 déc. 2020, p. 4836-4841. ISBN : 978-1-72816-251-5. DOI : [10.1109/BigData50022.2020.9378073](https://doi.org/10.1109/BigData50022.2020.9378073). URL : <https://ieeexplore.ieee.org/document/9378073/> (visité le 18/05/2021).

3.1 Méthodologie

3.1.1 Description des données

Le jeu de données utilisé pour construire le modèle provient du service des urgences du [CHU-AP](#). Il est constitué par les passages aux urgences entre janvier 2015 et juin 2019. Chaque passage est caractérisé par une orientation finale : une hospitalisation ou une sortie des urgences. Sur les quelques 260 000 passages disponibles à cette période, 35% des patients étaient hospitalisés ([tableau 3.1](#)). La durée moyenne de passage des patients sortants était de 4 heures et 23 minutes, soit environ 18% inférieure à celle des patients hospitalisés.

Année	Nombre de passages	Taux d'hospitalisation
2015	53 465	35.6%
2016	57 963	35.4%
2017	60 155	34.1%
2018	61 794	34.0%
2019	30 615	33.1%
$\Sigma = 263\ 992$		$\bar{x} = 35\%$

TABLE 3.1 – Jeu de données entraînant le modèle de prédiction du patient aux urgences[23]. Σ représente la somme et \bar{x} la moyenne sur l'ensemble du jeu de données.

Le jeu de données est constitué de 38 variables numériques, catégorielles et textuelles. Les variables numériques correspondent aux paramètres vitaux comme la température, la fréquence cardiaque, la pression artérielle, *etc.*, mais aussi à l'âge. Les variables catégorielles décrivent en général les caractéristiques du patient : le sexe, le mode d'arrivée, le statut familial, *etc.*. Les catégories sont définies par le système d'information hospitalier au travers un dictionnaire.

De plus, le jeu de données contient du texte libre : observation paramédicale de triage correspondant aux notes de l'[IOA](#), ainsi que les antécédents psychiatriques, chirurgicaux et médicaux du patient. Ces notes de texte sont rédigées en français, par les infirmières et les médecins du service pendant le passage du patient sans qu'il y ait de relecture. Il convient de mentionner que plusieurs milliers d'enregistrements contiennent du texte vide : environ 70 000 enregistrements ne contiennent aucun élément textuel.

3.1.2 Approche d'apprentissage

L'idée principale de notre approche est d'utiliser l'ensemble des données disponibles, quelle que soit leur nature, pour entraîner un modèle de classification visant à prédire l'hospitalisation du patient à l'issue de sa prise en charge. Plus précisément, nous avons cherché à utiliser les données numériques, catégorielles et textuelles dans le même modèle et durant la même session d'entraînement. D'une part un [RNP](#) est utilisé pour traiter les données standards (*i.e.*, numériques et catégorielles). D'autre part, un [RNC](#) est utilisé pour traiter les données textuelles. Les deux réseaux sont entraînés indépendamment et en parallèle.

Secondairement, une représentation globale des données est assemblée par la concaténation des sorties des deux précédents réseaux. Cette représentation est utilisée pour entraîner un nouveau [RNP](#) qui effectuera la tâche de classification. La [figure 3.1](#) représente l'articulation des différents réseaux.

3.1.3 Pré-traitement des données structurées

Le pré-traitement des données est un challenge puisque les données traitées sont de natures différentes : une donnée textuelle doit être pré-traitée différemment d'une donnée structurée. Le jeu de données structurées est constitué de 15 variables numériques et de neuf variables catégorielles ([tableau 3.2](#)).

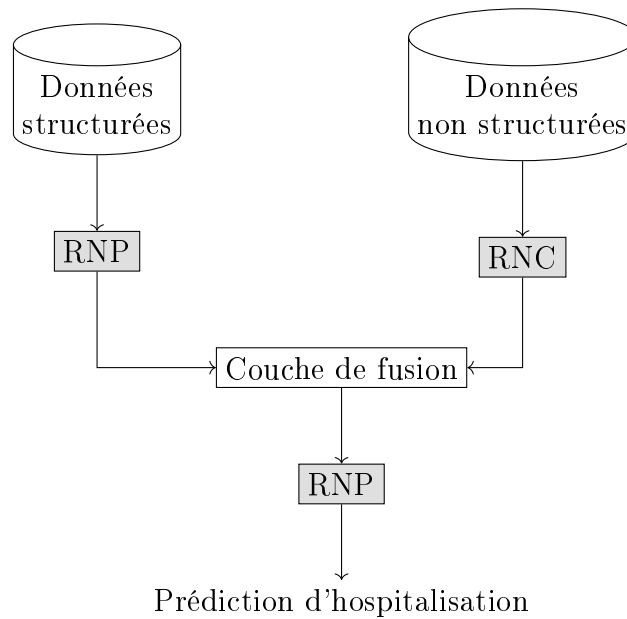


FIGURE 3.1 – Structuration des réseaux de neurones en vue d’une tâche de classification (i.e., hospitalisation vs. sortie des urgences) à partir des données structurées et non structurées[23]. RNP : réseau de neurones profond, RNC : réseau de neurones convolutionnel.

Le jeu de données initial contient un nombre important de données manquantes sur l’ensemble des variables. Les données manquantes numériques ont été remplacées par les données dites “normales” (*e.g.*, fréquence cardiaque à 80 battements par minute, pression artérielle systolique à 120 mmHg). Les données manquantes catégorielles ont été remplacées par la valeur la plus fréquente dans la base de données (*e.g.*, modalité d’attente : en brancard, présence d’accompagnant : aucun).

Les données numériques ont été linéairement normalisées entre $[0, 1]$ en utilisant la fonction min-max. Les données catégorielles ont été encodées en variables binaires indépendantes (i.e., one-hot encoding). Ces pré-traitements ont été réalisés en utilisant les bibliothèques Python Scikit-Learn[106] et Pandas[107]

3.1.4 Pré-traitement des données non structurées

Pour faciliter le traitement et l’apprentissage des données textuelles, la première étape consiste à concaténer les quatre variables en une seule. Le texte ainsi fusionné peut être considéré comme un document représentant l’ensemble des variables textuelles considérées pour le passage.

Le pré-traitement des données non structurées a été opéré conformément à l’état de l’art (*c.f.*, sous-section 2.1.5, figure 2.2)

3.2 Résultats expérimentaux

3.2.1 Données expérimentales

L’entraînement du modèle que nous avons construit nécessite la présence de texte pour chacun des enregistrements. Il a donc été nécessaire d’exclure l’ensemble des passages pour lesquels l’ensemble des données textuelles étaient vides. Au final, ce sont près de 190 000 passages qui ont été utilisés pour entraîner le modèle. Après avoir isolé 20% des données initiales pour constituer le jeu de validation, le jeu de données restant a été scindé en un jeu d’entraînement et un jeu de test suivant la méthode de validations croisées (sous-section 2.3.3) à trois itérations.

#	Variable	Type	Stratégie de valeur manquante
1	Jour / Heure d'arrivée	Catégorie	–
2	Genre	Catégorie	Fixe : Homme
3	Adressé par	Catégorie	Fixe : Autre
4	Mode d'arrivée	Catégorie	Fixe : Autre
5	Présence d'accompagnant	Catégorie	Fixe : Non
6	Statut familial	Catégorie	Fixe : Célibataire
7	Modalité d'attente	Catégorie	Plus fréquent : brancard
8	Motif de consultation	Catégorie	Fixe : Autre
9	Circonstances	Catégorie	Fixe : Autre
10	Âge	Numérique	–
11	Débit d'oxygène	Numérique	Fixe : 0 L/min
12	Fréquence cardiaque	Numérique	Fixe : 80 bt/min
13	Fréquence respiratoire	Numérique	Exclus
14	Pression artérielle systolique	Numérique	Fixe : 120 mmHg
15	Pression artérielle diastolique	Numérique	Fixe : 80 mmHg
16	Douleur	Numérique	Fixe : 0
17	Température	Numérique	Fixe : 37.4 °C
18	Saturation en oxygène	Numérique	Fixe : 100%
19	Glycémie capillaire	Numérique	Fixe : 5 mmol/L
20	Hémoglobininémie capillaire	Numérique	Fixe : 12 g/L
21	Volume vésical	Numérique	Fixe : 0 mL
22	Cétonémie capillaire	Numérique	Fixe : 0 mmol/L
23	Taux d'alcool expiré	Numérique	Fixe : 0 g/L
24	Tri infirmier	Numérique	Exclusion du patient
25	Notes infirmières	Texte	–
26	Antécédents psychiatriques	Texte	–
27	Antécédents chirurgicaux	Texte	–
28	Antécédents médicaux	Texte	–

TABLE 3.2 – Variables utilisées pour l'entraînement du classifieur visant à prédire l'hospitalisation du patient à la fin de la prise en charge[23].

3.2.2 Structure du modèle

L'architecture du modèle a été construite en utilisant deux procédures parallèles mais simultanées : l'une pour les données structurées et l'autre pour les données textuelles. D'une part le **RNP** constituant le traitement isolé des données structurées comprenait trois couches de neurones comprenant respectivement 128, 64 et 32 neurones. Cette architecture a été déterminée empiriquement après des expérimentations préalables.

D'autre part, le **RNC** constituant le traitement isolé des données non structurées comprenait tout d'abord trois couches convolutionnelles à une dimension ayant respectivement une taille de filtre définie à 16, 32 et 64, puis une couche de fusion de type *global-max* et enfin une couche dense. Les couches convolutionnelles à une dimension sont adaptées au traitement des vecteurs ayant une dimension (*e.g.*, des chaînes de caractères).

Enfin, les couches de sorties du **RNP** et du **RNC** ont été fusionnées et utilisées comme couche d'entrée d'un dernier **RNP** à trois couches comprenant respectivement 100, 50 et 25 neurones. Des couches de *Dropout* (*i.e.*, couche réinitialisant une proportion des poids afin de diminuer le risque de stagner sur un minimum local) ont été incluses par endroit pour limiter le risque de surapprentissage. La [figure 3.2](#) détaille la structuration du modèle. Le modèle a été implémenté en utilisant la librairie Python Keras[52] et la librairie Tensorflow[108].

3.2.3 Performances du modèle

Après l'exécution des trois itérations de la validation croisée sur 10 *epoch* ([sous-sous-section 2.2.2.1](#)), la performance du modèle **3P-U** a été mesurée sur les 20% des données initialement isolées. La [figure 3.3](#) représente l'évolution de la fonction de coût pendant l'entraînement et la validation. La [figure 3.4](#) montre la courbe **ROC**, témoignant d'une qualité de la classification puisque son aire est d'environ 83%. Plus précisément, la précision (*i.e.*, valeur prédictive positive) était de 72% et le rappel (*i.e.*, sensibilité) était de 62%.

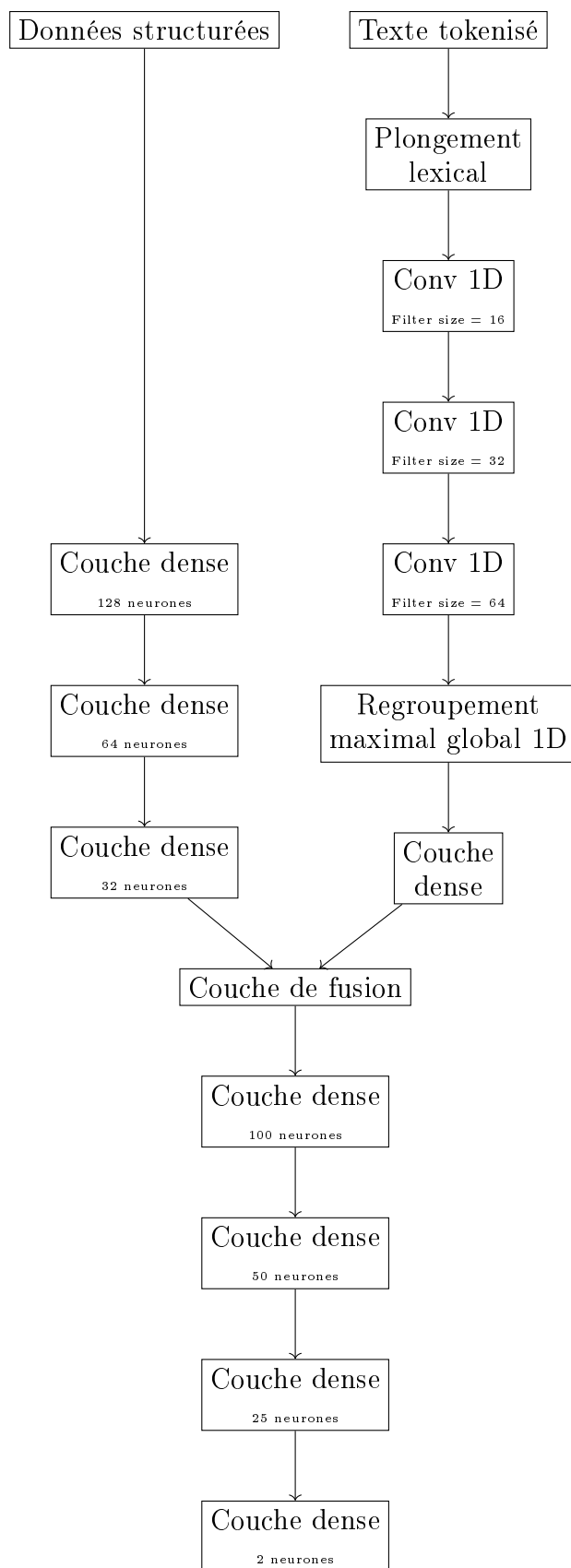


FIGURE 3.2 – Structure détaillée du modèle de prédiction du parcours patient aux urgences[23].

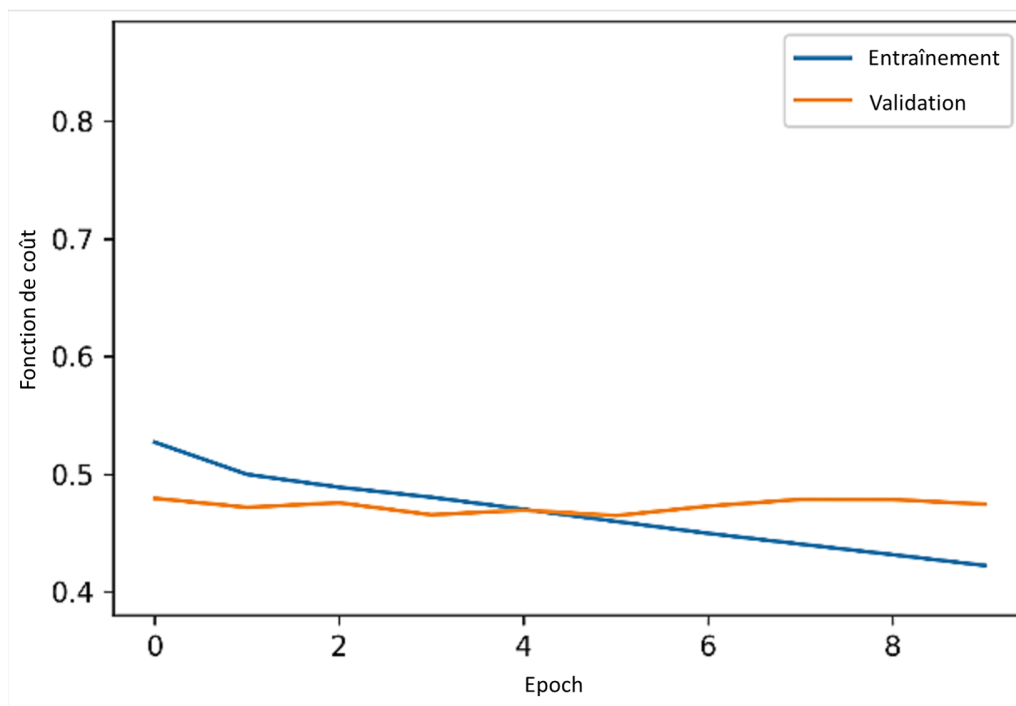


FIGURE 3.3 – Fonction de coût d’entraînement et de validation du modèle[23].

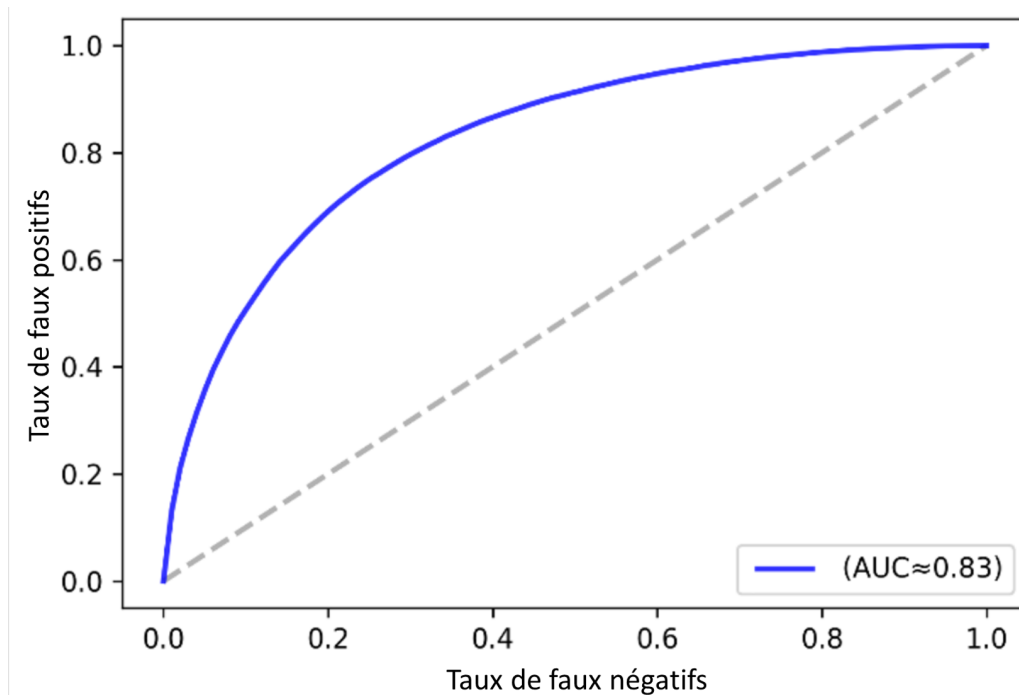


FIGURE 3.4 – Courbe ROC du modèle[23].

Chapitre 4

Gestion des données manquantes en médecine d'urgence

Points clés

Lors de la création de **3P-U**, le modèle de prédiction de parcours patient aux urgences, nous avons fait le choix de remplacer les données manquantes par une valeur physiologique dans la majorité des cas. Nous nous sommes demandés si cette stratégie était majoritaire dans les études du même type.

Cette contribution a été publiée sous la référence: Emilien ARNAUD *et al.* "Predictive Models in Emergency Medicine and Their Missing Data Strategies : A Systematic Review". In : *npj Digital Medicine* 6.1 (1 23 fév. 2023), p. 1-9. ISSN : 2398-6352. DOI : [10.1038/s41746-023-00770-6](https://doi.org/10.1038/s41746-023-00770-6). URL : <https://www.nature.com/articles/s41746-023-00770-6> (visité le 24/02/2023).

4.1 Introduction

4.1.1 Données manquantes en médecine d’urgence

Les patients consultent aux urgences pour des raisons très différentes. Certaines variables sont systématiquement collectées : les données administratives, les caractéristiques démographiques, la raison de consultation, *etc.* D’autres variables ne sont pas collectées systématiquement : pression artérielle, évaluation de la douleur, glycémie capillaire, *etc.* Dans les deux cas, ces variables peuvent être catégorielles (*e.g.*, l’évaluation de la douleur de 0 à 10 et peut inclure une catégorie additionnelle : “non remplie”) ou numérique (*e.g.*, la glycémie capillaire).

L’indication à la mesure de la glycémie capillaire diffère en fonction de la situation clinique et les raisons pour lesquelles la glycémie n’est pas disponible sont multiples (tableau 4.1). Ainsi, la mesure de la glycémie dépend d’autres variables et est donc considérée comme une donnée MAR (sous-sous-section 2.1.6.2). Parce que cette absence de données est liée à des considérations médicales et éthiques, nous considérons que ces données manquantes sont des “Données volontairement non collectées” (“*data not purposely collected*”) (DNPC), une sous-catégorie des données MAR.

Motif de consultation	Antécédents	Glycémie capillaire	Donnée manquante		Justification
			Médical	Statistique	
Traumatisme mineur de la cheville d’origine mécanique	Aucun	Non mesurée	Non	Non	Le motif de consultation ne requiert pas la mesure de la glycémie capillaire
Traumatisme mineur de la cheville d’origine mécanique	Diabète de type 1	Mesurée	Non	Non	Chez un patient diabétique la glycémie doit être mesurée quel que soit son motif de consultation
Traumatisme mineur de la cheville après un malaise	Aucun	Mesurée	Non	Non	La mesure de la glycémie est systématique sur un malaise
Traumatisme mineur de la cheville après un malaise	Aucun	Mesurée Non saisie	Non	Oui	La glycémie a été cliniquement évaluée mais non reportée dans le logiciel
Traumatisme mineur de la cheville après un malaise	Aucun	Non mesurée	Oui	Oui	La glycémie aurait dû être mesurée mais ne l’a pas été C’est une erreur et constitue une réelle donnée manquante

TABLE 4.1 – Exemples de situations où des données peuvent manquer chez un patient se présentant aux urgences[24].

4.1.2 Techniques de complétion

4.1.2.1 En général

Il existe plusieurs approches algorithmiques et chacune a des pré-requis différents concernant la complétude des données. Techniquement, la RL de base et les méthodes d’apprentissage profond n’acceptent pas des entrées contenant des données manquantes dans les algorithmes de bas niveau. Les arbres décisionnels et les méthodes dérivées (*e.g.*, arbres décisionnels dérivés, FA, *etc.*) ne requièrent pas un jeu de données complet, et les données manquantes sont interprétées comme une catégorie spécifique[109]. Des algorithmes plus complexes de RL (*e.g.*, la Version de l’approximation stochastique du maximum attendu, “*Stochastic Approximation version of*

Expectation-Maximization” (SAEM)[110]) acceptent également des données manquantes avec de faibles biais et une faible erreur standard. Cette méthode consiste à estimer les paramètres ($\Theta = \{\alpha, \beta, \dots\}$) d’une RL en maximisant l’approximation stochastique d’exemples simulés à partir du jeu de données, en trois étapes : (i) simuler une valeur initiale Θ_0 , (ii) réaliser k boucles d’une approximation stochastique, et (iii) maximiser l’estimation de Θ_k [110]..

Compléter des données manquantes est une stratégie connue et habituelle lors de la construction d’un modèle prédictif. Dans le domaine médical, plusieurs stratégies existent pour traiter les données manquantes, qu’elles soient numériques ou catégorielles :

- A) En utilisant un algorithme autonome sur la gestion des données manquantes (*i.e.*, les arbres décisionnels ou SARM : les données manquantes sont nativement gérées) : le chercheur ne réalise aucun pré-traitement sur le jeu de données initial et délègue cette gestion à l’algorithme d’apprentissage machine (que nous appellerons “no-op” ci-après).
- B) Sans utiliser d’algorithme autonome (*e.g.*, RL de base, et RNP), en appliquant un pré-traitement au jeu de données initial par le chercheur (figure 4.1) :
 - i) l’exclusion (*i.e.*, suppression des patients ayant des données manquantes),
 - ii) imputation par la moyenne (*i.e.*, remplacer les données manquantes par la moyenne des données existantes pour les variables numériques et le mode des variables catégorielles),
 - iii) remplissage par une valeur physiologique ou de référence choisie par l’expert,
 - iv) imputation par l’IA des données manquantes (*i.e.*, par des méthodes d’apprentissage machine telles que la RL et les équations chaînées)
- C) En adjoignant une stratégie supplémentaire telle que l’indicateur d’absence d’une valeur (*i.e.*, créer une variable dans le jeu de données indiquant pour chaque patient si la valeur est manquante ou non).

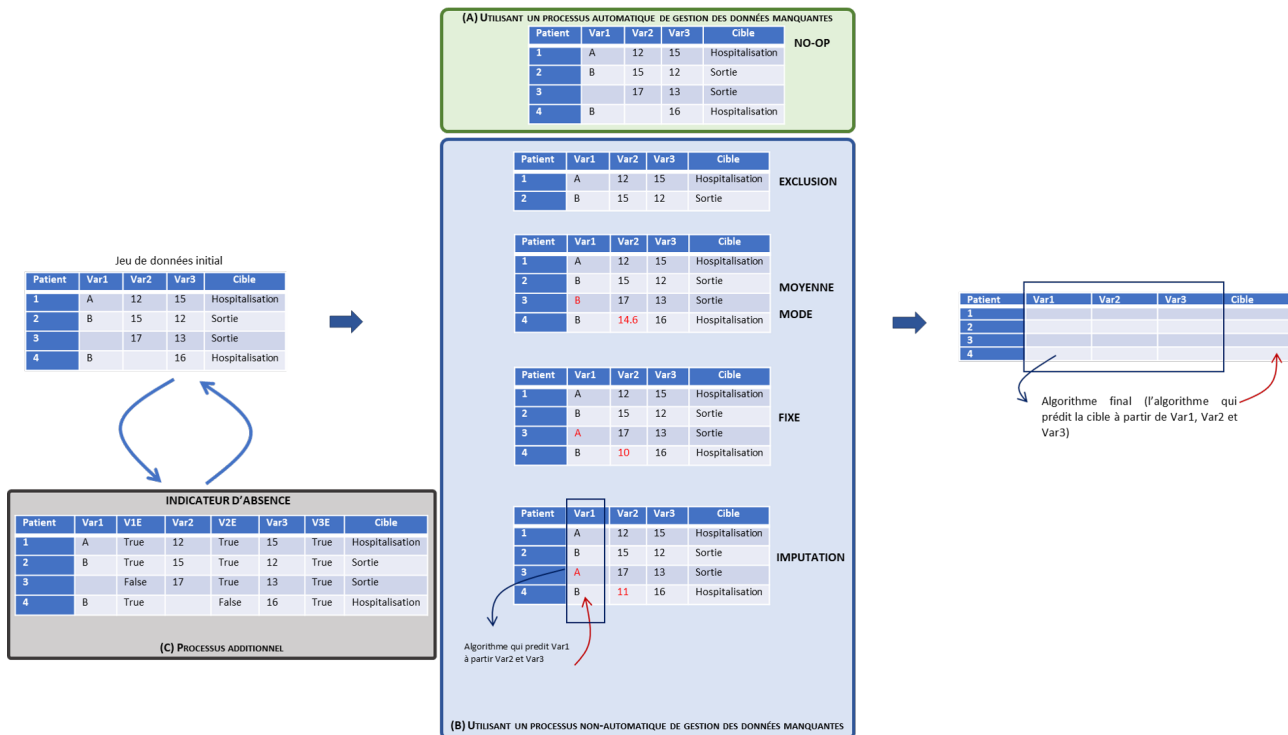


FIGURE 4.1 – Stratégies appliquées aux données manquantes avant le modèle prédictif final[24]. **A** Stratégies des données manquantes en utilisant un pré-traitement autonome lié à l’algorithme final. **B** Stratégies de données manquantes utilisant un pré-traitement non autonome. **C** Stratégies additionnelles de données manquantes.

NAEMI *et al.* [111] ont testé l’impact de différentes stratégies de gestion des données manquantes appliquées au jeu de données avant entraînement d’un modèle dans les mêmes conditions. L’AUROC du modèle d’arbre décisionnel boosté par le gradient descendant extrême augmentait de 84% à 94% après la complétion des données en utilisant un processus Gaussien et une génération synthétique des classes limites. Des méthodes complémentaires consistent à marquer les patients ayant des données manquantes dans une nouvelle variable, en parallèle de l’application d’une stratégie obligatoire. Cette information additionnelle peut augmenter les performances de l’apprentissage machine.

Les stratégies d’imputation doivent prendre en compte la connaissance du mécanisme de données manquantes. Le mécanisme à l’origine des données manquantes est considéré[112] comme pouvant être ignoré si la variable est MAR et que ψ et θ sont distincts :

$$p(\theta, \psi) = p(\theta) \cdot p(\psi) \quad (4.1)$$

Si le mécanisme à l’origine des données manquantes peut être ignoré, la distribution de Y est la même dans le groupe observé et dans le groupe non observé :

$$p(Y|Y_{obs}, R = 1) = p(Y|Y_{obs}, R = 0) \quad (4.2)$$

Dans ce cas, il n’est pas nécessaire de considérer le mécanisme à l’origine des données manquantes dans l’imputation, et un modèle *simple* d’imputation peut compléter la plupart des données manquantes sans introduire de biais. Au contraire, le mécanisme à l’origine des données manquantes doit être pris en compte dans un modèle d’imputation plus complexe (*e.g.*, [Modèle d’Imputation à Équations Chaînées \(MICE\)](#)) lorsque le mécanisme à l’origine des données manquantes est lié aux données elles-mêmes.

Considérant qu’il est complexe et technique de rapporter clairement ces éléments dans une étude de construction de modèle prédictif, et dans un souci de transparence nécessaire aux études scientifiques médicales, un groupe de méthodologistes a développé la méthode “[The Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis](#)” (TRIPOD). TRIPOD est une liste de 22 items essentiels à un rapport de qualité d’une étude développant ou validant un modèle prédictif de plusieurs variables[113].

4.1.2.2 En médecine d’urgence

Les services d’hospitalisation et les services d’urgences diffèrent sur plusieurs aspects. Premièrement, les types de patients et les motifs d’admission sont moins spécifiques aux urgences que dans les services d’hospitalisation : les patients consultent aux urgences pour des problèmes relevant de n’importe quelle spécialité. Ainsi, toutes les données ne sont pas systématiquement collectées pour tous les patients : la collecte dépend du motif de consultation. Deuxièmement, le flux de patients consultant aux urgences est continu (jusqu’à plusieurs centaines de patients par jour), instable (pics d’arrivées aux urgences en fin de matinée et fin d’après-midi) et sans limite (tous les patients sont accueillis, parfois jusqu’à 100 patients présents simultanément), alors que le flux de patients dans les services d’hospitalisation s’arrête lorsque tous les lits sont occupés (quelques entrées / sorties par jour) et est stable (les entrées se font le plus souvent en fin d’après-midi). En service d’hospitalisation, les patients restent le plus souvent plus d’une journée. En conséquence, le temps alloué au patient diffère aux urgences et dans les services, et cela impacte la collecte des données. Lorsqu’un patient consulte aux urgences, certaines variables sont systématiquement collectées (*e.g.*, données démographiques, dates d’entrée et de sortie, diagnostic final, *etc.*) et les autres sont uniquement collectées si le contexte le nécessite ou quand il est possible de collecter la donnée (*e.g.*, les infirmières ne collectent que les données strictement nécessaires dans un service d’urgence surchargé) : ce sont les DNPC.

Le contexte particulier que constitue le champ de la médecine d'urgence donne lieu à des jeux de données hétérogènes. Les **DNPC** ne constituent donc pas un biais en eux-mêmes puisque leur absence est délibérée. Ainsi, la complétion des **DNPC** n'est pas la même que la complétion des données manquantes d'une étude prospective standard où les données devraient être collectées mais ne le sont pas. La question qui se pose donc est : *comment le problème des **DNPC** est traité dans les études d'IA dans le champ de la médecine d'urgence ?*

Premièrement, nous allons effectuer une revue de la littérature analysant l'ensemble des stratégies traitant les données manquantes appliquées dans la construction de modèles prédictifs en médecine d'urgence. Deuxièmement, nous chercherons à identifier si une stratégie est meilleure qu'une autre dans ce contexte spécifique.

4.2 Matériels et méthodes

4.2.0.1 Construction

Nous avons extrait et rapporté nos résultats en accord avec les méthodes recommandées “Critical Appraisal and Data Extraction for Systematic Review of Prediction Modelling Studies” (CHARMS)[114] et “Preferred Reporting Items for Systematic Reviews and Meta-Analysis” (PRISMA)[115]¹

4.2.0.2 Stratégie de recherche

En janvier 2022, nous avons effectué la requête suivante sur la base de données PubMed en utilisant les mots clés “Medical Subject Headings” (MeSH) : “(emergency medicine OR emergency service) AND (artificial intelligence OR machine learning)”.

4.2.0.3 Critères d’inclusion et d’exclusion

Nous avons évalué toutes les publications en langue anglaise qui traitaient de la construction ou de l’application d’un modèle prédictif basé sur des données structurées de patients adultes dans les services d’urgences. Les études dupliquées et les revues de la littératures ont été exclues, ainsi que les publications antérieures à l’année 2000. Nous avons exclu les études pour lesquelles le modèle prédictif était uniquement basé sur l’analyse du langage naturel, puisque les données manquantes ne peuvent pas être traitées de la même manière que les données structurées. Enfin, les études non cliniques (*e.g.*, basées uniquement sur des données sociales) et les publications pour lesquelles le texte complet n’était pas disponible ont été également exclues.

4.2.0.4 Processus de sélection des études

Les résultats de la requête PubMed ont d’abord été analysées sur la base du titre. Puis les résumés ont été lus afin d’exclure les études inéligibles. Enfin, la lecture du texte complet du reste des publications permettait de déterminer si la méthodologie exposée correspondait aux critères d’inclusion. Ces différentes étapes sont décrites dans la [figure 4.2](#)

4.2.0.5 Processus de collecte de données

Les données issues des études incluses (références de la publication, les items de la liste CHARMS, les stratégies de données manquantes, le partitionnement du jeu de données, et la meilleure AUROC pour la phase d’entraînement et de validation) ont été entrées dans une feuille de calcul Excel[®] (Microsoft Corp, Cupertino, USA). Si l’étude ne spécifiait pas la stratégie appliquée pour les données manquantes, nous l’avons considérée comme étant “non spécifiée”. Si l’étude ne spécifiait pas comment le jeu de données avait été partitionné en jeu d’entraînement et jeu de test, nous avons considéré qu’ils avaient utilisé le ratio standard 80%-20%. Enfin, nous avons considéré que l’utilisation d’un jeu de données indépendant était obligatoire pour classer l’étude comme étant une étude de validation. Quand le jeu de données utilisé pour calculer les performances du modèle n’était pas indépendant, nous avons considéré que le modèle avait été testé mais non validé.

4.2.0.6 Analyse des données

Nous avons comptabilisé les stratégies des données manquantes appliquées dans les études revues en utilisant une matrice. Puisque plusieurs stratégies pouvaient être appliquées dans une même étude, une publication pouvait être comptée plusieurs fois. Nous avons considéré la moyenne des AUROC pour décrire la performance de la publication lorsqu’une unique stratégie avait été appliquée. Nous avons décrit les données manquantes dans une seconde table, en comptant, pour chaque variable, le nombre de publications, de patients et de données manquantes. Nous avons rapporté ces publications en utilisant la méthode “Prediction model Risk Of Bias ASsessment Tool” (PROBAST)[116].

1. La checklist PRISMA est accessible dans les [Supplementary Materials](#) de l’article original

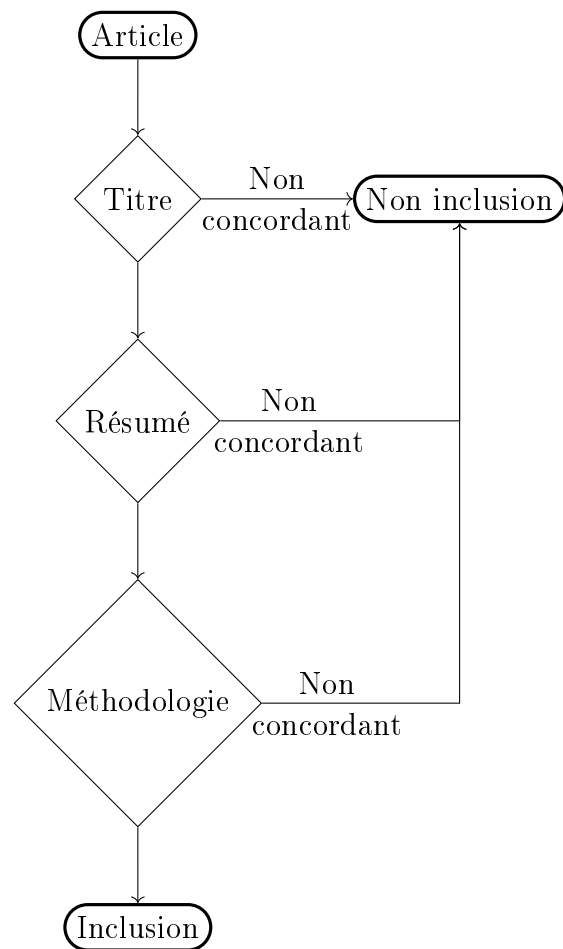


FIGURE 4.2 – Méthodologie d'inclusion des études

4.3 Résultats

4.3.1 Caractéristique des études

La revue de la littérature retrouve 628 publications. Après une première analyse des titres, nous avons sélectionné 213 publications pour l’analyse des articles. Au total, 141 publications ont été exclues sur la base des résumés (102) ou des textes complets (41). Ainsi, 72 textes complets² ont été inclus et analysés (figure 4.3). La moyenne (écart-type) du nombre de patients était 18 222 672 (1 119 747) au total, 5 635 600 (307 467) composent les jeux d’entraînement, et 699 702 (60 254) composent les jeux de validation. Seulement quatre publications ont suivi les recommandations **TRIPOD**. Aucune publication n’a déclaré un risque faible de biais.

Les variables cibles prédites par les modèles sont : un diagnostic dans 25 (35%), la mortalité dans 21 (29%), et la probabilité d’admission dans 21 (29%) publications (tableau 4.2). Les 12 (16%) publications restantes prédisent le score de tri ou l’administration de traitement. Huit (11%) publications prédisent plus d’une variable cible.

Cible	n. publications	Mortalité	Admission	Triage	Diagnostique	Traitement	\overline{AUROC}
Mortalité	21	14	3		2	2	0.88
Admission	21		17			1	0.83
Triage	4			4			0.92
Diagnostique	25				23		0.88
Traitement	8					5	0.85

TABLE 4.2 – Nombre de publications pour chaque type de variable cible (primaire ou secondaire)[24].

4.3.2 Modèles et stratégies d’imputation

En considérant le type de modèle utilisé pour produire la meilleure **AUROC**, la proportion d’études utilisant les algorithmes d’XGBoost et de la **RL** ont augmenté respectivement de 11% et 18% entre 2018 et 2021. L’utilisation des algorithmes d’apprentissage profond a augmenté de 39% de 2018 à 2019 puis a diminué de 52% en 2021 (figure 4.4). En 2021, 13 (44.8%) des 29 études ont utilisé XGBoost comme leur meilleur modèle et 10 (34.5%) ont utilisé la **RL**. Trente-sept (51.4%) études ont testé plusieurs algorithmes, et 35 (48.6%) n’en ont testé qu’un. Sur les 26 études ayant testé XGBoost, 16 (61.5%) l’ont testé de manière isolée, et 10 (38.5%) l’ont comparé à d’autres algorithmes. Il est intéressant de noter que sur les 16 études ayant testé seulement XGBoost, 9 (34.6%) ont utilisé une stratégie de gestion des données manquantes : une (11.1%) a remplacé les données manquantes par des valeurs fixes, trois (33.3%) ont supprimé les patients ayant des données manquantes, trois (33.3%) ont imputé les données manquantes par la moyenne, deux (22.2%) ont utilisé l’indicateur d’absence, et deux (22.2%) ont réalisé une imputation multiple automatique avant d’entraîner l’algorithme XGBoost.

Les stratégies appliquées aux données manquantes dans les études sont listées dans le tableau 4.3. Quarante-huit (67%) publications utilisent une seule stratégie (les chiffres en gras dans le tableau 4.3), 14 (19%) publications ont utilisé plus d’une stratégie, et 10 (14%) publications n’ont pas spécifié la stratégie utilisée.

La stratégie la plus fréquemment utilisée était l’exclusion de patients contenant des données manquantes (30 publications; 41%), suivie par l’imputation par la moyenne (17 publications; 23%), la prise en compte native (no-op; 13 publications; 18%), l’imputation algorithmique (6 publications; 8%) et enfin l’utilisation de constantes physiologiques/de référence (3 publications; 4%). Sept (10%) publications ont utilisé l’indicateur d’absence de valeur. L’utilisation de ces stratégies a objectivement évolué au cours du temps : la proportion d’études utilisant la méthode d’exclusion chute de 67% en 2017 à 33% en 2021, alors que la proportion d’études

2. La liste des textes complets est disponible dans les **Supplementary Materials** de l’article original

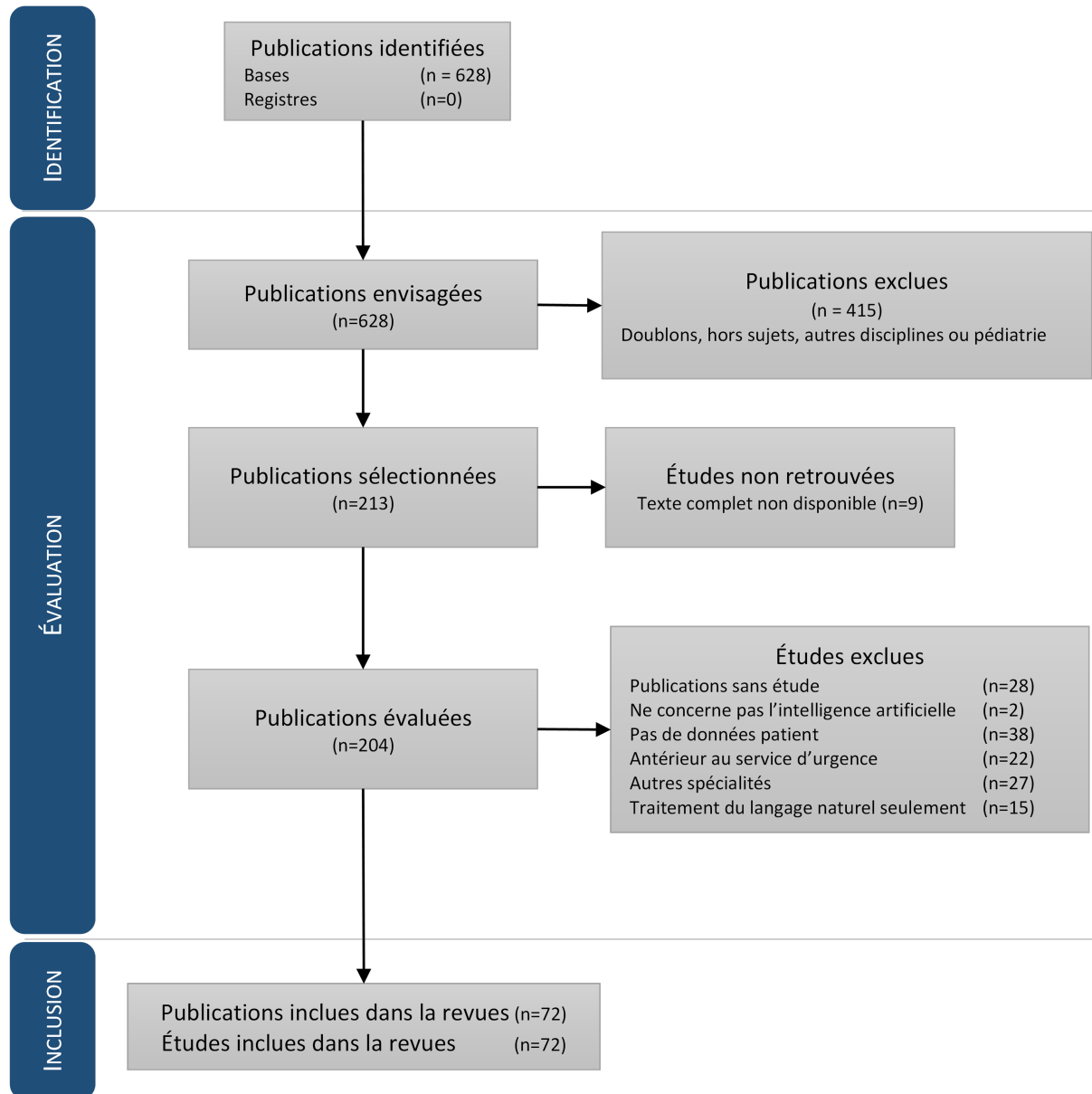


FIGURE 4.3 – Identification, analyse et inclusion des études via les bases de données et les registres[24].

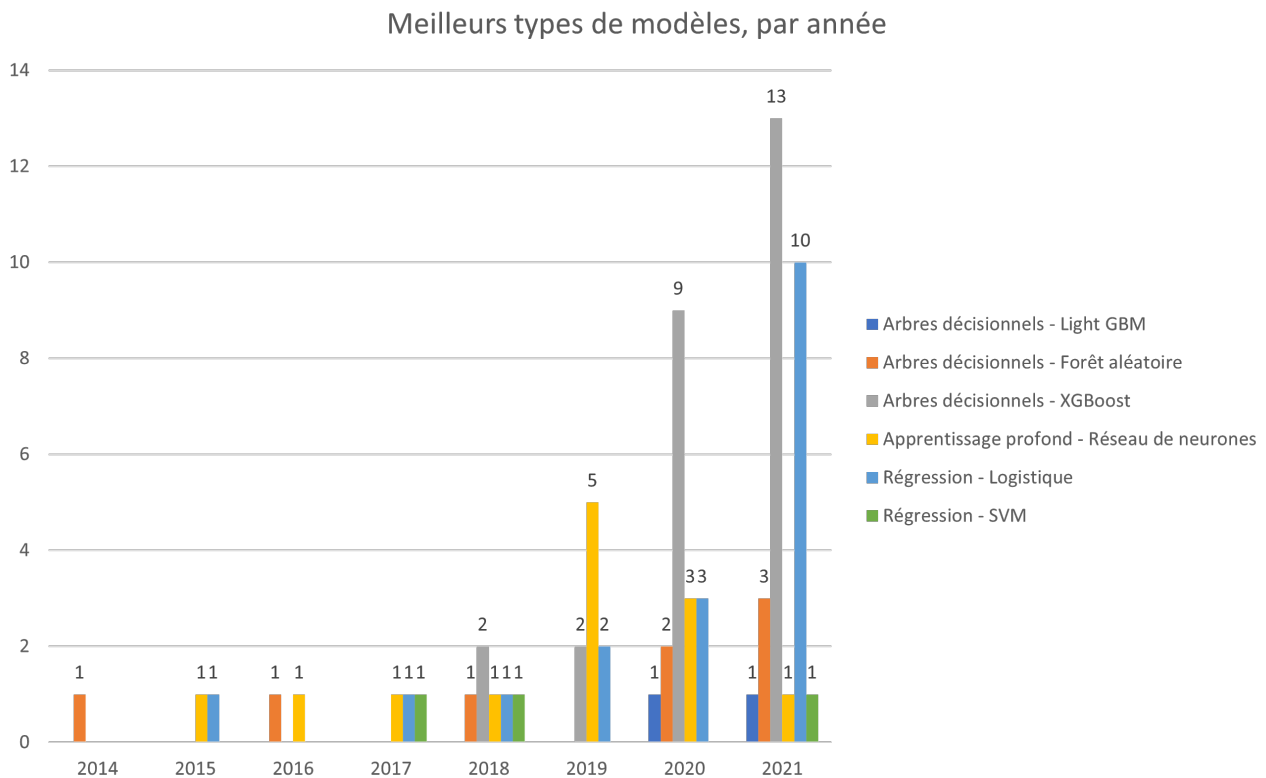


FIGURE 4.4 – Types d’algorithmes utilisés par année[24].

utilisant l’imputation augmente de 6% en 2020 à 17% en 2021 (figure 4.5 ; certaines publications utilisent plus d’une stratégie, le total excède donc 100%).

La variable la plus fréquemment utilisée était l’âge (70 publications ; 97%) et le sexe (66 publications ; 92%), alors que seulement 15 des deux groupes de publications (21% et 23%, respectivement) ont décrit leurs données manquantes. Les données manquantes pour les autres prédicteurs sont décrites dans le tableau 4.4.

Stratégie	n. publications	Exclusion	Moyenne	Indicateur d'absence	No-op	Constante	Imputation	Non disponible	\overline{AUROC}
Exclusion	30	23	2	2	1	1	1		0.87
Moyenne	17		11	2	1		1		0.84
Indicateur d'absence	7			0	2		1		0.85
No-op	13				9				0.86
Constante	3					2			0.89
Imputation	6						3		0.88
Non disponible	10							10	0.84

TABLE 4.3 – Stratégies appliquées aux données manquantes utilisées dans les publications relues[24].

Prédicteur	Publications n.(a)	Patients n. (b)	Description des données manquantes n.(c)	% (d)	Total des données manquantes n.(e)	% (f)
Age	70	24 554 988	15	21%	13 109	0.05%
Sexe	66	24 297 072	15	23%	1 614	0.01%
Origine	27	9 401 259	6	22%	30 691	0.33%
Mode d'attente	29	19 180 914	5	17%	15 509	0.08%
Motif de consultation	30	22 097 032	4	13%	10 319	0.05%
Temps de passage	12	16 046 251	0	0%	0	0.00%
Jour de consultation	13	14 327 915	0	0%	0	0.00%
Triage	29	17 241 664	3	10%	9 166	0.05%
GCS	35	16 776 293	5	14%	346 993	2.07%
FC	56	18 959 043	11	20%	82 330	0.43%
FR	56	18 959 885	11	20%	116 461	0.61%
PAS	56	18 960 674	9	16%	110 033	0.58%
PAD	47	16 887 450	7	15%	19 404	0.11%
Saturation en oxygène	45	7 678 163	7	16%	749 390	9.76%
Temperature	49	18 345 680	9	18%	336 436	1.83%
Échelle de douleur	13	1 132 598	2	15%	16 518	1.46%
Poids	10	238 607	3	30%	6 424	2.69%
Antériorité de consultation	16	8 226 874	1	6%	9 250	0.11%
Diagnostics CIM-10 antérieurs	38	6 369 720	4	11%	382	0.01%
Analyses hématologiques	24	1 595 185	6	25%	1 352	0.08%
Analyses biochimiques	25	1 837 740	3	12%	956	0.05%
Données textuelles	4	469 776	2	50%	6 777	1.44%
ECG	9	722 539	1	11%	0	0.00%

TABLE 4.4 – Description des données manquantes dans les publications[24]. (a) le nombre de publications dans lesquelles le prédicteur était utilisé. (b) le nombre total de patients inclus dans les publications utilisées par ce prédicteur. (c) et (d) le nombre et la proportion des publications où les données manquantes ont été décrites. (e) et (f) le nombre et la proportion de patients pour lesquelles au moins un prédicteur était absent. GCS : Glasgow Coma Scale; FC : fréquence cardiaque; FR : fréquence respiratoire; PAS : pression artérielle systolique; PAD : pression artérielle diastolique; SU : services d'urgences; CIM-10 : classification internationale des maladies, 10^{ème} version.

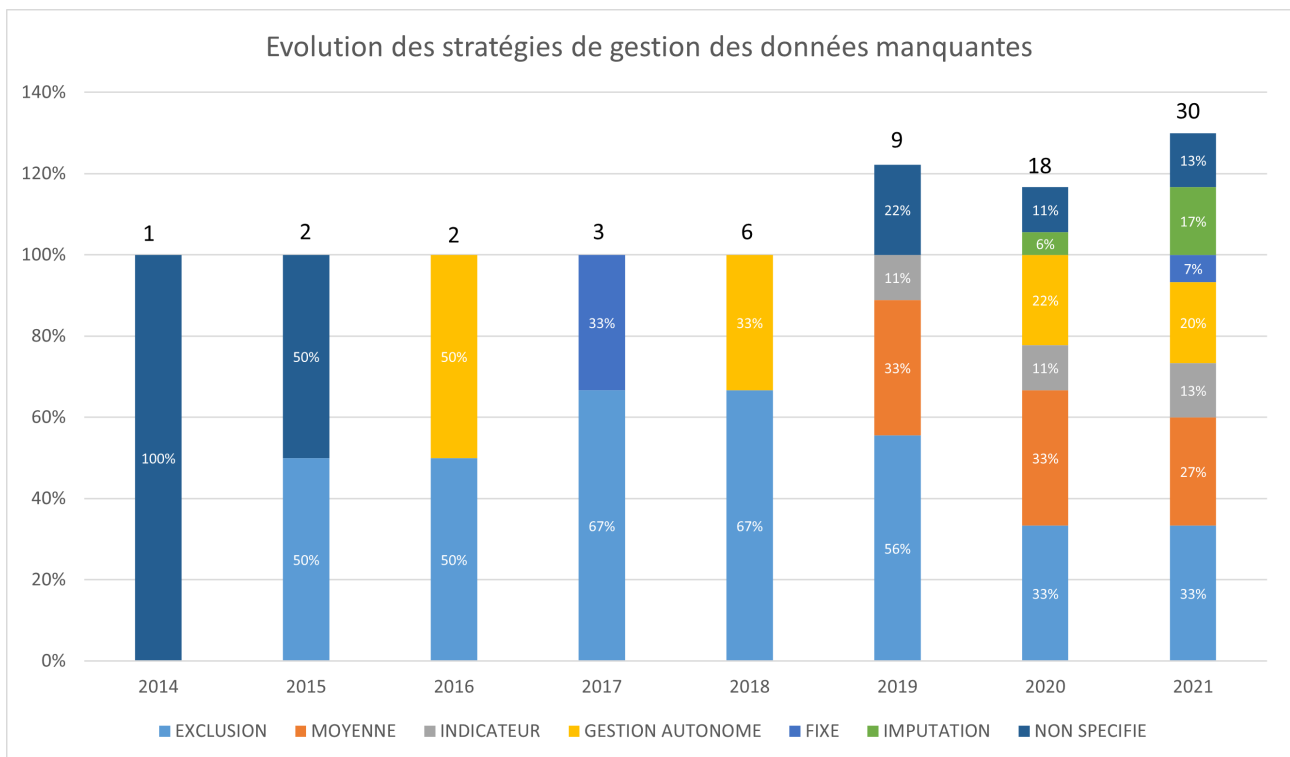


FIGURE 4.5 – Évolution de la fréquence d'utilisation des stratégies de données manquantes[24]. Certaines publications utilisant plus d'une stratégie, le total peut excéder 100%.

4.4 Discussion

L’objectif de cette revue de la littérature est d’étudier les stratégies de complétion de données appliquées au problème des **DNPC** dans les études d’apprentissage machine dans le domaine de la médecine d’urgence. Les modèles décrits dans les 72 publications prédisent une variété de variables cibles et diffèrent par leurs prédicteurs, leur type et leur taille de population. Ces études hétérogènes ne peuvent pas être comparées directement et toute inférence doit être formulée avec précaution. Néanmoins, la plupart des publications ont utilisé l’**AUROC** comme critère de jugement de la performance de leur modèle. En conséquence, nous avons utilisé l’**AUROC** comme comparateur dans notre étude, même si aucune conclusion ne peut être formulée sur ce critère.

4.4.1 Qualité de la description des stratégies

L’échec de description des données manquantes (plutôt que d’assumer d’avoir des données manquantes) peut être considéré comme une source importante de biais. Depuis 2014, plusieurs publications ont établi la méthodologie recommandée pour la description des données manquantes[117]. Pour tous les prédicteurs (autre que les données textuelles), moins de 25% des publications relues décrivent leur données manquantes. La quantité de données manquantes était probablement sous-estimée parce que sous-décrites, ce qui pourrait avoir introduit un biais significatif dans l’interprétation. Cette proportion est néanmoins plus élevée que celle reportées en 2019 par une revue de la littérature clinique en recherche pharmaceutique[118], où seulement 19.7% des publications sélectionnées décrivaient comment elles géraient leur données manquantes. Nous recommandons que les auteurs mentionnent explicitement la quantité de données manquantes et comment elles sont traitées. Cette information pourrait être résumée dans un schéma méthodologique explicite (figure 4.6).

4.4.2 Fréquence d’utilisation des stratégies

La stratégie de l’exclusion était la méthode la plus utilisée. Probablement parce que l’exclusion était la méthode la plus simple à implémenter : tous les patients contenant une donnée manquante étaient simplement exclus de l’étude. La technique de l’exclusion est la seule qui ne requiert aucune complétion. Son principal désavantage est de diminuer la taille de l’échantillon. De plus, cette technique peut conduire à un biais de sélection caché et pourrait signifier que la population n’est plus représentative de la population cible[62]. Si les patients étaient exclus à cause d’une variable **MCAR**, alors aucun biais n’a été inclus. Exclure des patients à cause de données **MCAR** est rare mais doit être documenté; et ce n’est pas souvent fait, particulièrement en médecine[120]. Quelques publications ont utilisé un seuil d’exclusion plus sophistiqué (*e.g.*, seulement les patients avec plus d’un certain nombre de données manquantes (*e.g.*, 50%) étaient exclus), ou une stratégie composite (figure 4.1). L’autre stratégie fréquente consiste à ne pas gérer directement les données manquantes (*i.e.*, stratégie “no-op”). La famille algorithmique des arbres décisionnels gère parfaitement les données manquantes en utilisant plusieurs méthodes (*e.g.*, le **Fractionnement par substitution**, le **fractionnement probabiliste**, la **propagation par blocs et l’incorporation d’attributs manquants (MIA)**)[121]. Cette stratégie est facile à implémenter et ne nécessite pas de pré-traitement mais oblige les chercheurs à utiliser un algorithme autonome sur la gestion des données manquantes.

Dans le contexte particulier des **DNPC**, la mesure de la glycémie capillaire à l’accueil d’un patient aux urgences est un très bon exemple à analyser. L’utilisation de la moyenne de la glycémie capillaire comme valeur d’imputation pourrait être influencée par un biais de sélection. Ce biais pourrait expliquer pourquoi la stratégie de l’utilisation de la “valeur moyenne” a la plus basse **AUROC** (figure 4.1). Les autres variables du même type incluent par exemple le taux d’alcool, le volume vésical, et quelques variables biologiques (*e.g.*, troponine, protéine C-réactive). Le biais de sélection associé aux **DNPC** est accepté en médecine d’urgence pour des raisons éthiques, organisationnelles et économiques. La mesure systématique et forcée de la glycémie capillaire pour tous les patients (particulièrement quand il s’agit d’une mesure invasive impliquant une ponction capillaire) n’est pas éthique lorsque qu’il est médicalement non justifié de la mesurer.

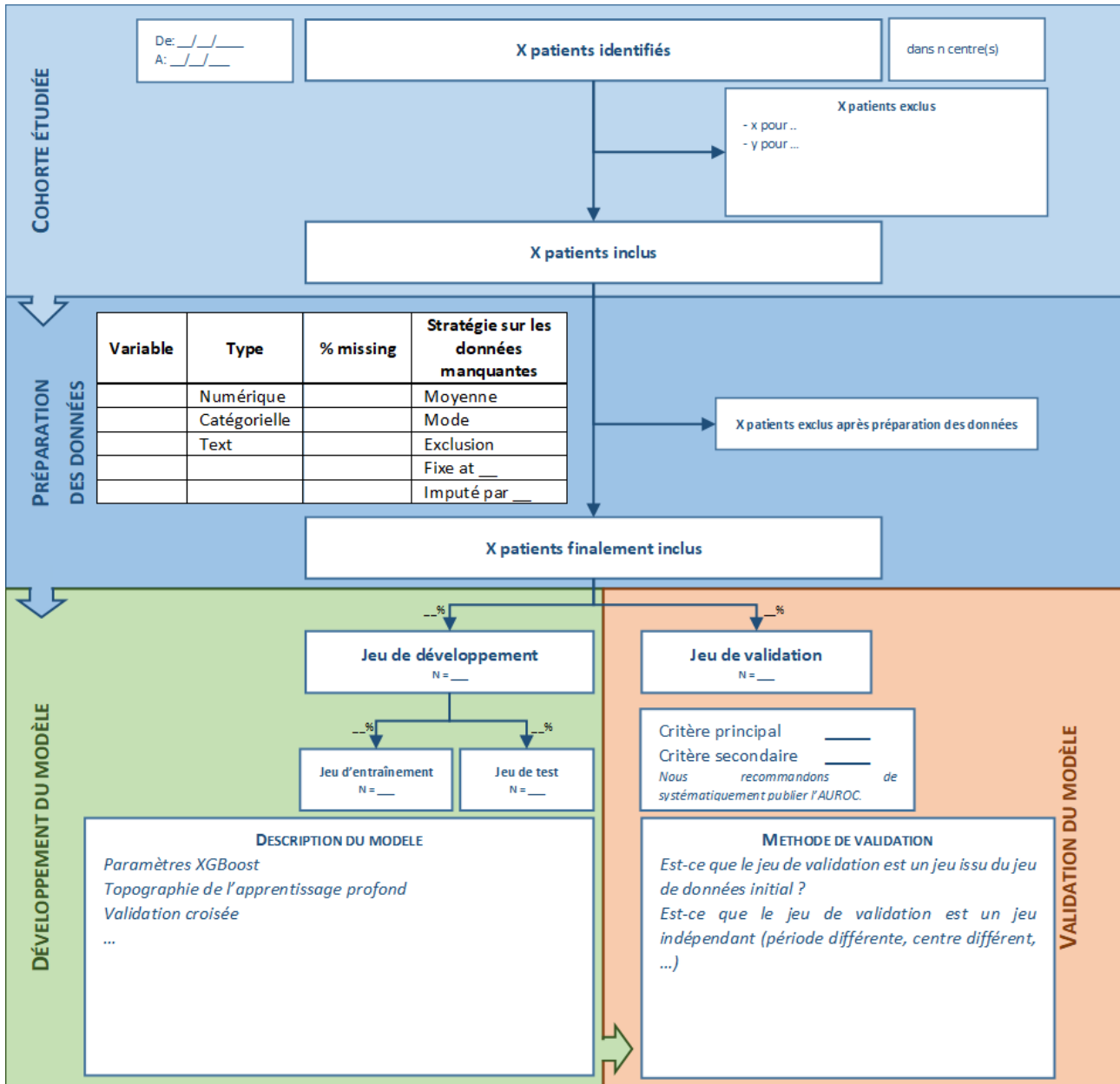


FIGURE 4.6 – Un schéma méthodologique explicite des données manquantes[24], adaptées de Martinez *et al.*[119]

L'imputation consiste à entraîner un modèle prédictif pour imputer les données manquantes à partir des données existantes dans le même jeu de données, en utilisant un algorithme d'équation chaînée. La technique la plus utilisée est [MICE](#)[122] qui a montré un avantage dans l'étude de FARIS *et al.* [123]. Cependant, plusieurs autres techniques existent, comme le traitement Gaussien ou encore le sur-échantillonnage minoritaire synthétique. La moyenne des [AUROC](#) des publications utilisant cette stratégie semblent confirmer les observations de FARIS *et al.* [123] Cependant, les stratégies d'imputation pourraient impliquer un pré-traitement plus complexe et requérir plus de temps et de ressources computationnelles, tout particulièrement lorsque le nombre de prédicteurs est important. Plusieurs options dans l'implémentation de [MICE](#) permettent d'accélérer l'imputation mais au prix d'une approximation plus importante. La stratégie de valeur fixée consiste à remplir les valeurs manquantes par une valeur constante (souvent une valeur physiologique ou de référence) pour chaque prédicteur. Il n'est pas attendu que cette méthode produise des résultats meilleurs que la technique d'imputation. La même valeur de remplissage est choisie pour tous les patients, sans considérer les autres variables. Ainsi, un patient diabétique ayant une glycémie capillaire manquante se verrait imputé d'une glycémie normale, ce qui n'est pas intuitif. La stratégie "non définie" correspond aux études n'ayant pas décrit la méthode qu'elles ont utilisée pour gérer leurs données manquantes.

La stratégie "non disponible" correspond aux études qui n'ont pas précisé la méthodologie utilisée pour gérer les données manquantes. Cependant, les algorithmes entraînés dans ces études sont parmi les moins bons, indiquant ainsi que ne pas décrire la méthode de traitement des données manquantes n'est pas un indicateur de qualité. La stratégie utilisée pour gérer les données manquantes était directement liée au type d'algorithme choisi. Plus spécifiquement, les chercheurs qui avaient choisi un algorithme autonome pour le pré-traitement des données manquantes n'avaient pas à gérer eux-même les données manquantes. Quelques chercheurs, qui avaient choisi d'utiliser XGBoost, ont préféré gérer eux-même les données manquantes (en fonction des données qui manquaient) ; ce choix est spécifique de la variable elle-même et de l'expertise du médecin-chercheur.

4.4.3 Évolution au cours du temps

La gestion des [DNPC](#) a changé au cours du temps, comme décrit dans la [figure 4.5](#). Nous avons distingué trois périodes :

1. Avant 2016 : sans stratégie spécifique
2. De 2016 à 2018 : avec une seule stratégie par étude
3. Après 2018 : combinaison de stratégies

Pendant la première période, les publications n'ont pas décrit leur stratégie ou ont simplement utilisé la méthode de l'exclusion. Pendant la seconde période, une seule et unique stratégie était utilisée : l'exclusion, l'utilisation d'une constante physiologique pour tous les patients, ou l'utilisation de l'algorithme natif (*i.e.*, no-op). Pendant la dernière période, les chercheurs ont choisi des méthodes plus complexes et les ont combinées : marquage des patients ayant des [DNPC](#), utilisation de la moyenne, ou imputation des données manquantes à partir des patients sans données manquantes. La proportion de publications utilisant des stratégies simples (*i.e.*, moyenne, no-op et exclusion) a diminué, et inversement la proportion de publications utilisant une stratégie plus compliquée (*i.e.*, l'imputation par algorithme indépendant) a augmenté à 17% en 2021.

4.4.4 Comparaison de la performance des stratégies

Les données collectées ne nous ont pas permis d'affirmer qu'une stratégie était toujours plus efficace qu'une autre. Cependant, en se basant sur l'[AUROC](#), la stratégie du no-op a produit les meilleurs résultats. [SAEM](#)[124] est un algorithme plus récemment développé et n'était pas utilisé dans les publications revues ; il pourrait réaliser une imputation meilleure que [MICE](#) avant une [RL](#)[110]. Malgré le fait que la stratégie basée sur une constante et l'exclusion produit de bons résultats, elle devient de moins en moins fréquemment utilisés.

Le comparatif de PEREZ-LEBEL *et al.* [125] des différentes stratégies pour gérer les données manquantes dans les bases de données médicales réelles montre que :

1. Les arbres décisionnels boostés par **MICE** donnent les meilleurs résultats à un coût faible (puisque c'est une stratégie no-op),
2. L'indicateur d'absence est informatif,
3. Utiliser **MICE** sur de larges bases de données est coûteux en ressources.

AUSTIN *et al.* [126] ont démontré qu'utiliser **MICE** sur des données cliniques était particulièrement rentable, notamment lorsque le mécanisme des données manquantes pouvait être considéré comme ignoré.

Dans les publications revues ici, les stratégies des données manquantes ont été appliquées sur l'ensemble du jeu de données. Nous faisons l'hypothèse que la meilleure stratégie est dépendante de la variable. Sur la base de cette revue, et sur notre expérience, nous recommandons de garder la totalité des patients (*i. e.*, ne pas appliquer la stratégie de l'exclusion) si l'étude en question couvre tous les patients se présentant aux urgences. Ensuite, les médecins experts de la population incluse pourraient choisir une stratégie spécifique variable par variable. Par exemple, quelques variables pourraient être complétées en suivant des règles logiques basées sur l'expertise médicale (*e.g.*, le score de Glasgow pourrait être évalué à 15/15 pour tous les patients présentant un traumatisme mineur). Si une **RL** peut être réalisée, nous recommandons d'utiliser **SAEM** au jeu de données contenant le reste des données manquantes. Dans les autres cas, nous recommandons d'utiliser les arbres décisionnels boostés qui gèrent nativement les données manquantes (*e.g.*, **XGBoost**). Si le mécanisme des données manquantes ne peut être considéré comme ignoré, l'algorithme **MICE** devrait être sérieusement considéré car il inclut le paramètre du mécanisme des données manquantes dans son modèle. Cette hypothèse mérite d'être testée spécifiquement dans une étude dédiée.

4.4.5 Améliorer le rapport des données manquantes et des stratégies appliquées

Les publications n'ont pas toujours décrit leurs données manquantes, et la qualité du report de leurs données varie d'une étude à l'autre ; ce problème pouvant être à l'origine d'un biais d'interprétation. Nous suggérons de formaliser la description des données manquantes dans la section "Matériels et méthodes" des études.

TRIPOD est la méthode de référence pour construire un modèle prédictif multivarié[113]. Sa section 9 recommande de décrire les données manquantes. Cependant, cette recommandation n'a pas été clairement formalisée. Nous recommandons alors d'inclure une table dédiée à la description des données manquantes et de la stratégie appliquée à chaque variable (figure 4.6).

Dans les études avec une variable cible déséquilibrée (*e.g.*, hospitalisation à 70% versus sortie des urgences à 30%), l'**AUROC** pourrait ne pas être le meilleur critère de jugement pour la performance du modèle. Dans ces cas, il pourrait être pertinent de calculer l'aire sous la courbe precision-rappel. Ceci étant dit, les chercheurs devraient indiquer systématiquement, en plus, l'**AUROC**, puisque ce critère est le plus commun.

La présentation de ces stratégies variait en fonction des publications. Même si toutes les publications ont rapporté le développement ou l'utilisation du modèle prédictif impliquant un apprentissage machine supervisé, les diagrammes de flux n'étaient pas standardisés. De plus, les données manquantes n'étaient pas systématiquement décrites, et le ratio de fractionnement entraînement-test n'était pas toujours mentionné. Sur ces constats, nous proposons d'ajouter à la liste de vérification de **TRIPOD** un diagramme basé sur l'étude de MARTINEZ *et al.* [119] (figure 4.6). Avoir ce type de diagramme dans une publication d'apprentissage machine (même en appendice) permettrait d'aider le lecteur.

4.4.6 Limites de l'étude

Cette étude a plusieurs limites. Premièrement, nous avons cherché uniquement dans la base PubMed ; les bases comme Google Scholar ou IEEE Xplore n'ont pas été utilisées. Néanmoins, PubMed indexe tous les journaux d'intérêt dans le champ de la médecine d'urgence, et le nombre et la variété des publications retournées par PubMed étaient suffisants pour une revue de cette nature. La requête similaire retourne plus de 200 000 résultats dans Google Scholar et plus de 1 000 dans IEEE Xplore. Deuxièmement, nous n'avons cherché que des publications rédigées en anglais et nous pourrions avoir raté quelques publications d'intérêt dans d'autres langues. Troisièmement, nous n'avons pas considéré les publications antérieures à l'année 2000. Cependant, les recherches sur la gestion des données manquantes n'ont pas été développées avant les années 2000, et la plupart des développements significatifs ont été fait sur les vingt dernières années. Quatrièmement, nous avons conduit une revue systématique des méthodes utilisées dans les études de médecine d'urgence, plutôt que des résultats. Nous nous sommes concentrés sur la gestion des données manquantes dans des études hétérogènes qui ne peuvent être directement comparées sans prendre un risque évident de biais d'analyse. Cette étude ne revoit pas toute la littérature sur les données manquantes, les données manquantes non ignorables ou toutes les méthodes pour apprendre les mécanismes aboutissant à l'absence de données. Cinquièmement, nous nous sommes limités au périmètre de la médecine d'urgence en raison des spécificités qu'ont les services d'urgence comparativement aux services d'hospitalisation conventionnelle. Plus spécifiquement, nous avons considéré le nombre de patients pris en charge, la variabilité du flux quotidien de patients, la durée de passage très courte dans un service d'urgence et la grande variabilité de motifs de consultation. Néanmoins, quelques mécanismes similaires de données manquantes pourraient être observés dans d'autres services. Sixièmement, nous avons utilisé l'**AUROC** comme comparateur, même si d'autres indicateurs[127] devraient être considérés (*e.g.*, exactitude, rappel, spécificité, précision, F-score, le coefficient de corrélation de Matthews, l'erreur absolue moyenne, ou le score de Brier).

4.4.7 Force de l'étude

Les **DNPC** constituent les “données manquantes” dans les études d'apprentissage machine dans le champ de la MU. Même si les **DNPC** ont été prises en compte rigoureusement depuis 2020, les stratégies utilisées ne sont pas exhaustives, systématiques ou homogènes. Les stratégies d'imputation semblent avoir permis d'obtenir les meilleurs résultats, mais qui requièrent plus de temps et plus de ressources de calcul. Cependant, compte-tenu de cette hétérogénéité des études revues, nous ne pouvons pas affirmer quelle stratégie serait meilleure. Pour améliorer la qualité et la comparabilité de ces études, nous recommandons de fournir la checklist **TRIPOD** pour toute nouvelle publication, de résumer le processus d'apprentissage machine dans un diagramme méthodologique spécifique, et de systématiquement reporter l'**AUROC** (même si ce n'est pas le critère de jugement principal).

Chapitre 5

Comment expliquer la prédiction de 3P-U ?

Points clés

Le modèle **3P-U** donne une prédiction, mais pas les raisons pour lesquelles il produit cette réponse précisément. Or dans le domaine de la médecine d'urgence, faire confiance à une prédiction sans pouvoir en appréhender les raisons peut diminuer l'adhésion à celui-ci.

Cette contribution a été publiée sous la référence: Émilien ARNAUD *et al.* "Development and Clinical Interpretation of an Explainable AI Model for Predicting Patient Pathways in the Emergency Department : A Retrospective Study". In : *Journal of Medical Internet Research* UNDER REVIEW (21 sept. 2023).

Dans le [chapitre 3](#), nous avons construit un modèle utilisant les données structurées et non structurées pour prédire l’hospitalisation du patient. Comme nous avons vu dans l’introduction sur l’explicabilité ([sous-section 2.4.2](#)), les techniques d’explicabilité des modèles prédictifs diffèrent pour les modèles incluant des données structurées et des données non structurées. L’objectif de cette étude est de construire un modèle explicable à partir des données structurées uniquement.

5.1 Matériel et méthodes

5.1.1 Modèle 3P-U

5.1.1.1 Design de l’étude, participants et sources de données

Nous avons conduit une étude rétrospective, monocentrique, sur tous les patients majeurs consultant au service d’accueil des urgences polyvalentes adultes du [CHU-AP](#) entre 2015 et 2016. Les données proviennent directement du dossier informatisé du patient (Résurgences[©], Berger Levrault, Boulogne-Billancourt, France). Nous avons exclu les patients n’ayant aucune donnée de triage, ou qui ont été transférés au [SAUV](#) immédiatement après leur arrivée. Conformément à l’organisation des urgences en France, le patient bénéficie d’une évaluation initiale par l’[IOA](#) ([sous-sous-section 1.2.0.1](#)) suivant la classification [FRENCH](#) ([tableau 1.1](#))

Les modèles ont été implémentés par une équipe interdisciplinaire d’urgentiste du [CHU-AP](#) et de chercheurs en [IA](#) de l’[Université Picardie Jules Verne \(UPJV\)](#). Seules les données structurées collectées à l’accueil des urgences pour chaque patient ([tableau 3.2](#), variables #1-24) ont été utilisées comme prédicteurs pour entraîner un modèle binaire visant à expliquer l’orientation à la sortie des urgences (*i.e.*, hospitalisation dans un service ou sortie des urgences). Les patients pour lesquels il existait des données manquantes n’ont pas été exclus. Nous avons considéré que les données d’accueil étaient des [DNPC](#) ([sous-section 4.4.7](#)) parce qu’elles apparaissaient avoir une faible pertinence clinique et/ou qu’il n’était pas éthique de les collecter. Nous avons donc décidé de remplacer les données manquantes par zéro ou une valeur physiologique ([tableau 3.2](#), variables #1-24)

5.1.1.2 Apprentissage machine et mesure de la performance des modèles

Pour valider les performances du meilleur modèle, 20% du jeu de données initial ont été réservés comme jeu de test, de manière randomisée et stratifiée sur la variable à expliquer. En utilisant les 80% restants, un ensemble de modèles ([RNP](#), [FA](#), [RL](#), et [Plus Proches Voisins \(PPV\)](#)) a été évalué en utilisant une validation croisée à cinq itérations. Le pré-traitement des données a consisté en une normalisation des variables numériques et de l’encodage binaire des variables catégorielles. Compte tenu de la faible quantité d’information médicale disponible à l’accueil des urgences, nous avons utilisé l’ensemble des données disponibles à cette étape de la prise en charge. L’architecture du [RNP](#) consiste en un perceptron multicouche composé de quatre couches denses et trois couches de réinitialisation. Cette architecture a été définie à partir de l’expérimentation du modèle complet ([sous-section 3.2.2](#)). Contrairement au modèle initial, les hyperparamètres (*i.e.*, nombre de neurones dans les trois premières couches denses et le taux de réinitialisation des poids) ont été optimisés en utilisant une recherche aléatoire sur la première itération, utilisant l’erreur quadratique moyenne comme fonction de coût. La fonction d’activation était “ReLU” (*i.e.*, rectified linear unit) dans les couches cachées, et une fonction sigmoïde dans la couche dense de sortie. Le processus d’entraînement était limité à dix *epochs* en utilisant l’optimiseur Adam[128]. L’architecture des [PPV](#) était entraînée jusqu’à dix voisins en utilisant la métrique Manhattan comme fonction de coût. L’architecture de [FA](#) était paramétrée pour utiliser 100 estimateurs, le critère Gini, “min_samples_split = 2”, et l’utilisation de la racine carré comme meilleur diviseur. L’architecture de [RL](#) a été paramétrée en utilisant une pénalité de type L2, et le solveur à mémoire limité Broyden-Fletcher-Goldfarb-Shanno[129]

La performance de chaque modèle était évaluée en utilisant la moyenne (écart-type) de l’[AUROC](#) des cinq itérations de la validation croisée du jeu d’entraînement (80%). Un test de Friedman[94] a été utilisé pour

comparer respectivement l'**AUROC** des modèles, utilisant $p < 0.05$ comme risque de première espèce. Enfin, les mesures de performances étaient évaluées une fois sur le meilleur modèle de chaque architecture en utilisant le jeu de test (20%). Les performances du meilleur modèle de chaque modèle ont été évaluées au regard du rappel, de la précision et de l'exactitude, en appliquant un seuil décisionnel (défini par le meilleur F1-score). Les modèles ont été entraînés et évalués en utilisant la librairie Python Scikit-Learn 0.21[106] et Keras 2.11.0[52]

5.1.2 Techniques d'explicabilité

Les techniques d'**XAI** utilisées dans cette études sont **SHAP** et **XAI** (sous-section 2.4.2). L'implémentation Python de **SHAP** 0.35.0[130] a été utilisée. L'implémentation Python **PDP**box 0.21[131] a été utilisée.

5.1.3 Interprétation médicale

Après l'application des techniques d'explicabilité, la pertinence médicale des résultats générés par **SHAP** et **PDP** a été évaluée par les médecins urgentistes participants à l'étude. Cette critique a été menée à partir de leur expertise clinique pour évaluer la conformité des résultats d'explicabilité avec la pratique médicale et la physiopathologie sous-jacente. Ces commentaires ont eu un rôle primordial dans la validation de la cohérence et de la pertinence clinique des résultats obtenus.

5.1.4 Autorisation réglementaire

Cette étude a été autorisée par la Commission Hors Lois Jardé du **CHU-AP** sous la référence PI2019_843_0066

5.2 Résultats

5.2.1 Modèle 3P-U

5.2.1.1 Design de l'étude, participants et sources de données

Entre 2015 et 2018, 351 019 patients ont consulté aux urgences du CHU-AP. Parmi eux, 48 053 (13.7%) ont été exclus car aucune donnée n'avait été enregistrée à l'accueil des urgences ou parce qu'ils ont été directement orientés vers la **SAUV**. Ainsi, 302 966 patients ont été inclus dans l'étude. La décision médicale finale est de 99 340 (32.8%) hospitalisations et 203 326 (67.2%) sortie des urgences. Les principales caractéristiques sont décrites dans le [tableau 5.1](#).

Caractéristique	Population	Imputation	Manquant
Caractéristiques démographiques			
Nombre de patients	302 966 (100%)		
Âge	51 (± 22)	–	0 (0%)
Genre		–	0 (0%)
Homme	156 621 (51.7%)		
Femme	146 345 (48.3%)		
Caractéristiques de triage			
Fréquence cardiaque (/min)	85 (± 18)	Fixé : 80	51 726 (17%)
Pression artérielle systolique (mmHg)	136 (± 24)	Fixé : 120	82 665 (27%)
Pression artérielle diastolique (mmHg)	77 (± 24)	Fixé : 80	82 665 (27%)
Saturation pulsée en oxygène (%)	99 (± 2)	Fixé : 100	95 (<1%)
Température corporelle (°C)	36.6 (± 0.8)	Fixé : 37.4	75 789 (25%)
Glycémie capillaire (mmol/L)	7.72 (± 4.8)	Fixé : 5.0	245 801 (81%)
Cétonémie capillaire (mmol/L)	1.13 (± 3.4)	Fixé : 0.0	300 153 (99%)
Débit d'oxygène administré (L/min)	0.65 (± 4.5)	Fixé : 0	237 405 (78%)

Caractéristique	Population	Imputation	Manquant
Hémoglobine capillaire (dg/dL)	11.4 (\pm 3.1)	Fixé : 12.0	290 510 (96%)
Taux d'alcool expiré corrigé (g/L)	1.87 (\pm 0.8)	Fixé : 0.0	295 898 (97%)
Volume vésical (mL)	366 (\pm 320)	Fixé : 0.0	301 939 (99%)
Niveau de douleur	3 (\pm 3)	Fixé : 0	49 422 (16%)
Échelle de triage FRENCH		Plus fréquent	4 508 (1%)
1	930 (0.1%)		
2	15 174 (5.1%)		
3	136 839 (45.9%)		
4	85 235 (28.7%)		
5	60 280 (20.2%)		
Tests urinaires			
Sang		Fixé : 0	301 020 (99%)
0	621 (31.9%)		
Traces	462 (23.7%)		
+	193 (9.9%)		
++	233 (12.0%)		
+++	429 (22.0%)		
++++	8 (0.5%)		
Nitrites		Fixé : 0	301 020 (99%)
0	1 683 (86.6%)		
+	218 (11.2%)		
++	25 (1.3%)		
+++	13 (0.7%)		
++++	5 (0.2%)		
Leucocytes		Fixé : 0	301 020 (99%)
0	1 224 (62.9%)		
Traces	227 (11.7%)		
+	260 (13.4%)		
++	92 (4.7%)		
+++	142 (7.3%)		
Glucose		Fixé : 0	301 020 (99%)
0	1 769 (91.0%)		
Traces	81 (4.2%)		
+	21 (1.1%)		
++	55 (2.8%)		
+++	16 (0.8%)		
++++	1 (<0.1%)		
Cétone		Fixé : 0	301 020 (99%)
0	1 447 (74.5%)		
Traces	184 (9.5%)		
+	129 (6.6%)		
++	99 (5.1%)		
+++	51 (2.6%)		
++++	31 (1.7%)		
Orientation à la sortie des urgences			
Hospitalisation	99 340 (32.8%)		
Sortie	203 626 (67.2%)		

Caractéristique	Population	Imputation	Manquant
-----------------	------------	------------	----------

TABLE 5.1: Principales caractéristiques de la population d’étude. L’échelle de tri [FRENCH](#)[11] est détaillée dans la [tableau 1.1](#). Les données sont exprimées en moyenne (écart-type) ou nombre (pourcentage).

5.2.1.2 Apprentissage machine et mesure de la performance des modèles

En utilisant le jeu d’entraînement (242 372 patients, 80%), la moyenne (écart-type) de l’**AUROC** était de 83.0% (0.1%) avec le modèle de **RNP**, 71.7% (0.2%) avec le modèle de **FA**, 67.2% (0.2%) avec le modèle des **PPV**, et 71.5% (<0.1%) avec le modèle de **RL** ([section 5.2.1.2](#)). En utilisant le jeu de test (60 593 patients, 20%) et le meilleur modèle de **RNP**, l’**AUROC** est de 83.2%. Pour les autres modèles, l’**AUROC** est de respectivement 71.5%, 67.1%, et 61.8% pour la **RL**, les **PPV**, et la **FA** ([figure 5.2.1.2](#) et [tableau 5.2.1.2](#)). En utilisant le meilleur seuil décisionnel (0.35) déterminé par le meilleur F1-score (0.67), le rappel est de 75.7%, la précision de 59.7%, et l’exactitude de 75.3% ([figure 5.2.1.2](#)). Le nombre optimisé de neurones pour les trois couches denses du RNP est de respectivement 192, 128 et 48. Le taux de réinitialisation optimisé est identique pour les trois couches à 0.2.

AUROC	RNP	RL	PPV	FA
Itération 1	83.1%	71.4%	66.9%	71.5%
Itération 2	82.8%	71.5%	67.2%	71.6%
Itération 3	83.1%	71.5%	67.6%	71.9%
Itération 4	83.2%	71.4%	67.3%	71.4%
Itération 5	82.9%	71.6%	67.0%	72.4%
Moyenne (écart-type)	83.0% (0.1%)	71.5% (< 0.1%)	67.2% (0.2%)	71.7% (0.2%)
p-value	0.002			

TABLE 5.2 – Résultat des différentes itérations d’apprentissage et comparaison des différentes moyennes à l’aide du test de Friedman.

5.2.2 Techniques d’explicabilité

La [figure 5.3](#) résume les variables explicatives les plus significatives détectées par **SHAP** et montre leur contribution additive à la probabilité d’être hospitalisé. Parmi les variables explicatives ayant les contributions les plus importantes, l’âge, la fréquence cardiaque, le niveau 2 de l’échelle de tri **FRENCH**, et le motif de consultation “traumatismes sévères multiples” augmentent significativement la probabilité d’être hospitalisé. Puisque l’âge et la fréquence cardiaque sont des variables quantitatives, le graphique indique à quel point ces variables sont impliquées dans la décision du modèle : un jeune patient (couleur bleu) a une probabilité d’admission plus faible (valeur **SHAP** négative, points situés à gauche), qu’un patient plus âgé (couleur rouge). Inversement, les variables explicatives comme le mode d’attente “en chaise”, le motif de consultation “douleur

	RNP	RL	PPV	FA
AUROC	83.2%	71.5%	67.1%	71.8%
Exactitude	77.5%	77.2%	73.0%	77.6%
Précision	68.9%	69.4%	60.6%	70.4%
Rappel	57.2%	54.7%	50.0%	54.8%

TABLE 5.3 – Performances des meilleurs modèles sur le jeu de test

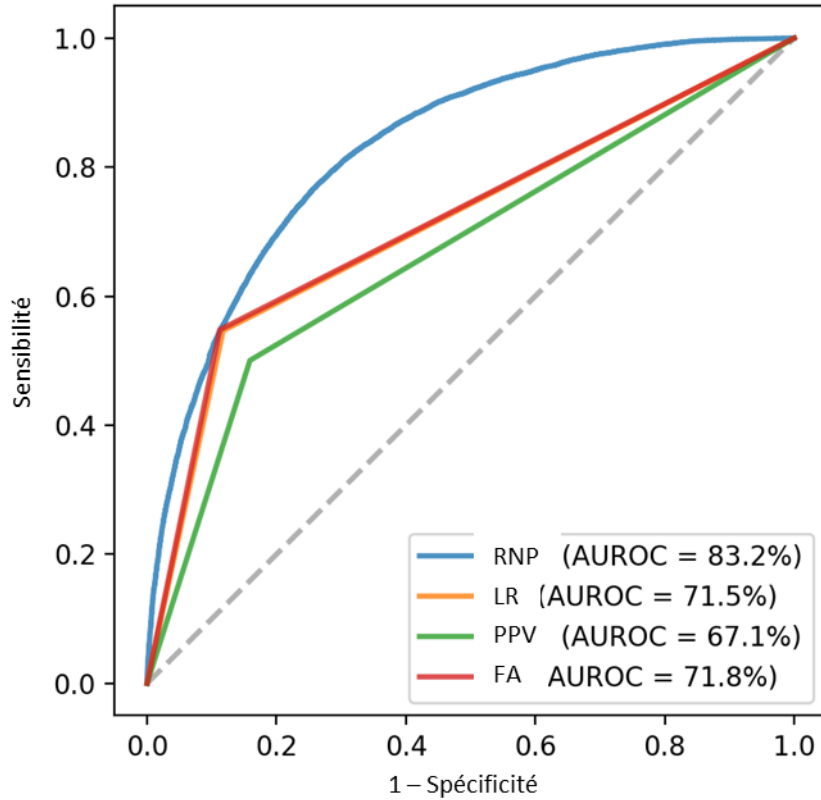


FIGURE 5.1 – Courbe ROC des modèles en utilisant le jeu de test

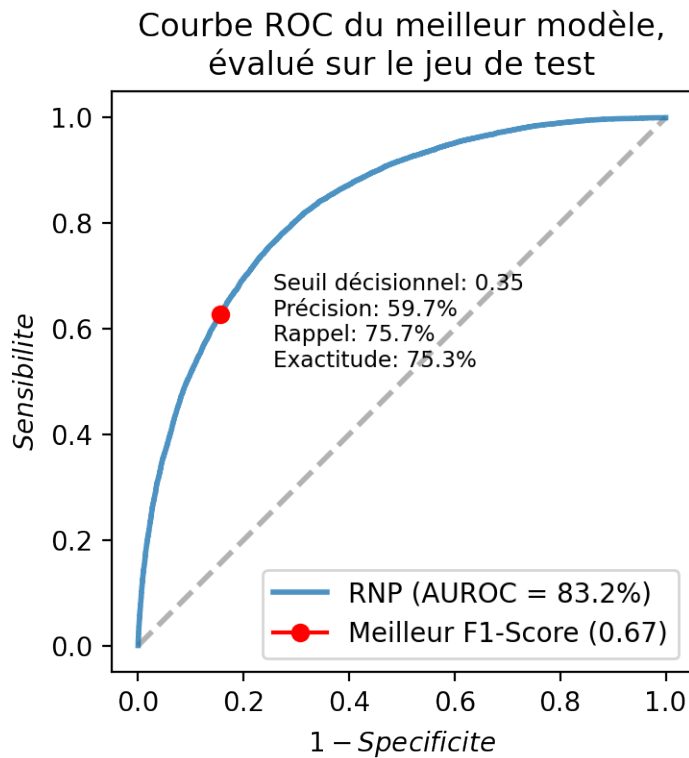


FIGURE 5.2 – Courbe ROC du meilleur modèle de RNP, et le seuil décisionnel pour le meilleur F1-score.

thoracique” ou “malaise” contribuent négativement de manière importante à la probabilité d’hospitalisation. Même si les variables (binaires) de mode d’arrivée “moyen personnel”, des niveaux 4 et 5 de l’échelle de tri **FRENCH**, et de la raison de consultation “traumatisme de la cheville” ont une contribution négative faible lorsqu’elles sont positives, leur valeur négative contribue positivement à la probabilité d’admission (*e.g.*, “ne pas venir aux urgences par ses moyens personnels” augmente la probabilité d’admission).

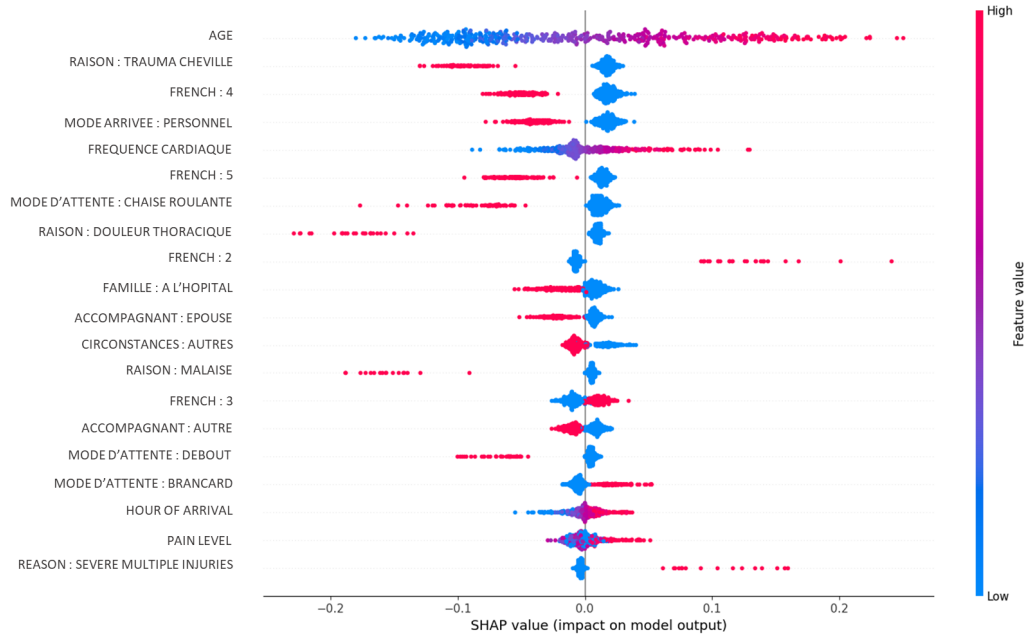


FIGURE 5.3 – Le graphique à point résumant l’explication globale de **SHAP** pour le modèle d’hospitalisation : plus la valeur **SHAP** est élevée (pour les variables quantitatives), plus la probabilité d’hospitalisation est élevée, et *vice versa*. Pour les variables binaires ou les variables catégorielles binarisées, la couleur rouge et bleue correspondent respectivement aux valeurs 1 et 0. L’échelle de tri **FRENCH** est détaillée dans le [tableau 1.1](#).

SHAP fournit également une explication locale de la prédiction des cas individuels en décrivant l’importance et la direction de la contribution de chaque variable explicative à la probabilité finale (rouge : augmentation de la probabilité d’hospitalisation, bleu : diminution de la probabilité d’hospitalisation), ainsi que la pertinence (*i.e.*, la longueur de la barre). Par exemple, la [section 5.2.2](#) montre respectivement l’explication d’un cas réellement positif et d’un cas réellement négatif. Dans les deux cas, la prédiction commence à une valeur Shapley de base à 0.378. Ce plancher correspond à la moyenne de probabilité d’hospitalisation dans le jeu d’entraînement. D’après les contributions générales dans le graphique général ([figure 5.3](#)), l’âge du patient était une des variables les plus importantes pour les deux cas. Dans l’exemple (a) du vrai négatif (*i.e.*, une prédiction correcte de non hospitalisation), l’âge et, moins significativement, le motif de consultation “dyspnée à moins de 32 respirations par minute” sont décisifs : leur contribution sont respectivement -0.2 et -0.12. Dans l’exemple (b) du vrai positif (*i.e.*, une prédiction correcte d’hospitalisation), le niveau 2 de l’échelle de tri **FRENCH** le plus largement, l’âge de 79 ans, et le motif de consultation “altération de l’état général” ont contribué positivement à la probabilité d’hospitalisation respectivement de +0.11, +0.1 et +0.12.

Dans le cas spécifique des variables quantitatives, **PDP** quantifie la contribution à la probabilité comme une fonction d’amplitude. En générant un graphique pour chaque variable explicative quantitative (dans l’optique d’évaluer la tendance d’impact sur la probabilité), nous avons appris que les variables explicatives qui augmentent le plus la probabilité d’hospitalisation sont l’âge ([section 5.2.2](#)) et la fréquence cardiaque ([figure 5.2.2](#)) ; ce qui confirme les résultats de **SHAP**. Les autres variables (*e.g.*, la cétonémie capillaire ou la glycémie capillaire) contribuent aussi largement positivement à la probabilité d’admission quand les valeurs extrêmes (non imputées) sont considérées : 59 mmol/L de cétonémie capillaire et 108 mmol/L de glycémie capillaire correspondent à

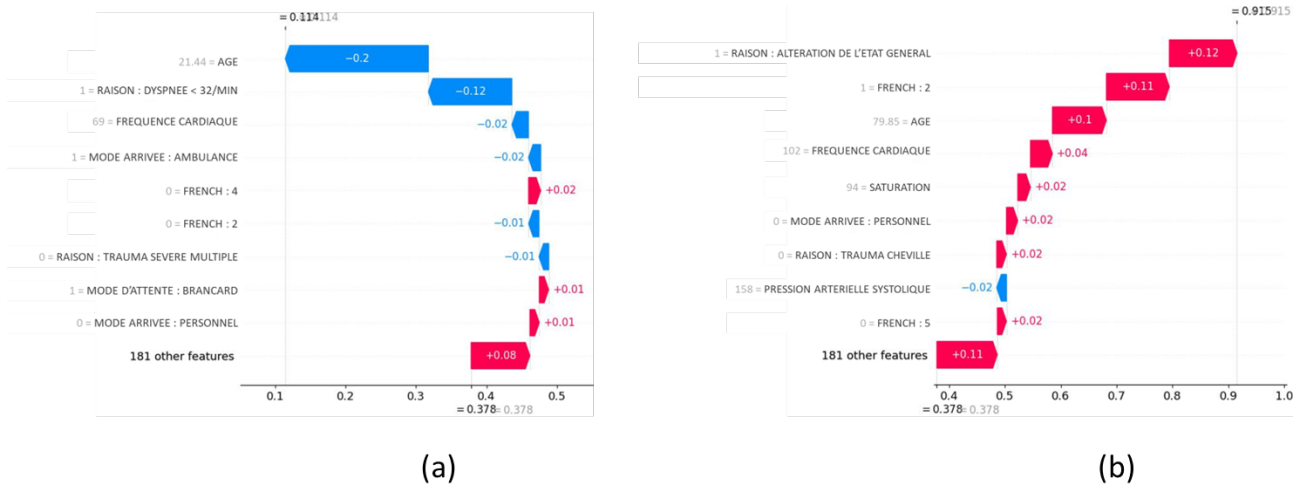


FIGURE 5.4 – Explication individuelle d’un cas réellement positif (a) et d’un cas réellement négatif (b) en utilisant SHAP.

une hospitalisation quasi-certaine (figure 5.2.2 et figure 5.2.2). En conséquence, considérer ces variables reste informatif alors même qu’elles sont majoritairement manquantes.

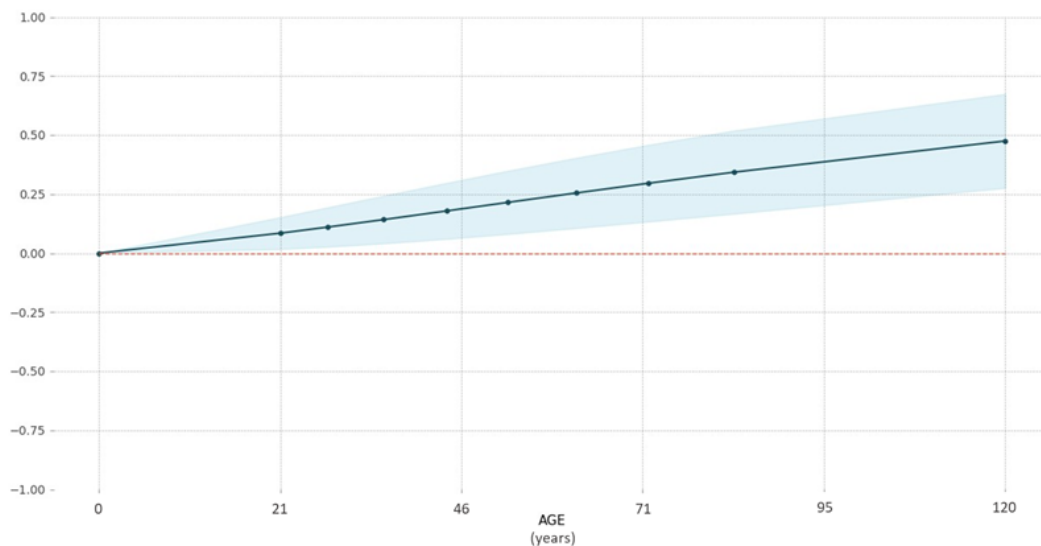


FIGURE 5.5 – Graphique PDP pour la variable quantitative de l’âge.

Les augmentations de la probabilité d’hospitalisation identifiées à l’aide de SHAP pour les autres variables explicatives catégorielles sont également confirmées par les graphiques PDP : le motif de consultation pour “traumatismes sévères multiples” (figure 5.2.2), le niveau 2 de l’échelle de tri FRENCH (figure 5.2.2) et le motif de consultation “altération de l’état général” (figure 5.2.2) augmentent la probabilité d’hospitalisation respectivement de 25%, 12.5% et 9%. Il convient de noter que plus de variables explicatives ayant une contribution positive (de l’ordre de 15 à 20%) à la probabilité d’hospitalisation sont détectées par PDP que par SHAP : motifs de consultation “glycémie supérieure à 13 mmol/L et cétonémie capillaire supérieure à 1.5mmol/L” (section 5.2.2), “ictère” (figure 5.2.2), “ascite” (figure 5.2.2), “anémie inférieure à 8g/L” (figure 5.2.2), “score de Glasgow inférieur à 8” (figure 5.2.2), “score de Glasgow entre 9 et 12” (figure 5.2.2), constat d’une dyspnée majeure (figure 5.2.2) ou d’une hypoxémie inférieure à 85% (figure 5.2.2). Par ailleurs, le taux d’alcool expiré

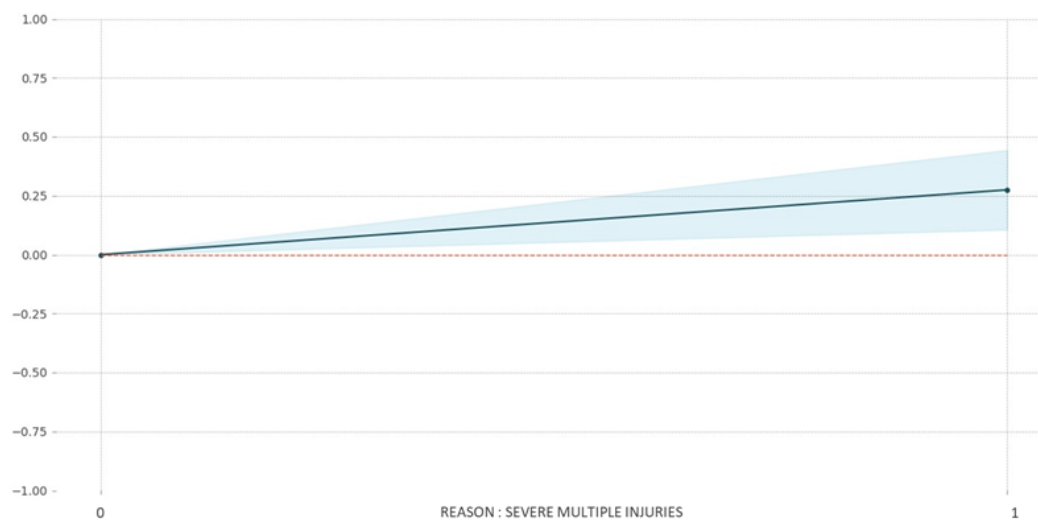


FIGURE 5.6 – Graphique PDP pour la variable binaire du motif de consultation “traumatisé sévère multiple”.

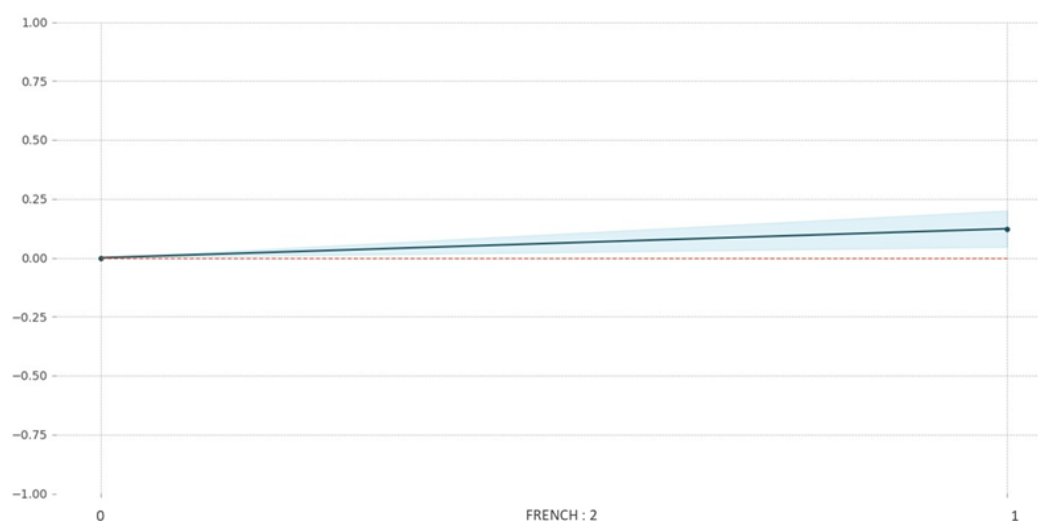


FIGURE 5.7 – Graphique PDP pour la variable binaire du niveau 2 de l'échelle French.

(figure 5.2.2) et le débit d'oxygène administré (figure 5.2.2) sont associés à une augmentation substantielle de la probabilité d'hospitalisation, respectivement 26% et 30%.

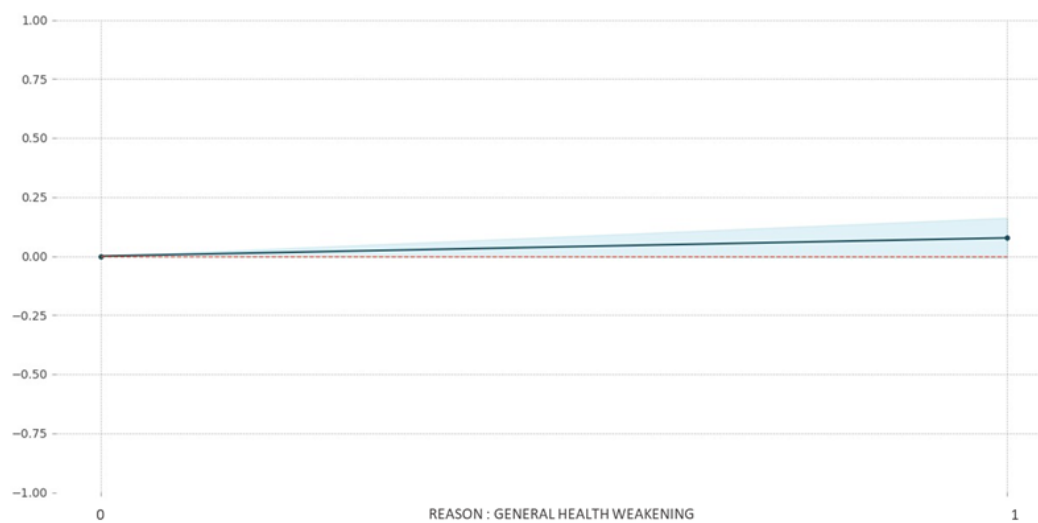


FIGURE 5.8 – Graphique PDP pour la variable binaire du motif de consultation “altération de l’état général”.

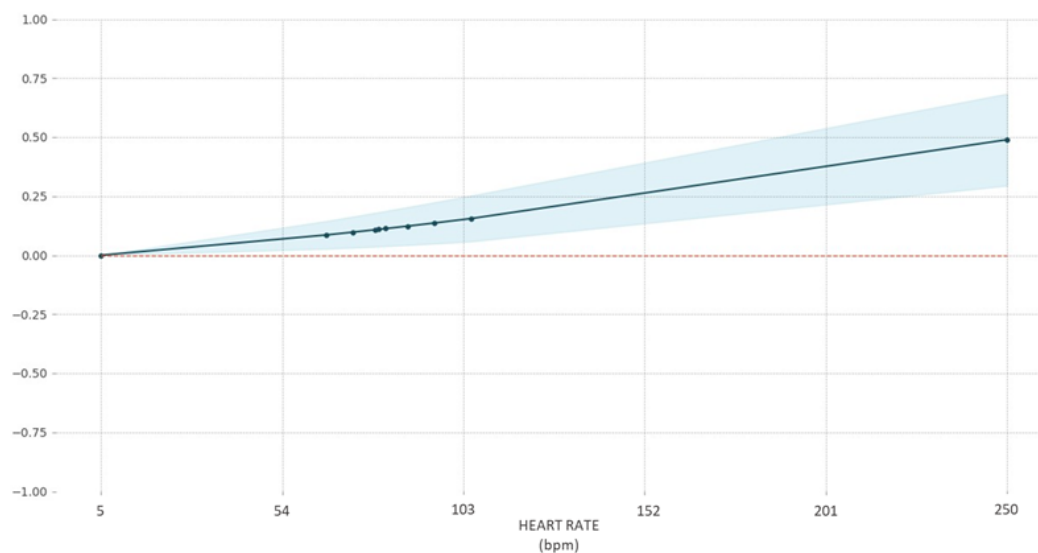


FIGURE 5.9 – Graphique PDP pour la variable quantitative de la fréquence cardiaque.

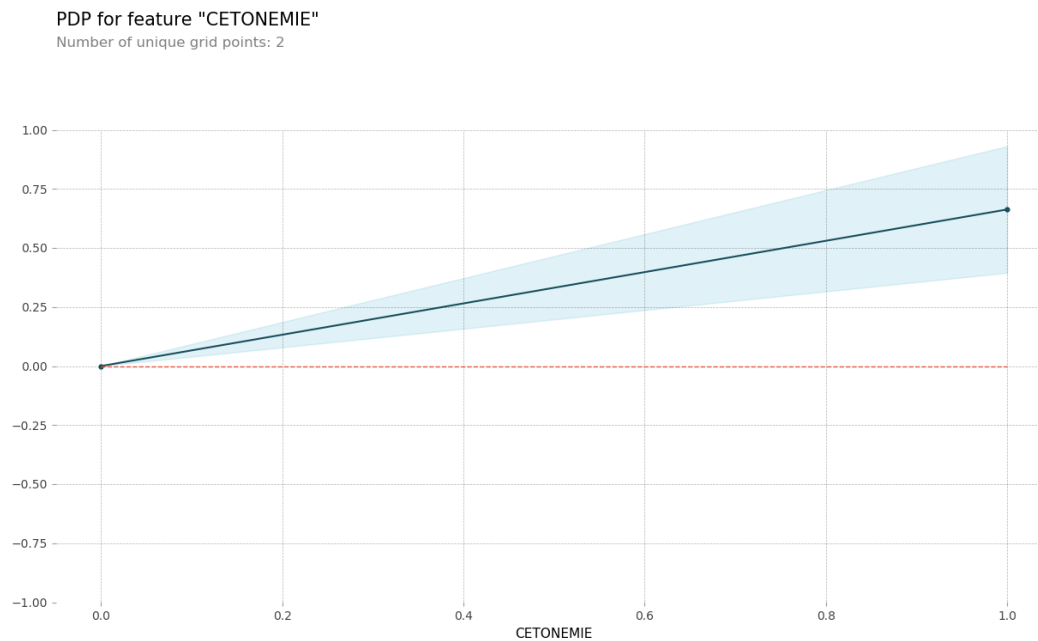


FIGURE 5.10 – Graphique PDP pour la variable quantitative de la cétonémie capillaire.

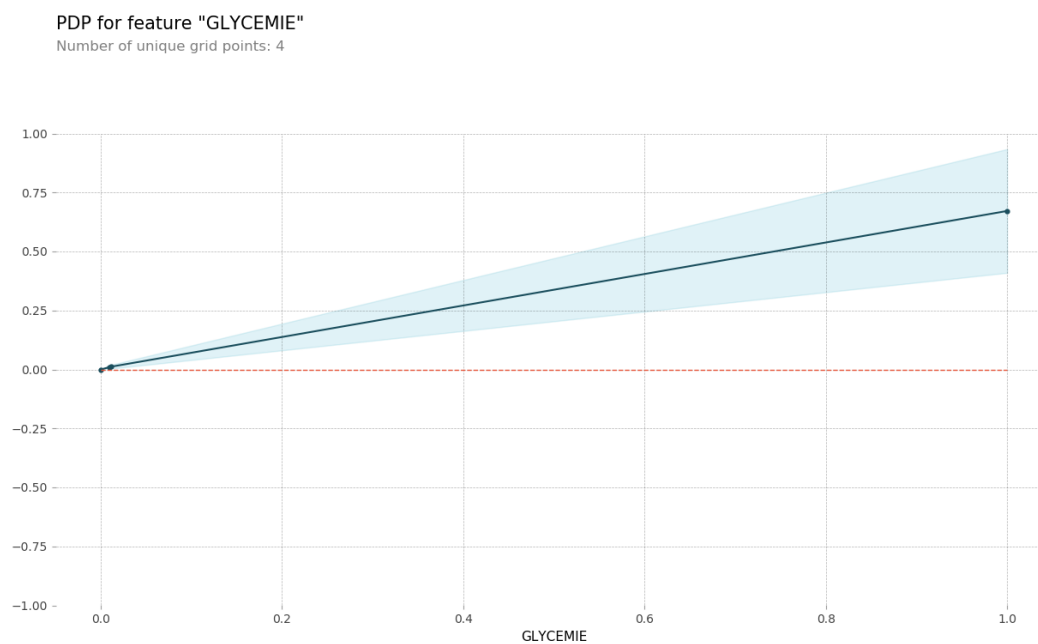


FIGURE 5.11 – Graphique PDP pour la variable quantitative de la glycémie capillaire.

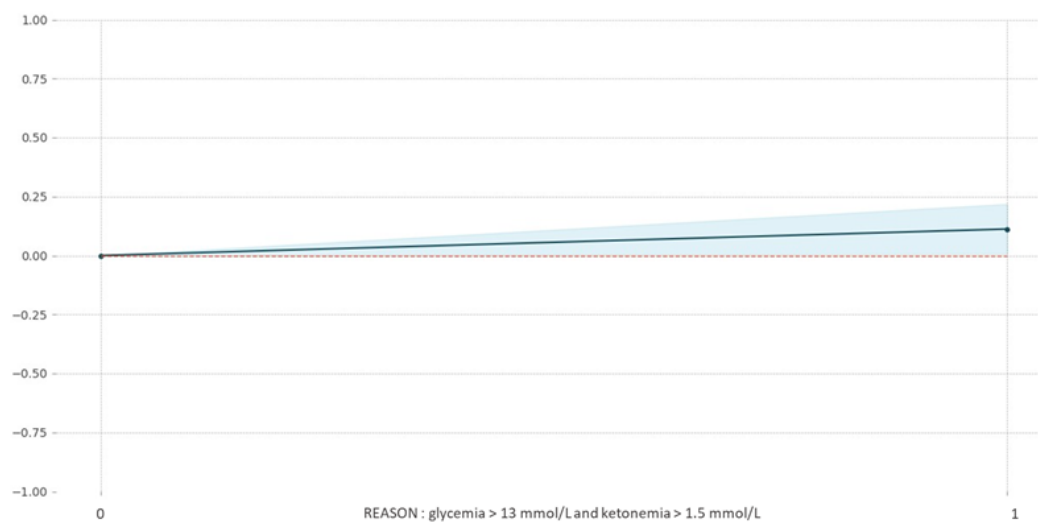


FIGURE 5.12 – Graphique PDP pour la variable binaire du motif de consultation "glycémie supérieure à 13 mmol/L et cétonémie capillaire supérieure à 1.5mmol/L".

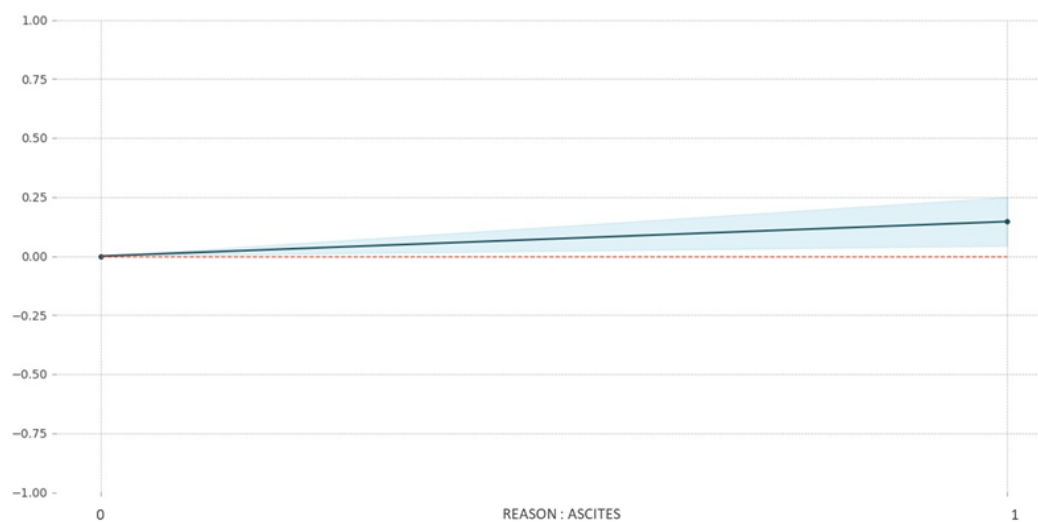


FIGURE 5.13 – Graphique PDP pour la variable binaire du motif de consultation "ascite".

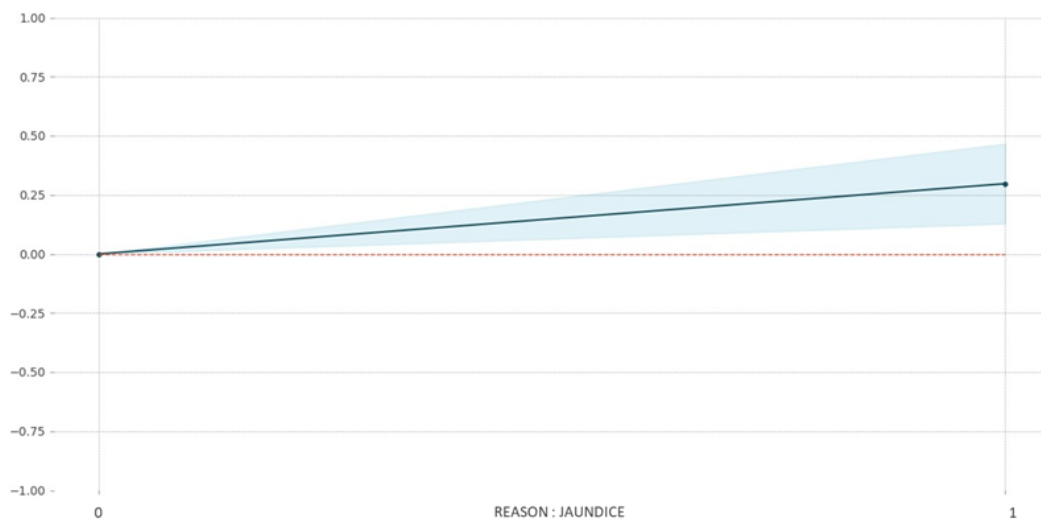


FIGURE 5.14 – Graphique PDP pour la variable binaire du motif de consultation “ictère”.

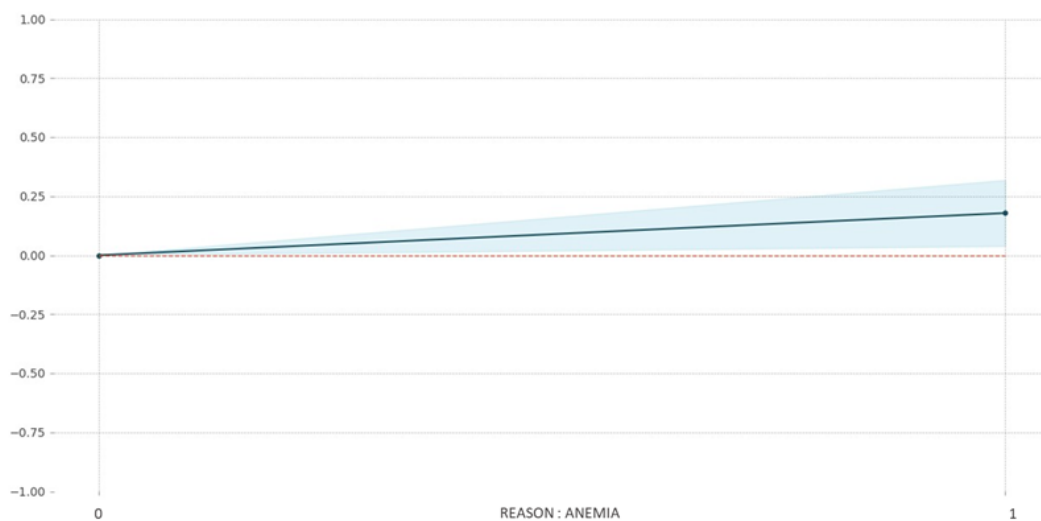


FIGURE 5.15 – Graphique PDP pour la variable binaire du motif de consultation “anémie”.

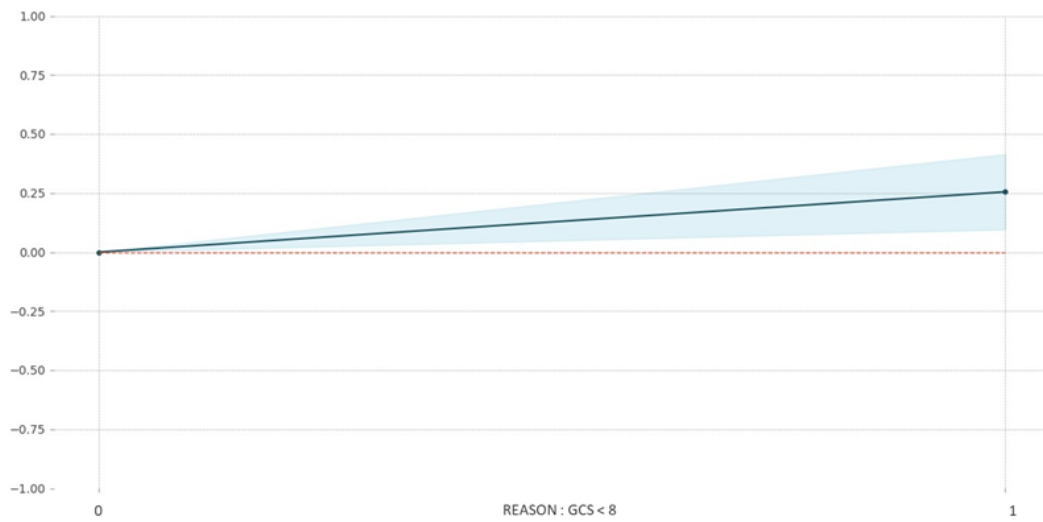


FIGURE 5.16 – Graphique PDP pour la variable binaire du motif de consultation “score de Glasgow inférieur à 8”.

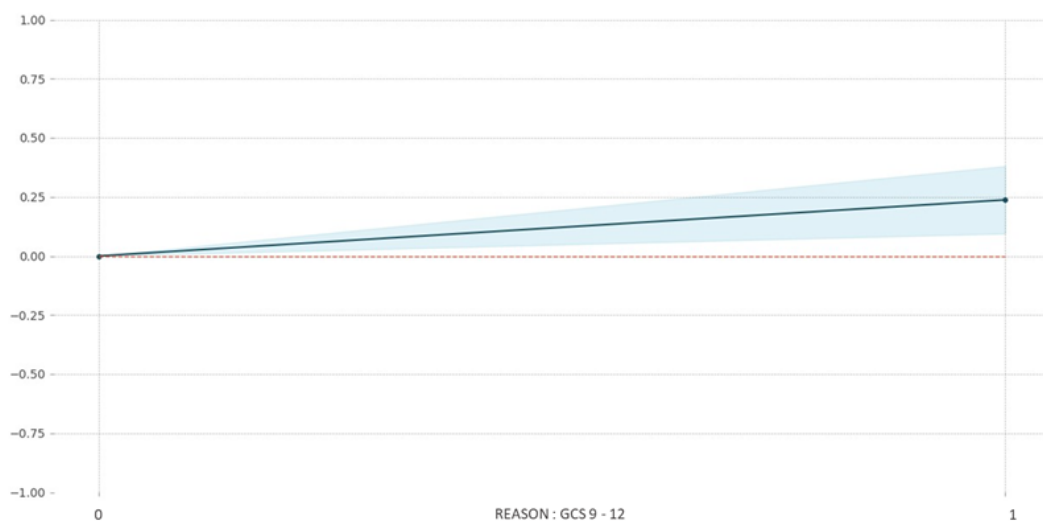


FIGURE 5.17 – Graphique PDP pour la variable binaire du motif de consultation “score de Glasgow entre 9 et 12”.

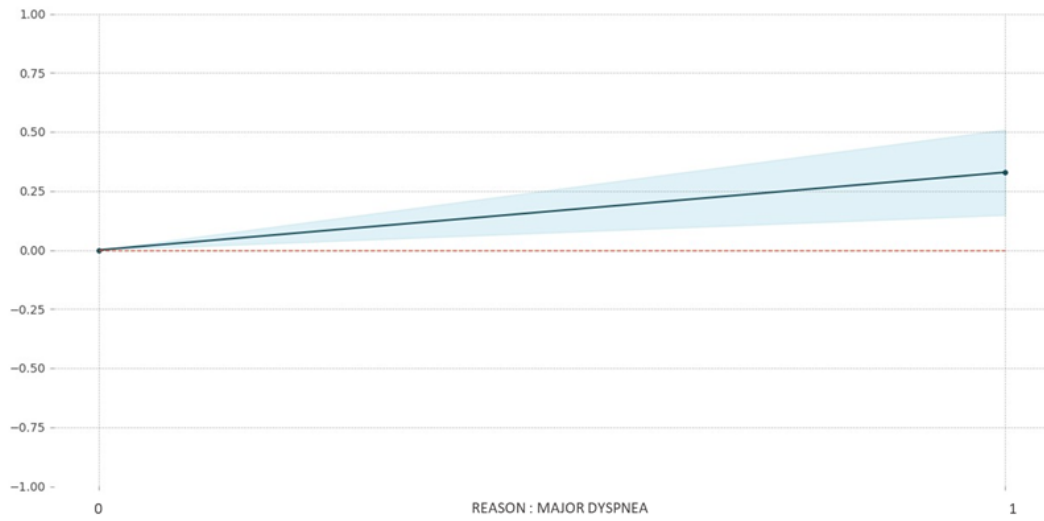


FIGURE 5.18 – Graphique PDP pour la variable binaire du constat d'une dyspnée majeure.

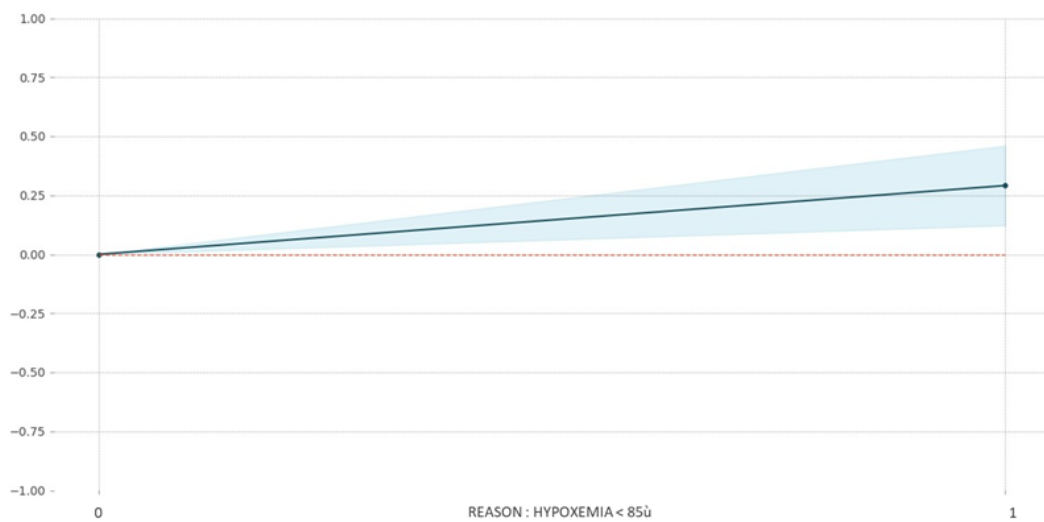


FIGURE 5.19 – Graphique PDP pour la variable binaire du constat d'une hypoxémie inférieure à 85%.

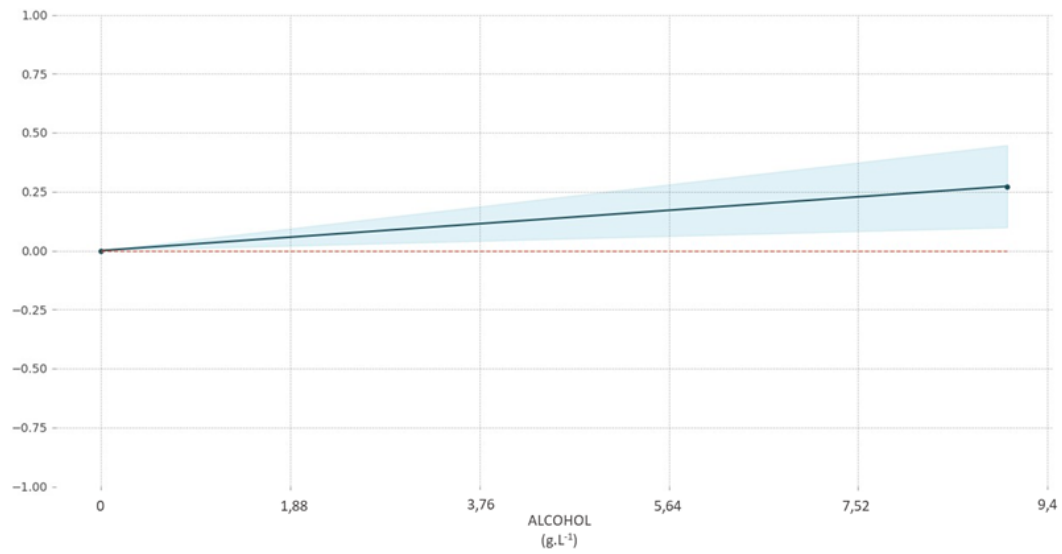


FIGURE 5.20 – Graphique PDP pour la variable du taux d'alcool expiré.

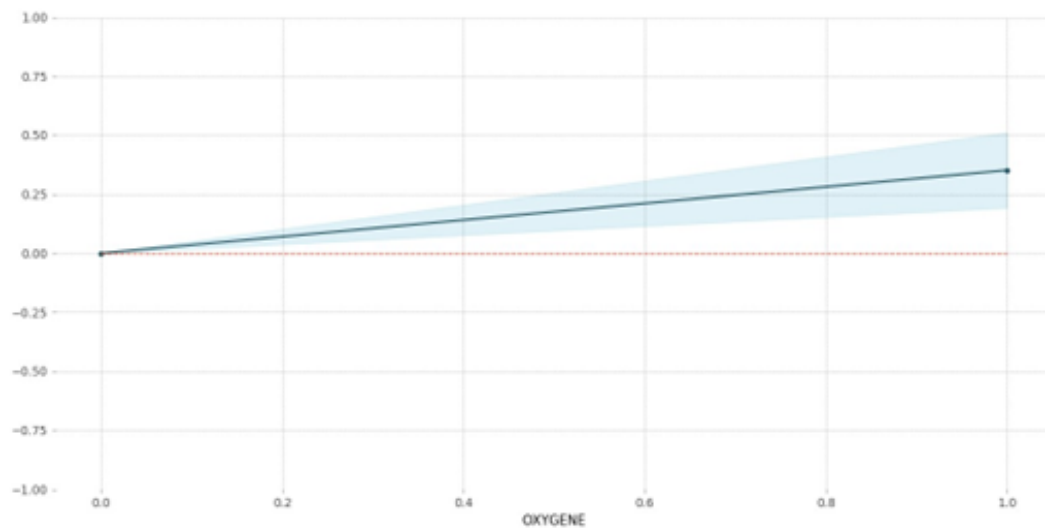


FIGURE 5.21 – Graphique PDP pour la variable du débit d'oxygène administré.

5.2.3 Interprétation médicale

L'application des techniques d'explicabilité au modèle **3P-U** met en valeur des variables explicatives qui influencent significativement la probabilité d'hospitalisation à partir des urgences, comme la fréquence cardiaque, l'âge, le motif de consultation pour des traumatismes sévères multiples, et un niveau 2 de tri sur l'échelle **FRENCH**. Ces explications sont en concordance avec des connaissances médicales bien établies : (i) la probabilité d'hospitalisation plus importante pour les patients âgés peut être imputée à leur plus grande fragilité et un retour au domicile plus complexe à organiser à court/moyen terme[132], (ii) les patients consultant pour des traumatismes sévères multiples requièrent souvent une surveillance d'au moins 24 heures même si les premiers éléments cliniques et paracliniques sont normaux (il a été démontré que le premier scanner corps entier pouvait rater des atteintes sévères dans 10% des cas dont 6.7% avec un impact clinique et 3.7% sans impact clinique[133]), et (iii) une fréquence cardiaque anormale est un indicateur commun d'un état clinique précaire dans les maladies cardiovasculaires[134], le sepsis[135], les saignements post-opératoires[136].

Inversement, les variables explicatives comme "attendre sur une chaise" ou "arriver aux urgences par ses propres moyens" comptent comme une part substantielle de la prédiction de sortie des urgences. Il est raisonnable de poser l'hypothèse que les patients attendant leur examen médical en étant assis ou debout ont une probabilité de sortir des urgences plus élevée. Dans ces cas, il n'était pas nécessaire d'allonger le patient sur un brancard, ce qui aurait rendu la sortie du patient moins probable.

Les explications fournies par **PDP** identifient d'autres variables explicatives comme clairement contributives de la prédiction d'hospitalisation : les taux élevés de cétonémie capillaire, de glycémie capillaire ou d'alcool expiré, le niveau 2 de l'échelle de tri **FRENCH**, et la présence d'un ictère. Ces résultats sont cohérents avec les explications cliniques : (i) les taux élevés de cétonémie capillaires et de glycémie capillaires tendent à indiquer que le patient souffre d'un diabète décompensé possiblement compliqué d'une acido-cétose[137] et implique quasi-certainement une hospitalisation pour une prise en charge adaptée, (ii) le niveau 2 de l'échelle de tri **FRENCH** différencie les patients les plus urgents et requiert l'intervention paramédicale infirmière et médicale rapide alors que les niveaux 3, 4 et 5 sont pour les patients moins urgents; plus un patient est urgent, plus sa probabilité d'hospitalisation est élevée[11], et (iii) l'ictère indique une pathologie hépato-biliaire requérant une prise en charge hospitalière..

La confirmation de ces résultats renforce l'importance et la validité des explications données par **SHAP** et **PDP** en vue d'une utilisation de **3P-U** par les médecins urgentistes pour déterminer si un patient a une forte probabilité de rester hospitalisé. En mettant en lumière l'importance clinique de ces prédicteurs, l'explicabilité automatique renforce la crédibilité du modèle et son alignement sur les pratiques de prise de décision médicale aux urgences.

5.3 Discussion

Le modèle **3P-U** a démontré une performance acceptable dans la prédiction du parcours de soin du patient au sein des urgences et les explications automatiques de ses prédictions ont été validées par les médecins urgentistes impliqués dans l'étude. Nous allons examiner les différents aspects de la performance du modèle et évaluer sa pertinence dans son utilisation.

5.3.1 Comparaison avec les travaux existants

Il y a une différence significative entre les modèles de RNP, FA, PPV et RL. Nous avons choisi le meilleur modèle (celui du RNP) comme référence. À partir des données de triage, **3P-U** prédit l'hospitalisation ou la sortie des urgences avec une exactitude de 75%, un rappel de 76%, une précision de 60%, et une **AUROC** de 83%. Le modèle de GRAHAM *et al.* [138] prédit aussi l'hospitalisation du patient avec une performance meilleure (**AUROC** : 85%). Cependant, **3P-U** et le modèle de Graham *et al.* n'utilisent pas exactement les mêmes types de

données. Les modèles qui s'intéressent à un type spécifique de patient (*e.g.*, dans un département de neurologie ou de cardiologie) tendent à avoir de meilleures performances. Par exemple, KLANG *et al.* [139] prédit l'admission dans un service de soin critique sur la base de données non structurées (AUROC : 92%); données que nous n'avons pas utilisées dans ce modèle.

5.3.2 Méthodes d'explicabilité et alternatives

LONGONI, BONEZZI et MOREWEDGE [140] ont étudié différents aspects de l'acceptabilité des outils d'IA dans le domaine médical : les modèles d'IA sont plus facilement acceptés lorsque : (i) ils considèrent le patient dans son individualité, et (ii) ils produisent un support à la prise en charge ou à la décision (plutôt que de chercher à remplacer le médecin).

En IA, il existe un conflit inévitable entre la performance et l'explicabilité : les modèles les plus complexes (*e.g.*, ensembles d'arbres décisionnels et réseaux de neurones) sont habituellement plus performants, mais moins explicables. La plupart des études sur l'apprentissage machine qui affichent des résultats raisonnablement acceptables pour la prédiction d'hospitalisation sont souvent basés sur des arbres décisionnels et des approches ensemblistes. Les modèles basés sur les RNP ont été sous-exploités, probablement pour leur difficulté d'interprétation du résultat et de leur coût supérieur d'architecture et d'entraînement. Néanmoins, l'apprentissage profond est une méthode très pratique lorsque les données ne sont pas structurées (*e.g.*, le texte dans les notes cliniques) or lorsque le nombre de variables est très élevé[37, 38, 70].

Le compromis exactitude/explicabilité est spécifique à chaque domaine. Dans le domaine médical, l'utilisateur final (*i.e.*, le clinicien et souvent les patients) et les experts en méga-données doivent se mettre d'accord sur ce compromis. Plusieurs études sur l'explicabilité de la prédiction d'hospitalisation se sont attachées à décrire l'importance des variables explicatives. Par exemple, ZHU *et al.* [141] a appliqué SHAP à un modèle prédisant la réadmission en unité psychiatrique à partir de sources variées d'information (démographiques, pharmacologiques, psychologiques, et cliniques). L'apprentissage profond et les techniques d'explicabilité ont été appliquées récemment dans la COVID-19, en explicitant les éléments pouvant conduire au diagnostic et la compréhension de la pathologie[142]. Par ailleurs, ESTIRI, STRASSER et MURPHY [143] ont utilisé la sélection des variables explicatives pour améliorer l'interprétabilité de modèle centré sur le patient prédisant le risque d'admission en unité conventionnelle, en unité de soins critiques, le risque de recours à la ventilation mécanique et de mort. De plus, les capacités de prise de décision rapide des modèles d'IA offertes aux unités de soins critiques ont été appliqués aux enfants atteints de pneumopathie; SHAP a été utilisé pour évaluer le modèle de FA (AUROC : 98.7%) et obtenir la liste de l'importance des variables secondairement validées par les cliniciens[144]. La pertinence des variables explicatives fournies par les techniques d'explicabilité est utilisée pour aider les modèles prédictifs de l'hospitalisation dans d'autres champs de la médecine : en cardiologie[145], en transplantation d'organe[146], en chirurgie médullaire[147], en gastro-entérologie[148], et en général sur la réadmission hospitalière[149].

5.3.3 Implications

La prédiction de l'hospitalisation ou de la sortie depuis les urgences est importante pour les patients, pour les services d'urgence et pour l'hôpital. Premièrement, la connaissance de la probabilité d'hospitalisation devrait permettre au patient d'informer ses proches en conséquence. Deuxièmement, les Médecins d'Accueil et d'Orientation (MOAs) cherchent à optimiser le flux des patients aux urgences. En utilisant les données du triage, les MOA peuvent (i) orienter les patients dans les secteurs adéquats en fonction de leur charge et de leur capacité, (ii) anticiper les examens complémentaires (*e.g.*, biologie, imagerie, avis d'autres spécialités), et (iii) anticiper la recherche de lits lorsque la probabilité d'admission est élevée. Troisièmement, connaître le nombre de lits nécessaires en temps réel aide le gestionnaire de lits à adopter une stratégie de ciblage des lits disponibles en fonction des besoins pour les patients des urgences. Tous ces mécanismes devraient aider à réduire le temps de passage aux urgences et diminuer la mortalité dans les services d'urgence[5, 7].

À présent, **3P-U** prédit en routine la probabilité d’hospitalisation sur la base de données collectées pendant le triage paramédical peu de temps après l’arrivée du patient. Puis, cette prédiction est utilisée pour catégoriser le patient comment ayant une “forte probabilité d’hospitalisation” (*i.e.*, si la probabilité est supérieure à 0.6) ou “une non forte probabilité d’hospitalisation” sinon. Le gestionnaire de lits peut accéder à la liste des patients ayant une forte probabilité d’hospitalisation et ainsi, avec l’accord du médecin spécialiste responsable de l’unité cible, préparer un plan d’attribution des lits. Néanmoins, l’approche de **3P-U** n’a pas été largement adoptée par les cliniciens responsables des unités d’hospitalisation. Ceci peut s’expliquer par (i) un manque de littérature, (ii) le scepticisme à l’utilisation de l’**IA**, et (iii) la peur de la perte du contrôle des admissions dans leur unité. Nous avons cependant utilisé **3P-U** pour estimer le besoin en lits lorsque de nombreux patients arrivaient aux urgences (pendant la pandémie de COVID-19)[26]. Puisque le modèle est maintenant explicable et affiche des performances comparables à des modèles similaires, nous pouvons espérer que les cliniciens adoptent **3P-U** plus largement[149].

Avant cette étude, l’explicabilité de **3P-U** n’a pas été présentée aux médecins urgentistes. Avec l’objectif d’améliorer l’adoption de **3P-U**, nous envisageons de valider **3P-U** avec chaque spécialité dans les mois à venir. L’adoption devrait devenir meilleure à mesure que les cliniciens sont inclus dans l’expérimentation concrète.

5.3.4 Limites de l’étude

Cette étude n’est pas sans limitation. Premièrement, même si les résultats d’explicabilité ont été produits de manière indépendante, le modèle a été créé par les cliniciens et ceci pourrait entraîner un biais dans leur raisonnement. Bien que notre validation soit conforme à la littérature, une validation indépendante par d’autres cliniciens devrait être envisagée. Ultérieurement, il pourrait être utile d’analyser l’impact clinique de ces résultats sur les patients dans notre service d’urgence et dans d’autres hôpitaux.

Secondairement, le jeu de données vient d’un seul centre, et en conséquence, une étude multicentrique devrait être conduite.

Troisièmement, la variable à expliquer (*i.e.*, hospitalisation ou sortie) est binaire; il serait plus pertinent de désigner l’unité dans laquelle le patient devrait être hospitalisé. Ce dernier objectif est plus difficile à prédire car beaucoup de patients sont hospitalisés dans des unités inadaptées à leur pathologie (*e.g.*, un patient avec un diabète décompensé admis dans une unité de néphrologie, plutôt que dans une unité d’endocrinologie à cause d’un manque de lits). Aux urgences, l’unité dans laquelle le patient a été admis est documentée de manière non fiable.

Enfin, nous n’avons pas investigué les potentiels biais liés aux **DNPC** parce que les performances du modèle affiche des performances acceptables.

5.4 Conclusions

Le modèle **3P-U** développé aux urgences du CHU-AP utilise un **RNP** pour prédire la probabilité d’hospitalisation. Les techniques d’explicabilité automatique (**SHAP** et **PDP**) mises en oeuvre dans cette étude mettent en évidence l’âge, la fréquence cardiaque, le mode d’arrivée, les traumatismes multiples sévères et l’échelle de tri **FRENCH** comme variables à expliquer les plus importantes. Ces résultats ont été cliniquement validés par les développeurs de **3P-U**. Nous posons l’hypothèse que le déploiement de **3P-U** devrait aider à diminuer la durée de passage aux urgences avant l’hospitalisation et ainsi devrait aider à mieux gérer le flux de patients. De manière générale, l’adoption de l’explicabilité automatique devrait améliorer l’acceptabilité et la confiance dans le modèle **3P-U**.

Chapitre 6

Application concrète du modèle de prédiction du parcours patient aux urgences dans le contexte de la COVID-19

Points clés

Après avoir entraîné **3P-U**, et l'avoir rendu explicable, nous avons utilisé les résultats de **3P-U** pour aider à l'adaptation de l'organisation hospitalière dans le contexte de la COVID-19.

Cette contribution a été publiée sous la référence: Emilien ARNAUD *et al.* "Use of Artificial Intelligence to Manage Patient Flow in Emergency Department during the COVID-19 Pandemic : A Prospective, Single-Center Study". In : *International Journal of Environmental Research and Public Health* 19.15 (5 août 2022), p. 9667. ISSN : 1660-4601. DOI : [10.3390/ijerph19159667](https://doi.org/10.3390/ijerph19159667).

6.1 Méthodologie

6.1.1 Construction de l'étude

Nous avons construit une étude prospective, monocentrique, des patients consultant aux urgences du CHUAP pour valider les performances du modèles **3P-U** dans le contexte spécifique de la crise COVID-19. Cette étude a été menée en conformité avec le “Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis statement”[113] et a été autorisée par la Commission Hors-Loi Jardé sous la référence PI2019_843_0066.

Pour rappel, le SAU du **CHU-AP** est constitué de trois zones : (i) un circuit de **CC**, (ii) un circuit de **CL**, et (iii) une **SAUV**. Pendant l'épidémie de COVID-19, les **CC** et **CL** ont été dupliqués pour accueillir les patients suspects d'être COVID-19 dans un secteur nommé “Haute Charge Virale” (**HCV**), et deux box de la **SAUV** ont été dédiés aux patients **HCV**.

6.1.2 Participants

Nous avons inclus tous les patients consultant aux urgences après le triage infirmier, du 1^{er} décembre 2019 au 31 décembre 2021. Nous avons exclu les patients directement admis en **SAUV** à leur arrivée et les patients n'ayant pas de donnée de triage.

Nous avons utilisé le modèle **3P-U**, déployé au **CHU-AP**. Ce modèle permet de prédire la décision médicale d'orientation (*i.e.*, admission ou sortie des urgences) en se basant sur les données de triage. Le modèle **3P-U** a été entraîné sur 233 814 patients ayant consulté entre 2015 et 2018. En considérant un seuil décisionnel de 0.5, l'aire sous la courbe **ROC** est de 83%, la précision (*i.e.*, valeur prédictive positive) est de 72% et le rappel (*i.e.*, la sensibilité) est de 62%[23]. Les prédicteurs sont : les caractéristiques démographiques (age et sexe), les caractéristiques cliniques du tri (fréquence cardiaque, pression artérielle, saturation en oxygène, température, glycémie capillaire, cétonémie capillaire, débit en oxygène, hémoglobine capillaire, taux d'alcool expiré, niveau de douleur, résultats d'une éventuelle bandelette urinaire, niveau de triage de l'échelle **FRENCH** par l'infirmière d'accueil), les caractéristiques de triage non clinique (le mode d'arrivée, le mode d'adressage), si le patient était accompagné, sa modalité d'attente (debout, assis, en brancard), son contexte familial, son heure d'arrivée, et des données non structurées (notes de l'infirmière d'accueil et antécédents du patients).

Dans le contexte particulier de la médecine d'urgence, les données peuvent ne pas être collectées de manière volontaire (*i.e.*, **DNPC**[24]) et donc ne pas correspondre à une donnée manquante (*e.g.*, la glycémie capillaire n'est pas mesurée pour les patients se présentant pour un traumatisme simple de la cheville). La méthode de complétion des données manquantes dépend du type de variable (**tableau 3.2**).

6.1.3 Intervention

À leur arrivée aux urgences, une **IOA** priorisait les patients en suivant l'échelle de tri **FRENCH** et triait les patients suivant s'ils étaient suspects d'être atteints de la COVID-19 ou non. Toutes les données saisies dans le dossier patient informatisé (Resurgences[©], Berger-Levrault, Boulogne-Billancourt, France) ont été automatiquement extraites par **3P-U**, et ont été ensuite pré-traitées suivant la même procédure que pendant le développement du modèle. La prédiction était calculée et enregistrée dans la base de données (**figure 6.1**).

À partir d'un prélèvement nasopharyngé, les patients suspects d'être atteints de la COVID-19 étaient testés avec un test de flux latéral (Panbio COVID-19Ag (Nasal Version), Abbott Chicago, IL, USA) confirmé ensuite par un test **Réaction en Chaîne Polymérase** (“*polymerase chain reaction*”) (**PCR**) (TaqPath COVID 19 CE-IVD RT-PCR Kit, Thermofisher, Waltham, MA, USA). Ni l'information suspicion COVID-19, ni le test rapide, ni

INFIRMIÈRE D'ACCUEIL ET D'ORIENTATION

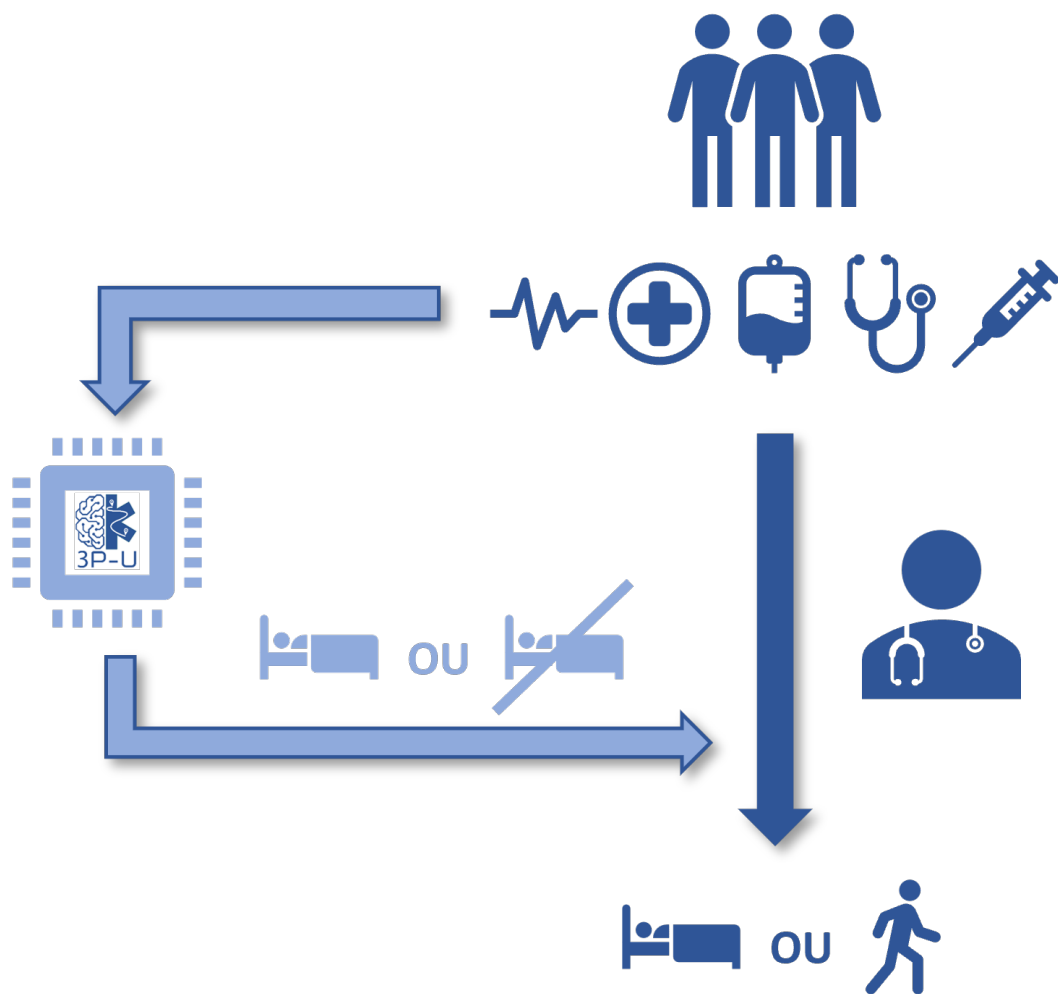


FIGURE 6.1 – Parcours du patient (et de ses données) au cours d'une utilisation réelle de 3P-U.

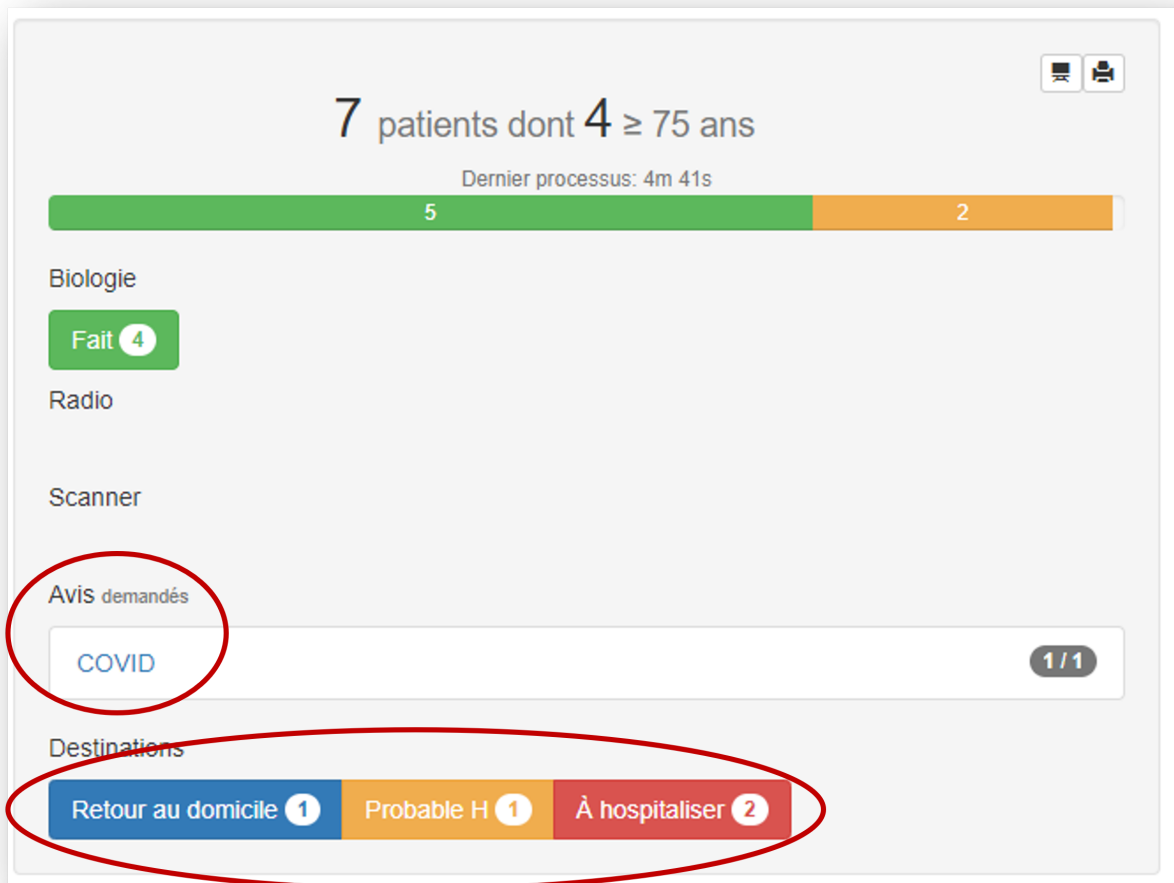


FIGURE 6.2 – Tableau de bord disponible à la cellule de gestion des lits.

la confirmation par **PCR** n'étaient incluses comme prédicteur du modèle puisque ces variables étaient absentes de la phase d'entraînement.

La cellule de gestion des lits avait accès à un tableau de bord (figure 6.2) avec le statut réel des patients : "orientation inconnue", "à hospitaliser" (*i.e.*, décision médicale définitive), "probablement à hospitaliser" (*i.e.*, recommandation de **3P-U**), "probable sortie" (*i.e.*, recommandation **3P-U**), "retour au domicile" (*i.e.*, décision médicale définitive). La phase de développement et la phase de validation diffèrent au regard du contexte (*i.e.*, avant versus pendant l'épidémie de COVID-19) et de l'interprétation de la probabilité d'admission. Nous avons défini un patient comme "évident" lorsque la probabilité prédite d'admission était inférieure à 0.2 (*i.e.*, interprété comme "probable sortie") ou supérieure à 0.6 (*i.e.*, interprété comme "probablement à hospitaliser"). Entre 0.2 et 0.6, nous avons considéré que le patient était dans une "zone grise" et l'avons indiqué par le statut : "orientation inconnue".

6.1.4 Analyses statistiques

Les variables continues ont été décrites en utilisant la moyenne (écart-type), et les variables catégorielles ont été décrites en utilisant le nombre de patients (pourcentage). Pour décrire la population, nous avons récupéré le résultat des tests **PCR** effectués aux urgences. Les tests **PCR** pour la COVID-19 étaient réalisés pour tous les patients hospitalisés et les patients symptomatiques avant l'admission aux urgences. Les performances du modèle ont été évaluées en utilisant l'**AUROC**. Nous avons formé deux jeux de patients : tous les patients (*i.e.*, pour déterminer le nombre maximum de lits nécessaires) et les "patients évidents" (*i.e.*, pour évaluer le nombre

minimum de lits nécessaires). La recommandation par **3P-U** et la décision médicale finale étaient comparées en utilisant la précision, le rappel et l’exactitude pour les deux jeux de patients ([sous-section 2.2.1](#)). La précision est définie par le rapport entre les vraies hospitalisations détectées et toutes les hospitalisations détectées. Le rappel est défini par le rapport entre les vraies hospitalisations détectées et toutes les hospitalisations. L’exactitude est définie par la proportion de patients avec une recommandation conforme à la décision médicale finale.

6.2 Résultats

6.2.1 Caractéristiques des patients de l’étude

Entre le 1^{er} décembre 2019 et le 31 décembre 2021, 136 139 patients ont consulté aux urgences du **CHU-AP**. Après l’exclusion de 30 682 patients (4 044 immédiatement admis en **SAUV**, 21 991 pour absence de données de triage et 4 647 partis sans attendre), nous avons inclus 105 457 patients. Suivant la décision médicale finale, 33 841 (32%) patients ont été hospitalisés (21 470 (21%) dans un service de médecine, 7 604 (7%) dans un service de chirurgie, et 4 641 (4%) dans un service de soins critiques), et 71 616 (68%) sont sortis des urgences. Les caractéristiques des patients dont l’échelle de tri **FRENCH** sont synthétisés dans le [tableau 6.1](#). La [figure 6.3](#) présente l’évolution des résultats **PCR** dans le département de la Somme dans lequel le **CHU-AP** est situé pendant la période d’inclusion.

6.2.2 Performances du modèle

En se basant sur l’exactitude de la classification, l’**AUROC** du modèle pour tous les patients était de 0.82 et de 0.90 pour les patients évidents. En se basant sur les règles d’interprétation définies ci-avant, 38 353 (36.4%) patient ont été identifiés par **3P-U** comme étant “probablement sortants”, 18 815 (17.8%) comme étant “probablement à hospitaliser” et 48 297 (45.8%) n’étaient pas catégorisés. En se basant sur l’exactitude (*i.e.*, définie comme étant le nombre de patients correctement classés parmi les patients classés), 49 377 (86.4%) des 57 168 patients catégorisés l’ont été correctement. Parmi ces 49 377 patients, 13 840 (24.2%) ont été correctement catégorisés comme “probablement à hospitaliser” et 35 537 (62.2%) ont été correctement catégorisés comme “probablement sortants”. Inversement, 7 791 (13.6%) patient ont été mal catégorisés : 4 975 (8.7%) patients ont été mal catégorisés comme “probablement à hospitaliser” mais sont finalement sortis, et 2 816 (4.9%) patients ont été mal catégorisés comme “probablement sortants” mais ont été finalement hospitalisés.

Ainsi pour les patients évidents, le rappel (*i.e.*, la sensibilité) était de 83.1%, la précision (*i.e.*, la valeur prédictive positive) de 73.6%, la spécificité de 87.7% et la valeur prédictive négative de 92.7% ([figure 6.4](#)). En se basant sur le nombre minimum d’admissions (*i.e.*, pour les patients évidents seulement) et le nombre maximum d’admissions (*i.e.*, pour tous les patients), la cellule de gestion des lits a permis de coordonner la transformation des unités d’hospitalisation en unité COVID-19 ([figure 6.5](#)). Cet indicateur a également été suivi pour la désescalade et le retour à une activité conventionnelle. Le tableau de bord proposé par **3P-U** est considéré comme plus efficace qu’une estimation manuelle par les médecins en cours de soin.

6.2.3 Prédiction individuelle

A l’étape du triage, les infirmières d’accueil ont catégorisé 4 031 (3.8%) patients comme étant probablement atteints de la COVID-19 et 101 426 (96.2%) comme étant probablement non atteints par la COVID-19. Dans le groupe COVID-19, **3P-U** a catégorisé 768 (0.7%) probables hospitalisations et 1 420 (1.3%) probables sorties. Dans le groupe non-COVID-19, **3P-U** a catégorisé 18 018 (17.1%) probables hospitalisations et 36 886 (35.1%) probables sorties ([figure 6.6](#)).

Caractéristiques démographiques	Population	Données manquantes	Stratégie d'imputation
Nombre de patients	105 457 (100%)		
Age	51 (± 22)	0 (00%)	Aucune
Sexe		0 (00%)	Aucune
Homme	50 639 (48%)		
Femme	54 818 (52%)		
Caractéristiques cliniques de tri			
Fréquence cardiaque (/min)	86 (± 18)	20 253 (19%)	Fixé : 80
Pression artérielle systolique (mmHg)	138 (± 24)	25 558 (24%)	Fixé : 120
Pression artérielle diastolique (mmHg)	79 (± 23)	25 558 (24%)	Fixé : 80
Saturation en oxygène (%)	99 (± 2)	38 (<1%)	Fixé : 100
Température corporelle ($^{\circ}$ C)	36.4 (± 0.7)	19 769 (18%)	Fixé : 37.4
Glycémie capillaire (mmol/L)	7.33 (± 4.19)	83 606 (79%)	Fixé : 5.0
Cétonémie capillaire (mmol/L)	0.98 (± 1.97)	104 207 (98%)	Fixé : 0.0
Débit d'oxygène (L/min)	0.6 (± 3.7)	73 410 (69%)	Fixé : 0
Hémoglobininémie capillaire (dg/dL)	11.72 (± 2.96)	102 533 (97%)	Fixé : 12.0
Taux d'alcool expiré (g/L)	1.81 (± 0.78)	102 625 (97%)	Fixé : 0.0
Volume vésical (mL)	334 (± 305)	105 078 (99%)	Fixé : 0.0
Niveau de douleur	3 (± 3)	17 095 (16%)	Fixé : 0
Échelle de tri FRENCH		1 356 (1%)	Plus fréquent
1	235 (< 1%)		
2	3 975 (4%)		
3	56 679 (54%)		
4	28 363 (27%)		
5	14 849 (14%)		
Tests urinaires			
Sang		104 696 (99%)	Fixé : 0
0	223 (29%)		
Traces	167 (22%)		
+	93 (12%)		
++	93 (12%)		
+++	184 (24%)		
++++	1 (<1%)		
Nitrites		104 696 (99%)	Fixé : 0
0	661 (87)		
+	89 (11%)		
++	5 (<1%)		
+++	1 (<1%)		
++++	2 (<1%)		
Leucocytes		104 696 (99%)	Fixé : 0
0	475 (63%)		
Traces	93 (12%)		
+	93 (12%)		
++	94 (12%)		
+++	37 (5%)		
++++	60 (8%)		
Orientation des patients			
Hospitalisation dans un service de médecine	21 470 (21%)		
Hospitalisation dans un service de chirurgie	7 604 (7%)		
Hospitalisation dans un service de soins critiques	4 641 (4%)		
Sortie des urgences	71 616 (68%)		

TABLE 6.1 – Caractéristiques de la population d'étude[26]. Les variables continues sont représentés par leur moyenne (\pm écart-type) et les variables catégorielle par le nombre de patients (pourcentage).

Statistiques du COVID-19 dans le département de la Somme

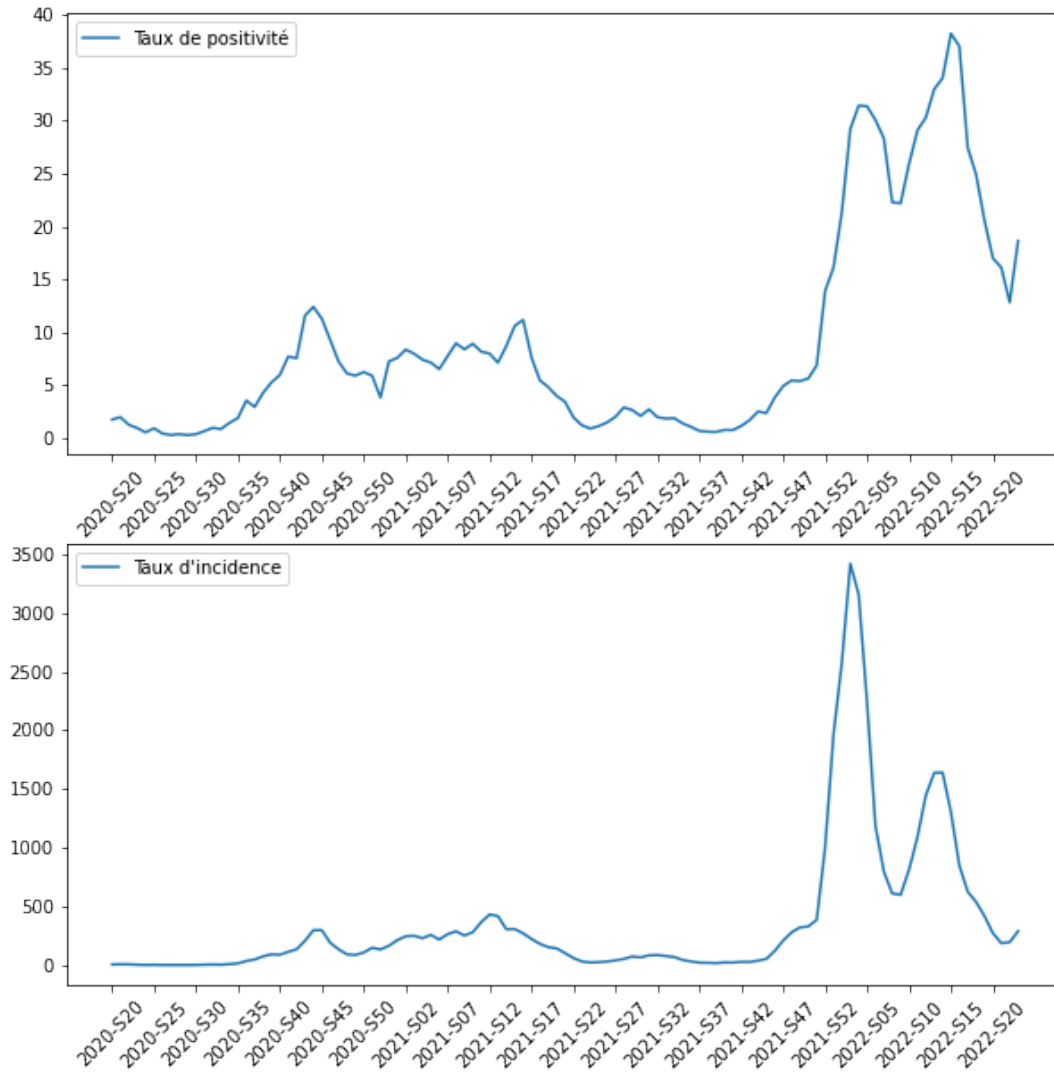


FIGURE 6.3 – Évolution des résultats PCR du département de la Somme. Les données ont été extraites depuis data.gouv.fr, le 1^{er} juillet 2022[150]

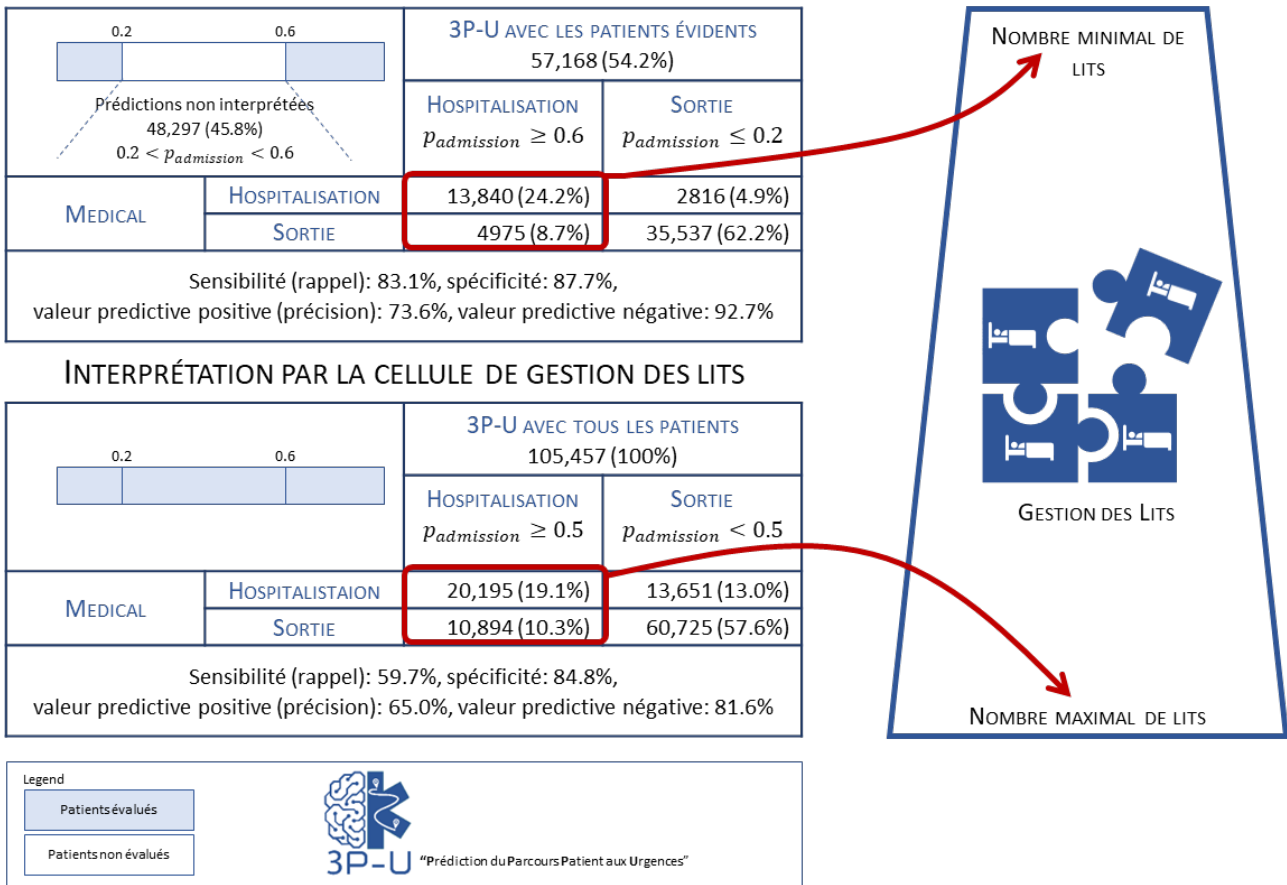


FIGURE 6.4 – Performances du modèle 3P-U et interprétation par la cellule de gestion des lits[26].

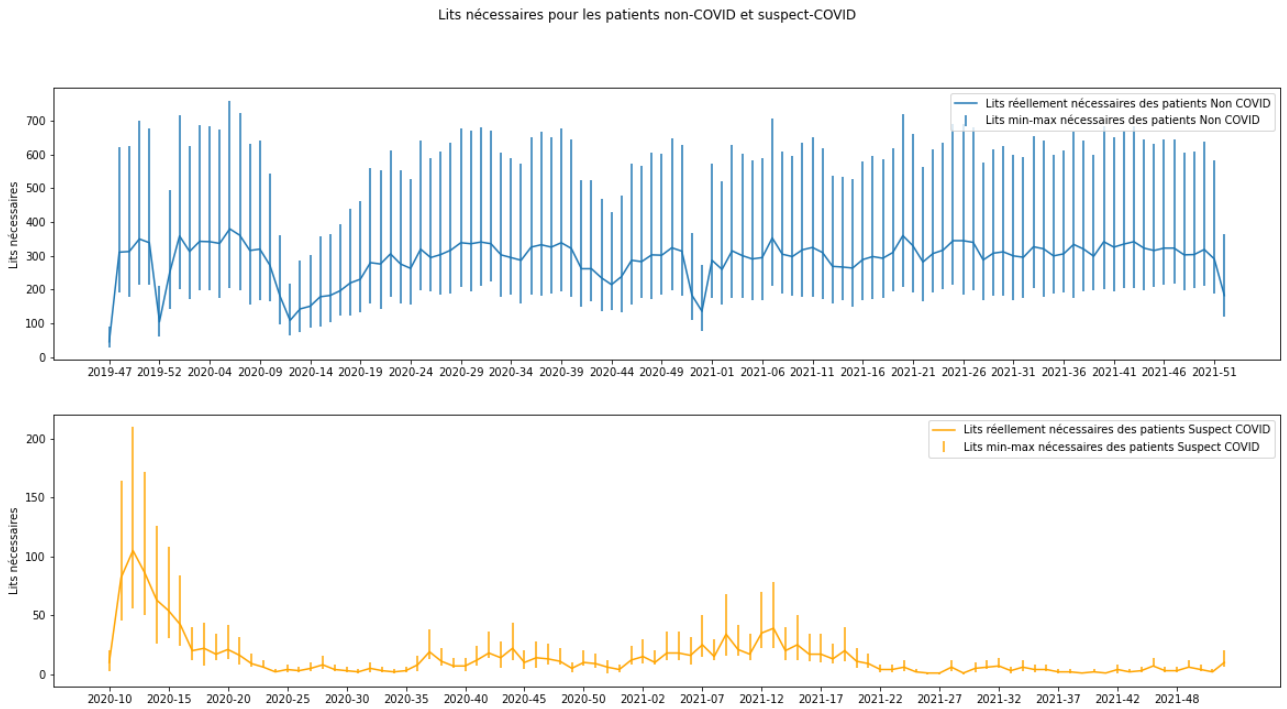


FIGURE 6.5 – Lits nécessaires pour les patients présumés non-COVID-19 et les patients suspects COVID-19 par semaine, tel qu'utilisé par la cellule de gestion des lits pour optimiser les unités d'hospitalisation[26].

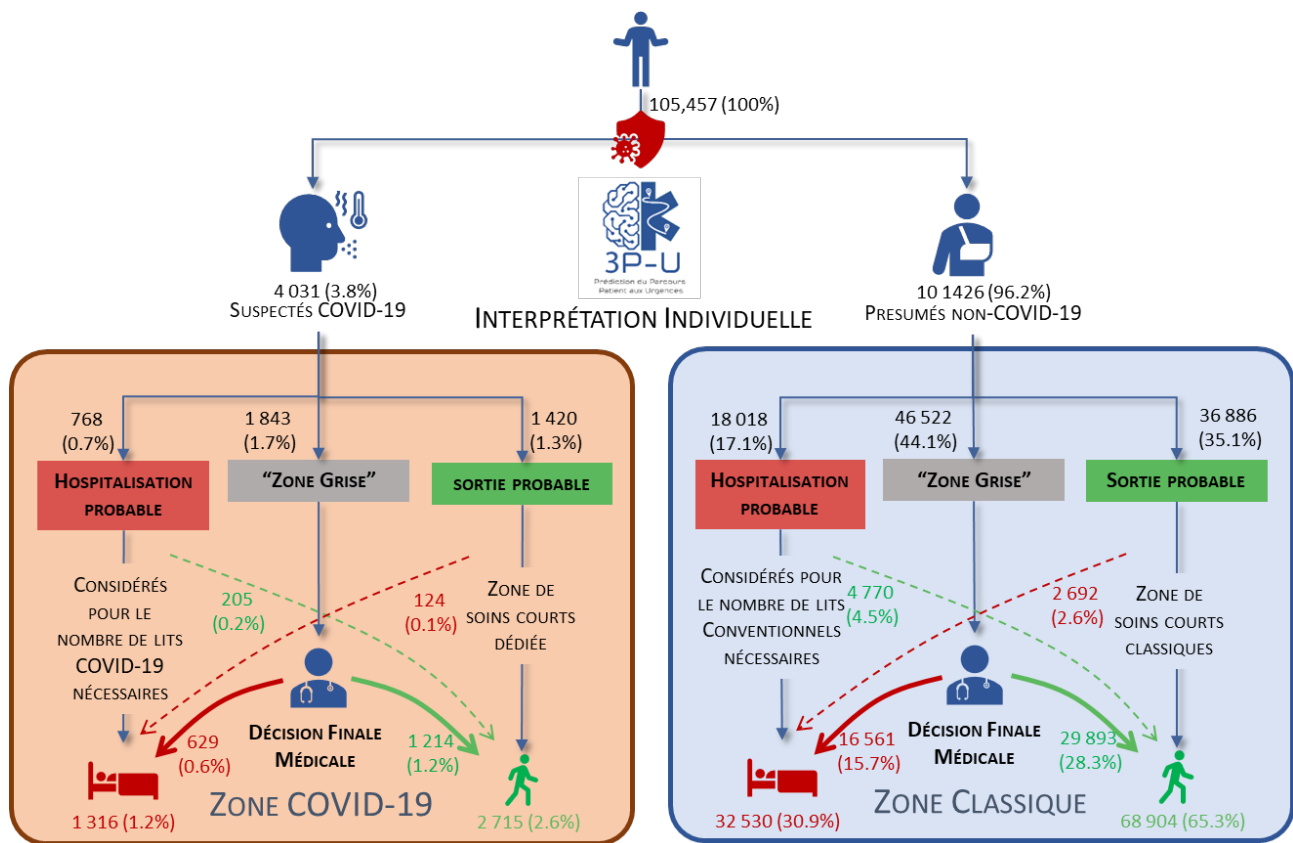


FIGURE 6.6 – Prédiction individuelle de triage aux urgences du CHU-AP pendant la période COVID-19[26].

6.3 Discussion

6.3.1 Seuil décisionnel de 3P-U

L'**AUROC** du modèle **3P-U** était de 82%, une valeur moyenne dans l'intervalle des performances retrouvées dans les études similaires[151]. Le choix du seuil décisionnel est complexe. Dans notre précédente étude[23], nous avons utilisé le seuil de 0.5 parce qu'il présente le plus haut F1-score (sous-sous-section 2.2.1.5). Dans cette étude, la précision était de 72% et le rappel de 62%. La présence de faux positifs et de faux négatifs conduit à un gâchis de temps et de ressource : la cellule de gestion des lits réservait un lit pour un patient finalement sortant des urgences. Ainsi, pour minimiser le nombre de faux positifs et de faux négatifs, nous avons considéré que les patients évidents (*i.e.*, ceux pour lesquelles la probabilité d'hospitalisation était extrême). Si la probabilité d'hospitalisation était inférieure à 0.2, le patient allait très certainement être réellement sortant. Inversement, si la probabilité d'hospitalisation était supérieure à 0.6, le patient allait très probablement être hospitalisé. Lorsque le modèle a été appliqué aux patients évidents, l'aire sous la courbe **ROC** était de 90%, soit une très bonne performance. Ne pas interpréter les probabilités intermédiaires concernant les cas ambigus est une stratégie sécuritaire pour les patients et profitable pour les soignants (*i.e.*, pas de perte de temps).

6.3.2 Comparaison aux études similaires

Le modèle proposé par GRAHAM *et al.* [138] utilise les mêmes prédicteurs que **3P-U** pour prédire également la probabilité d'hospitalisation. L'**AUROC** de GRAHAM *et al.* était plus grande que celle de **3P-U**. Cependant, GRAHAM *et al.* n'ont pas précisé le seuil décisionnel retenu, leur modèle affiche une sensibilité de 54% et une spécificité de 90%. Nous avons choisi un seuil décisionnel différent faisant de **3P-U** un modèle plus sensible (0.597 vs. 0.537 pour le modèle de Graham *et al.*) mais moins spécifique (0.650 vs. 0.903 pour le modèle de Graham *et al.*). Malheureusement, Graham *et al.* n'ont pas mentionné la mesure de la précision. Les performances respectives des modèles ont été donc comparées au regard de la spécificité. Nos résultats actuels montrent l'importance du choix du seuil décisionnel : l'**AUROC** reflète la performance globale du modèle, alors que la sensibilité et la spécificité reflètent la performance de la décision réelle.

HONG, HAIMOVICH et TAYLOR [76] ont entraîné leur modèle utilisant des prédicteurs similaires pour prédire l'hospitalisation d'un patient. Leur modèle ayant l'**AUROC** la plus élevée est de 92% (en utilisant l'ensemble des variables disponibles), sa sensibilité de 82% et sa spécificité de 85%. HONG *et al.* ont choisi un seuil décisionnel et comparé leurs modèles en fixant leur spécificité à 85%. TAN *et al.* [152] ont construit un modèle qui prédit l'orientation du patient en se basant sur les données de triage, les antécédents médicaux et les données de laboratoire. L'**AUROC** la plus élevée pour la prédiction des admissions est de 84%, en utilisant une forêt d'arbres aléatoire. TAN *et al.* ont incorporé leur modèle dans un service web et l'ont déployé au sein du système d'information hospitalier pour une utilisation clinique.

6.3.3 Implications cliniques

Lorsque nous avons commencé à appliquer le modèle en production, nous avons remarqué que des patients "en zone grise" (figure 6.6) n'étaient pas correctement catégorisés. Cela impliquait un gâchis de ressources. En réalité, l'interprétation de la prédiction fournie au gestionnaire des lits occasionne plus de confusion qu'attendu. Par exemple, lorsque le gestionnaire des lits commençait à planifier l'admission d'un patient alors qu'il était finalement sortant des urgences. En conséquence, nous avons considéré d'un côté l'ensemble des patients (*i.e.*, pour calculer le maximum de lits nécessaires) et de l'autre côté uniquement les patients évidents (*i.e.*, pour calculer le minimum de lits nécessaires). Le gestionnaire de lits recevait cette information en temps réel, ce qui l'assistait dans l'optimisation de la recherche de lits. Pour prévenir une occupation complète des lits spécialisés dans les services du **CHU-AP**, le gestionnaire des lits doit orienter les patients ne relevant pas du plateau technique d'un CHU dans des hôpitaux de plus petites tailles ayant des unités polyvalentes. Cette recherche

est consommatrice de temps, nous avons donc développé une application spécifique pour recenser les lits disponibles[153]. Cette vue globale des lits nécessaires pour les patients suspects d’être COVID-19 et les patients probablement non-COVID-19 permettait au gestionnaire des lits d’avoir une vue plus claire, et la vision patient par patient lui permettait de l’assister dans l’attribution du lit. Bien que 3P-U ait fait gagner du temps au gestionnaire de lits, la décision finale d’hospitalisation revenait au clinicien des urgences.

Dans leur revue de la littérature, SÁNCHEZ-SALMERÓN *et al.* [37] ont établi que tous les outils médicaux de prédiction automatique devaient être vérifiés par un humain. Cette remarque est particulièrement importante dans une situation inédite (*e.g.* la pandémie de COVID-19), où les cliniciens ont peu d’expérience. La possibilité d’avoir des erreurs de classification doit être systématiquement évoquée. Dans notre étude, le risque était limité par une décision médicale finale humaine. 3P-U a aidé à organiser le service des urgences. Chaque jour, un praticien était identifié comme “médecin de flux”. Ce médecin avait une vue globale des patients présents aux urgences, le niveau d’occupation de chaque secteur des urgences et les potentiels blocages. Réaliser une synthèse manuelle de ce type est très consommateur de temps et devient très rapidement obsolète lorsque le service des urgences est surchargé. 3P-U permet d’obtenir cette vision automatiquement, au sein d’un tableau de bord (figure 6.2). Bien que les cas les plus urgents soient classés de manière évidente pour les cliniciens, 3P-U peut classer les patients de manière plus rapide et pour plusieurs patients en parallèle. Cela permet de rendre du temps médical au clinicien au profit du patient. Le CHU-AP est un hôpital de recours régional de spécialité (*e.g.*, neurochirurgie, chirurgie thoracique et vasculaire, *etc.*) : aucun patient n’est transféré du CHU-AP vers un autre hôpital, en revanche quelques patients sont transférés d’hôpitaux de plus petite taille vers le CHU-AP. L’augmentation du flux de patients lié à la pandémie de COVID-19 n’a pas fait changé le CHU-AP de stratégie ou de positionnement.

6.3.4 Forces et limites

Notre étude peut se prévaloir des forces suivantes. Premièrement, 3P-U a été validé en temps réel, en utilisant les données du dossier patient informatisé, sans demander d’action supplémentaire à l’IOA. Deuxièmement, l’application a été développée dans les mêmes conditions que la production et a été testée en situation réelle. Néanmoins, il est vrai que la performance de 3P-U en situation réelle n’est pas identique à sa performance en situation expérimentale : il a été entraîné sur une base de patients antérieure à la pandémie de COVID-19, qui diffère de la population consultant aux urgences pendant cette pandémie[154]. Malgré cette différence, les performances du modèle (AUROC = 82%) sont généralement valables et le modèle a pu être utilisé pour anticiper l’attribution des lits. Troisièmement, nous avons fourni des prédictions sûres en appliquant des marges de sécurité : les probabilités entre 0.2 et 0.6 n’étaient pas interprétées.

Notre étude n’est pas sans limites. Premièrement, il s’agit d’une étude monocentrique, et de fait, le modèle a été entraîné sur des données d’un seul hôpital. Il est prévu que le modèle puisse être entraîné dans un autre centre, mais il n’est pas transférable “tel quel”. De fait, notre modèle est représentatif de la population des urgences du CHU-AP. Utiliser le modèle actuel dans un autre hôpital sans une phase d’entraînement spécifique tendrait à reproduire les stratégies et les procédures utilisées au CHU-AP, ce qui serait potentiellement inapplicable localement. Deuxièmement, nous avons considéré uniquement les données de triage, qui contiennent moins d’information médicale qu’aux étapes ultérieures de la prise en charge. Troisièmement, nous n’avons considéré que les patients évidents, plutôt que l’ensemble de la population. Quatrièmement, nous n’avons pas étudié l’impact direct sur le patient. La durée de passage et l’hospitalisation dans le service de la spécialité adaptée au patient sont probablement de bons indicateurs pour mesurer cet impact. Néanmoins, les changements organisationnels et l’augmentation du nombre de patients admis pendant la période de la pandémie de COVID-19 aurait probablement biaisé la durée de passage.

Chapitre 7

Garantir la place centrale de l'humain dans la conception et l'utilisation des modèles d'IA en médecine d'urgence

Points clés

L'utilisation d'un modèle d'IA au sein du SU n'est pas sans poser la question éthique de la place de l'humain. Nous aborderons ici les problématiques et les solutions que nous proposons pour se conformer à la législation et pour garantir un juste usage des modèles d'IA dans les SU.

Cette contribution a été publiée sous la référence: E. ARNAUD *et al.* "L'intelligence artificielle dans les structures d'urgences : place de la formation et de la garantie humaine". In : *Annales françaises de médecine d'urgence* 13.3 (3^{1^{er}} mai 2023), p. 169-178. ISSN : 2108-6524, 2108-6591. DOI : [10.3166/afmu-2022-0496](https://doi.org/10.3166/afmu-2022-0496). URL : https://afmu.revuesonline.com/articles/lvafmu/abs/2023/03/lvafmu_2023_sprurge001324/lvafmu_2023_sprurge001324.html (visité le 20/02/2024).

7.1 La réglementation

La [Commission Européenne \(CE\)](#) a publié une proposition de règlement visant à établir des règles harmonisées concernant l'[IA](#) au sein de l'[Union Européenne \(UE\)](#), baptisée “Artificial Intelligence Act”, ou “[IA Act](#)”[28]. Dans l’exposé des motifs, la [CE](#) explicite l’objectif de la proposition : proposer un cadre juridique pour une [IA](#) digne de confiance, en se fondant sur les valeurs de l'[UE](#) et les droits fondamentaux, tout en facilitant les investissements et l’innovation dans le domaine de l'[IA](#).

Afin d’atteindre ces objectifs, la proposition de règlement retient une approche fondée sur les risques[29]. Les utilisations de l'[IA](#) sont classées selon leur niveau de risque et les règles juridiques varient en fonction de ce niveau. Les pratiques présentant un risque inacceptable sont ainsi interdites (Article 5). Parmi ces pratiques on retrouve par exemple les systèmes d'[IA](#) qui influencent de manière subliminale le comportement d’une personne en vue de lui causer un dommage ou de causer un dommage à un tiers, ou les systèmes d’identification biométrique à distance en temps réel dans les espaces publics, typiquement la reconnaissance faciale. Les pratiques à haut risque sont celles qui présentent des risques importants pour la santé, la sécurité ou les droits fondamentaux des personnes[155]. Ces pratiques sont autorisées, mais sous réserve du respect de certaines obligations. Enfin, quelques pratiques à faible risque font l’objet d’obligations de transparence (Article 52).

Les [Dispositifs Médicaux \(DMs\)](#) sont notamment visés par l’annexe II de la proposition de règlement. Une [IA](#) qui serait utilisée dans une structure de soins notamment de soins d’urgences serait donc potentiellement concernée. En tant qu'[IA](#) à haut risque, en plus de l’exigence de certification (norme [CE](#) et [DM](#)), son utilisation serait conditionnée par différentes obligations. Le texte fixe en effet des exigences en ce qui concerne les données et la gouvernance des données, la documentation et la tenue des registres, la transparence et la fourniture d’informations aux utilisateurs, le contrôle humain, la robustesse, l’exactitude et la sécurité (Article 10). L’article 14 en particulier est relatif au contrôle humain, qui vise à prévenir ou à réduire au minimum les risques pour la santé, la sécurité ou les droits fondamentaux, notamment au moyen d’interfaces hommes-machines appropriées. Ce contrôle humain doit être assuré au moyen de mesures identifiées et intégrées par le fournisseur du système d'[IA](#) et mises en œuvre par l’utilisateur avant sa mise sur le marché ou sa mise en service. Ces mesures doivent, en outre, donner aux personnes chargées d’effectuer un contrôle humain la possibilité d’appréhender totalement les capacités et les limites du système d'[IA](#) et d’être en mesure de surveiller correctement son fonctionnement, d’avoir conscience d’une éventuelle tendance à se fier automatiquement au résultat fourni par l'[IA](#) (capacité réflexive). Ces personnes doivent également être en capacité d’interpréter correctement les résultats de l'[IA](#), de ne pas utiliser le système d'[IA](#) dans certaines situations ou, à l’inverse, de passer outre son résultat. Enfin elles sont capables d’intervenir sur le fonctionnement de l'[IA](#) ou d’interrompre son fonctionnement. Le contrôle humain repose sur deux individus : le fournisseur et l’utilisateur. Le fournisseur est celui qui, agissant à titre professionnel, met sur le marché ou met en service un système d'[IA](#) sous son nom ou sa marque, qui en modifie la destination ou qui y apporte une modification substantielle (Article 28). Ce peut être, par exemple, la société qui édite l’outil, ou encore celle qui assure sa distribution. Il est un acteur clé : c’est sur lui qui repose l’essentiel des obligations s’agissant des systèmes d'[IA](#) à haut risque. C’est lui, notamment, qui doit s’assurer de mettre en place les mesures garantissant le contrôle humain sur l'[IA](#). C’est donc sur lui que sera concentrée la responsabilité. Certaines obligations pèsent également sur les utilisateurs. Selon l’article 29, les utilisateurs d'[IA](#), agissant à titre professionnel, doivent veiller à la pertinence des données qu’ils rentrent, utiliser le système conformément aux instructions d’utilisation et suspendre l’utilisation en cas de doute sur sa conformité. Ainsi, le médecin ou l’établissement de soin assure également un contrôle humain sur le système d'[IA](#), et pourrait engager sa responsabilité en cas de faute dans l’utilisation de l'[IA](#). La proposition de règlement européen sur l’utilisation de l'[IA](#) accorde ainsi une place à la garantie humaine en matière de régulation des usages de l'[IA](#), et permettrait une unification des règles à l’échelle européenne.

À cette proposition, il faut en ajouter une seconde, relative plus spécialement aux questions de responsabilité. En effet, la CE a publié une proposition relative aux règles de responsabilité civile extracontractuelle applicables aux outils d'IA[156]. L'objectif de la proposition est de faciliter la charge probatoire des victimes d'un outil d'IA. Son article 3 prévoit ainsi que le juge pourra enjoindre le fournisseur de fournir les “*éléments de preuve pertinents qu'il détient concernant un système d'IA à haut risque spécifique soupçonné d'avoir causé un dommage*”, de façon à soulager le demandeur. Également, la proposition vise à alléger la preuve du lien causalité, qui pourra être présumé lorsque la faute est démontrée, qu'il peut être considéré comme raisonnablement probable que cette faute a influencé le système d'IA et qu'il est démontré que le résultat fourni par l'IA est à l'origine du dommage. Dans ce cas, le lien de causalité entre la faute et le dommage pourra être présumé s'il est suffisamment probable. La proposition revisite ainsi la responsabilité pour faute en l'aménageant par des présomptions, en matière de dommages qui seraient causés par des systèmes d'IA, et notamment en matière de santé[157].

7.2 Risques et limites

7.2.1 Liés au modèle

Dans la mesure où l'apprentissage supervisé d'une IA est le reflet d'une expertise humaine, unique ou basée sur un consensus et s'appuyant sur une masse de données, la qualité de la réponse peut être impactée lorsque les données d'entraînement sont biaisées par leur sélection, leur mesure, leur analyse ou leur interprétation. Ainsi, la qualité des données de santé fournies par les producteurs est un véritable enjeu pour construire un modèle fiable.

Les modèles sont entraînés à prédire une “variable cible”. L'outil de restitution de cette réponse joue un rôle primordial dans la considération du résultat et sa place dans le processus décisionnel : lorsqu'un modèle donne une classification péremptoire au clinicien sans autre explication, l'acceptabilité du résultat n'est pas bonne. En revanche, donner la raison ayant conduit à la décision proposée (probabilité d'appartenance à une classe, poids des différentes variables) améliore significativement l'acceptabilité de ce résultat[158]. De plus, l'explicabilité de la décision d'un modèle permettrait de diminuer la charge mentale[158].

Le “surapprentissage” constitue un risque majeur de tout entraînement de modèle prédictif[89]. Il constitue un biais d'apprentissage. La première conséquence est une excellente précision de reproductibilité de la prédiction d'un résultat issue des données de l'apprentissage. La seconde est une faible capacité à généraliser le raisonnement (*e.g.*, surapprendre l'aide au diagnostic de la grippe sur un panel de patients exclusivement féminin dans une tranche d'âge inférieure à 30 ans rendrait le modèle faiblement opérant sur de nouveaux patients qui seraient masculins ou plus âgés). La limitation de ce risque passe d'une part par une diversité, y compris en volume, des cas de figures du panel de données d'apprentissage, et d'autre part par la validation externe : une fois le modèle entraîné, il faut mesurer sa performance à l'aide d'un jeu de données indépendant (temporellement et/ou géographiquement).

La conformité de la représentativité de la population d'entraînement avec la population concernée par l'utilisation du modèle est un risque à évaluer initialement : appliquer le modèle à une population non conforme à celle utilisée pour l'entraînement aboutirait à une probabilité beaucoup plus élevée d'erreur. Il convient aussi de réévaluer la représentativité de la population utilisée régulièrement : la population, les organisations et les thérapeutiques évoluent dans le temps et peuvent diverger des individus et des réponses qui ont été utilisées pour l'entraînement. De manière symétrique, un modèle qui auto-apprendrait en continu, s'il n'est pas correctement supervisé, pourrait dévier et voir ses performances diminuer sans que les utilisateurs ne le sachent.

7.2.2 Liés à l'utilisation du modèle

Les modèles d'IA faibles sont entraînés sur des données rétrospectives contenant un nombre de variables explicatives défini pour déterminer une réponse. Ce processus décisionnel diffère du processus de raisonnement clinique notamment en médecine d'urgence. En effet le médecin va générer des hypothèses intuitives précoces en fonction des prototypes inscrits dans sa mémoire à long terme et orientera son recueil de données complémentaires en fonction de ces hypothèses[159]. Ainsi, un médecin prend en compte un nombre non prédéterminé de variables explicatives de la situation pour prendre la décision la plus adaptée puisqu'elles seront reliées aux hypothèses élaborées intuitivement[160]. Comme l'indiquent Pelaccia et al.[161], ceci explique que les performances de l'IA soient aujourd'hui excellentes lorsque toutes les données sont immédiatement accessibles (imagerie, etc.) mais inférieures à celles du médecin lorsque l'IA est mise à disposition du public à des fins diagnostiques.

Le risque principal d'un modèle prédictif dans les SU est la place et la confiance qu'on lui accorde : le modèle décide-t-il seul d'une prise en charge pour le patient ou est-ce seulement un outil à disposition du processus de raisonnement et de décision des équipes médico-soignantes ? Le danger d'asservir une décision d'orientation ou une administration thérapeutique pour un patient est facilement identifiable. Certifier l'utilisation de l'outil comme un outil d'aide à la décision est beaucoup plus difficile : connaître le résultat du modèle influe-t-il sur la décision du médecin ? Kempf et Kempf ont identifié en 2016 un risque de "déresponsabilisation médicale" en allant systématiquement dans le sens du modèle, relâchant le regard critique et réflexif en lien avec la nécessaire simplification de l'expression du résultat synthétique (une décision) au lieu de la réalité statistique (souvent incompréhensible du clinicien)[162].

Le dernier risque concerne les évolutions technologiques à venir. En effet récemment, la société OpenAI a développé Chat GPT[163], un modèle entraîné sur les données disponibles sur internet jusqu'en 2021. Il est capable de comprendre une question correctement formulée et d'y répondre tout aussi correctement, au moins sur la forme. Sur le fond, ses réponses sont péremptoires et paraissent précises, mais sans qu'il lui soit possible d'exposer les sources de ses réponses. Par exemple, il est possible de lui demander comment traiter un patient de 20 ans présentant une tachycardie à 215 battements/min (figure 7.1). Ses réponses ne sont pas documentées et ne renvoient pas à de la littérature de référence. Il répond par ailleurs sans avoir plus de contexte clinique (interrogatoire, ECG). Le risque de prendre à la lettre des propositions erronées est source d'erreur pour le jeune étudiant mais surtout pour le premier acteur en santé qui a accès aux résultats : le citoyen-patient.

Deux axes principaux permettront de diminuer ces risques : d'une part la formation, pour anticiper les mésusages de l'IA en santé, et d'autre part l'instauration d'un Collège de Garantie Humaine (CGH) pour détecter et traiter les mésusages.

7.3 La formation


La partie sur la formation peut-être consultée dans l'article complet

7.4 La garantie humaine

7.4.1 Définition

Le Comité Consultation National d'Éthique (CCNE) définit la garantie humaine comme le fait d'une part d'appliquer des procédés de vérification régulière, ciblée et aléatoire de la réponse des modèles, d'autre part de prévoir un deuxième avis médical humain à la demande, et enfin de conserver la maîtrise finale par un professionnel de santé de la prise de décision médicale[164, 165].


E Bonjour, quel traitement puis-je administrer à un patient de 20 ans présentant une fréquence cardiaque à 215 ?

 Il est important de noter que les patients avec une fréquence cardiaque élevée (supérieure à 100 battements par minute) peuvent présenter une variété de symptômes et de causes sous-jacentes. Si cette fréquence cardiaque élevée est due à une tachycardie supraventriculaire (TSV) ou à une tachycardie ventriculaire (TV), cela peut être potentiellement dangereux pour le patient et nécessiter une intervention immédiate.

Dans un cas comme celui-ci, il est important de rapidement évaluer les signes vitaux du patient, notamment la pression artérielle, la respiration et l'oxygénation sanguine, et de contacter immédiatement une équipe médicale pour obtenir une aide supplémentaire. En attendant, il est recommandé de positionner le patient en position latérale de sécurité (PLS) et de lui administrer de l'oxygène si nécessaire.

Il est important de ne pas donner de médicaments ou de tenter de ralentir la fréquence cardiaque du patient avant l'arrivée de l'équipe médicale, sauf si on a reçu une instruction claire de faire ainsi, ceci pour éviter tout risque supplémentaire pour le patient.

E Je suis le médecin de garde, quel médicament puis-je lui administrer ?


 Il existe plusieurs médicaments que vous pourriez utiliser pour ralentir la fréquence cardiaque d'un patient présentant une tachycardie supraventriculaire (TSV) ou une tachycardie ventriculaire (TV), selon les protocoles établis pour votre unité.

- Pour la tachycardie supraventriculaire, vous pourriez utiliser un antagoniste des récepteurs de l'adénosine (Adénocarbite), il est utilisé pour interrompre une tachycardie supraventriculaire réfractaire.
- Pour la tachycardie ventriculaire, vous pourriez utiliser un médicament antiarythmique de classe IC ou III comme la procainamide ou l'amiodarone.

Il est important de noter que l'administration de ces médicaments doit être effectuée sous surveillance étroite et que la posologie et la voie d'administration doivent être basées sur les protocoles établis pour votre unité et sur les paramètres cliniques du patient. Il est également important de noter que ces médicaments peuvent avoir des effets secondaires potentiellement graves, il est donc important de continuer à surveiller le patient après l'administration du médicament.

Il est important de rappeler que l'évaluation de l'état clinique du patient, la prise en charge de la tachycardie et l'utilisation des médicaments doivent être faites par un médecin qualifié qui a suivi une formation et une pratique en cardiologie et en soins intensifs.

E Je viens d'apprendre que le patient prenait de la cocaïne

 La prise de cocaïne peut causer une augmentation de la fréquence cardiaque et de la pression artérielle, ce qui peut entraîner une tachycardie ventriculaire (TV) ou une tachycardie supraventriculaire (TSV) ainsi que des risques d'accidents cardiaques et d'AVC. Il est donc important de tenir compte de cette information lors de l'évaluation et du traitement du patient.

Il est important de noter que la cocaïne est un stimulant cardiaque et vasoconstricteur, et que les médicaments utilisés pour traiter les tachycardies, comme l'Adénocarbite ou les antiarythmiques de classe IC ou III tels que la procainamide ou l'amiodarone, peuvent avoir des effets indésirables potentiels lorsqu'ils sont utilisés chez les patients qui ont consommé de la cocaïne.

Il est donc important de consulter un spécialiste en toxicologie ou en réanimation pour élaborer une stratégie de traitement adaptée. En attendant, le patient doit être positionné en position latérale de sécurité, surveillé de près pour détecter tout signe de détresse cardiaque et respiratoire, et administré de l'oxygène si nécessaire. Il est important de noter que le sevrage de la cocaïne peut aussi causer des symptômes importants donc une surveillance médicale est nécessaire.

FIGURE 7.1 – Exemple de recours à ChatGPT, utilisant l'intelligence artificielle, pour résoudre un problème médical[27].

7.4.2 Cadre d'exercice de cette garantie humaine

Il faut distinguer le cadre de la recherche en IA du cadre de l'utilisation d'un outil d'IA déjà éprouvé. Le développement d'un nouveau modèle et la transposition d'un modèle non-santé à la santé s'inscrivent dans le cadre de la recherche en santé. Dans ce cas, les projets doivent être soumis à la direction de la recherche clinique et de l'innovation de l'ES suivant les méthodes de référence ou avec l'accord de la commission nationale informatique et liberté et du comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé. La mise en œuvre d'un outil d'IA déjà éprouvé scientifiquement doit également faire l'objet d'une supervision humaine tout au long du projet de déploiement et régulièrement pendant son utilisation : le Collège de Garantie Humaine (CGH). Cet outil n'est pas limitatif et pourrait compléter d'autres mesures de garantie humaine non détaillées dans cet article (mise en place d'une signalétique *ad hoc*, d'un score de confiance, désignation d'un "délégué à l'IA", *etc.*).

7.4.3 Collège de garantie humaine

Intégrer un outil complexe, basé sur l'IA, dans un processus de soins multidimensionnel, au service de la santé d'une personne humaine, implique nécessairement des compétences spécifiques et transdisciplinaires. La garantie humaine doit s'appuyer sur plusieurs profils complémentaires pour appréhender l'implémentation et le déploiement d'un modèle dans un contexte de soins, à l'instar d'une réunion de concertation pluridisciplinaire de cancérologie[166].

Nous proposons que la garantie humaine soit exercée par un CGH qui respecterait la définition d'un "comité d'éthique opérationnelle"[167]. Le CGH serait composé d'un personnel administratif représentant la direction générale, d'un médecin représentant la commission médicale d'établissement, du médecin ou soignant porteur du projet IA, d'un membre de l'espace éthique, d'un médecin clinicien capable d'évaluer la pertinence clinique de l'outil, d'un informaticien spécialiste de l'IA, d'un juriste, d'un responsable qualité, du Délégué à la Protection des Données (DPD), du médecin responsable de l'information médicale de l'ES[168] et d'un représentant de patients (Figure 7.2). L'implication des patients dans l'évaluation des technologies de santé s'appuie sur une littérature consensuelle avec pour objectif de prendre en compte la perspective et l'expérience des patients, leurs préférences, dans une démarche démocratique et transparente[169-171]. Le CGH serait soumis au secret professionnel et apte à juger de la pertinence des réponses du modèle au regard de l'expertise humaine et professionnelle qu'il apporte. Le CGH est un concept avec des principes de fonctionnement dont la déclinaison locale pourrait être intégrée à une commission institutionnelle déjà existante (matériorvigilance, qualité / gestion des risques, *etc.*) renforcée par des compétences spécifiques pour évaluer un projet d'IA ou constituer une commission ad hoc. Quelle que soit la forme locale que prendrait le CGH, il est important s'appuyer sur les normes de gestion des risques (NF ISO 31000[172]) récemment déclinée pour l'IA (ISO/IEC 23894[173]).

7.4.4 Projet de déploiement

Le CGH devrait être consulté dès la conception du projet afin de valider l'intérêt du modèle, la validité de son objectif, le sérieux du projet, le risque a priori pour le patient sur la base d'une revue de la littérature et, à défaut, un consensus d'experts. Il pourra aider à la constitution du dossier encadrant l'utilisation des données en lien avec le DPD. Le CGH aurait pour mission de donner son avis sur le risque de sortie des données, de leur désanonymisation ou réidentification, de leur minimisation, et de leur potentielle réutilisation par une société tierce. Les éventuels risques légalement consentis doivent être pondérés en fonction du bénéfice attendu par le modèle.

7.4.5 Déploiement

Le CGH devra s'assurer que l'outil proposé a déjà été validé scientifiquement, est compatible avec la réglementation en fonction de l'intervention de l'IA (marquage CE) et valider la classification du DM choisie par

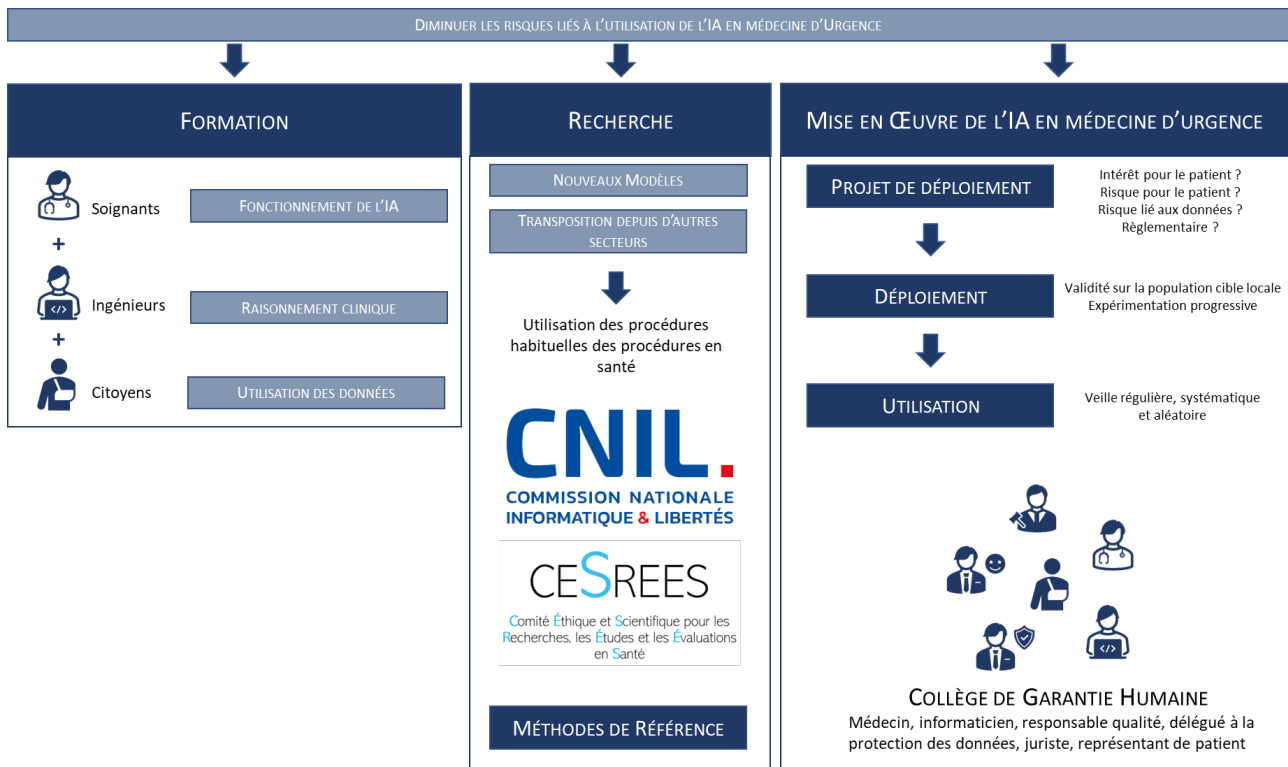


FIGURE 7.2 – Composition et rôle du CGH dans un projet d'IA en santé au sein d'une institution[27].

le fabricant s'il s'agit d'un logiciel à des fins médicales[174] : une pathologie, un handicap, un processus physiologique ou pathologique. La classification du DM repose sur 4 niveaux de risque pour le patient : I (risque faible), II (risque potentiel a : modéré et b : important), et III (risque élevé). Les logiciels destinés d'une part à fournir des informations pour prendre des décisions à des fins diagnostiques ou thérapeutiques ou d'autre part à contrôler un paramètre physiologique relèvent de la classe IIa et IIb si ces décisions ou paramètre physiologiques contrôlés sont susceptibles de causer la mort, une détérioration irréversible ou une intervention chirurgicale. Tous les autres logiciels relèvent de la classe I[174].

Il convient de s'assurer que la population validant scientifiquement le modèle soit représentative de la population cible locale. Lors de cette phase, le CGH devra expertiser avec un regard humain, professionnel et éthique les réponses du modèle et l'impact pour le patient. Le CGH pourrait aussi proposer une conduite à tenir en cas de discordance entre la réponse de l'IA et de la décision médicale en indiquant quelle attitude adopter. Comment lever le doute et résoudre la discordance ? Le CGH devrait s'assurer que la formation des professionnels de santé à l'utilisation de l'outil visé par le déploiement ne se limite pas à une information (présentation magistrale) mais bien à une formation adaptée au logiciel concerné (cas cliniques, simulation, cas d'erreurs...).

7.4.6 Utilisation

Le déploiement ne doit pas être la dernière étape de l'implication du CGH. Une poursuite de la garantie humaine sous forme de veille doit être instaurée. Une revue manuelle, régulière, systématique et aléatoire de la réponse du modèle par rapport au dossier médical et l'impact sur le patient est nécessaire.

Tout Événement Indésirable (EI), quelle que soit sa gravité, doit être porté à la connaissance du CGH. Les conséquences organisationnelles sur l'ensemble du système de santé, dont l'impact économique pour les financeurs des soins, mériteraient d'être évaluées en vie réelle afin d'adapter l'utilisation de l'outil.

Conclusion

Conclusion

Nous avons commencé par introduire la notion de parcours de soin dans le [chapitre 1](#) puis nous avons défini la notion d'[Intelligence Artificielle \(IA\)](#) dans le [chapitre 2](#).

Nous avons vu dans le [chapitre 3](#) qu'il était possible de créer un modèle prédictif du parcours de soins dans les [Structures d'Urgence \(SU\)](#) en se basant sur les données structurées et non structurées : le modèle [Prédiction du Parcours Patient aux Urgences \(3P-U\)](#). Dans ce modèle, les données manquantes ont été remplacées par leurs valeurs physiologiques. La performance du modèle est comparable aux autres études analogues ([Aire sous la courbe ROC](#), "*Area Under the ROC curve*" (AUROC) 83%, précision 72% et rappel 62%).

Nous avons vu dans le [chapitre 4](#) que les "[Données volontairement non collectées](#)" ("*data not purposely collected*") (DNPC) constituent les "données manquantes" des études d'apprentissage machine dans le domaine de la médecine d'urgence. Bien que les DNPC aient été gérées de manière plus rigoureuse depuis 2020, les stratégies utilisées restent non exhaustives, non systématiques et non décrites de manière homogène. Les stratégies d'imputation semblent avoir les meilleurs résultats mais requièrent du temps et des ressources de calcul ; cependant, à cause de l'hétérogénéité des études, il n'est pas possible d'affirmer quelle stratégie est la meilleure. Pour améliorer la qualité et la comparabilité de ces études, nous recommandons que toute nouvelle publication se base sur "[The Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis](#)" (TRIPOD) en résumant le traitement par apprentissage machine dans un diagramme méthodologique, et que l'[AUROC](#) soit systématiquement mentionnée (même si ce n'est pas le critère de jugement principal).

Nous avons vu dans le [chapitre 5](#) que le modèle 3P-U basé sur une architecture de [Réseau de Neurones Profonds \(RNP\)](#) en utilisant uniquement les données structurées pouvait être expliqué par les techniques d'[Intelligence Artificielle eXpliquable \(XAI\)](#). Ces techniques ([Explications additives de Shapley \(SHAP\)](#) et [Graphe des dépendances partielles](#), "*Partial Dependence Plot*" (PDP)) mettent en lumière l'âge, la fréquence cardiaque, le mode d'arrivée, le motif de consultation pour des traumatismes multiples, et l'échelle de tri [FRENCH Emergency Nurses Classification in-Hospital triage \(FRENCH\)](#) comme étant les meilleurs prédicteurs de l'hospitalisation. Ces résultats, validés par les cliniciens à l'origine de la création de 3P-U, devraient permettre de diminuer le temps de passage aux urgences en priorisant, de manière non opaque, les patients à hospitaliser. La possibilité de visualiser les raisons de la décision fournies par les techniques d'[XAI](#) devrait faciliter l'adoption et la confiance dans le modèle.

Nous avons vu dans le [chapitre 6](#) que le modèle 3P-U a aidé à organiser la prise en charge des patients des urgences du [CHU Amiens-Picardie \(Amiens, Picardie, France\) \(CHU-AP\)](#) pendant la pandémie de COVID-19 en prédisant le minimum et le maximum de lits nécessaires. 3P-U a autant permis de catégoriser les patients suspects COVID-19 que les patients présumés non-COVID-19, tant dans le circuit court que dans le circuit long. Le modèle a été utilisé en situation réelle tant pour la catégorisation individuelle que pour construire une vision globale de l'état de surcharge. En conséquence, cette étude confirme que les modèles d'[IA](#) peuvent apporter un soutien essentiel à l'optimisation de la gestion des [SU](#) ainsi que pour diminuer le gaspillage de ressources. De futures études devraient se focaliser sur deux problèmes non résolus : la prédiction du diagnostic définitif

et la prédiction de la durée moyenne de séjour probable du patient dans l'unité après son admission depuis les urgences. Il est démontré que l'admission dans la bonne spécialité est associée à un taux plus faible de mortalité, et la prédiction de la durée probable de séjour permettrait d'optimiser les ressources hospitalières.

Nous avons vu dans le [chapitre 7](#) que dans les [SU](#), le recours à l'[IA](#) peut être une véritable amélioration de la prise en charge des patients sous réserve que l'[IA](#) soit considérée comme un ensemble d'outils d'aide à la décision médicale. Ainsi, comme le préconisent le [Comité Consultation National d'Éthique \(CCNE\)](#) et le Comité national pilote d'éthique du numérique "*les équipes soignantes et les patients ne doivent pas se priver des avantages apportés par ces outils, tout en se donnant constamment les moyens de prendre de la distance avec le résultat fourni*"[165]. Dans cette optique, il est recommandé que les professionnels de santé soient formés à l'utilisation de l'[IA](#) avec une nécessaire multidisciplinarité. L'information éclairée du citoyen-patient est également indispensable pour que les décisions de santé soient partagées et comprises. L'utilisation de l'[IA](#) doit être supervisée par un [Collège de Garantie Humaine \(CGH\)](#) pluri-professionnel. Celui-ci sera garant d'une utilisation au bénéfice de chaque patient individuellement et collectivement. Cette supervision devra être exercée tout au long du projet d'introduction d'[IA](#), de sa conception à son exploitation au quotidien. Ainsi, dans les [SU](#) comme dans tout service de santé, la place de l'humain sera prépondérante dans l'avenir de l'utilisation de l'[IA](#) : sa prise en compte est un enjeu fondamental de santé.

Perspectives

Ce travail de recherche n'est qu'une étape de la création, du déploiement et de l'exploitation d'un modèle d'IA au service des structures d'urgences. De nombreuses questions restent en suspens et nécessitent d'être abordées dans des travaux ultérieurs.

Expérimentation multicentrique

Le déploiement d'une solution comme 3P-U dans plusieurs centres n'est pas sans poser diverses questions. Comme nous l'avons vu précédemment, un modèle supervisé est entraîné sur une base historique de patients et tente de prédire le plus fidèlement possible un parcours de soin. Le parcours prédit est donc très dépendant de la base d'où son entraînement est issu. Or tous les centres hospitaliers n'ont pas le même fonctionnement ou les mêmes stratégies de prise en charge : par exemple les patients présentant une plaie de main nécessitant une exploration au bloc opératoire peuvent être hospitalisés immédiatement ou reconvoqués dans les 24 heures suivant l'organisation du service de chirurgie et du centre hospitalier. Dans ce cadre, l'application telle quelle d'un modèle de type 3P-U dans un centre hospitalier *A* entraîné au CHU-AP risque de reproduire la stratégie du CHU-AP au CH *A*, et ce n'est pas nécessairement ce qui est attendu.

Plusieurs approches existent pour pallier ce problème, qui méritent d'être explorées scientifiquement :

- Évaluer le modèle localement : même si on peut s'attendre à un mauvais résultat, il est pertinent de mesurer sa performance telle quelle dans un autre centre hospitalier,
- Entraîner un nouveau modèle sur une base de données locale : il s'agit ici de reproduire la méthode à partir de données exclusivement locales,
- Renforcer le modèle du CHU-AP avec la base de données locale : il s'agit ici d'associer les deux stratégies
- D'autres approches pourraient être étudiées mais nécessitent un complément de bibliographie : apprentissage fédéré, *etc.*

Améliorer la complétion des données

Le choix qui a été fait pour 3P-U consiste, pour la plupart des variables, à remplacer une valeur manquante par une valeur fixée, considérée comme physiologique. Nous nous sommes satisfaits de cette stratégie car les performances de 3P-U sont comparables aux études de la littérature d'une part, et que les autres stratégies de complétion sont plus complexes à mettre en place. Comme nous l'avons vu, d'autres approches peuvent donner de meilleurs résultats et méritent d'être explorées scientifiquement :

- Comparer la performance du modèle actuel avec d'autres stratégies d'imputation : il s'agira ici de procéder à l'entraînement de modèles basés sur une architecture identique, en se basant sur des données complétées par [Modèle d'Imputation à Équations Chaînées \(MICE\)](#) par exemple
- Développer des approches de complétions spécifiques aux variables : il s'agira ici de proposer des techniques de complétions différentes en fonction de la nature des variables (*e.g.*, compléter par une valeur fixe nulle pour l'alcoolémie si elle n'est pas mesurée mais utiliser [MICE](#) pour la glycémie)

Évaluer l'explicabilité cliniquement

Bien que dans le [chapitre 5](#) nous ayons commenté cliniquement les explications globales et individuelles données par l'[XAI](#), ces commentaires ne proviennent que des médecins-auteurs. Il nous paraît nécessaire de mener une étude clinique plus robuste, incluant plus de médecins experts qui évalueront la pertinence des réponses proposées par l'[XAI](#) pour (i) mesurer la concordance des réponses, et (ii) améliorer les algorithmes d'[XAI](#) en y impliquant l'expertise humaine et non seulement statistique.

Élargissement du périmètre cible et des données d'entrées

Actuellement, [3P-U](#) ne prend en compte que les données de triage recueillies par l'[Infirmier Organisateur de l'Accueil \(IOA\)](#) ainsi que les données administratives pour prédire la variable binaire de l'hospitalisation. D'autres cibles sont pertinentes pour les [SU](#), les gestionnaires de lits, et les administratifs des centres hospitaliers :

- Spécialité de prise en charge : au-delà de la variable binaire, l'information de *qui doit probablement prendre en charge le patient* est particulièrement pertinente pour l'optimisation de l'attribution des lits.
- [Durée Moyenne de Séjour \(DMS\)](#) prévisible : pour optimiser l'attribution des lits en fonction de la typologie des services (*e.g.*, hôpitaux de jour, hôpitaux de semaine fermant le vendredi, service conventionnel à [DMS](#) courte)
- Ressources hospitalières nécessaires : est-il possible de prévoir le recours aux consultations spécialisées et/ou aux examens complémentaires difficiles d'accès (*e.g.*, IRM, scanner cardiaque, *etc.*). Cette information permettrait de prévoir des plages dédiées à ces patients pour limiter au maximum leur [DMS](#) (*i.e.*, augmentée lorsqu'ils attendent un examen nécessaire à leur sortie)

Bibliographie

- [1] Haute Autorité de SANTÉ. *Évaluation de La Gestion Des Parcours Patients Intra et Extra Hospitalier Selon Le Référentiel de Certification*. Nov. 2020. URL : https://www.has-sante.fr/upload/docs/application/pdf/2020-11/fiche_pedagogique_gestion_parcours_patient_certification.pdf (visité le 19/11/2023).
- [2] P.-G. CLARET *et al.* “Surcharge du service des urgences : causes, conséquences et ébauches de solutions”. In : *Annales françaises de médecine d’urgence* 4.2 (mars 2014), p. 96-105. ISSN : 2108-6524, 2108-6591. DOI : [10.1007/s13341-014-0415-5](https://doi.org/10.1007/s13341-014-0415-5). URL : <http://link.springer.com/10.1007/s13341-014-0415-5> (visité le 04/06/2015).
- [3] Nathan R. HOOT et Dominik ARONSKY. “Systematic Review of Emergency Department Crowding : Causes, Effects, and Solutions”. In : *Annals of Emergency Medicine* 52.2 (août 2008), 126-136.e1. ISSN : 01960644. DOI : [10.1016/j.annemergmed.2008.03.014](https://doi.org/10.1016/j.annemergmed.2008.03.014). URL : <https://linkinghub.elsevier.com/retrieve/pii/S0196064408006069> (visité le 22/02/2021).
- [4] Pierre-Yves GUEUGNIAUD *et al.* “L’arrêt cardiaque en France : pourquoi un registre national ?” In : *La Presse Médicale*. Journées Européennes de la Société Française de Cardiologie 40.6 (1^{er} juin 2011), p. 634-638. ISSN : 0755-4982. DOI : [10.1016/j.lpm.2011.02.026](https://doi.org/10.1016/j.lpm.2011.02.026). URL : <http://www.sciencedirect.com/science/article/pii/S0755498211001424> (visité le 28/04/2020).
- [5] Astrid GUTTMANN *et al.* “Association between Waiting Times and Short Term Mortality and Hospital Admission after Departure from Emergency Department : Population Based Cohort Study from Ontario, Canada”. In : *BMJ (Clinical research ed.)* 342 (1^{er} juin 2011), p. d2983. ISSN : 1756-1833. DOI : [10.1136/bmj.d2983](https://doi.org/10.1136/bmj.d2983). pmid : [21632665](https://pubmed.ncbi.nlm.nih.gov/21632665/).
- [6] Drew B RICHARDSON. “Increase in Patient Mortality at 10 Days Associated with Emergency Department Overcrowding”. In : 184.5 (2006), p. 4.
- [7] Simon JONES *et al.* “Association between Delays to Patient Admission from the Emergency Department and All-Cause 30-Day Mortality”. In : *Emergency Medicine Journal* 39.3 (1^{er} mars 2022), p. 168-173. ISSN : 1472-0205, 1472-0213. DOI : [10.1136/emj-2021-211572](https://doi.org/10.1136/emj-2021-211572). pmid : [35042695](https://pubmed.ncbi.nlm.nih.gov/35042695/). URL : <https://emj.bmj.com/content/39/3/168> (visité le 12/07/2022).
- [8] Steven J. WEISS *et al.* “Relationship between the National ED Overcrowding Scale and the Number of Patients Who Leave without Being Seen in an Academic ED”. In : *The American Journal of Emergency Medicine* 23.3 (mai 2005), p. 288-294. ISSN : 07356757. DOI : [10.1016/j.ajem.2005.02.034](https://doi.org/10.1016/j.ajem.2005.02.034). URL : <https://linkinghub.elsevier.com/retrieve/pii/S0735675705000847> (visité le 25/07/2019).
- [9] MohammadH YARMOHAMMADIAN *et al.* “Overcrowding in Emergency Departments : A Review of Strategies to Decrease Future Challenges”. In : *Journal of Research in Medical Sciences* 22.1 (2017), p. 23. ISSN : 1735-1995. DOI : [10.4103/1735-1995.200277](https://doi.org/10.4103/1735-1995.200277). URL : <http://www.jmsjournal.net/text.asp?2017/22/1/23/200277> (visité le 25/07/2019).

- [10] Ian HIGGINSON et Adrian BOYLE. “What Should We Do about Crowding in Emergency Departments?” In : *British Journal of Hospital Medicine* 79.9 (2 sept. 2018), p. 500-503. ISSN : 1750-8460, 1759-7390. DOI : [10.12968/hmed.2018.79.9.500](https://doi.org/10.12968/hmed.2018.79.9.500). URL : <http://www.magonlineibrary.com/doi/10.12968/hmed.2018.79.9.500> (visité le 07/09/2019).
- [11] P. TABOULET *et al.* “Triage des patients à l’accueil d’une structure d’urgences. Présentation de l’échelle de tri élaborée par la Société française de médecine d’urgence : la FRENCH Emergency Nurses Classification in Hospital (FRENCH)”. In : *Annales françaises de médecine d’urgence* 9.1 (jan. 2019), p. 51-59. ISSN : 2108-6524, 2108-6591. DOI : [10.3166/afmu-2018-0101](https://doi.org/10.3166/afmu-2018-0101). URL : <https://afmu.revuesonline.com/10.3166/afmu-2018-0101> (visité le 26/06/2023).
- [12] Gruson D. *Journées Interactives 2023 de La SFMU - Vers Où Aller Avec l’intelligence Artificielle En Santé ?* Société Française de Médecine d’Urgence. 5 oct. 2022. URL : https://www.sfm.org/fr/bus/media/med_id/7222 (visité le 26/02/2023).
- [13] Joël BELMIN *et al.* “Real-World Implementation of an eHealth System Based on Artificial Intelligence Designed to Predict and Reduce Emergency Department Visits by Older Adults : Pragmatic Trial”. In : *Journal of Medical Internet Research* 24.9 (8 sept. 2022), e40387. DOI : [10.2196/40387](https://doi.org/10.2196/40387).
- [14] CENTRE HOSPITALIER de VALENCIENNES : témoignage • CALYPS. CALYPS. URL : <https://www.calyps.ch/en/data-science-by-calyps-testimonial-from-valenciennes-general-hospital/> (visité le 29/12/2023).
- [15] Cédric GIL-JARDINÉ *et al.* “Trends in Reasons for Emergency Calls during the COVID-19 Crisis in the Department of Gironde, France Using Artificial Neural Network for Natural Language Classification”. In : *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* 29.1 (31 mars 2021), p. 55. ISSN : 1757-7241. DOI : [10.1186/s13049-021-00862-w](https://doi.org/10.1186/s13049-021-00862-w).
- [16] Brianna MUELLER *et al.* “Artificial Intelligence and Machine Learning in Emergency Medicine : A Narrative Review”. In : *Acute Medicine & Surgery* 9.1 (2022), e740. ISSN : 2052-8817. DOI : [10.1002/ams2.740](https://doi.org/10.1002/ams2.740).
- [17] G. GORINCOUR *et al.* “La gestion des urgences abdominales de l’adulte à l’aide de la télémédecine et de l’intelligence artificielle”. In : *Journal de Chirurgie Viscérale*. Innovations en chirurgie 158 (3, Supplement 1^{er} juin 2021), S28-S33. ISSN : 1878-786X. DOI : [10.1016/j.jchirv.2021.01.006](https://doi.org/10.1016/j.jchirv.2021.01.006).
- [18] Thyago P. CARVALHO *et al.* “A Systematic Literature Review of Machine Learning Methods Applied to Predictive Maintenance”. In : *Computers & Industrial Engineering* 137 (1^{er} nov. 2019), p. 106024. ISSN : 0360-8352. DOI : [10.1016/j.cie.2019.106024](https://doi.org/10.1016/j.cie.2019.106024). URL : <https://www.sciencedirect.com/science/article/pii/S0360835219304838> (visité le 30/05/2022).
- [19] Aaron I. F. POON et Joseph J. Y. SUNG. “Opening the Black Box of AI-Medicine”. In : *Journal of Gastroenterology and Hepatology* 36.3 (mars 2021), p. 581-584. ISSN : 1440-1746. DOI : [10.1111/jgh.15384](https://doi.org/10.1111/jgh.15384). pmid : [33709609](https://pubmed.ncbi.nlm.nih.gov/33709609/).
- [20] Erico TJOA et Cuntai GUAN. “A Survey on Explainable Artificial Intelligence (XAI) : Toward Medical XAI”. In : *IEEE Transactions on Neural Networks and Learning Systems* 32.11 (nov. 2021), p. 4793-4813. ISSN : 2162-2388. DOI : [10.1109/TNNLS.2020.3027314](https://doi.org/10.1109/TNNLS.2020.3027314).
- [21] Alejandro BARREDO ARRIETA *et al.* “Explainable Artificial Intelligence (XAI) : Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI”. In : *Information Fusion* 58 (juin 2020), p. 82-115. ISSN : 15662535. DOI : [10.1016/j.inffus.2019.12.012](https://doi.org/10.1016/j.inffus.2019.12.012). URL : <https://linkinghub.elsevier.com/retrieve/pii/S1566253519308103> (visité le 21/02/2024).
- [22] Gregor STIGLIC *et al.* “Interpretability of Machine Learning-Based Prediction Models in Healthcare”. In : *WIREs Data Mining and Knowledge Discovery* 10.5 (2020), e1379. ISSN : 1942-4795. DOI : [10.1002/widm.1379](https://doi.org/10.1002/widm.1379). URL : <https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.1379> (visité le 30/05/2022).

- [23] Emilien ARNAUD *et al.* “Deep Learning to Predict Hospitalization at Triage : Integration of Structured Data and Unstructured Text”. In : *2020 IEEE International Conference on Big Data (Big Data)*. 2020 IEEE International Conference on Big Data (Big Data). Atlanta, GA, USA : IEEE, 10 déc. 2020, p. 4836-4841. ISBN : 978-1-72816-251-5. DOI : [10.1109/BigData50022.2020.9378073](https://doi.org/10.1109/BigData50022.2020.9378073). URL : <https://ieeexplore.ieee.org/document/9378073/> (visité le 18/05/2021).
- [24] Emilien ARNAUD *et al.* “Predictive Models in Emergency Medicine and Their Missing Data Strategies : A Systematic Review”. In : *npj Digital Medicine* 6.1 (1 23 fév. 2023), p. 1-9. ISSN : 2398-6352. DOI : [10.1038/s41746-023-00770-6](https://doi.org/10.1038/s41746-023-00770-6). URL : <https://www.nature.com/articles/s41746-023-00770-6> (visité le 24/02/2023).
- [25] Émilien ARNAUD *et al.* “Development and Clinical Interpretation of an Explainable AI Model for Predicting Patient Pathways in the Emergency Department : A Retrospective Study”. In : *Journal of Medical Internet Research* UNDER REVIEW (21 sept. 2023).
- [26] Emilien ARNAUD *et al.* “Use of Artificial Intelligence to Manage Patient Flow in Emergency Department during the COVID-19 Pandemic : A Prospective, Single-Center Study”. In : *International Journal of Environmental Research and Public Health* 19.15 (5 août 2022), p. 9667. ISSN : 1660-4601. DOI : [10.3390/ijerph19159667](https://doi.org/10.3390/ijerph19159667).
- [27] E. ARNAUD *et al.* “L’intelligence artificielle dans les structures d’urgences : place de la formation et de la garantie humaine”. In : *Annales françaises de médecine d’urgence* 13.3 (3 1^{er} mai 2023), p. 169-178. ISSN : 2108-6524, 2108-6591. DOI : [10.3166/afmu-2022-0496](https://doi.org/10.3166/afmu-2022-0496). URL : https://afmu.revuesonline.com/articles/lvafmu/abs/2023/03/lvafmu_2023_sprurge001324/lvafmu_2023_sprurge001324.html (visité le 20/02/2024).
- [28] *Proposition de règlement du Parlement Européen et du Conseil établissant des règles harmonisées concernant l’intelligence artificielle (législation sur l’intelligence artificielle) et modifiant certains actes législatifs de l’Union*. Avec la coll. de Commission EUROPÉENNE. 21 avr. 2021. URL : <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:52021PC0206> (visité le 19/02/2023).
- [29] Crichton C. *Projet de règlement sur l’IA (II) : une approche fondée sur les risques - Intelligence artificielle*. Dalloz Actualité. 4 mai 2021. URL : https://www.dalloz-actualite.fr/flash/projet-de-reglement-sur-l-ia-ii-une-approche-fondée-sur-les-risques#.Y_HvYnrMK9I (visité le 19/02/2023).
- [30] DREES. *Enquête Urgences 2023 | Direction de La Recherche, Des Études, de l’évaluation et Des Statistiques*. Direction de la recherche, des études, de l’évaluation et des statistiques, 2023. URL : <https://drees.solidarites-sante.gouv.fr/sources-outils-et-enquetes/enquete-urgences-2023> (visité le 22/11/2023).
- [31] Manon FEDER. “Description des motifs de transfert du service d’accueil des urgences polyvalentes adultes vers la salle d’accueil des urgences vitales au sein des urgences du CHU d’Amiens”. 25 juin 2021. URL : <https://dumas.ccsd.cnrs.fr/dumas-03340755> (visité le 26/02/2023).
- [32] SAMU URGENCES DE FRANCE. *Les Ressources Médicales et Non Médicales Nécessaires Au Bon Fonctionnement Des Structures d’urgence*. Nov. 2011.
- [33] FEDORU. *Chiffres Clés Urgences*. 2022. URL : <https://fedoru.fr/chiffres-cles-urgences-2022> (visité le 22/11/2023).
- [34] Gérard BLEICHNER et François BRAUN. *ARCHITECTURE DES SERVICES D’URGENCE*. Recommandations. Société Francophone de Médecine d’Urgence, 16 juill. 2004. URL : https://www.sfm.org/upload/referentielsSFMU/recom_sfm_archi_service_urg.pdf.
- [35] D. PATERON. “Une organisation des flux au sein des urgences”. In : *Annales françaises de médecine d’urgence* 3.2 (mars 2013), p. 69-70. ISSN : 2108-6524, 2108-6591. DOI : [10.1007/s13341-013-0294-1](https://doi.org/10.1007/s13341-013-0294-1). URL : <http://link.springer.com/10.1007/s13341-013-0294-1> (visité le 04/06/2015).

- [36] Melanie ROUSSEL *et al.* “Overnight Stay in the Emergency Department and Mortality in Older Patients”. In : *JAMA Internal Medicine* 183.12 (1^{er} déc. 2023), p. 1378-1385. ISSN : 2168-6106. DOI : [10.1001/jamainternmed.2023.5961](https://doi.org/10.1001/jamainternmed.2023.5961). URL : <https://doi.org/10.1001/jamainternmed.2023.5961> (visité le 12/04/2024).
- [37] Rocío SÁNCHEZ-SALMERÓN *et al.* “Machine Learning Methods Applied to Triage in Emergency Services : A Systematic Review”. In : *International Emergency Nursing* 60 (1^{er} jan. 2022), p. 101109. ISSN : 1755-599X. DOI : [10.1016/j.ienj.2021.101109](https://www.sciencedirect.com/science/article/pii/S1755599X21001476). URL : <https://www.sciencedirect.com/science/article/pii/S1755599X21001476> (visité le 15/01/2022).
- [38] Nick M MURRAY *et al.* “Artificial Intelligence to Diagnose Ischemic Stroke and Identify Large Vessel Occlusions : A Systematic Review”. In : *Journal of NeuroInterventional Surgery* 12.2 (fév. 2020), p. 156-164. ISSN : 1759-8478, 1759-8486. DOI : [10.1136/neurintsurg-2019-015135](http://jn.is.bmj.com/lookup/doi/10.1136/neurintsurg-2019-015135). URL : <http://jn.is.bmj.com/lookup/doi/10.1136/neurintsurg-2019-015135> (visité le 15/01/2022).
- [39] Ki Hong KIM *et al.* “Emergency Department Routine Data and the Diagnosis of Acute Ischemic Heart Disease in Patients with Atypical Chest Pain”. In : *PLOS ONE* 15.11 (5 nov. 2020), e0241920. ISSN : 1932-6203. DOI : [10.1371/journal.pone.0241920](https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0241920). URL : <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0241920> (visité le 08/02/2022).
- [40] Le ROBERT. *Intelligence*. Définition. 2023. URL : <https://dictionnaire.lerobert.com/definition/intelligence> (visité le 09/01/2023).
- [41] Stuart J. RUSSELL *et al.* *Artificial Intelligence : A Modern Approach*. Fourth edition, global edition. Pearson Series in Artificial Intelligence. Harlow : Pearson, 2022. 1166 p. ISBN : 978-1-292-40113-3.
- [42] Jacques PITRAT. “Artificial Beings : The Conscience of a Conscious Machine”. In : *Artificial Beings : The Conscience of a Conscious Machine* (3 fév. 2010). ISSN : 9781848211018. DOI : [10.1002/9780470611791](https://roia.centre-mersenne.org/articles/10.5802/roia.19/).
- [43] Henri PRADE et Jean-Paul HATON. “L’œuvre scientifique de Jacques Pitrat (1934-2019) – Une perspective historique”. In : *Revue Ouverte d’Intelligence Artificielle* 3.1-2 (23 mars 2022), p. 69-80. DOI : [10.5802/roia.19](https://roia.centre-mersenne.org/articles/10.5802/roia.19/). URL : <https://roia.centre-mersenne.org/articles/10.5802/roia.19/> (visité le 06/03/2024).
- [44] Éditions LAROUSSE. *Algorithmes*. 2023. URL : <https://www.larousse.fr/dictionnaires/francais/algorithmes/2238> (visité le 09/01/2023).
- [45] Leo BREIMAN *et al.* *Classification and Regression Trees*. Taylor & Francis, 1^{er} jan. 1984. 372 p. ISBN : 978-0-412-04841-8.
- [46] Christopher M. BISHOP. *Neural Networks for Pattern Recognition*. Clarendon Press, 23 nov. 1995. 501 p. ISBN : 978-0-19-853864-6. Google Books : [TOS0BgAAQBAJ](https://books.google.fr/books?id=TOS0BgAAQBAJ).
- [47] Christine FROIDEVAUX, Marie-Claude GAUDEL et Michèle SORIA. *Types de données et algorithmes*. Ediscience internationale, 1993. 577 p. ISBN : 978-2-84074-023-0. Google Books : [r0FGAAACAAJ](https://books.google.fr/books?id=r0FGAAACAAJ).
- [48] Société Française de MÉDECINE D’URGENCE (SFMU). *Intervalle de Confiance / Société Française de Médecine d’Urgence - SFMU*. URL : <https://www.sfm.org/fr/vie-professionnelle/outils-professionnels/ebm/intervalledeconfiance> (visité le 29/12/2023).
- [49] Peter BUNEMAN *et al.* “Adding Structure to Unstructured Data”. In : *Database Theory — ICDT ’97*. Sous la dir. de Foto AFRATI et Phokion KOLAITIS. Lecture Notes in Computer Science. Berlin, Heidelberg : Springer, 1997, p. 336-350. ISBN : 978-3-540-49682-3. DOI : [10.1007/3-540-62222-5_55](https://doi.org/10.1007/3-540-62222-5_55).
- [50] *Natural Language Processing with Python [Book]*. URL : <https://www.oreilly.com/library/view/natural-language-processing/9780596803346/> (visité le 30/10/2022).
- [51] Jonathan J. WEBSTER et Chunyu KIT. “Tokenization as the Initial Phase in NLP”. In : *Proceedings of the 14th Conference on Computational Linguistics -*. The 14th Conference. T. 4. Nantes, France : Association for Computational Linguistics, 1992, p. 1106. DOI : [10.3115/992424.992434](https://portal.acm.org/citation.cfm?doid=992424.992434). URL : [http://portal.acm.org/citation.cfm?doid=992424.992434](https://portal.acm.org/citation.cfm?doid=992424.992434) (visité le 22/09/2023).

- [52] François CHOLLET *et al.* *Keras*. 2015. URL : <https://keras.io>.
- [53] Pau RODRÍGUEZ *et al.* “Beyond One-Hot Encoding : Lower Dimensional Target Embedding”. In : *Image and Vision Computing* 75 (1^{er} juill. 2018), p. 21-31. ISSN : 0262-8856. DOI : [10.1016/j.imavis.2018.04.004](https://doi.org/10.1016/j.imavis.2018.04.004). URL : <https://www.sciencedirect.com/science/article/pii/S0262885618300623> (visité le 22/09/2023).
- [54] Bin WANG *et al.* “Evaluating Word Embedding Models : Methods and Experimental Results”. In : *AP-SIPA Transactions on Signal and Information Processing* 8 (jan. 2019), e19. ISSN : 2048-7703. DOI : [10.1017/ATSIP.2019.12](https://doi.org/10.1017/ATSIP.2019.12). URL : <https://www.cambridge.org/core/journals/apsipa-transactions-on-signal-and-information-processing/article/evaluating-word-embedding-models-methods-and-experimental-results/EDF43F837150B94E71DBB36B28B85E79> (visité le 22/09/2023).
- [55] STUDENT. “The Probable Error of a Mean”. In : *Biometrika* 6.1 (1908), p. 1-25. ISSN : 0006-3444. DOI : [10.2307/2331554](https://doi.org/10.2307/2331554). JSTOR : [2331554](https://www.jstor.org/stable/2331554). URL : <https://www.jstor.org/stable/2331554> (visité le 22/09/2023).
- [56] H. B. MANN et D. R. WHITNEY. “On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other”. In : *The Annals of Mathematical Statistics* 18.1 (mars 1947), p. 50-60. ISSN : 0003-4851, 2168-8990. DOI : [10.1214/aoms/1177730491](https://doi.org/10.1214/aoms/1177730491). URL : <https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-18/issue-1/On-a-Test-of-Whether-one-of-Two-Random-Variables/10.1214/aoms/1177730491.full> (visité le 29/12/2023).
- [57] J. C. W. RAYNER et D. J. BEST. “Smooth Tests of Goodness of Fit : An Overview”. In : *International Statistical Review / Revue Internationale de Statistique* 58.1 (1990), p. 9-17. ISSN : 0306-7734. DOI : [10.2307/1403470](https://doi.org/10.2307/1403470). JSTOR : [1403470](https://www.jstor.org/stable/1403470). URL : <https://www.jstor.org/stable/1403470> (visité le 22/09/2023).
- [58] Graham J. G. UPTON. “Fisher’s Exact Test”. In : *Journal of the Royal Statistical Society : Series A (Statistics in Society)* 155.3 (1992), p. 395-402. ISSN : 1467-985X. DOI : [10.2307/2982890](https://doi.org/10.2307/2982890). URL : <https://onlinelibrary.wiley.com/doi/abs/10.2307/2982890> (visité le 30/12/2023).
- [59] Quinn MCNEMAR. “Note on the Sampling Error of the Difference between Correlated Proportions or Percentages”. In : *Psychometrika* 12.2 (1^{er} juin 1947), p. 153-157. ISSN : 1860-0980. DOI : [10.1007/BF02295996](https://doi.org/10.1007/BF02295996). URL : <https://doi.org/10.1007/BF02295996> (visité le 30/12/2023).
- [60] Donald B RUBIN. “Inference and Missing Data”. In : *Biometrika* 63.3 (1976), p. 581-592. DOI : [10.1093/biomet/63.3.581](https://doi.org/10.1093/biomet/63.3.581).
- [61] Stef van BUUREN. *Flexible Imputation of Missing Data, Second Edition*. 2^e éd. New York : Chapman and Hall/CRC, 12 juill. 2018. 444 p. ISBN : 978-0-429-49225-9. DOI : [10.1201/9780429492259](https://doi.org/10.1201/9780429492259).
- [62] Roderick J. A. LITTLE et Donald B. RUBIN. *Statistical Analysis with Missing Data*. John Wiley & Sons, 23 avr. 2019. 462 p. ISBN : 978-0-470-52679-8. Google Books : [BemMDwAAQBAJ](https://books.google.com/books?id=BemMDwAAQBAJ).
- [63] W. J. YOUDEN. “Index for Rating Diagnostic Tests”. In : *Cancer* 3.1 (1950), p. 32-35. ISSN : 1097-0142. DOI : [10.1002/1097-0142\(1950\)3:1<32::AID-CNCR2820030106>3.0.CO;2-3](https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3). URL : <https://onlinelibrary.wiley.com/doi/abs/10.1002/1097-0142%281950%293%3A1%3C32%3A%3AAID-CNCR2820030106%3E3.0.CO%3B2-3> (visité le 20/10/2023).
- [64] David M. W. POWERS. *Evaluation : From Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation*. 10 oct. 2020. DOI : [10.48550/arXiv.2010.16061](https://doi.org/10.48550/arXiv.2010.16061). arXiv : [2010.16061](https://arxiv.org/abs/2010.16061) [cs, stat]. URL : <http://arxiv.org/abs/2010.16061> (visité le 20/10/2023). preprint.
- [65] David G. KLEINBAUM et Mitchel KLEIN. *Logistic Regression*. Statistics for Biology and Health. New York, NY : Springer New York, 2010. ISBN : 978-1-4419-1741-6 978-1-4419-1742-3. DOI : [10.1007/978-1-4419-1742-3](https://doi.org/10.1007/978-1-4419-1742-3). URL : <http://link.springer.com/10.1007/978-1-4419-1742-3> (visité le 01/11/2022).

- [66] Muhammad FAISAL *et al.* “A Comparison of Logistic Regression Models with Alternative Machine Learning Methods to Predict the Risk of In-Hospital Mortality in Emergency Medical Admissions via External Validation”. In : *Health Informatics Journal* (29 nov. 2018), p. 146045821881360. ISSN : 1460-4582, 1741-2811. DOI : [10.1177/1460458218813600](https://doi.org/10.1177/1460458218813600). URL : <http://journals.sagepub.com/doi/10.1177/1460458218813600> (visité le 29/03/2019).
- [67] Yoshihiko RAITA *et al.* “Emergency Department Triage Prediction of Clinical Outcomes Using Machine Learning Models”. In : *Critical Care* 23.1 (22 fév. 2019), p. 64. ISSN : 1364-8535. DOI : [10.1186/s13054-019-2351-7](https://doi.org/10.1186/s13054-019-2351-7). URL : <https://doi.org/10.1186/s13054-019-2351-7> (visité le 08/02/2022).
- [68] WIKISTAT. *Réseaux de Neurones – Wikistats*. URL : <https://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-m-app-rn.pdf> (visité le 24/12/2022).
- [69] David STUTZ. *Illustrating (Convolutional) Neural Networks in LaTeX with TikZ • David Stutz*. David Stutz. 2 juin 2020. URL : <https://davidstutz.de/illustrating-convolutional-neural-networks-in-latex-with-tikz/> (visité le 23/12/2022).
- [70] Amin NAEMI *et al.* “Machine Learning Techniques for Mortality Prediction in Emergency Departments : A Systematic Review”. In : *BMJ open* 11.11 (2 nov. 2021), e052663. ISSN : 2044-6055. DOI : [10.1136/bmjopen-2021-052663](https://doi.org/10.1136/bmjopen-2021-052663). pmid : [34728454](https://pubmed.ncbi.nlm.nih.gov/34728454/).
- [71] Yann LECUN et Yoshua BENGIO. “Convolutional Networks for Images, Speech, and Time Series”. In : *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA, USA : MIT Press, 1^{er} oct. 1998, p. 255-258. ISBN : 978-0-262-51102-5.
- [72] Zewen LI *et al.* “A Survey of Convolutional Neural Networks : Analysis, Applications, and Prospects”. In : *IEEE Transactions on Neural Networks and Learning Systems* 33.12 (déc. 2022), p. 6999-7019. ISSN : 2162-237X, 2162-2388. DOI : [10.1109/TNNLS.2021.3084827](https://doi.org/10.1109/TNNLS.2021.3084827). URL : <https://ieeexplore.ieee.org/document/9451544/> (visité le 27/09/2023).
- [73] Leo BREIMAN. “Random Forests”. In : *Machine Learning* 45.1 (1^{er} oct. 2001), p. 5-32. ISSN : 1573-0565. DOI : [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324). URL : <https://doi.org/10.1023/A:1010933404324> (visité le 06/12/2022).
- [74] Tianqi CHEN et Carlos GUESTRIN. “XGBoost : A Scalable Tree Boosting System”. In : *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, California, USA). KDD '16. New York, NY, USA : ACM, 2016, p. 785-794. ISBN : 978-1-4503-4232-2. DOI : [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785). URL : <http://doi.acm.org/10.1145/2939672.2939785>.
- [75] Guolin KE *et al.* “LightGBM : A Highly Efficient Gradient Boosting Decision Tree”. In : *Advances in Neural Information Processing Systems*. Sous la dir. d’I. GUYON *et al.* T. 30. Curran Associates, Inc., 2017. URL : https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.
- [76] Woo Suk HONG, Adrian Daniel HAIMOVICH et R. Andrew TAYLOR. “Predicting Hospital Admission at Emergency Department Triage Using Machine Learning”. In : *PLOS ONE* 13.7 (20 juil. 2018), e0201016. ISSN : 1932-6203. DOI : [10.1371/journal.pone.0201016](https://doi.org/10.1371/journal.pone.0201016). URL : <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0201016> (visité le 12/02/2022).
- [77] Tadahiro GOTO *et al.* “Machine Learning–Based Prediction of Clinical Outcomes for Children During Emergency Department Triage”. In : *JAMA Network Open* 2.1 (11 jan. 2019), e186937. ISSN : 2574-3805. DOI : [10.1001/jamanetworkopen.2018.6937](https://doi.org/10.1001/jamanetworkopen.2018.6937). URL : <http://jamanetworkopen.jamanetwork.com/article.aspx?doi=10.1001/jamanetworkopen.2018.6937> (visité le 03/08/2019).
- [78] M. Emre CELEBI et Kemal AYDIN, éd. *Unsupervised Learning Algorithms*. Cham : Springer International Publishing, 2016. ISBN : 978-3-319-24209-5 978-3-319-24211-8. DOI : [10.1007/978-3-319-24211-8](https://doi.org/10.1007/978-3-319-24211-8). URL : <http://link.springer.com/10.1007/978-3-319-24211-8> (visité le 31/12/2023).

- [79] Libo ZHANG *et al.* “Sequential Three-Way Decision Based on Multi-Granular Autoencoder Features”. In : *Information Sciences* 507 (1^{er} jan. 2020), p. 630-643. ISSN : 0020-0255. DOI : [10.1016/j.ins.2019.03.061](https://doi.org/10.1016/j.ins.2019.03.061). URL : <https://www.sciencedirect.com/science/article/pii/S0020025519302750> (visité le 04/10/2023).
- [80] David CHARTE *et al.* “An Analysis on the Use of Autoencoders for Representation Learning : Fundamentals, Learning Task Case Studies, Explainability and Challenges”. In : *Neurocomputing* 404 (3 sept. 2020), p. 93-107. ISSN : 0925-2312. DOI : [10.1016/j.neucom.2020.04.057](https://doi.org/10.1016/j.neucom.2020.04.057). URL : <https://www.sciencedirect.com/science/article/pii/S092523122030624X> (visité le 04/10/2023).
- [81] JANOSH RIEBESELL. *Janosh/Tikz : V0.1.0*. Version v0.1.0. Zenodo, 27 déc. 2022. DOI : [10.5281/ZENODO.7486911](https://doi.org/10.5281/ZENODO.7486911). URL : <https://zenodo.org/record/7486911> (visité le 28/09/2023).
- [82] V7 LABS. *Autoencoders in Deep Learning : Tutorial & Use Cases [2023]*. 4 juin 2021. URL : <https://www.v7labs.com/blog/autoencoders-guide,%20https://www.v7labs.com/blog/autoencoders-guide> (visité le 08/01/2024).
- [83] Ashish VASWANI *et al.* “Attention Is All You Need”. In : *Advances in Neural Information Processing Systems*. T. 30. Curran Associates, Inc., 2017. URL : https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html (visité le 23/11/2023).
- [84] Tom KELDENICH. *Le Mécanisme de l'Attention en Deep Learning - Comprendre Rapidement*. Inside Machine Learning, 19 oct. 2021. URL : <https://inside-machinelearning.com/mecanisme-attention/> (visité le 23/11/2023).
- [85] Tianyang LIN *et al.* “A Survey of Transformers”. In : *AI Open* 3 (1^{er} jan. 2022), p. 111-132. ISSN : 2666-6510. DOI : [10.1016/j.aiopen.2022.10.001](https://doi.org/10.1016/j.aiopen.2022.10.001). URL : <https://www.sciencedirect.com/science/article/pii/S2666651022000146> (visité le 05/10/2023).
- [86] Qi WANG *et al.* “A Comprehensive Survey of Loss Functions in Machine Learning”. In : *Annals of Data Science* 9.2 (avr. 2022), p. 187-212. ISSN : 2198-5804, 2198-5812. DOI : [10.1007/s40745-020-00253-5](https://doi.org/10.1007/s40745-020-00253-5). URL : <https://link.springer.com/10.1007/s40745-020-00253-5> (visité le 23/12/2022).
- [87] KOHONEN, BARNA et CHRISLEY. “Statistical Pattern Recognition with Neural Networks : Benchmarking Studies”. In : *IEEE International Conference on Neural Networks*. Proceedings of 1993 IEEE International Conference on Neural Networks (ICNN '93). San Diego, CA, USA : IEEE, 1988, 61-68 vol.1. ISBN : 978-0-7803-0999-9. DOI : [10.1109/ICNN.1988.23829](https://doi.org/10.1109/ICNN.1988.23829). URL : <http://ieeexplore.ieee.org/document/23829/> (visité le 23/12/2022).
- [88] Peter J. HUBER. “Robust Estimation of a Location Parameter”. In : *The Annals of Mathematical Statistics* 35.1 (1964), p. 73-101. DOI : [10.1214/aoms/1177703732](https://doi.org/10.1214/aoms/1177703732). URL : <https://doi.org/10.1214/aoms/1177703732>.
- [89] Douglas M. HAWKINS. “The Problem of Overfitting”. In : *Journal of Chemical Information and Computer Sciences* 44.1 (1^{er} jan. 2004), p. 1-12. ISSN : 0095-2338. DOI : [10.1021/ci0342472](https://doi.org/10.1021/ci0342472).
- [90] S.E BLEEKER *et al.* “External Validation Is Necessary in Prediction Research :” in : *Journal of Clinical Epidemiology* 56.9 (sept. 2003), p. 826-832. ISSN : 08954356. DOI : [10.1016/S0895-4356\(03\)00207-5](https://doi.org/10.1016/S0895-4356(03)00207-5). URL : <https://linkinghub.elsevier.com/retrieve/pii/S0895435603002075> (visité le 06/08/2019).
- [91] Muhammad FAISAL *et al.* “A Comparison of Logistic Regression Models with Alternative Machine Learning Methods to Predict the Risk of In-Hospital Mortality in Emergency Medical Admissions via External Validation”. In : *Health Informatics Journal* 26.1 (mars 2020), p. 34-44. ISSN : 1741-2811. DOI : [10.1177/1460458218813600](https://doi.org/10.1177/1460458218813600). pmid : 30488755.
- [92] Miranda MOURBY, Katharina Ó CATHAOIR et Catherine Bjerre COLLIN. “Transparency of Machine-Learning in Healthcare : The GDPR & European Health Law”. In : *Computer Law & Security Review* 43 (1^{er} nov. 2021), p. 105611. ISSN : 0267-3649. DOI : [10.1016/j.clsr.2021.105611](https://doi.org/10.1016/j.clsr.2021.105611). URL : <https://www.sciencedirect.com/science/article/pii/S0267364921000844> (visité le 30/05/2022).

- [93] Scott M. LUNDBERG *et al.* “Explainable Machine-Learning Predictions for the Prevention of Hypoxaemia during Surgery”. In : *Nature Biomedical Engineering* 2.10 (10 oct. 2018), p. 749-760. ISSN : 2157-846X. DOI : [10.1038/s41551-018-0304-0](https://doi.org/10.1038/s41551-018-0304-0). URL : <https://www.nature.com/articles/s41551-018-0304-0> (visité le 30/05/2022).
- [94] Jerome H. FRIEDMAN. “Greedy Function Approximation : A Gradient Boosting Machine.” In : *The Annals of Statistics* 29.5 (oct. 2001), p. 1189-1232. ISSN : 0090-5364, 2168-8966. DOI : [10.1214/aos/1013203451](https://doi.org/10.1214/aos/1013203451). URL : <https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-5/Greedy-function-approximation-A-gradient-boosting-machine/10.1214/aos/1013203451.full> (visité le 30/05/2022).
- [95] Marta FERNANDES *et al.* “Clinical Decision Support Systems for Triage in the Emergency Department Using Intelligent Systems : A Review”. In : *Artificial Intelligence in Medicine* 102 (jan. 2020), p. 101762. ISSN : 09333657. DOI : [10.1016/j.artmed.2019.101762](https://doi.org/10.1016/j.artmed.2019.101762). URL : <https://linkinghub.elsevier.com/retrieve/pii/S0933365719301265> (visité le 15/01/2022).
- [96] Clare Allison PARKER *et al.* “Predicting Hospital Admission at the Emergency Department Triage : A Novel Prediction Model”. In : *The American Journal of Emergency Medicine* 37.8 (août 2019), p. 1498-1504. ISSN : 07356757. DOI : [10.1016/j.ajem.2018.10.060](https://doi.org/10.1016/j.ajem.2018.10.060). URL : <https://linkinghub.elsevier.com/retrieve/pii/S073567571830891X> (visité le 03/08/2019).
- [97] Fatemeh RAHIMIAN *et al.* “Predicting the Risk of Emergency Admission with Machine Learning : Development and Validation Using Linked Electronic Health Records”. In : *PLOS Medicine* 15.11 (20 nov. 2018). Sous la dir. d’Aziz SHEIKH, e1002695. ISSN : 1549-1676. DOI : [10.1371/journal.pmed.1002695](https://doi.org/10.1371/journal.pmed.1002695). URL : <http://dx.plos.org/10.1371/journal.pmed.1002695> (visité le 03/08/2019).
- [98] Yuval BARAK-CORREN, Shlomo Hanan ISRAELIT et Ben Y REIS. “Progressive Prediction of Hospitalisation in the Emergency Department : Uncovering Hidden Patterns to Improve Patient Flow”. In : *Emergency Medicine Journal* 34.5 (mai 2017), p. 308-314. ISSN : 1472-0205, 1472-0213. DOI : [10.1136/emermed-2014-203819](https://doi.org/10.1136/emermed-2014-203819). URL : <http://emj.bmj.com/lookup/doi/10.1136/emermed-2014-203819> (visité le 03/08/2019).
- [99] Nicholas W. STERLING *et al.* “Prediction of Emergency Department Patient Disposition Based on Natural Language Processing of Triage Notes”. In : *International Journal of Medical Informatics* 129 (sept. 2019), p. 184-188. ISSN : 1872-8243. DOI : [10.1016/j.ijmedinf.2019.06.008](https://doi.org/10.1016/j.ijmedinf.2019.06.008). pmid : 31445253.
- [100] Xingyu ZHANG *et al.* “Prediction of Emergency Department Hospital Admission Based on Natural Language Processing and Neural Networks”. In : *Methods of Information in Medicine* 56.5 (2017), p. 377-389. ISSN : 0026-1270, 2511-705X. DOI : [10.3414/ME17-01-0024](https://doi.org/10.3414/ME17-01-0024). URL : <http://www.thieme-connect.de/DOI/DOI?10.3414/ME17-01-0024> (visité le 21/02/2022).
- [101] Chien-Hua CHEN *et al.* “Emergency Department Disposition Prediction Using a Deep Neural Network with Integrated Clinical Narratives and Structured Data”. In : *International Journal of Medical Informatics* 139 (juill. 2020), p. 104146. ISSN : 13865056. DOI : [10.1016/j.ijmedinf.2020.104146](https://doi.org/10.1016/j.ijmedinf.2020.104146). URL : <https://linkinghub.elsevier.com/retrieve/pii/S1386505620300538> (visité le 05/02/2022).
- [102] David CHANG, Woo Suk HONG et Richard Andrew TAYLOR. “Generating Contextual Embeddings for Emergency Department Chief Complaints”. In : *JAMIA Open* 3.2 (15 juill. 2020), p. 160-166. ISSN : 2574-2531. DOI : [10.1093/jamiaopen/ooaa022](https://doi.org/10.1093/jamiaopen/ooaa022). pmid : 32734154. URL : <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7382638/> (visité le 21/03/2023).
- [103] République FRANÇAISE. *Code Civil - Article 1240*. 2016. URL : https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000032041571 (visité le 28/03/2023).
- [104] République FRANÇAISE. *Code Civil - Article 1242*. 2016. URL : https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000032041559/ (visité le 20/01/2023).

-
- [105] Henri CAPITANT *et al.* *Les grands arrêts de la jurisprudence civile T2*. Dalloz. T. 2. Grands arrêts. Paris : Dalloz, sept. 2015. ISBN : 978-2-247-15358-9.
- [106] Fabian PEDREGOSA. “Scikit-Learn : Machine Learning in Python”. In : *MACHINE LEARNING IN PYTHON* (2011), p. 6.
- [107] The pandas development TEAM. *Pandas-Dev/Pandas : Pandas*. Version latest. Zenodo, fév. 2020. DOI : [10.5281/zenodo.3509134](https://doi.org/10.5281/zenodo.3509134). URL : <https://doi.org/10.5281/zenodo.3509134>.
- [108] Martin ABADI *et al.* “TensorFlow : A System for Large-Scale Machine Learning”. In : 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16). Savannah, GA, USA, 2016. ISBN : 978-1-931971-33-1.
- [109] Shangzhi HONG et Henry S. LYNN. “Accuracy of Random-Forest-Based Imputation of Missing Data in the Presence of Non-Normality, Non-Linearity, and Interaction”. In : *BMC Medical Research Methodology* 20.1 (25 juill. 2020), p. 199. ISSN : 1471-2288. DOI : [10.1186/s12874-020-01080-1](https://doi.org/10.1186/s12874-020-01080-1). URL : <https://doi.org/10.1186/s12874-020-01080-1> (visité le 24/02/2022).
- [110] Wei JIANG, Julie JOSSE et Marc LAVIELLE. “Logistic Regression with Missing Covariates—Parameter Estimation, Model Selection and Prediction within a Joint-Modeling Framework”. In : *Computational Statistics & Data Analysis* 145 (mai 2020), p. 106907. ISSN : 01679473. DOI : [10.1016/j.csda.2019.106907](https://doi.org/10.1016/j.csda.2019.106907). URL : <https://linkinghub.elsevier.com/retrieve/pii/S0167947319302622> (visité le 05/08/2022).
- [111] Amin NAEMI *et al.* “Quantifying the Impact of Addressing Data Challenges in Prediction of Length of Stay”. In : *BMC Medical Informatics and Decision Making* 21.1 (30 oct. 2021), p. 298. ISSN : 1472-6947. DOI : [10.1186/s12911-021-01660-1](https://doi.org/10.1186/s12911-021-01660-1). URL : <https://doi.org/10.1186/s12911-021-01660-1> (visité le 02/02/2022).
- [112] Roderick J. LITTLE, Donald B. RUBIN et Sahar Z. ZANGENEH. “Conditions for Ignoring the Missing-Data Mechanism in Likelihood Inferences for Parameter Subsets”. In : *Journal of the American Statistical Association* 112.517 (2 jan. 2017), p. 314-320. ISSN : 0162-1459, 1537-274X. DOI : [10.1080/01621459.2015.1136826](https://doi.org/10.1080/01621459.2015.1136826). URL : <https://www.tandfonline.com/doi/full/10.1080/01621459.2015.1136826> (visité le 12/10/2022).
- [113] Gary S. COLLINS *et al.* “Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) : The TRIPOD Statement”. In : *BMC Medicine* 13.1 (6 jan. 2015), p. 1. ISSN : 1741-7015. DOI : [10.1186/s12916-014-0241-z](https://doi.org/10.1186/s12916-014-0241-z). URL : <https://doi.org/10.1186/s12916-014-0241-z> (visité le 21/10/2023).
- [114] Karel G. M. MOONS *et al.* “Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies : The CHARMS Checklist”. In : *PLoS Medicine* 11.10 (14 oct. 2014), e1001744. ISSN : 1549-1676. DOI : [10.1371/journal.pmed.1001744](https://doi.org/10.1371/journal.pmed.1001744). URL : <https://dx.plos.org/10.1371/journal.pmed.1001744> (visité le 18/01/2022).
- [115] Matthew J PAGE *et al.* “The PRISMA 2020 Statement : An Updated Guideline for Reporting Systematic Reviews”. In : *BMJ* (29 mars 2021), n71. ISSN : 1756-1833. DOI : [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71). URL : <https://www.bmj.com/lookup/doi/10.1136/bmj.n71> (visité le 24/01/2022).
- [116] Robert F. WOLFF *et al.* “PROBAST : A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies”. In : *Annals of Internal Medicine* 170.1 (1^{er} jan. 2019), p. 51. ISSN : 0003-4819. DOI : [10.7326/M18-1376](https://doi.org/10.7326/M18-1376). URL : <http://annals.org/article.aspx?doi=10.7326/M18-1376> (visité le 18/01/2022).

- [117] Tianjing LI *et al.* “Standards Should Be Applied in the Prevention and Handling of Missing Data for Patient-Centered Outcomes Research : A Systematic Review and Expert Consensus”. In : *Journal of Clinical Epidemiology* 67.1 (jan. 2014), p. 15-32. ISSN : 08954356. DOI : [10.1016/j.jclinepi.2013.08.013](https://doi.org/10.1016/j.jclinepi.2013.08.013). URL : <https://linkinghub.elsevier.com/retrieve/pii/S089543561300351X> (visité le 01/03/2022).
- [118] Sujita W NARAYAN *et al.* “Missing Data Reporting in Clinical Pharmacy Research”. In : *American Journal of Health-System Pharmacy* 76.24 (2 déc. 2019), p. 2048-2052. ISSN : 1079-2082, 1535-2900. DOI : [10.1093/ajhp/zxz245](https://doi.org/10.1093/ajhp/zxz245). URL : <https://academic.oup.com/ajhp/article/76/24/2048/5649007> (visité le 01/03/2022).
- [119] Diego A. MARTINEZ *et al.* “Early Prediction of Acute Kidney Injury in the Emergency Department With Machine-Learning Methods Applied to Electronic Health Record Data”. In : *Annals of Emergency Medicine* 76.4 (oct. 2020), p. 501-514. ISSN : 01960644. DOI : [10.1016/j.annemergmed.2020.05.026](https://doi.org/10.1016/j.annemergmed.2020.05.026). URL : <https://linkinghub.elsevier.com/retrieve/pii/S0196064420303978> (visité le 05/02/2022).
- [120] Olawale F. AYILARA *et al.* “Impact of Missing Data on Bias and Precision When Estimating Change in Patient-Reported Outcomes from a Clinical Registry”. In : *Health and Quality of Life Outcomes* 17.1 (déc. 2019), p. 106. ISSN : 1477-7525. DOI : [10.1186/s12955-019-1181-2](https://doi.org/10.1186/s12955-019-1181-2). URL : <https://hqlo.biomedcentral.com/articles/10.1186/s12955-019-1181-2> (visité le 23/03/2022).
- [121] Julie JOSSE *et al.* *On the Consistency of Supervised Learning with Missing Values*. 3 juill. 2020. arXiv : [1902.06931](https://arxiv.org/abs/1902.06931) [cs, math, stat]. URL : <http://arxiv.org/abs/1902.06931> (visité le 05/08/2022). preprint.
- [122] Ian R. WHITE, Patrick ROYSTON et Angela M. WOOD. “Multiple Imputation Using Chained Equations : Issues and Guidance for Practice”. In : *Statistics in Medicine* 30.4 (20 fév. 2011), p. 377-399. ISSN : 02776715. DOI : [10.1002/sim.4067](https://doi.org/10.1002/sim.4067). URL : <http://doi.wiley.com/10.1002/sim.4067> (visité le 29/04/2021).
- [123] Peter D FARIS *et al.* “Multiple Imputation versus Data Enhancement for Dealing with Missing Data in Observational Health Care Outcome Analyses”. In : *Journal of Clinical Epidemiology* 55.2 (fév. 2002), p. 184-191. ISSN : 08954356. DOI : [10.1016/S0895-4356\(01\)00433-4](https://doi.org/10.1016/S0895-4356(01)00433-4). URL : <https://linkinghub.elsevier.com/retrieve/pii/S0895435601004334> (visité le 16/12/2021).
- [124] E. KUHN et M. LAVIELLE. “Maximum Likelihood Estimation in Nonlinear Mixed Effects Models”. In : *Computational Statistics & Data Analysis* 49.4 (juin 2005), p. 1020-1038. ISSN : 01679473. DOI : [10.1016/j.csda.2004.07.002](https://doi.org/10.1016/j.csda.2004.07.002). URL : <https://linkinghub.elsevier.com/retrieve/pii/S0167947304002221> (visité le 17/10/2022).
- [125] Alexandre PEREZ-LEBEL *et al.* “Benchmarking Missing-Values Approaches for Predictive Models on Health Databases”. In : *GigaScience* 11 (15 avr. 2022), giac013. ISSN : 2047-217X. DOI : [10.1093/gigascience/giac013](https://doi.org/10.1093/gigascience/giac013). URL : <https://academic.oup.com/gigascience/article/doi/10.1093/gigascience/giac013/6568998> (visité le 05/08/2022).
- [126] Peter C. AUSTIN *et al.* “Missing Data in Clinical Research : A Tutorial on Multiple Imputation”. In : *Canadian Journal of Cardiology* 37.9 (sept. 2021), p. 1322-1331. ISSN : 0828282X. DOI : [10.1016/j.cjca.2020.11.010](https://doi.org/10.1016/j.cjca.2020.11.010). URL : <https://linkinghub.elsevier.com/retrieve/pii/S0828282X20311119> (visité le 12/10/2022).
- [127] Mingxing GONG. “A Novel Performance Measure for Machine Learning Classification”. In : *International Journal of Managing Information Technology* 13.1 (28 fév. 2021), p. 11-19. ISSN : 09755926. DOI : [10.5121/ijmit.2021.13101](https://doi.org/10.5121/ijmit.2021.13101). URL : <https://aircconline.com/ijmit/V13N1/13121ijmit01.pdf> (visité le 23/03/2022).
- [128] D. P. KINGMA et L. J. BA. “Adam : A Method for Stochastic Optimization”. In : (2015). URL : <https://dare.uva.nl/search?identifier=a20791d3-1aff-464a-8544-268383c33a75> (visité le 29/10/2022).

- [129] R. H. BYRD, L. PEIHUANG et J. NOCEDAL. *A Limited-Memory Algorithm for Bound-Constrained Optimization*. Other Information : PBD : [1996]. 3 jan. 1996. DOI : [10.2172/204262](https://doi.org/10.2172/204262). URL : <https://digital.library.unt.edu/ark:/67531/metadc666315/> (visité le 01/11/2022).
- [130] *Shap/Shap : A Game Theoretic Approach to Explain the Output of Any Machine Learning Model*. URL : <https://github.com/shap/shap> (visité le 20/10/2023).
- [131] Li JIANGCHUN. *PDPbox*. 15 oct. 2023. URL : <https://github.com/SauceCat/PDPbox> (visité le 20/10/2023).
- [132] Marie-Josée SIROIS *et al.* “Measuring Frailty Can Help Emergency Departments Identify Independent Seniors at Risk of Functional Decline After Minor Injuries”. In : *The Journals of Gerontology Series A : Biological Sciences and Medical Sciences* 72.1 (jan. 2017), p. 68-74. ISSN : 1079-5006, 1758-535X. DOI : [10.1093/gerona/glv152](https://doi.org/10.1093/gerona/glv152). URL : <https://academic.oup.com/biomedgerontology/article-lookup/doi/10.1093/gerona/glv152> (visité le 05/07/2022).
- [133] Lucas L GEYER *et al.* “Incidence of Delayed and Missed Diagnoses in Whole-Body Multidetector CT in Patients with Multiple Injuries after Trauma”. In : *Acta Radiologica* 54.5 (1^{er} juin 2013), p. 592-598. ISSN : 0284-1851. DOI : [10.1177/0284185113475443](https://doi.org/10.1177/0284185113475443). URL : <https://doi.org/10.1177/0284185113475443> (visité le 05/07/2022).
- [134] Sripal BANGALORE *et al.* “The Association of Admission Heart Rate and In-Hospital Cardiovascular Events in Patients with Non-ST-segment Elevation Acute Coronary Syndromes : Results from 135 164 Patients in the CRUSADE Quality Improvement Initiative”. In : *European Heart Journal* 31.5 (1^{er} mars 2010), p. 552-560. ISSN : 0195-668X. DOI : [10.1093/eurheartj/ehp397](https://doi.org/10.1093/eurheartj/ehp397). URL : <https://doi.org/10.1093/eurheartj/ehp397> (visité le 05/07/2022).
- [135] Douglas BARNABY *et al.* “Heart Rate Variability in Emergency Department Patients with Sepsis”. In : *Academic Emergency Medicine* 9.7 (2002), p. 661-670. ISSN : 1553-2712. DOI : [10.1197/aemj.9.7.661](https://doi.org/10.1197/aemj.9.7.661). URL : <https://onlinelibrary.wiley.com/doi/abs/10.1197/aemj.9.7.661> (visité le 05/07/2022).
- [136] Alexander OLAUSSEN *et al.* “Review Article : Shock Index for Prediction of Critical Bleeding Post-Trauma : A Systematic Review”. In : *Emergency Medicine Australasia* 26.3 (2014), p. 223-228. ISSN : 1742-6723. DOI : [10.1111/1742-6723.12232](https://doi.org/10.1111/1742-6723.12232). URL : <https://onlinelibrary.wiley.com/doi/abs/10.1111/1742-6723.12232> (visité le 05/07/2022).
- [137] Abbas E. KITABCHI et Barry M. WALL. “Management of Diabetic Ketoacidosis”. In : *American Family Physician* 60.2 (1^{er} août 1999), p. 455-464. URL : <https://www.aafp.org/pubs/afp/issues/1999/0801/p455.html> (visité le 10/10/2022).
- [138] Byron GRAHAM *et al.* “Using Data Mining to Predict Hospital Admissions From the Emergency Department”. In : *IEEE Access* 6 (2018), p. 10458-10469. ISSN : 2169-3536. DOI : [10.1109/ACCESS.2018.2808843](https://doi.org/10.1109/ACCESS.2018.2808843). URL : <http://ieeexplore.ieee.org/document/8300528/> (visité le 25/05/2022).
- [139] Eyal KLANG *et al.* “Predicting Adult Neuroscience Intensive Care Unit Admission from Emergency Department Triage Using a Retrospective, Tabular-Free Text Machine Learning Approach”. In : *Scientific Reports* 11.1 (déc. 2021), p. 1381. ISSN : 2045-2322. DOI : [10.1038/s41598-021-80985-3](https://doi.org/10.1038/s41598-021-80985-3). URL : <http://www.nature.com/articles/s41598-021-80985-3> (visité le 25/01/2022).
- [140] Chiara LONGONI, Andrea BONEZZI et Carey K MOREWEDGE. “Resistance to Medical Artificial Intelligence”. In : *Journal of Consumer Research* 46.4 (1^{er} déc. 2019), p. 629-650. ISSN : 0093-5301, 1537-5277. DOI : [10.1093/jcr/ucz013](https://doi.org/10.1093/jcr/ucz013). URL : <https://academic.oup.com/jcr/article/46/4/629/5485292> (visité le 05/07/2022).
- [141] Ting ZHU *et al.* “Individualized Prediction of Psychiatric Readmissions for Patients with Major Depressive Disorder : A 10-Year Retrospective Cohort Study”. In : *Translational Psychiatry* 12.1 (1 23 avr. 2022), p. 1-18. ISSN : 2158-3188. DOI : [10.1038/s41398-022-01937-7](https://doi.org/10.1038/s41398-022-01937-7). URL : <https://www.nature.com/articles/s41398-022-01937-7> (visité le 30/05/2022).

- [142] Amril NAZIR et Hyacinth Kwadwo AMPADU. “Interpretable Deep Learning for the Prediction of ICU Admission Likelihood and Mortality of COVID-19 Patients”. In : *PeerJ Computer Science* 8 (17 mars 2022), e889. ISSN : 2376-5992. DOI : [10.7717/peerj-cs.889](https://doi.org/10.7717/peerj-cs.889). URL : <https://peerj.com/articles/cs-889> (visité le 30/05/2022).
- [143] Hossein ESTIRI, Zachary H. STRASSER et Shawn N. MURPHY. “Individualized Prediction of COVID-19 Adverse Outcomes with MLHO”. In : *Scientific Reports* 11.1 (15 mars 2021), p. 5322. ISSN : 2045-2322. DOI : [10.1038/s41598-021-84781-x](https://doi.org/10.1038/s41598-021-84781-x). URL : <https://www.nature.com/articles/s41598-021-84781-x> (visité le 30/05/2022).
- [144] Yun-Chung LIU *et al.* “Evaluation of the Need for Intensive Care in Children With Pneumonia : Machine Learning Approach”. In : *JMIR Medical Informatics* 10.1 (27 jan. 2022), e28934. DOI : [10.2196/28934](https://doi.org/10.2196/28934). URL : <https://medinform.jmir.org/2022/1/e28934> (visité le 30/05/2022).
- [145] Selcuk KUCUKSEYMEN *et al.* “Noncontrast Cardiac Magnetic Resonance Imaging Predictors of Heart Failure Hospitalization in Heart Failure With Preserved Ejection Fraction”. In : *Journal of Magnetic Resonance Imaging* 55.6 (2022), p. 1812-1825. ISSN : 1522-2586. DOI : [10.1002/jmri.27932](https://doi.org/10.1002/jmri.27932). URL : <https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.27932> (visité le 30/05/2022).
- [146] Michael O KILLIAN *et al.* “Machine Learning-Based Prediction of Health Outcomes in Pediatric Organ Transplantation Recipients”. In : *JAMIA Open* 4.1 (1^{er} jan. 2021), ooab008. ISSN : 2574-2531. DOI : [10.1093/jamiaopen/ooab008](https://doi.org/10.1093/jamiaopen/ooab008). URL : <https://doi.org/10.1093/jamiaopen/ooab008> (visité le 30/05/2022).
- [147] Michael L. MARTINI *et al.* “Drivers of Prolonged Hospitalization Following Spine Surgery : A Game-Theory-Based Approach to Explaining Machine Learning Models”. In : *JBJS* 103.1 (6 jan. 2021), p. 64-73. ISSN : 0021-9355. DOI : [10.2106/JBJS.20.00875](https://doi.org/10.2106/JBJS.20.00875). URL : https://journals.lww.com/jbjsjournal/Abstract/2021/01060/Drivers_of_Prolonged_Hospitalization_Following.10.aspx (visité le 30/05/2022).
- [148] Farah DESHMUKH et Shamel S. MERCHANT. “Explainable Machine Learning Model for Predicting GI Bleed Mortality in the Intensive Care Unit”. In : *Official journal of the American College of Gastroenterology / ACG* 115.10 (oct. 2020), p. 1657-1668. ISSN : 0002-9270. DOI : [10.14309/ajg.0000000000000632](https://doi.org/10.14309/ajg.0000000000000632). URL : https://journals.lww.com/ajg/Abstract/2020/10000/Explainable_Machine_Learning_Model_for_Predicting.22.aspx (visité le 30/05/2022).
- [149] Donghee SHIN. “The Effects of Explainability and Causability on Perception, Trust, and Acceptance : Implications for Explainable AI”. In : *International Journal of Human-Computer Studies* 146 (1^{er} fév. 2021), p. 102551. ISSN : 1071-5819. DOI : [10.1016/j.ijhcs.2020.102551](https://doi.org/10.1016/j.ijhcs.2020.102551). URL : <https://www.sciencedirect.com/science/article/pii/S1071581920301531> (visité le 05/07/2022).
- [150] *Données de laboratoires pour le dépistage (A COMPTEUR DU 18/05/2022) - SI-DEP - data.gouv.fr*. URL : <https://www.data.gouv.fr/fr/datasets/donnees-de-laboratoires-pour-le-depistage-a-compter-du-18-05-2022-si-dep/> (visité le 16/06/2022).
- [151] Marta FERNANDES *et al.* “Predicting Intensive Care Unit Admission among Patients Presenting to the Emergency Department Using Machine Learning and Natural Language Processing”. In : *PLOS ONE* 15.3 (3 mars 2020), e0229331. ISSN : 1932-6203. DOI : [10.1371/journal.pone.0229331](https://doi.org/10.1371/journal.pone.0229331). URL : <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0229331> (visité le 04/02/2022).
- [152] Tian-Hoe TAN *et al.* “Predicting Outcomes in Older ED Patients with Influenza in Real Time Using a Big Data-Driven and Machine Learning Approach to the Hospital Information System”. In : *BMC geriatrics* 21.1 (27 avr. 2021), p. 280. ISSN : 1471-2318. DOI : [10.1186/s12877-021-02229-3](https://doi.org/10.1186/s12877-021-02229-3). pmid : [33902485](https://pubmed.ncbi.nlm.nih.gov/33902485/).
- [153] Benjamin TERRASI *et al.* “French ICUs Fight Back : An Example of Regional ICU Organisation to Tackle the SARS-CoV-2 Outbreak”. In : *Anaesthesia, Critical Care & Pain Medicine* (30 avr. 2020). ISSN : 2352-5568. DOI : [10.1016/j.accpm.2020.03.018](https://doi.org/10.1016/j.accpm.2020.03.018). pmid : [32360980](https://pubmed.ncbi.nlm.nih.gov/32360980/).

- [154] Shu Hui LEOW *et al.* “The Attend Study : A Retrospective Observational Study of Emergency Department Attendances During the Early Stages of the COVID-19 Pandemic”. In : *Cureus* (21 juill. 2020). ISSN : 2168-8184. DOI : [10.7759/cureus.9328](https://doi.org/10.7759/cureus.9328). URL : <https://www.cureus.com/articles/34364-the-attend-study-a-retrospective-observational-study-of-emergency-department-attendances-during-the-early-stages-of-the-covid-19-pandemic> (visité le 16/05/2022).
- [155] Brunessen BERTRAND. “Chronique Droit européen du numérique – La proposition de régulation générale pour l’intelligence artificielle dans l’Union européenne : l’IA Act”. In : *RTDeur. Revue trimestrielle de droit européen* 03 (24 oct. 2022), p. 473. URL : <https://shs.hal.science/halshs-03798713> (visité le 19/02/2023).
- [156] *Proposition de directive 496 relative à l’adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l’intelligence artificielle (Directive sur la responsabilité en matière d’IA)*. Avec la coll. de Parlement EUROPÉEN et Conseil EUROPÉEN. 2022. URL : <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:52022PC0496> (visité le 28/03/2023).
- [157] Margo BERNELIN. “Intelligence artificielle : une proposition de directive sur la responsabilité civile extracontractuelle”. In : *Dalloz actualités* (22 nov. 2022). URL : <https://www.dalloz-actualite.fr/flash/intelligence-artificielle-une-proposition-de-directive-sur-responsabilite-civile-extracontract#ZCMpg3ZByHs> (visité le 28/03/2023).
- [158] Antoine HUDON *et al.* “Explainable Artificial Intelligence (XAI) : How the Visualization of AI Predictions Affects User Cognitive Load and Confidence”. In : *Information Systems and Neuroscience*. Sous la dir. de Fred D. DAVIS *et al.* Lecture Notes in Information Systems and Organisation. 2021, p. 237-246. DOI : [10.1007/978-3-030-88900-5_27](https://doi.org/10.1007/978-3-030-88900-5_27).
- [159] Amy FRANKLIN *et al.* “Opportunistic Decision Making and Complexity in Emergency Care”. In : *Journal of Biomedical Informatics*. Biomedical Complexity and Error 44.3 (1^{er} juin 2011), p. 469-476. ISSN : 1532-0464. DOI : [10.1016/j.jbi.2011.04.001](https://doi.org/10.1016/j.jbi.2011.04.001).
- [160] T. PELACCIA *et al.* “Comment les médecins urgentistes raisonnent-ils ? Synthèse des principaux résultats d’une recherche qualitative multicentrique et multidisciplinaire sur la prise de décision en médecine d’urgence”. In : *Annales françaises de médecine d’urgence* 7.3 (3 1^{er} juin 2017), p. 153-158. ISSN : 2108-6524, 2108-6591. DOI : [10.1007/s13341-017-0729-1](https://doi.org/10.1007/s13341-017-0729-1).
- [161] T. PELACCIA, G. FORESTIER et C. WEMMERT. “Une intelligence artificielle raisonne-t-elle de la même façon que les cliniciens pour poser des diagnostics ?” In : *Rev. Med. Interne* 41.3 (1^{er} mars 2020), p. 192-195. ISSN : 0248-8663. DOI : [10.1016/j.revmed.2019.12.014](https://doi.org/10.1016/j.revmed.2019.12.014).
- [162] Aude KEMPF et Emmanuelle KEMPF. “L’informatisation de l’aide à la décision : la décision médicale est-elle indemne ? L’exemple d’un outil prédictif en cancérologie”. In : *Revue française d’éthique appliquée* 1.1 (2016), p. 59-70. ISSN : 2494-5757. DOI : [10.3917/rfeap.001.0059](https://doi.org/10.3917/rfeap.001.0059).
- [163] OPENAI. *ChatGPT : Optimizing Language Models for Dialogue*. OpenAI. 30 nov. 2022. URL : <https://openai.com/blog/chatgpt/> (visité le 11/01/2023).
- [164] Comité Consultatif National D’ETHIQUE. *Avis 129 Contribution Du Comité Consultatif National d’éthique à La Révision de La Loi de Bioéthique 2018-2019*. 129. 18 sept. 2018. URL : <https://www.ccne-ethique.fr/node/174> (visité le 10/01/2023).
- [165] Comité Consultatif National D’ETHIQUE. *Avis 141 Du CCNE et 4 Du CNPEN "Diagnostic Médical et Intelligence Artificielle : Enjeux Ethiques"*. 141. 23 nov. 2022. URL : <https://www.ccne-ethique.fr/node/531> (visité le 16/02/2023).
- [166] David GRUSON et Claude KIRCHNER. “La « garantie humaine », levier majeur de régulation des enjeux éthiques associés au numérique et à l’intelligence artificielle en santé”. In : *Haut Conseil de la Santé Publique*. Actualité et dossier en santé publique 108 (sept. 2019), p. 41-42. URL : <https://www.hcsp.fr/Explore.cgi/adsp?clef=170> (visité le 13/02/2023).

-
- [167] AFNOR. *XP Z77-101 - Guide de Bonnes Pratiques En Matière de Gouvernance Des Démarches Éthiques Au Sein Des Organisations*. Août 2021. URL : <https://norminfo.afnor.org/norme/xp-z77-101/guide-de-bonnes-pratiques-en-matiere-de-gouvernance-des-demarches-ethiques-au-sein-des-organisations/192346> (visité le 28/03/2023).
- [168] Code de SANTÉ PUBLIQUE. *Articles R6113-1 à R6113-11*. 2016. URL : https://www.legifrance.gouv.fr/codes/section_lc/LEGITEXT000006072665/LEGISCTA000032473589?idSecParent=LEGISCTA000006190793&anchor=LEGISCTA000032473592#LEGISCTA000032473592 (visité le 19/02/2023).
- [169] Karen FACEY *et al.* “Patients’ Perspectives in Health Technology Assessment : A Route to Robust Evidence and Fair Deliberation”. In : *International Journal of Technology Assessment in Health Care* 26.3 (juill. 2010), p. 334-340. ISSN : 1471-6348, 0266-4623. DOI : [10.1017/S0266462310000395](https://doi.org/10.1017/S0266462310000395).
- [170] Julia ABELSON *et al.* “Public and Patient Involvement in Health Technology Assessment : A Framework for Action”. In : *International Journal of Technology Assessment in Health Care* 32.4 (2016/ed), p. 256-264. ISSN : 0266-4623, 1471-6348. DOI : [10.1017/S0266462316000362](https://doi.org/10.1017/S0266462316000362).
- [171] Hervé NABARETTE. “Le processus d’implication des patients dans l’évaluation des technologies de santé à l’HAS”. In : *Sante Publique* 30.6 (2018), p. 767-775. ISSN : 0995-3914.
- [172] AFNOR. *NF ISO 31000 - Management Du Risque — Lignes Directrices*. Juin 2018. URL : <https://norminfo.afnor.org/norme/nf-iso-31000/management-du-risque-lignes-directrices/111640> (visité le 28/03/2023).
- [173] AFNOR. *ISO/IEC 23894 - Technologies de l’information — Intelligence Artificielle — Recommandations Relatives Au Management Du Risque*. Mars 2023. URL : <https://norminfo.afnor.org/norme/nf-isoiec-23894/technologies-de-linformation-intelligence-artificielle-recommandations-relatives-au-management-du-risque/188524> (visité le 28/03/2023).
- [174] Parlement EUROPÉEN et Conseil EUROPÉEN. *Règlement (UE) 2017/745 du 5 avril 2017 relatif aux dispositifs médicaux, modifiant la directive 2001/83/CE, le règlement (CE) n° 178/2002 et le règlement (CE) n° 1223/2009 et abrogeant les directives du Conseil 90/385/CEE et 93/42/CEE (Texte présentant de l’intérêt pour l’EEE.)* 5 avr. 2017. URL : <http://data.europa.eu/eli/reg/2017/745/oj/fra> (visité le 11/02/2023).

Résumé

D'une part, les structures d'urgence sont régulièrement surchargées, conduisant à une diminution de la sécurité et de la qualité des soins : une durée moyenne de séjour plus élevée, une mortalité plus élevée, *etc.* À son arrivée, le patient est vu par l'**Infirmier Organisateur de l'Accueil (IOA)** qui collecte des informations contextuelles et cliniques pour effectuer un *triage* afin de prioriser la prise en charge médicale des patients. D'autre part, les modèles d'**Intelligence Artificielle (IA)** à apprentissage supervisés proposent de prédire une variable cible de manière rapide, reproductible, systématique et infatigable quelle que soit l'heure du jour ou de la nuit, sur la base de données explicatives.

Ainsi, nous avons d'abord entraîné **Prédiction du Parcours Patient aux Urgences (3P-U)**, un modèle d'**IA** basé sur un réseau de neurones profond comprenant trois couches de 128, 64, et 32 neurones, et un réseau de neurones convolutionnel utilisant les notes de l'**IOA**, permettant de prédire de l'hospitalisation du patient à partir des données d'accueil (aire sous la courbe ROC à 83%). Nous avons effectué une revue systématique de la littérature des méthodes utilisées pour gérer les données manquantes sur des données provenant du contexte de médecine d'urgence. Sur les 72 études incluses, l'exclusion de patients contenant des données manquantes (30 publications; 41%) était la stratégie la plus courante, suivie par l'imputation par la moyenne (17 publications; 23%). Ensuite, basé sur les données structurées uniquement, nous avons entraîné un modèle prédictif explicable en utilisant les techniques des explications additives de Shapley et des graphes de dépendance partielle : l'âge, la fréquence cardiaque et la présentation pour de multiples traumatismes ont le plus influencé la réponse du modèle sur l'admission du patient, confirmant ainsi les mécanismes de réflexion cognitif des urgentistes. Puis, utilisé pendant la pandémie de COVID-19, **3P-U** a été utilisé pour optimiser les lits : sur les 105 457 patients inclus, l'aire sous la courbe ROC était de 82% pour tous les patients et de 90% pour les patients évidents ($P_{admission} \in [0; 0.2] \cup [0.6; 1]$), permettant à la cellule de crise hospitalière de coordonner la conversion des services en unités dédiées aux patients COVID-19. Enfin, vu la proposition de règlement de la Commission Européenne visant à établir des règles harmonisées concernant l'**IA** baptisée "Artificial Intelligence Act", nous nous posons la question de la mise en place d'un collège de garantie humaine pluriprofessionnel visant à garantir la supervision des outils d'**IA** de la conception au développement, au déploiement et à l'utilisation quotidienne permettra ainsi d'assurer durablement la sécurité des patients.

Abstract

On one side, emergency departments are often overcrowded, leading to decrease quality and safety in care : length of stay and mortality increase, *etc.* At his arrival, the **Triage Nurse (TN)** evaluates the vital status based on contextual and clinical variables : *triage* stage. On the other hand, **Artificial Intelligence (AI)** supervised models offer to predict an outcome variable in a fast, reproductible, systematic, and tireless maner, independently of the day or night time, based on explicatives variables.

Thus, we first trained **Patients's Preditive Pathway in Emergency (3P-U)**, an **AI** model based on a deep neural network composed of three layers of 128, 64, and 32 neurones, and a convolutive neural network using the **TN** notes. It provides the prediction of admission from its triage data (area under the ROC curve at 83%). Then, we performed a systematic review on methods to manage missing data in **AI** models in emergency medicine. Among the 72 included studies, dropping patients with missing data (30 studies; 41%) was the most used strategy, followed by the mean imputation (17 publications; 23%). Then, based on structured data only, we trained an explicable predictive model using additive Shapley explanation, and partial dependence plot. Age, heart frequency and coming for multiple sever injuries mostly influenced the model, confirming the reflexive cognitif mechanisms of emergency physicians. Next, used during the COVID-19 outbreak, **3P-U** has been used to optimize beds management : among the 105 457 included patients, the area under the curve was 82% for all patients and 90% for obvious patients ($P_{admission} \in [0; 0.2] \cup [0.6; 1]$). At last, the European Commission proposal to harmonize the rules of the " **AI Act**", we raised the question of creating a human oversight college multi-professionnal to guarantee the **AI** based tools from the conception, to the deployment and the daily use offering a guarantee of safety.