



Titre: Modèles d'apprentissage automatique pour prédire et expliquer la congestion à l'urgence
Title: congestion à l'urgence

Auteur: Ilitea Kina
Author:

Date: 2024

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Kina, I. (2024). Modèles d'apprentissage automatique pour prédire et expliquer la congestion à l'urgence [Master's thesis, Polytechnique Montréal]. PolyPublie.
Citation: <https://publications.polymtl.ca/58356/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/58356/>
PolyPublie URL:

Directeurs de recherche: Louis-Martin Rousseau, & Nadia Lahrichi
Advisors:

Programme: Maîtrise recherche en mathématiques appliquées
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Modèles d'apprentissage automatique pour prédire et expliquer la congestion à
l'urgence**

ILITEA KINA

Département de mathématiques et de génie industriel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
Mathématiques appliquées

Avril 2024

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

**Modèles d'apprentissage automatique pour prédire et expliquer la congestion à
l'urgence**

présenté par **Ilitea KINA**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
a été dûment accepté par le jury d'examen constitué de :

Antoine LEGRAIN, président

Louis-Martin ROUSSEAU, membre et directeur de recherche

Nadia LAHRICHI, membre et codirectrice de recherche

Andrea LODI, membre

DÉDICACE

Aux travailleurs de la santé qui portent le système sur leurs épaules

A bad system will beat a good person everytime

RÉSUMÉ

La congestion des services d'urgence est un problème mondial pressant exacerbé par une demande croissante de soins de santé et des ressources limitées, surtout dans le contexte de la pandémie de COVID-19. Générée par un déséquilibre entre les besoins et les ressources, la congestion accentue à son tour la demande en augmentation la morbidité et la mortalité chez les patients, en plus de déléter les ressources en épuisant le personnel soignant. C'est dans ce contexte qu'est née la collaboration avec l'équipe du Centre Hospitalier de l'Université de Montréal.

Avec l'objectif principal de prédire la congestion 24 heures et 48 heures d'avance, nous introduisons une approche novatrice en utilisant le score EDWIN comme mesure objective de la congestion et comme variable dépendante à prédire. Les approches traditionnelles pour évaluer la congestion se basent souvent sur des proxy, tels que l'achalandage ou le nombre d'admissions, plutôt que d'évaluer directement le phénomène de congestion lui-même.

En intégrant les notions de temporalité et de logistique hospitalière, notre étude s'est concentrée sur les délais logistiques pour obtenir les services nécessaires à l'urgence, tels les examens diagnostiques et délais de visite médicales, comme variables indépendantes. Reconnaisant la nature dynamique dans le temps de ces variables, des délais spécifiques à chaque variable ont été ajustés à travers une fonction de corrélation croisée pour prédire la congestion.

En exploitant des modèles d'apprentissage automatique et d'optimisation, notamment des arbres de décision, forêts aléatoires, arbres boostés par gradient et arbres de classification optimaux, nous prédisons la congestion dans 24 et 48 heures, atteignant des taux de précision élevés de 90,1 % et 89,9 %, respectivement ainsi que des R^2 ajustés de 84,9%. De plus, l'intégration de délais spécifiques aux variables à travers une fonction de corrélation croisée améliore significativement les performances des modèles.

Notre analyse révèle le rôle pivot des score de congestion passés dans la prédiction des occurrences futures et met en lumière la nature dynamique des étiologies de la congestion. Grâce aux valeurs SHAP et aux valeurs contrefactuelles, nous montrons une variabilité entre les variables importantes à un niveau global et local ainsi qu'une variabilité dynamique dans le temps. Nous présentons un outil prédictif couplé à des méthodes d'intelligence artificielle explicables, offrant des aperçus sur les causes quotidiennes de congestion et aidant les administrateurs dans l'allocation et la planification des ressources. En comblant un vide critique dans la prédiction directe de la congestion et en fournissant des insights exploitables, notre étude contribue à améliorer les pratiques de gestion des services d'urgence et, ultimement,

les soins aux patients.

ABSTRACT

Emergency department overcrowding is a pressing global problem exacerbated by growing healthcare demand and limited resources, especially in the aftermath of the COVID-19 pandemic. Generated by an imbalance between needs and resources, overcrowding, in turn, exacerbates demand by increasing patient morbidity and mortality, as well as depleting resources by exhausting healthcare personnel. Against this backdrop, collaboration with the Centre Hospitalier de l'Université de Montréal team was born.

With the primary objective of predicting overcrowding 24 and 48 hours in advance, we are introducing an innovative approach using the EDWIN score as an objective measure of overcrowding and as the dependent variable to be predicted. Traditional approaches to assessing overcrowding often rely on proxies, such as ridership or number of admissions, rather than directly assessing the overcrowding phenomenon itself.

By integrating temporality and hospital logistics notions, our study focused on logistical delays in obtaining necessary emergency services, such as diagnostic tests and medical visit times, as independent variables. Recognizing the time-dynamic nature of these variables, variable-specific delays were fitted through a crosscorrelation function to predict overcrowding.

By exploiting machine learning and optimization models, including decision trees, random forests, gradient-boosted trees, and optimal classification trees, we predict overcrowding in 24 and 48 hours, achieving high accuracy rates of 90.1% and 89.9%, respectively, as well as adjusted R^2 of 84.9%. In addition, the integration of variable-specific lags through a crosscorrelation function was found.

Our analysis reveals the pivotal role of past overcrowding scores in predicting future occurrences and highlights the dynamic nature of overcrowding etiologies. Using SHAP and counterfactual values, we show variability between important variables at a global and local level, as well as dynamic variability over time. We present a predictive tool coupled with explainable artificial intelligence methods, offering insights into the day-to-day causes of overcrowding and assisting administrators in resource allocation and planning. By filling a critical gap in the direct prediction of overcrowding and providing actionable insights, our study contributes to improving emergency department management practices and, ultimately, patient care.

TABLE DES MATIÈRES

DÉDICACE	iii
RÉSUMÉ	iv
ABSTRACT	vi
TABLE DES MATIÈRES	vii
LISTE DES TABLEAUX	x
LISTE DES FIGURES	xi
LISTE DES SIGLES ET ABRÉVIATIONS	xiii
LISTE DES ANNEXES	xiv
CHAPITRE 1 INTRODUCTION	1
1.1 Modèle de fonctionnement de l'urgence	1
1.1.1 Entrée (input)	1
1.1.2 Débit (throughput)	2
1.1.3 Sortie (output)	2
1.2 Problématique	2
1.3 Objectifs et hypothèses	3
1.4 Domaine d'application	3
1.5 Réalisation de la recherche	4
1.6 Structure du mémoire	5
CHAPITRE 2 REVUE DE LITTÉRATURE	6
2.1 Impacts médicaux	6
2.1.1 Qualité des soins	6
2.1.2 Mortalité	6
2.1.3 Morbidité	7
2.1.4 Satisfaction	7
2.1.5 Coûts	8
2.2 Causes de la congestion	8
2.3 Variables étudiées	9

2.3.1	Achalandage de l'urgence	9
2.3.2	Admission	10
2.3.3	Durée de séjour et temps d'attente	10
2.4	Scores de congestion	11
2.5	Modèles de prédiction	12
2.5.1	Arbres de classification et de régression	12
2.5.2	Forêts aléatoires	13
2.5.3	Arbres boostés par le gradient	13
2.5.4	Arbres de classification optimaux	14
CHAPITRE 3 DÉMARCHE DE RECHERCHE		16
3.1	Lacunes de recherche	16
3.2	Nature temporelle du phénomène	16
3.3	Prédiction	17
3.4	Méthodes	17
3.4.1	Population	17
3.5	Sécurité et considérations éthiques	17
3.6	Cohérence des travaux	18
CHAPITRE 4 ARTICLE 1: FORECASTING AND EXPLAINING EMERGENCY DEPARTMENT OVERCROWDING USING MACHINE LEARNING MODELS .		19
4.1	Abstract	19
4.2	Introduction	20
4.3	Literature review and related work	21
4.4	Methodology	25
4.4.1	Data preparation	25
4.4.2	Variable specific lags	27
4.4.3	Modeling	28
4.4.4	Local interpretation	29
4.5	Results	31
4.5.1	Variable specific lags	31
4.5.2	Models and performance	32
4.5.3	Local interpretation	36
4.6	Managerial Insights	37
4.7	Conclusion	39
CHAPITRE 5 ANALYSES COMPLÉMENTAIRES		42

5.1	Portrait de l'urgence	42
5.2	Gestion des variables	42
5.2.1	Sélection des variables	42
5.2.2	Relations non-linéaires	44
5.2.3	Variable dépendante - Score EDWIN	46
5.3	Analyses temporelles	46
5.4	Analyses comparatives - Méthodes de régression	48
5.4.1	Régression linéaire	48
5.4.2	Arbres de régression	49
5.4.3	Forêts aléatoires	50
5.4.4	Arbres boostés par le gradient	50
5.5	Arbres de classification optimaux	51
5.5.1	Idée générale	51
5.5.2	Modèle théorique	52
5.5.3	Modèle pratique	54
5.6	Modèle de librairie	56
CHAPITRE 6	DISCUSSION	58
6.1	Analyses temporelles	58
6.2	Performance des analyses comparatives	59
6.3	Score EDWIN	59
CHAPITRE 7	CONCLUSION	62
7.1	Synthèse des travaux	62
7.2	Limitations de la solution proposée	62
7.3	Améliorations futures	63
RÉFÉRENCES		66
ANNEXES		71

LISTE DES TABLEAUX

Table 4.1	Summary of the characteristics of the theoretical ideal modelization of overcrowding versus models in literature	26
Table 4.2	Comparison of performance between ML models for classification . .	34
Table 4.3	Comparison of performance between ML models for classification . .	36
Tableau 5.1	Modèles de régression linéaires pour différents délais de prédiction . .	49
Tableau 5.2	Modèles d'arbres de régression pour différents délais de prédiction . .	50
Tableau 5.3	Modèles de forêts aléatoires de régression pour différents délais de prédiction	50
Tableau 5.4	Modèles de d'arbres de régression boostés par le gradient pour différents délais de prédiction	51
Tableau 5.5	Modèles d'arbres de classification optimaux pour différents délais de prédiction	56
Tableau 5.6	Modèles d'arbre de classification optimaux ajustés pour la régression pour différents délais de prédiction	56
Tableau A.1	Variables indépendantes intégrées aux modèles	71

LISTE DES FIGURES

Figure 4.1	A representation of the flow in the ED. Patients can exit the ED by either being reoriented by a triage nurse to a walk-in clinic, being discharged by the ER doctor of a specialist consulted, or being admitted to the hospital. While in the ED, patients continue to require ED resources, including ED stretchers and investigations.	22
Figure 4.2	A summary of the methodology used in this study	31
Figure 4.3	Comparison between crosscorrelation function results in a 96-hour time frame for pending consultations (number) and pending admissions (delay)	32
Figure 4.4	Pareto curve showing the ten variables having the most impact on variance in the logistic regression model using variable-specific lags forecasting overcrowding in 48 hours	33
Figure 4.5	A classification tree forecasting overcrowding in 24 hours with a variable-specific lag	34
Figure 4.6	Feature importance of variable-specific GBM models for predictions in 24 hours and 48 hours	35
Figure 4.7	Optimal classification tree using variable-specific lags to predict overcrowding in 24 hours.	35
Figure 4.8	Highest SHAP values for 24 hrs forecasting using the variable-specific GBM model for a random overcrowded moment (August 13th, 2021, at 1 :15 AM)	36
Figure 4.9	Highest SHAP values for 24 hrs forecasting using variable-specific GBM model for another random overcrowded moment (September 5th, 2021, at 9 PM)	37
Figure 4.10	SHAP values trended every 6 hours over 120 hours for five key features	38
Figure 4.11	Highest Counterfactual values for 24 hrs forecasting using GBM model for a random overcrowded moment	39
Figure 4.12	Highest Counterfactual values for 24 hrs forecasting using variable-specific GBM model for another random overcrowded moment (September 5th, 2021, at 9 PM)	40
Figure 4.13	A virtual dashboard summarizing the results of the study.	41
Figure 5.1	Orientation des patients après le triage	43
Figure 5.2	Distribution des catégories de triage des patients, où P1 représente la catégorie la plus urgente et P5 la moins urgente	43

Figure 5.3	Prise en charge (PEC) des patients à partir du département d'urgence	44
Figure 5.4	Score EDWIN normalisé selon le jour de la semaine et l'année	45
Figure 5.5	Relation entre le score EDWIN et le nombre de consultations en attente	45
Figure 5.6	Valeur numérique du score EDWIN sur la plage de temps étudiée . .	47
Figure 5.7	Distribution des quartiles du score EDWIN	47
Figure 5.8	Décomposition multiplicative des séries temporelles	48
Figure 5.9	Méthode de fenêtre coulissante pour éviter une contamination entre les ensembles d'entraînement et les ensembles tests pour des données temporellement liées	49
Figure 5.10	Arbre de régression avec délais spécifiques aux variables pour prédire la congestion dans 24 heures	51
Figure 5.11	Variables les plus importantes dans les modèles d'arbres boostés par le gradient avec délais spécifiques aux variables pour des prédictions à 24h et 48h	52
Figure 5.12	Arbre de régression optimal pour une prédiction dans 24h en utilisant des délais spécifiques aux variables	57
Figure 6.1	Variation du score EDWIN sur 24 heures	61
Figure B.1	Raisons de consultations des patients ayant visité l'urgence pendant la période étudiée	72

LISTE DES SIGLES ET ABRÉVIATIONS

EDWIN	Emergency Department Work Index
COVID-19	Coronavirus Disease 2019
CTAS	Canadian Triage Acuity Scale
CART	Classification And Regression Tree
CHUM	Centre Hospitalier de l'Université de Montréal
CITADEL	Centre d'intégration et d'analyse des données médicales
LSSS	Loi sur les services de santé et les services sociaux
IQR	Interquantile Range
ARIMA	autoregressive integrated moving average
OCT	Optimal Classification Tree
CART	Classification And Regression Tree
RF	Random Forest
GBM	Gradient Boosting Machines
LR	Logistic Regression
CCF	Crosscorrelation Function
RMSE	Residual Mean Squared Error
NEDOCS	National Emergency Department Overcrowding Scale
READI	Rapid Emergency Department Intensity
LOS	Length of stay (durée de séjour hospitalier)
SHAP	Shapley Additive Explanations

LISTE DES ANNEXES

Annexe A	Variables indépendantes intégrées aux modèles	71
Annexe B	Raisons de consultation à l'urgence	72

CHAPITRE 1 INTRODUCTION

Dans un système de santé avec des ressources limitées, le phénomène de congestion à l’urgence est un véritable fardeau, au Canada comme ailleurs dans le monde [1]. La congestion à l’urgence représente un phénomène où les demandes de soins surpassent la capacité d’une institution à fournir ces soins [2]. Ce phénomène a pour effet d’entraîner un ralentissement de l’accessibilité aux soins.

1.1 Modèle de fonctionnement de l’urgence

La fluidité des patients à l’urgence a été souvent décrite à travers un modèle de *input-throughput-output*, qu’on peut traduire entrée-débit-sortie [3]. Le modèle entrée-débit-sortie sert de cadre conceptuel pour comprendre et optimiser le flux de patients dans les services d’urgence. Ce modèle fournit une approche structurée pour analyser les dynamiques impliquées dans la gestion de l’afflux, du traitement et de la sortie des patients, dans le but ultime d’améliorer l’efficacité et les résultats pour les patients.

1.1.1 Entrée (input)

La phase d’entrée représente l’arrivée des patients en quête de soins d’urgence. Cet afflux peut varier considérablement en fonction de facteurs tels que l’heure de la journée, la saisonnalité, la situation géographique et la gravité des incidents survenus dans la communauté. Les départements d’urgence se situent dans des centres hospitaliers. Les patients peuvent se présenter dans le département d’urgence de plusieurs manières : par eux-mêmes (ambulatoire), en ambulance ou via la recommandation d’un autre médecin. Lorsque les patients arrivent à l’urgence, ils sont enregistrés dans le système de l’urgence par un agent administratif. Ils sont ensuite évalués par un infirmier de triage. À ce moment, le patient explique ses symptômes et est examiné. Au triage, le patient est classifié selon l’échelle de priorité Canadian Acuity Triage Scale (CTAS) de 1 à 5, 1 étant des symptômes qui mettent imminemment sa survie en danger et nécessitent l’attention immédiate d’un médecin en salle de réanimation, et 5, des symptômes mineurs ou non-urgents qui nécessitent l’attention d’un médecin dans les prochaines heures.

1.1.2 Débit (throughput)

Après le triage, les patients sont orientés selon l'évaluation du triage : dans la salle d'attente, dans l'aire des civières, dans l'aire de réanimation. La prochaine étape pour les patients est alors d'attendre de voir le médecin d'urgence. Lorsque le patient voit le médecin d'urgence, il se peut que le médecin d'urgence ait besoin d'informations supplémentaires avant de conclure à la disposition finale du patient.

La phase de débit comprend donc les processus qui, au sein de l'urgence, effectuent l'évaluation, le diagnostic, le traitement et la prise en charge des patients. Cette phase englobe diverses activités, notamment l'évaluation médicale, les tests de diagnostic, des essais thérapeutiques tels l'administration de médicaments ou des procédures médicales et les consultations avec des spécialistes.

1.1.3 Sortie (output)

Une fois la phase de débit résolue, le médecin d'urgence a suffisamment d'information pour décider de la disposition finale du patient. La phase de sortie comprend le congé hospitalier, le transfert dans un autre établissement ou l'admission des patients. Le congé peut être ralenti par des facteurs de transport pour des patients en perte d'autonomie, le transfert peut être ralenti par un délai dans l'acceptation de l'autre établissement à recevoir le patient et l'admission peut être ralenti par le délai dans l'accueil de l'unité d'hospitalisation.

1.2 Problématique

Au fil des années, le portrait des soins de santé a connu des changements significatifs, principalement en raison des progrès de la médecine et de l'amélioration de l'espérance de vie. Ces progrès ont entraîné une augmentation notable des besoins de soins de santé dans le monde entier. L'augmentation de l'espérance de vie cause une augmentation du nombre de visites dans les services d'urgence, ayant pour effet d'amplifier la phase d'entrée du modèle de fluidité.

De même, la démographie des patients dans les services d'urgence a évolué, témoignant une population plus âgée avec des conditions médicales de plus en plus complexes. Cette évolution démographique est attribuée à l'allongement de la durée de vie et à une meilleure compréhension de la médecine, ce qui engendre que les patients se présentent avec des problèmes de santé plus complexes. Par conséquent, la phase de débit s'est allongée, car les modalités de diagnostic et d'options de traitement se sont multipliées. Dans le même sens, la phase

de sortie s'est complexifiée en raison de l'expansion des investigations et des modalités de traitement disponibles.

L'incapacité des établissements de santé à faire face à l'augmentation et la complexification des demandes de soins de santé représente la congestion vécue à l'urgence et à l'hôpital.

À travers ce problème de gestion s'introduisent les notions de planification et de prédiction de la congestion afin d'atténuer ou d'éviter les méfaits de la congestion. Peu d'études se sont réellement questionnées sur la prédiction de la congestion à l'urgence en elle-même.

C'est dans le contexte d'un problème prédictif complexe dans une situation où les décideurs n'ont pas de littérature mathématiques où il devient crucial de développer des modèles très interprétables. C'est dans cet objectif que les arbres de classification optimaux ont été créés et c'est pourquoi nous avons choisi de les étudier dans ce mémoire.

1.3 Objectifs et hypothèses

Notre objectif primaire est de développer des modèles d'algorithmes d'apprentissage automatique qui pourraient prédire la congestion de l'urgence, en se concentrant sur des modèles qui sont plus facilement interprétables soit les arbres de classification optimaux. Nous pensons que la congestion de l'urgence est un phénomène graduel qui prend plusieurs heures à s'installer et qui pourrait donc être prédit plusieurs heures à l'avance.

Notre objectif secondaire est de déterminer la nature des différents délais dans la prise en charge et l'orientation d'un patient qui causent la congestion à l'urgence. Notre hypothèse est que les différents départements qui font ralentir le roulement de l'urgence ne sont pas toujours les mêmes (parfois le département de radiologie ou le laboratoire, parfois les hospitalisations en chirurgie, etc.). Il deviendrait donc important de déterminer de manière dynamique quel département ou service engendrerait une congestion de l'urgence.

Comme des méthodes de régression et de classification seront utilisées, l'évaluation sera faite à partir des valeurs de R^2 ajusté, RMSE, score F1 et la précision.

1.4 Domaine d'application

Les résultats attendus sont un modèle d'apprentissage automatique conçu pour prédire et expliquer les causes de la congestion des établissements de santé, à la fois à l'échelle globale et à court terme, dans les 24 et 48 heures qui suivent. Cet outil présenterait un potentiel d'application important dans divers domaines du secteur des soins de santé.

L'un des principaux domaines d'application de cet outil de recherche est la gestion des éta-

blissements de santé. En s'appuyant sur ces modèles, les administrateurs des établissements de santé peuvent anticiper les situations de congestion imminentes et prendre des mesures proactives pour en atténuer l'impact. Ces informations quotidiennes et dynamiques peuvent permettre d'effectuer des décisions adaptées à la réalité ponctuelle. Il peut s'agir d'ajuster les niveaux de personnel selon le débit anticipé, de réaffecter les ressources aux services en nécessité, d'optimiser le flux des patients et de mettre en œuvre des interventions ciblées pour s'attaquer aux causes sous-jacentes de la congestion.

Bien que des interventions ponctuelles soient un domaine d'application principal, des interventions plus globales appuyées par des données peuvent également améliorer la prise en charge et la gestion du système hospitalier. Il est essentiel de comprendre les facteurs qui contribuent à la congestion des établissements de santé pour éclairer les initiatives de planification et d'élaboration des politiques institutionnelles. En analysant les prédictions et les explications fournies par le modèle d'apprentissage automatique, les décideurs peuvent identifier les problèmes systémiques contribuant à la congestion. Ces informations peuvent éclairer le développement d'interventions ciblées, de stratégies d'allocation des ressources et de réformes politiques visant à s'attaquer aux causes de la congestion et à améliorer les systèmes de prestation de soins de santé à plus grande échelle.

Le dernier élément envisagé dans les avancées possibles liées à ce travail sont la recherche et l'innovation dans le domaine de la gestion et des opérations des soins de santé.

L'évaluation et la validation continues des performances d'un outil de prédiction et d'explication de la congestion peuvent contribuer à l'avancement des pratiques fondées sur des données probantes pour la gestion et la prévention de la congestion dans les établissements de santé. Une compréhension dynamique des facteurs causant la congestion de l'urgence ouvre la voie à la poursuite de la recherche et de l'innovation dans ce domaine.

1.5 Réalisation de la recherche

L'étude a été menée à partir de données rétrospectives du Centre Hospitalier de l'Université de Montréal (CHUM), un hôpital universitaire tertiaire au Québec. Contrairement aux études précédentes qui s'appuyaient sur des indicateurs de congestion, tels que les taux d'achalandage ou les volumes d'admission, cette étude a utilisé un score de congestion validé EDWIN. Ce score fournit une représentation plus nuancée et plus précise de la congestion. Avec un focus sur les facteurs logistiques et les variables influençant le fonctionnement de l'hôpital, des modèles prédictifs ont été développés. En s'appuyant sur l'analyse des séries temporelles, la modélisation de la régression et les méthodes d'ensemble, les algorithmes prédictifs ont

été entraînés et validés à l'aide de l'ensemble de données du CHUM. De multiples modèles prédictifs ont été explorés, chacun incorporant différents ensembles de caractéristiques et d'algorithmes, afin d'identifier l'approche la plus efficace.

En outre, l'étude a adopté une analyse comparative des modèles prédictifs afin d'évaluer leur performance et leur interprétabilité. En évaluant des paramètres tels que la précision, le score F1, le R2 ajusté et le RMSE, le modèle optimal a été identifié.

Les progrès réalisés démontrent la faisabilité de la prédiction de la congestion dans l'environnement hospitalier du CHUM. En s'appuyant sur des données chronologiques et un score de congestion validé, les modèles prédictifs développés ont le potentiel d'améliorer la planification de la capacité hospitalière, l'allocation des ressources et la gestion des flux de patients.

1.6 Structure du mémoire

Le chapitre 2 présente la littérature concernant les causes et conséquences de la congestion et les méthodes d'évaluation de la congestion. Ce chapitre inclut également une sélection d'études qui se sont penchées sur la prédiction de proxys de la congestion et sur les modèles de prédiction utilisés dans cette étude. Le chapitre 3 explique le raisonnement derrière les choix de recherche et contextualise la méthode utilisée dans l'article. Le chapitre 4 consiste en l'article sur la prédiction de la congestion à l'urgence, incluant seulement les modèles de classification. Le chapitre 5 intègre des statistiques sur le portrait global de l'urgence ainsi que les modèles de régression. Le chapitre 6 apporte un complément à la discussion incluse dans l'article, avec un focus sur les résultats présentés dans le chapitre 5. Le chapitre 7 conclut ce mémoire avec la synthèse des travaux, les limitations et les améliorations futures.

CHAPITRE 2 REVUE DE LITTÉRATURE

2.1 Impacts médicaux

La congestion est un phénomène très répandu et délétère dans les urgences au Québec [4]. Elle est difficile à prédire à l'heure actuelle et les mesures pour la corriger sont souvent tardives et mettent de la pression sur le personnel. De nombreuses études ont démontré l'impact négatif de la congestion à l'urgence sur la qualité des soins fournis aux patients, augmentant ainsi la mortalité [5] et la morbidité [6] des patients qui ont été pris en charge à l'urgence. Par ailleurs, le taux de satisfaction des patients ainsi que du personnel soignant est plus faible lorsque l'urgence est congestionnée [7].

2.1.1 Qualité des soins

Dans une urgence congestionnée, la pression sur le personnel soignant ainsi que le ralentissement d'accès aux services nécessaires cause une diminution de la qualité des soins. Des études ont démontré que dans une urgence où le nombre de patients est supérieur à celui pour lequel l'urgence est préparée, le risque d'obtenir un délai dans les soins essentiels et un écartement entre la prise en charge et les lignes directrices en syndrome coronarien aigu [8]. Dans les cas de sepsis sévères et de chocs septiques, des infections mettant la survie en danger, la congestion à l'urgence est associée à un délai à administrer des antibiotiques, à une moins bonne adhérence aux manœuvres de réanimation et à une moins bonne compliance aux lignes directrices en sepsis [9]. Cette diminution de la qualité des soins entraîne souvent une détérioration des *outcome* des patients, dont la mortalité et la morbidité.

2.1.2 Mortalité

Le risque de mortalité augmentée en conséquence de délais dans l'administration des soins est très bien documenté dans la littérature médicale. Une augmentation de la mortalité très claire a été liée aux délais de traitements pour les maladies mettant imminemment la survie en danger : infarctus du myocarde ("crise cardiaque") [10], traumatismes majeurs [11] et arrêts cardio-respiratoires [12]. Ces conséquences sont si bien documentées que la majorité des hôpitaux et des systèmes ambulanciers sont munis de codes pour avertir plusieurs départements de l'hôpital de se préparer à la venue d'un patient souffrant d'une de ces conditions.

Même pour les patients qui ne souffrent pas d'une de ces conditions, ceux qui sont gravement

malades et qui ont besoins de soins intensifs ont également un risque de mortalité augmenté lorsqu'il y a un délai d'admission à l'unité de soins intensifs. Une étude a conclu que cette augmentation de la mortalité correspondait à 1.5% de mortalité supplémentaire par heure supplémentaire passée à attendre à l'urgence [13]. Il s'agit d'un pourcentage extrêmement significatif en médecine. Outre l'augmentation de la mortalité, les patients qui ont connus une diminution de la qualité des soins en raison de délais hospitaliers sont également plus à risque de complications (morbidity).

2.1.3 Morbidity

Dû à la diminution de la qualité des soins et aux délais d'accès aux traitements causés par la congestion, le taux de complications est supérieur chez les patients ayant vécu la congestion de l'urgence. Une augmentation de la durée de séjour à l'urgence avant l'admission aux soins intensifs est associée à une augmentation de la durée de séjour aux soins intensifs et à la durée de recours à de la ventilation mécanique [14, 15]. De plus, la congestion est associée à une hausse marquée des besoins en hémodialyse dû à une augmentation des complications rénales ainsi qu'à une augmentation du nombre de choc requérant un recours à des vasopresseurs [16]. Ces complications requièrent toutes un traitement intra-hospitalier, augmentant ainsi la durée de séjour des patients. En combinant ces facteurs, la congestion à l'urgence crée un cercle vicieux où des délais supplémentaires dans le traitement médical essentiel entraînent des complications potentiellement évitables, impactant ainsi la qualité globale des soins prodigués. La diminution de la qualité des soins entraînant des conséquences chez les patients, le taux de satisfaction à la fois des patients et du personnel soignant est diminué lorsqu'ils font face à la congestion de l'urgence.

2.1.4 Satisfaction

En raison des longs délais et de la diminution de la qualité perçue des soins, la congestion des urgences diminue la satisfaction des patients à l'égard des soins, ce qui affecte non seulement leur expérience lors de leur visite aux urgences, mais aussi leur satisfaction pendant le reste de leur hospitalisation [17].

D'une part, la congestion des services d'urgence entraîne une augmentation de la demande de soins; d'autre part, il conduit à un épuisement des ressources. La littérature indique que la congestion des urgences non seulement affecte le moral du personnel, mais accroît également l'anxiété des professionnels de la santé [18]. Le déséquilibre entre les ratio de patients :infirmières ainsi que l'environnement de travail surchargé sont des facteurs de perte de rétention de personnel [19]. Pour le personnel médical apprenant, tels les étudiants et

résidents, une diminution de la congestion peut améliorer l'exposition clinique et la qualité de l'apprentissage des résidents [20]. Au niveau hospitalier, le roulement de personnel représente une conséquence administrative importante de la congestion, mais n'est pas la seule. La congestion cause également des coûts financiers à l'hôpital.

2.1.5 Coûts

La congestion au service d'urgence exerce un impact significatif sur l'efficacité du flux de patients, se traduisant par une prolongation moyenne d'une journée de leur séjour. Pour un hôpital de 600 lits avec une durée de séjour moyenne de 6 jours (au lieu de 5), ceci représente une diminution nette de 100 lits disponibles si la durée de séjour avait été de 5 jours [21]. La congestion persistante au sein du service d'urgence engendre des répercussions économiques significatives pour un hôpital. En prolongeant la durée moyenne de séjour des patients d'au moins une journée, cette congestion entraîne inévitablement une augmentation des coûts opérationnels liés aux soins, à la logistique et aux ressources médicales nécessaires [22]. De plus, cette situation précaire peut entraîner la perte de profits potentiels, notamment lorsque des patients décident de quitter l'établissement contre l'avis médical en raison des délais et des désagréments engendrés par la congestion ou à cause de la diversion des ambulances vers d'autres hôpitaux moins congestionnés [23]. Ces visites manquées peuvent non seulement compromettre la santé des patients, mais également priver l'hôpital de revenus attendus provenant de traitements et d'interventions supplémentaires.

2.2 Causes de la congestion

La recherche menée sur le phénomène de congestion des départements d'urgence met en lumière les nombreux facteurs contribuant à ce défi omniprésent. Avec les avancées de la médecine et de la technologie au fil des ans, notre compréhension des diverses maladies et de leurs modalités de traitement s'est considérablement approfondie. Cette augmentation des connaissances a entraîné des améliorations significatives dans les résultats cliniques et une augmentation correspondante de l'espérance de vie. Cependant, ce progrès a également entraîné des changements démographiques, notamment le vieillissement de la population ainsi que la complexification du dossier médical des patients. Cette population plus âgée et plus complexe nécessite souvent une attention médicale plus intensive et prolongée, ce qui exerce une pression supplémentaire sur des ressources déjà limitées dans les départements d'urgence [4, 24].

Les défis logistiques découlant de cette transition démographique sont devenus de plus en plus

apparents. Les retards dans l’obtention de résultats diagnostiques critiques, tels que des tests de laboratoire ou des études d’imagerie, peuvent entraîner des séjours prolongés des patients à l’urgence, exacerbant la congestion [25]. De plus, les taux d’occupation élevés dans les hôpitaux signifient moins de lits disponibles pour les patients admis, ce qui se traduit par des séjours prolongés à l’urgence en attendant le transfert ou l’admission [26]. L’*access-block*, un phénomène où les patients ne peuvent pas avoir accès aux structures de soins nécessaires (salle d’opération, laboratoire de cathéterisation, etc.) en raison de divers goulots d’étranglement, aggrave encore la situation en entravant le flux des patients dans les départements d’urgence [27].

2.3 Variables étudiées

Malgré que la congestion semble découler de phénomènes logistiques dans les études portant sur le sujet, la majorité des études qui tentent de prédire un proxy de la congestion n’intègrent que très peu de données logistiques. La majorité des études utilisent également des proxy qui se concentrent sur les phases d’entrée ou de sortie du modèle de fluidité plutôt que de voir la congestion comme un problème de débit [28].

2.3.1 Achalandage de l’urgence

Au fil des ans, les chercheurs ont exploré différentes approches pour prédire avec précision les futures visites aux urgences. Ils ont utilisé divers facteurs tels que les conditions météorologiques, les événements de calendrier et la disponibilité des ressources de soins de santé dans leur modélisation prédictive. Le nombre de visites passées ainsi que les tendances de variations du nombre de visites peuvent également être utilisées pour prédire les visites futures [29]. Certaines études ont montré qu’en utilisant uniquement les conditions météorologiques et les données de calendrier comme caractéristiques, des modèles complexes tels que les réseaux de neurones Long Short-Term memory, les Gated Recurrent Networks et les réseaux neuronaux récurrents simples peuvent prédire les visites quotidiennes aux urgences avec un taux de précision très élevé de 94 % [30].

Pour améliorer la précision de ces modèles, les chercheurs ont également intégré davantage de données extrahospitalières, telles que la disponibilité des établissements de soins secondaires et les événements publics locaux. Ils ont constaté que des méthodes plus simples, notamment ARIMA et les moindres carrés récurrents, peuvent prédire avec une bonne précision les visites aux urgences pour le jour suivant en utilisant ces sources de données supplémentaires [31].

Ces études ont démontré que les visites quotidiennes aux urgences sont un résultat prévi-

sible des sources de données préhospitalières et extrahospitalières, et non de la logistique intrahospitalière. Bien que la congestion des urgences soit un problème important et que de nombreuses visites aux urgences puissent y contribuer, ces deux mesures ne décrivent pas nécessairement le même problème. Pour se pencher davantage sur les processus hospitaliers, les chercheurs tentent souvent de prédire les admissions à l'hôpital.

2.3.2 Admission

Ces dernières années, de nombreux chercheurs ont utilisé des techniques d'apprentissage automatique pour prédire les admissions de patients à partir des dossiers médicaux électroniques. Ces modèles intègrent un large éventail de données cliniques, notamment les symptômes, l'âge, le sexe et la catégorie de triage d'un patient [32]. Les méthodologies d'analyse prédictive comme la régression logistique et les algorithmes de gradient boosting ont montré une grande précision dans l'estimation des probabilités d'admission.

De plus, des études récentes ont révélé que les données textuelles non structurées provenant des notes de triage constituent un facteur important qui peut améliorer de manière significative la capacité prédictive, en particulier pour les admissions dans les 24 heures [33]. Un modèle qui a attiré l'attention est celui qui utilise les dossiers médicaux électroniques en temps réel et les classificateurs XGBoost pour prédire les admissions de patients dans les services d'urgence [34]. En tirant parti de ce modèle, les établissements de soins de santé peuvent allouer leurs ressources plus efficacement et planifier leur capacité, ce qui, en fin de compte, contribue à améliorer les résultats pour les patients.

Cependant, il est important de noter que ces modèles se concentrent généralement sur les trajectoires individuelles des patients et ne tiennent pas compte de la dynamique plus large des flux hospitaliers. Par conséquent, de nombreuses études sont encore nécessaires pour comprendre l'interaction complexe entre la prévision des admissions et le flux hospitalier. Pour combler l'écart entre la visite d'un patient (entrée) et l'admission (sortie), les chercheurs se sont penchés sur la durée de séjour et le temps d'attente à l'urgence (débit).

2.3.3 Durée de séjour et temps d'attente

De nombreuses études ont visé à prédire les temps d'attente pour les consultations en médecine d'urgence et la durée des séjours dans des services d'urgence très fréquentés. Ces études ont fait preuve d'une précision prometteuse dans leurs prédictions. Cependant, la plupart d'entre elles se concentrent sur l'analyse de facteurs individuels tels que les caractéristiques des patients et les attributs cliniques pour faire ces prédictions, au lieu de prendre en compte

le paysage logistique global du service d'urgence.

Par exemple, une étude s'est concentrée sur la prédiction de la durée de séjours des patients des services d'urgence pédiatriques. L'étude a analysé une série de données, notamment des données démographiques sur l'âge, la répartition des sexes, les pathologies diagnostiquées et la nécessité d'examen diagnostiques ultérieurs tels que des analyses de laboratoire, des radiographies, des échographies et des tomodensitométries [35]. Une autre étude s'est attachée à prédire la durée de séjour des patients dans un service d'urgence général. Cette étude a pris en compte des variables telles que la répartition de l'âge et du sexe, l'urgence indiquée par les catégories de triage, les modes d'arrivée et même les nuances temporelles des arrivées [36].

De nombreuses études ont également visé à prédire les temps d'attente pour les patients qui consultent un médecin d'urgence dans une urgence générale. Ces études utilisent des modèles qui prennent en compte des variables telles que l'achalandage de patients aux urgences et la disponibilité des médecins [37,38].

Bien que ces modèles aient permis de prédire l'expérience des patients dans les services d'urgence, leur capacité à fournir une vue d'ensemble de la manière dont les goulots d'étranglement logistiques affectent les niveaux de congestion et, en fin de compte, les temps d'attente et la durée des séjours des patients dans les services de santé d'urgence, reste insuffisante.

2.4 Scores de congestion

L'évolution et l'application des systèmes d'évaluation de la congestion des services d'urgence sont devenues des outils essentiels pour les professionnels de la santé qui doivent faire face à la complexité des soins d'urgence. Parmi ces systèmes, le National Emergency Department Overcrowding Scale (NEDOCS) s'est imposée comme un modèle important [39]. Elle a été affinée dans des versions spécialisées telles que PEDOCS [40] pour les urgences pédiatriques et CEDOCS pour les hôpitaux communautaires [41], reconnaissant les défis distincts et les caractéristiques démographiques des patients dans ces environnements.

Le Emergency Department Work Index (EDWIN), introduit en 2003, est un autre score important dans ce domaine. Le score EDWIN se distingue par le fait qu'il a été ajusté par des évaluations subjectives du personnel soignant et intègre des mesures objectives, ce qui permet d'avoir une vision plus complète de l'état opérationnel du service d'urgence [42].

Les analyses comparatives de ces systèmes de cotation, notamment NEDOCS, EDWIN et d'autres comme le score READI (Rapid Emergency Department Intensity), révèlent leurs forces et leurs faiblesses respectives. NEDOCS nécessite des données détaillées sur le volume

de patients, leur acuité et les ressources disponibles, qu'il peut être difficile de rassembler en temps utile. À l'inverse, le score EDWIN utilise des données très facilement accessibles et moins granulaires, bien qu'il manque potentiellement la profondeur des données saisies par NEDOCs [43].

Les recherches comparant ces outils ont mis en évidence les différences de leur efficacité. Le score EDWIN s'est avéré fiable pour surveiller la congestion et identifier les changements dans la dynamique des urgences [44], avec des scores élevés qui corrèlent avec des résultats négatifs tels que la diminution de la satisfaction des patients, l'augmentation de l'insatisfaction du personnel et une incidence plus élevée d'erreurs médicales [45–47]. En outre, il est associé à une augmentation du nombre de patients qui quittent l'hôpital sans avoir été vus [48], ce qui souligne l'impact critique de la congestion sur les soins et la sécurité des patients.

2.5 Modèles de prédiction

2.5.1 Arbres de classification et de régression

Les arbres de classification et de régression (CART) sont une technique de modélisation prédictive utilisée dans l'analyse des données et l'apprentissage automatique. Ils sont utiles pour faire des prédictions basées sur des variables indépendantes en partitionnant les données en sous-ensembles, puis en divisant récursivement les sous-ensembles en groupes plus petits et plus homogènes [49].

L'arbre divise les données en branches, chaque branche représentant une classe possible. À chaque nœud de l'arbre, l'algorithme sélectionne la variable prédictive qui sépare le mieux les données en classes distinctes sur la base de certains critères, tels que l'impureté de Gini. Les arbres de régression fonctionnent avec le même processus, mais au lieu de prédire des classes, ils prédisent la valeur moyenne de la variable de résultat au sein de chaque sous-ensemble.

Les modèles CART fournissent des résultats intuitifs et faciles à interpréter car ils résultent en un arbre visualisable avec des variables claires. La structure de l'arbre de décision représente visuellement le processus de prise de décision, ce qui permet d'identifier clairement les variables les plus influentes dans la prédiction. Pour cette raison, c'est une méthode fréquemment utilisée dans des contextes médicaux [50]. L'utilisation de la méthode "rpart" en R donne d'ailleurs des résultats très rapides [51].

Parce que les divisions sont itératives et basées sur les classes dominantes, les modèles CART peuvent capturer des relations entre les variables prédictives et le résultat qui échappent aux méthodes de régression linéaires [52]. Cela peut représenter un avantage si la relation entre les variables n'est pas linéaire.

Les modèles CART effectuent automatiquement une sélection des variables en identifiant les variables prédictives les plus informatives dans le modèle. Cela permet de rationaliser le processus de modélisation et de l'utiliser plus simplement dans des applications pratiques [53].

2.5.2 Forêts aléatoires

Les forêts aléatoires sont une méthode d'apprentissage d'ensemble utilisée dans l'apprentissage automatique pour les tâches de classification et de régression. Elles sont conçues pour construire plusieurs arbres de décision et produire une prédiction moyenne ou modale des arbres individuels [54].

Les forêts aléatoires ont permis d'obtenir une grande précision de prédiction pour différents types de données, supérieure à celle des arbres de classification et de régression [55]. En agrégeant les prédictions de plusieurs arbres de décision, les forêts aléatoires améliorent les performances de généralisation, ce qui en fait des modèles robustes pour capturer les relations complexes dans de grands ensembles de données, et en font des modèles fréquemment utilisés dans des contextes médicaux [56]. Elles sont particulièrement performante pour gérer de grands ensembles de données avec une haute dimensionalité [57].

Les forêts aléatoires sont intrinsèquement robustes au bruit et aux valeurs aberrantes des données. En construisant plusieurs arbres de décision basés sur des sous-ensembles aléatoires de données et de caractéristiques d'apprentissage, les forêts aléatoires réduisent l'impact des points de données individuels bruyants ou des valeurs aberrantes sur la prédiction globale, ce qui permet d'obtenir des prédictions plus fiables dans les ensembles de données du monde réel présentant une variabilité inhérente [54].

Les forêts aléatoires fournissent une mesure de l'importance des variables, indiquant la contribution relative de chaque variable prédictive à la tâche de prédiction [58]. Ce classement de l'importance des variables peut aider à identifier les facteurs clés d'une problématique, facilitant ainsi la compréhension des causes sous-jacentes et éclairant les interventions ciblées visant à l'atténuer.

2.5.3 Arbres boostés par le gradient

Les arbres boostés par le gradient, également connus sous le nom de "gradient boosting machines" (GBM), sont de puissants algorithmes d'apprentissage automatique qui appartiennent à la famille de l'apprentissage par ensemble [59]. Contrairement aux forêts aléatoires, qui construisent plusieurs arbres de décision de manière indépendante, les arbres boostés par le gradient construisent une séquence d'arbres de décision de manière itérative, chaque arbre se

concentrant sur l'apprentissage et la correction des erreurs commises par l'arbre précédent.

Les arbres boostés par le gradient sont similaires aux forêts aléatoires dans leur interprétabilité par l'importance des variables et par leur précision [60]. En effet, les arbres boostés par le gradient sont réputés pour leur grande précision de prédiction, surpassant souvent d'autres algorithmes d'apprentissage automatique dans divers domaines, faisant de ces modèles des méthodes de choix dans la littérature médicale [61]. En optimisant itérativement les performances du modèle par la minimisation des erreurs de prédiction, les arbres boostés par le gradient peuvent capturer des relations et des modèles complexes dans les grands ensembles de données, conduisant à des prédictions plus précises. À l'instar des forêts aléatoires, les arbres boostés par le gradient fournissent une mesure de l'importance des variables, soulignant la contribution relative de chaque variable prédictive à la tâche de prédiction.

Les arbres boostés par le gradient peuvent capturer des relations non linéaires entre les variables prédictives et le résultat. La flexibilité du gradient boosting lui permet de s'adapter à un large éventail de distributions de données et de saisir des nuances subtiles dans les données que les modèles linéaires peuvent négliger [59].

Les arbres boostés par le gradient utilisent des techniques de régularisation telles que le rétrécissement et les contraintes de profondeur de l'arbre pour éviter l'ajustement excessif et améliorer la généralisation du modèle. Cette régularisation permet d'atténuer le risque de complexité du modèle et de s'assurer que le modèle capture des modèles significatifs dans les données sans mémoriser le bruit ou les détails non pertinents [60].

2.5.4 Arbres de classification optimaux

En 2017, Bertsimas et al. proposait un nouveau modèle d'arbre comme alternative aux arbres de régression et de classification (CART) traditionnels, soit les arbres de classification optimaux [62]. Les arbres CART sont générés avec une méthode "top-down", c'est-à-dire que chaque noeud est généré itérativement en débutant par le premier noeud. Il a été suggéré que cette approche peut mener à des arbres sous-optimaux qui ne sont pas représentatifs des données. En effet, la sélection d'un noeud ne prend pas en compte les prochains noeuds et les prochains embranchements qui vont être générés en conséquence. Comme les arbres générés sont généralement très profonds, ils nécessitent souvent un émondage. Le dernier problème qui a été soulevé par les créateurs des arbres de classification optimaux est que les arbres de décision prennent souvent comme mesure principale l'impureté (souvent à travers le score de Gini) plutôt que la méclassification.

L'idée des arbres de classification optimaux est de traiter l'arbre comme un problème d'op-

timisation mixte, où tous les noeuds doivent être résolus en même temps. Ceci implique que les noeuds sont générés en tenant compte des noeuds subséquents.

Dans les dernières années, la capacité computationnelle des solveurs d'optimisation a vu une croissance importante, permettant d'utiliser des modèles d'optimisation sur des données d'envergure.

Lorsque comparé à des méthodes comme des arbres de classification et de régression ou des forêts aléatoires, les arbres de classification optimaux ont démontré une bonne performance prédictive qui rivalise avec celle des forêts aléatoires. Ceci est particulièrement vraiment pour les ensembles de données ayant une haute dimensionalité. Un arbre de classification optimal ayant la même performance prédictive qu'une forêt aléatoire serait donc supérieur dans son utilisation potentielle, en raison de sa grande interprétabilité. Dans un contexte médical, l'interprétabilité des modèles est une caractéristique cruciale pour la prise de décision et les arbres de classification optimaux présentent donc une méthode judicieuse [63]

CHAPITRE 3 DÉMARCHE DE RECHERCHE

3.1 Lacunes de recherche

Ce travail de recherche aborde la question cruciale de la congestion des services d'urgence, qui représente la phase de débit (*throughput*) du modèle de fluidité des services d'urgence. Alors que les études précédentes se sont principalement concentrées sur l'évaluation de la congestion à l'aide d'indicateurs au niveau individuel, tel les facteurs influençant le temps d'attente individuel d'un patient, il existe une lacune dans la compréhension du fonctionnement global des urgences. La congestion des urgences découle souvent de problèmes logistiques intra-hospitaliers [64], qui n'ont pas été pris en compte de manière adéquate dans les modèles prédictifs existants qui prédisent l'achalandage de l'urgence ou les admissions à l'hôpital.

Pour combler cette lacune, ce travail adopte une nouvelle approche. Au lieu de s'appuyer sur des proxys, un score de congestion est utilisé pour fournir une représentation plus précise et directe de la congestion des urgences. De plus, des variables logistiques sont incorporées comme variables de modélisation, reconnaissant l'importance des facteurs intra-hospitaliers dans la compréhension et la prévision de la congestion. Également, l'étude met l'accent sur la prédiction de la congestion au niveau institutionnel plutôt qu'au niveau individuel.

3.2 Nature temporelle du phénomène

La congestion aux urgences est un phénomène dynamique qui évolue constamment en raison de sa dépendance à divers facteurs temporels. L'achalandage des urgences fluctue de manière significative, variant de minute en minute, et est potentiellement impactée par plusieurs éléments temporels, tels que l'heure de la journée, le jour de la semaine et la saison.

Cette variation temporelle se manifeste également dans les différentes demandes des patients, qui évoluent en fonction de leurs besoins individuels. Par exemple, alors qu'un patient présentant une fracture peut nécessiter une radiographie, un autre patient avec une infection urinaire peut avoir besoin d'une analyse d'urine. Même si l'achalandage que ces deux patients représentent est le même, le poids sur la congestion de l'urgence est différent.

De plus, les ressources en personnel subissent des ajustements en fonction des quarts de travail, avec des horaires réduits à certains moments. De même, les ressources en matériel, comme certains départements, peuvent être fermées à certains moments, entraînant une variation dans la disponibilité des équipements médicaux. Cette fluctuation temporelle complexe

souligne la dynamique constante de la congestion aux urgences.

3.3 Prédiction

Afin de pouvoir mettre en place des mesures préventives lorsque nécessaire, une prédiction de la congestion dans 24h et dans 48h a été convenue. Notre hypothèse étant que la congestion à l'urgence est un phénomène qui se développe progressivement dans le temps, il est d'autant plus important de faire des analyses temporelles pour révéler des éléments précurseurs. Ces prédictions seront calculées à la fois pour une variable catégorique de congestion (score EDWIN > 1.50) ainsi que la valeur numérique du score de congestion.

Comme variables indépendantes de prédiction, notre étude se concentrera sur des variables qui représentent la demande de soins et les délais à fournir ces soins, afin de mettre de l'emphase sur les problèmes logistiques. Une liste exhaustive des variables intégrées est annexée à l'Annexe A.1.

3.4 Méthodes

3.4.1 Population

Il s'agit d'une étude rétrospective sur l'état de congestion de l'urgence du CHUM, en prenant comme variables d'études les différentes caractéristiques des cohortes de patients s'étant présentées à l'urgence ainsi que l'état du reste de l'hôpital.

Les données ont été recueillies du 1er juillet 2018 au 31 décembre 2021. Tous les patients ayant visité l'urgence du CHUM ont été considérés, ainsi que tous les patients ayant nécessité des soins durant cette période de temps. Il y a 70 000 visites par an à l'urgence du CHUM. Ainsi pour la période d'inclusion de l'étude, nous prévoyons une taille d'échantillon d'environ 315 000 visites de patients.

Les données ont été recueillies à partir de l'urgence du Centre Hospitalier de l'Université de Montréal (CHUM). Il s'agit d'un hôpital académique de soins tertiaires, qui connaît une population plus lourde en greffe, toxicomanie, troubles de santé mentale, neurologie et oncologie.

3.5 Sécurité et considérations éthiques

Les données sont stockées sur la plateforme CITADEL et sont accessibles sur les serveurs sécurisés de CITADEL depuis lesquels les codes informatiques sont produits et roulés. Puis-

qu'il s'agit d'une étude sur dossier médical et que l'accès aux données sera local au CHUM, nous ne demandons pas de consentement individuel de patient, mais bien un consentement à l'accès aux dossiers par le Directeur des Services Professionnels en conformité avec la LSSS et la loi 25. Bien qu'il y aura transfert des connaissances apprises par ce projet, aucune donnée patient n'a été partagée et seules des statistiques agrégées seront présentées.

3.6 Cohérence des travaux

L'article présenté à la prochaine section recoupe l'ensemble du travail de recherche pour les modèles de classification issus de librairies disponibles.

Des analyses complémentaires pour la régression ont été faites et sont disponibles au chapitre suivant, bien que ces analyses aient moins de potentiel d'impact et d'utilisation dans le système hospitalier.

Une analyse de classification pour les arbres de classification optimaux a été programmée de toute pièce et est discutée dans le chapitre d'analyses complémentaires. Cette analyse a une performance inférieure au modèle de librairie et n'a pas été incluse dans l'article.

CHAPITRE 4 ARTICLE 1: FORECASTING AND EXPLAINING EMERGENCY DEPARTMENT OVERCROWDING USING MACHINE LEARNING MODELS

Authors : Ilitea Kina; Louis-Martin Rousseau and Nadia Lahrichi

Submitted on May 24, 2024 to Health Care Management

4.1 Abstract

Emergency department (ED) overcrowding is a pressing global issue exacerbated by increasing demand and constrained resources, especially after the COVID-19 pandemic. Traditional approaches to assessing overcrowding often rely on proxies, such as the number of ED visits, admissions, or waiting time in the ED, rather than directly evaluating the balance between demand and resources. These proxies can also be used to predict overcrowding. While most studies in the literature use these proxies to predict overcrowding, this study presents a novel approach. By extending the use of a well-established overcrowding score into a dynamic model, we model and predict overcrowding based on real-time logistical delays in the hospital setting. Leveraging machine learning models, we predict overcrowding 24 and 48 hours ahead, achieving high accuracy rates of 90.1% and 89.9%, respectively. Additionally, incorporating variable-specific lags through a crosscorrelation function significantly improves model performance. Our analysis reveals that past overcrowding is a major factor influencing future overcrowding events. Additionally, the explanations for overcrowding, as revealed by SHAP values and counterfactual analysis, vary significantly over time. This dynamic nature suggests that targeted solutions, based on the root causes identified by our real-time analysis, are crucial for effective overcrowding management. We present a predictive tool coupled with explainable AI methods, offering insights into daily overcrowding causes and assisting administrators in resource allocation and planning. By addressing a critical gap in predicting overcrowding directly and providing actionable insights, our study enhances ED management practices and ultimately could improve patient care outcomes.

Keywords : Overcrowding, Emergency department, Prediction, Machine learning

Highlights

- We propose ML models that predict overcrowding, quantified through the validated EDWIN overcrowding score, using real-time logistical features, with good accuracy (90.1% at 24 hours and 89.9% at 48 hours)
- We propose variable-specific lags for each feature, established through a crosscorrelation function, increasing accuracy by up to 4.6% and increasing the F1-score by up to

7.5%

- We use feature importance to illustrate the important inertia in the ED, with past overcrowding the main factor in predicting future overcrowding
- We leverage SHAP Values and counterfactuals to explain the daily causes of overcrowding

4.2 Introduction

Overcrowding of emergency departments is a worldwide problem that has persisted for decades. Following the COVID-19 global pandemic, a massive shortage of acute care healthcare workers continues to impact the emergency departments. The situation has worsened in Canada over the last few years, with emergency medicine associations decrying the unacceptable situations the patients face in the ED [1]. Overcrowding is defined as a state where the demand for healthcare services surpasses the capacity of a healthcare institution to provide these services [2]. In recent years, the increase in demand has faced reduced resources, a precarious situation calling for optimization efforts.

In medical literature, extensive evidence shows that delays in accessing care due to overcrowded emergency departments lead to higher mortality and morbidity rates [5, 6]. This is particularly critical for patients with life-threatening conditions like myocardial infarction [8, 10], cardiopulmonary arrest [12], or traumatic injuries [?]. Additionally, even for patients with other critical illnesses, each extra hour spent in the emergency department before admission to intensive care units increases their mortality risk by 1.5% [13]. This consequently increases the workload of hospitals. On the one hand, overcrowding in emergency departments leads to increased demand; on the other hand, it leads to a depletion of resources. Research indicates that ED overcrowding not only diminishes staff morale but also heightens anxiety among healthcare workers. Additionally, ED overcrowding diminishes patient satisfaction with care, affecting their experience during their ED visit and their satisfaction throughout the remainder of their hospitalization [7, 17].

Most studies rely on proxies to assess overcrowding. These include the volume of emergency department visits, admissions, length of stay in the ED, and waiting time [28]. Overcrowding in itself is seldom predicted. While these measures often appear interchangeable, overcrowding hinges on the delicate equilibrium between demand and available resources and is, therefore, significantly caused and affected by logistical problems [64]. Conversely, the number of ED visits is unaffected by intra-hospital logistics and measures only demand that precedes temporally ED overcrowding. The likelihood of hospital admission depends more on a patient’s clinical condition than on logistical aspects and represents the pressure on the hospital rather

than the ED. In this study, we use the EDWIN overcrowding score, a validated measure in medical literature, to represent the balance between resources and demand on the emergency department. By leveraging the ratio of patients in the ED compared to doctors and available beds, this score allows us to evaluate overcrowding and tackle the problem objectively, providing a comprehensive assessment of ED workflow dynamics. This score correlates with staff frustration, enabling us to address overcrowding directly and concentrate on its impact on healthcare resources.

Our study makes multiple key contributions to real-time prediction and explanation of ED overcrowding. First, we develop machine learning models trained on real-time logistical features from the ED and other hospital departments to predict overcrowding as measured by the validated EDWIN score. These models achieve high accuracy rates of 90.1% and 89.9% for 24 and 48-hour forecasts respectively. Second, we incorporate variable-specific time lags identified through cross-correlation analysis, significantly improving prediction accuracy, up to 4.6%, and F1-score, up to 7.5%. Through feature importance analysis, we highlight the significant impact of past overcrowding on future overcrowding, shedding light on the inertia within the hospital environment. Employing Explainable AI methods, we elucidate the day-to-day variability in the causes of overcrowding, providing valuable insights for decision-makers. Finally, our development of a predictive tool for overcrowding in the next 24 and 48 hours, coupled with explanations of daily contributing factors, equips administrators with a practical resource for proactive management and decision-making.

This paper is structured as follows. Section 3 describes the knowledge surrounding the causes of overcrowding, the variables targeted in prediction models, and the overcrowding scores developed. Section 4 explains our methodology, focusing on the different lags applied to the variables, the various prediction models tested, and their global feature importance. Section 5 compares local predictions and interpretation through explainable AI models for random overcrowded moments. Finally, section 6 discusses the results and the considerations for hospital applications.

4.3 Literature review and related work

Many studies have attempted to model the workflow in emergency departments (EDs) and investigate the various factors contributing to overcrowding. These studies include efforts to model the flow dynamics within EDs, identify the different causes of overcrowding, and predict key measures within the ED ecosystem, such as waiting time, length of stay, ED visits, and admissions. Furthermore, researchers have created scoring systems designed to accurately measure the severity and impact of overcrowding. In this review, we will first outline the

known causes of overcrowding in the ED, then compare them to different variables used in prediction models that use proxies to measure overcrowding, and finally present overcrowding scores developed to quantify overcrowding.

Causes of overcrowding The flow in the ED is often described as an input-throughput-output model [3] as represented in Figure 4.1. A patient enters the ED through registration and triage (input). The patient is then seen by an emergency physician, investigated, and treated in the ED (throughput). The patient then exits the ED either through admission to a hospital bed or discharge (output). As illustrated in Figure 4.1, the flow model emphasizes the ED’s lack of control over *input*, which precedes ED management. Furthermore, the flow model highlights how an inability to achieve *output* has the potential of significantly expanding the workload in the phase *throughput* by continuing investigations and treatments in the ED. Logistics of *throughput* and *output* phases then become key factors in ED crowding.

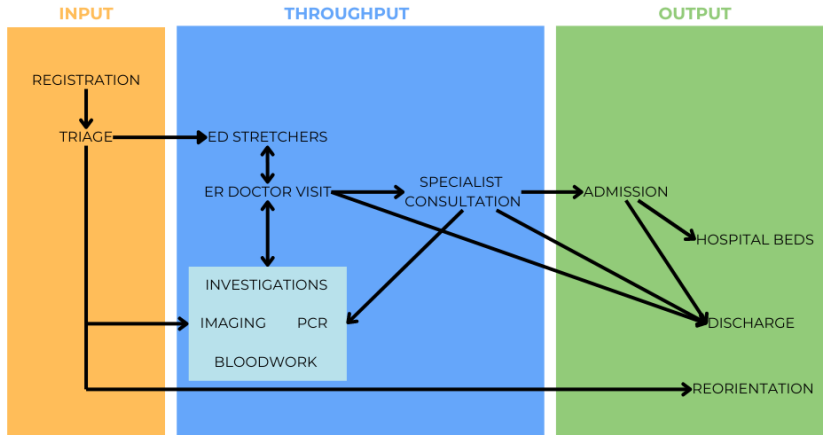


FIGURE 4.1 A representation of the flow in the ED. Patients can exit the ED by either being reoriented by a triage nurse to a walk-in clinic, being discharged by the ER doctor or a specialist consulted, or being admitted to the hospital. While in the ED, patients continue to require ED resources, including ED stretchers and investigations.

Over the years, the patient population of the ED has become older and medically more complex, requiring more meticulous care and inpatient resources [4, 24]. The hospitals’ lack of ability to adapt to this change stems from logistic issues. Delays in obtaining diagnostic results are believed to prolong patient stays in the ED, exacerbating overcrowding [25]. Hospital occupancy rates further compound the problem by reducing bed availability, thereby increasing ED length of stay [26]. Lastly, access block, where patients cannot transition to

necessary care structures such as operating rooms or specialized interventions, contributes significantly to overcrowding by impeding patient flow [27]. These factors are logistical.

In the following, we focus on the most popular overcrowding proxies used in prediction models (i.e. input, throughput and output), and how the features used in these accurate prediction models do not necessarily fit the known causes of overcrowding.

Forecasting input Previous studies explore various predictive modeling approaches utilizing factors such as calendar events, weather conditions, and the availability of healthcare resources to forecast future ED visits accurately. Using only weather conditions and calendar data as features, models including Long Short-Term Memory networks, Gated Recurrent Units, and simple Recurrent Neural Networks can predict daily ED visits with high accuracy (94%) [30]. By incorporating more out-of-hospital data such as calendar events, loads of secondary care facilities, and local public events, simpler methods, including ARIMA and recursive least squares, show good accuracy in predicting ED visits for the next day [31]. These studies demonstrate that daily ED visits are a predictable result of prehospital and out-of-hospital data sources, not intra-hospital logistics. While many ED visits could theoretically contribute to the phenomenon of overcrowding, these two measures do not describe the same problem.

Forecasting throughput Numerous studies focus on predicting waiting times to see an emergency department doctor and the length of stay (LOS) in the ED, demonstrating promising accuracy in forecasting these metrics. However, a common observation across these studies is the reliance on individual-level factors, such as patient characteristics and clinical attributes, rather than ED logistics in making these predictions. One study successfully predicts a patient’s length of stay in a pediatric emergency department using clinical data such as age, sex, diagnosis, and the need for further investigations such as laboratory studies, radiographs, ultrasounds, and scanners [35]. Similarly, another study predicts a patient’s length of stay in a general ED by looking at their age, sex, triage category, mean of arrival, and time of arrival [36]. Other studies have successfully predicted how long a patient needs to wait before seeing an ED doctor by looking at the number of patients in the ED and the number of doctors available [37, 38]. While these models show encouraging results in understanding and anticipating individual patient experiences, they do not capture the impact of logistical delays on ED overcrowding at an institutional level.

Forecasting Output Utilizing machine learning techniques, multiple studies have leveraged electronic medical records to forecast the likelihood of admissions within specific time-

frames. These models typically integrate clinical data sources, including symptoms, age, sex, and triage acuity, to inform predictions [32]. While techniques such as logistic regression, random forest, and gradient boosting have been proven to yield good accuracy, the integration of text data from the triage notes [33] has increased accuracy for a prediction in 24 hours. Another model focused on predicting the admissions for all ED patients, using their live electronic health records and XGBoost classifiers, to predict the number of admissions [34]. Models that predict a patient’s likelihood of being admitted to the hospital mainly focus on the clinical signs that participate in the physician’s decision to admit a patient to a hospital ward. These studies do not capture how the admissions impact the hospital flow or vice versa.

General overcrowding scores Because proxies do not reflect the phenomenon of overcrowding, multiple overcrowding scores have been developed in medical literature to assess overcrowding objectively. Scores like NEDOCS, with derivatives like PEDOCS for pediatric EDs and CEDOCS for community hospitals, have been used in literature in recent years [39–41]. The EDWIN score was described in 2003 as an objective assessment of overcrowding in the ED. It compares the demand in the ED, related to patients and their acuity, to the resources, including doctors and stretchers in the ED. It is modeled based on the nurse and physician assessment of overcrowding with good interrater agreement [42]. Compared to the NEDOCS score, both have a good AUC for predicting overcrowding, with NEDOCS being marginally better but utilizing data not readily available [43]. In a discrete-event simulation comparison between NEDOCS, EDWIN, and the READI score, the EDWIN score has a good measure of ED crowdedness. It quickly identifies changing ED conditions [44]. This score is shown to be correlated to reduced patient satisfaction [45] and to the feeling of frustration of ED workers [46]. An increased EDWIN score is also correlated with an increased frequency of medical errors in the ED [47] and an increased number of patients leaving the ED without being seen [48], denoting a reduced quality of care related to this score.

Research gap While the features affecting ED visits stem from out-of-hospital phenomena, such as weather conditions and calendar events, and the features affecting admission stem from clinical phenomena, such as the patient’s symptoms, vital signs, and medical history, overcrowding is a logistical problem, likely impacted by logistical factors. While the sum of the patient’s length of stay might translate to overcrowding, literature is often focused on predicting an individual’s length of stay due to their clinical attributes. As summarized in Table 4.1, our understanding of overcrowding is that it is a phenomenon impacting the *throughput* phase caused by intra-hospital logistics and directly impacting the global flow of the ED. These elements are in green in Table 4.1, with some of these elements found in the

studies that attempt to predict overcrowding, but not all. The different studies conducted on proxies of overcrowding do not focus on intra-hospital logistics and studies focusing on *throughput* predict the impact of logistics on individuals instead of the institution. In itself, overcrowding is seldom described in literature through a validated overcrowding score, and the logistical factors that cause overcrowding are rarely used as predictors in prediction models. To align with our understanding of overcrowding, our methodology employs a scoring system incorporating logistical rather than clinical features, as logistical factors are the primary drivers behind its occurrence.

4.4 Methodology

Considering the limitations of existing research, which often focuses on individual patients or proxies for overcrowding, our methodology is specifically designed to address overcrowding on an institutional level. We achieve this by utilizing a time series analysis. This approach allows us to analyze logistical factors, the true drivers of overcrowding within the ED, as predictors. Furthermore, we employ a quantitative overcrowding score, the EDWIN score, to assess this phenomenon objectively. This combined approach aligns with our understanding of overcrowding as a logistical issue impacting hospital throughput.

The methodology is structured as follows : we cleaned the data and determined a cut-off value for classification according to the literature. We used a crosscorrelation function to apply different lags on the variables and find the best lag for each variable. We converted this data into time series and fitted five different models using a sliding window approach. This leads to global interpretations for each model. We selected the best model and made local interpretations for random overcrowded moments. This approach is described in Figure 4.2. In what follows, we describe each step in detail, beginning with the methods used to clean the data and create time series.

4.4.1 Data preparation

This section outlines the steps for data preparation to analyze factors contributing to emergency department (ED) overcrowding. Here, we detail the process of transforming raw patient data into a format suitable for time series analysis. Time series datasets capturing real-time and lagged information about service delays were created. Finally, the established EDWIN score was employed as the dependent variable, representing the level of ED overcrowding.

Table 4.1 Summary of the characteristics of the theoretical ideal modelization of over-crowding versus models in literature

Measure	Flow model phase	Features	Level
<i>Theoretical understanding of overcrowding</i>			
Overcrowding score	Throughput	Intra-hospital logistics	Institutional
<i>Literature review of overcrowding</i>			
ED visits	Input	Weather, extra-hospital logistics Number of ED physicians	Institutional
Waiting Time	Throughput		Individual
ED LOS	Throughput	Clinical features, Testing needs	Individual
Admissions	Output	Clinical features	Individual

Data cleansing All the patients present in the CHUM between July 2018 and December 2021 were included. For the variables related to the emergency department, only patients who were registered for the ED were included. Patients who were registered to be directly admitted by different specialties were excluded. The predictors chosen are calendar data and services required in the ED and comprise the more significant categories of imaging, laboratory testing, nursing and physician visits, consultations, admissions, and hospitalizations. The variables selected are specified in Table A.1. The remaining data was then cleaned by removing duplicate values, filling missing values with similar data entries, and removing values related to departments other than the ED. Finally, a correlation matrix was obtained to identify and remove the correlated variables.

The time between the request of a service and its completion was calculated as the delay to complete a service and converted to seconds. Negative values were removed, and excessively long values corresponding to the definition of an outlier point were removed. This created a loss of less than 4% of the data points overall.

EDWIN score - the dependent variable The EDWIN score [42] is an overcrowding criterion. The scoring formula is as follows :

$$\sum \frac{n_i t_i}{N_a(B_T - B_A)}$$

Where n_i = number of patients in the triage category i , t_i = weight associated with the

triage category i (P1, most acute, = 5, P2 = 4, P3 = 3, P4 = 2, P5, least acute, = 1), N_a = number of emergency physicians available, B_T = number of stretchers in the ED, B_A = number of stretchers in the ED occupied by an admitted patient. The readily available data needed to calculate the EDWIN score, alongside its established alignment with healthcare worker experiences of overcrowding, made it a suitable choice as the objective measure for this analysis [42].

The dependent variable was calculated using the EDWIN score formula without any modification in how the different variables were calculated. All the patients who were physically present in the emergency department were included. As described in the validation study of the score [42], we chose a cut-off value of 1.50 to separate between not-crowded and busy or overcrowded moments. We divided the classification into two categories, merging average and high crowding into one crowded category.

Time series The cleaned data was converted into time series to capture the temporal dynamics and trends related to overcrowding. For the dependent variable, the EDWIN score was calculated every 15 minutes. For the predictors, the number of pending services (e.g., number of x-rays pending) and the mean delay for the pending services were calculated every 15 minutes. We also integrated the numerical and categorical EDWIN scores to represent overcrowding 12, 24, and 48 hours ago in the predictors described in Table A.1. This created a first-time series of independent and dependent variables, with no lag that we will refer to as T_0 . Two other time series were created by applying a fixed lag on the dependent variable of 24 hours, making T_{24}^{FIX} , and of 48 hours, creating T_{48}^{FIX} . Creating these lagged time series raises the question : Do different factors influencing overcrowding have distinct moments of peak influence? In other words, is the time delay between a specific factor (independent variable) and overcrowding (dependent variable) unique for each factor? We employ the crosscorrelation function in the next section to explore this question.

4.4.2 Variable specific lags

We investigated crosscorrelations between the independent and dependent variables to improve the fixed lag time series. This technique allows us to identify the time lag at which each independent variable exhibits the strongest correlation with the dependent variable (ED overcrowding score). We compared the correlation time for each independent variable with the dependent variable through the crosscorrelation function with the following formula :

$$\rho_k(x, y) = \text{corr}(x_{i-k}, y_i) = \text{corr}(x_i, y_{i+k})$$

with the correlation function r

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

Where :

1. x and y represent an independent variable and the dependent variable
2. y_i is the value of the time series at datapoint i
3. n is the number of data points in the time series
4. k is the delay inflicted on the time series
5. y_{i+k} is the value of the time series at data point i + the delay inflicted

Using the highest peak of correlation for each variable, with a minimum of 24 hours and a minimum of 48 hours in advance, we created two new time series also used for model fitting.

The datasets used for fitting the models are therefore :

1. No lag - T_0
2. 24 hours fixed lag for all variables - T_{24}^{FIX}
3. 48 hours fixed lag for all variables - T_{48}^{FIX}
4. 24 hours variable-specific lag - T_{24}^{CCF}
5. 48 hours variable-specific lag - T_{48}^{CCF}

4.4.3 Modeling

In the interest of predicting overcrowding in 24 hours and 48 hours, a cross-validation method with 10 folds was used with a sliding window approach [65], which consists of taking a window of consecutive data representing 10% of the overall data for training and testing the model generated on the following 2% of the successive data. This ensures that the model is tested on data that the model has yet to see during training and that is not temporally before the data on which the model was trained. Multiple prediction models have been trained, beginning with logistic regression.

Logistic regression was used as a baseline comparator for the machine learning models. This statistical method uses the following formula ¹

$$\text{logit}(p) = \ln \left(\frac{p}{1-p} \right) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \epsilon$$

1. implemented in R, using the library `caret` with the method "glm" and the family "binomial".

Classification trees² were the first machine learning model trained in our tree-based model comparison. Classification trees iteratively select the cut-offs that best discriminate between a high likelihood of being overcrowded and a high likelihood of being not overcrowded [49].

Random forests were fitted with our data. Random forests use partial data in the form of bagging and feature randomness to generate multiple trees [54].³

Gradient-boosted classification trees were investigated. Gradient boosting uses the misclassified data to fit future trees and create a gradient-boosted forest [59].⁴

Optimal classification trees were proposed for the first time in 2017 [62] to increase a classification tree's accuracy but preserve its interpretability. This method builds a classification tree by solving an optimization problem instead of generating nodes iteratively. This implies that instead of finding local minima iteratively, the optimal classification tree finds a global minimum, like classic decision trees.⁵ For all the models, the same evaluation criteria were applied.

For performance, two primary metrics are used : accuracy and F1-score. Since the data was split into ten folds, the mean value of the ten folds was used. We chose accuracy as an intuitive measure to compare the models' overall performance. We decided on the F1 score as defined by the following formula :

$$2 \times \frac{precision \times recall}{precision + recall}$$

The F1-score is an essential metric to balance the risks of false negatives (underestimating overcrowding and being less prepared) and false positives (overestimating overcrowding and overusing resources).

4.4.4 Local interpretation

In addition to extracting the importance of features for the trained models and identifying which variables are likely to contribute to overcrowding on a global level, we also investigated local interpretation. This means we must analyze the data more granularly to fully understand the factors driving overcrowding on a given day and time. As there is significant variability in the delays in care stemming from multiple departments, punctual explanations could vary daily. To investigate explanations, we start by describing SHAP values.

2. In R, the library `caret` was used with the method "rpart2" to fit a classification tree.

3. The method "ranger" was used in the library `caret`

4. The method "gbm" was used in the library `caret`

5. The library `IAI` was used in R.

SHAP values, short for SHapley Additive exPlanations, seek to find the marginal contribution of each variable in the prediction made [66].

The following formula is used :

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! \times (|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)]$$

Where :

1. ϕ represents the Shapley value for variable i
2. N represents the set of variables, with $|N|$ being the total number of variables
3. S represents a subset of variables, excluding the variable i , with $|S|$ being the number of variables in subset S
4. $f(S)$ represents the prediction of the model using the variables in S
5. $f(S \cup \{i\})$ represents the prediction of the model using the variables in S and i .

The higher the absolute SHAP value, the more significant its contribution to the prediction. Knowing the distribution of the SHAP values allows us to know which variables contribute to the prediction made for a local data point. To investigate further what modifications should be applied, we evaluated counterfactuals.

Counterfactuals seek to find the minimal change needed to the features to be able to make the opposite prediction [67]. They represent the minimization of a multi-objective loss function. The formula used is the following :

$$L(x, x', y', X^{obs}) = (o_1(\hat{f}(x'), y'), o_2(x, x'), o_3(x, x'), o_4(x, X^{obs}))$$

With each objective representing either having the desired prediction (o_1), minimizing the change in values (o_2), minimizing the number of variables changed (o_3), and choosing realistic values compared to the data set (o_4). In our context, counterfactuals represent what features need to change to predict that the ED will not be overcrowded if the prediction is heading for overcrowding.⁶ With this logic, non-actionable features were removed, including past EDWIN scores and features related to dates, as those features cannot be changed.

6. Counterfactuals resulted from using the "MOCClassif" function of the MOC library in R.

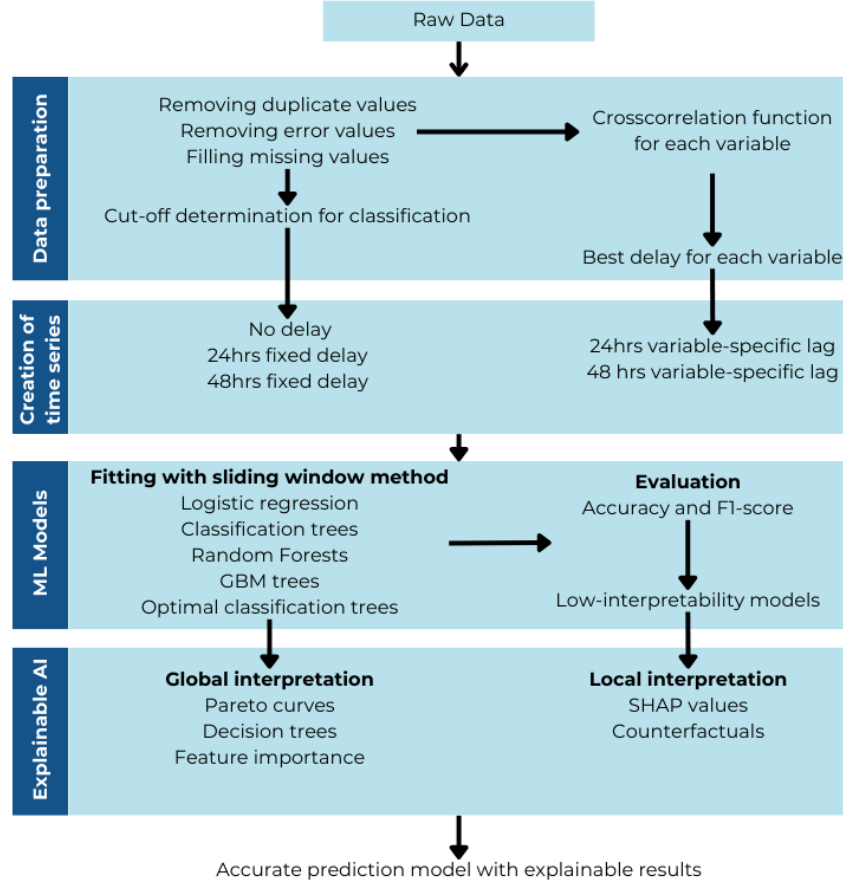


FIGURE 4.2 A summary of the methodology used in this study

4.5 Results

We applied the methodology described to the CHUM data, an academic tertiary center in Montreal. Our data had lower EDWIN scores than the validation study of the EDWIN score, with a median of 1.24 instead of 1.50. Separating the data into not-overcrowded and crowded categories resulted in a not-overcrowded ED 64% of the time, representing relatively balanced data. We cross-correlated this dependent variable to the independent variables to assess optimal lags.

4.5.1 Variable specific lags

We generated crosscorrelation graphs for each independent variable. The graphs have a sinusoidal shape for all variables except day, month, and hour. In Figure 4.3, we compare the CCF values in a 96-hour time frame for the number of consultations pending and the mean delay for admissions. Each graph has four peaks over the four days. The cycle between

the two peaks is precisely 24 hours, evoking a seasonality in the variables. Consultations, admissions, hospitalization teams, and the radiology department work with reduced staff in the evenings and nights. Most hospital departments function in cycles of 24 hours, which explains the seasonality of the CCF graphs. Moreover, the peaks in the graphs do not occur at the same time. The maximum correlation between 24 and 48 hours earlier is at -46 hours for consultations and -44 hours for admissions. This demonstrates that each variable has a specific lag that best correlates with the outcome variable.

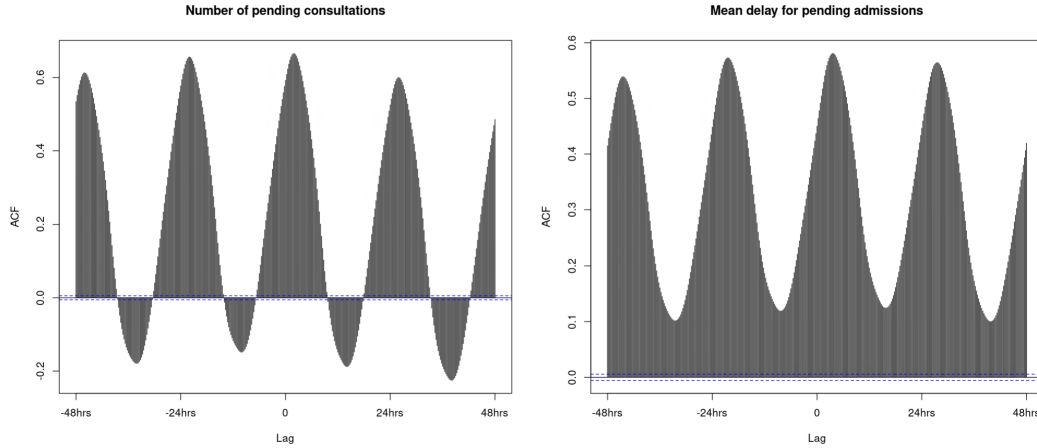


FIGURE 4.3 Comparison between crosscorrelation function results in a 96-hour time frame for pending consultations (number) and pending admissions (delay)

4.5.2 Models and performance

We applied the generated time series to the five different models discussed.

A Pareto curve was extracted following the model fitting for the logistic regression model, showing the ten most essential coefficients. The Pareto curve for the prediction model 48 hours ahead using variable-specific lags is shown in Figure 4.4. Some critical features to explain variance are the numerical past EDWIN scores 12, 24, and 48 hours ago, the occupancy of ED stretchers, LOS in medical and OBGYN beds (hospitalization units), and calendar data.

Fitted classification trees were produced for the five different time series. The tree predicting overcrowding 24 hours ahead is visualized in Figure 5.2, using variable-specific lags. The first and second levels are nodes related to the EDWIN score 24 hours ago. The first node sets a cut-off value of 1.5, which perfectly matches the theoretical cut-off chosen to split our data. The first node effectively represents the overcrowding class 24 hours ago. The second level doubles down on the overcrowding score, with a 96% chance of being non-overcrowded in 24

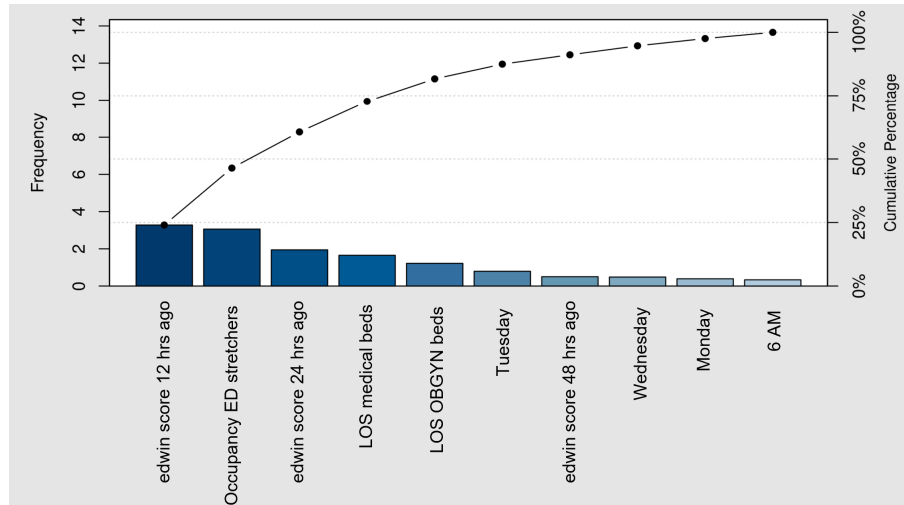


FIGURE 4.4 Pareto curve showing the ten variables having the most impact on variance in the logistic regression model using variable-specific lags forecasting overcrowding in 48 hours

hours if the overcrowding 24 hours ago was low (i.e., < 1.2) and a 94% of being overcrowded in 48 hours if 24 hours ago the overcrowding was high (i.e., > 1.8). The factors increasing the likelihood of overcrowding for the cases in between are an increased consultation delay, increased LOS in surgical and gynecology beds, and increased number of admissions, showing the importance of incorporating logistical features in the prediction model.

We trained random forests and gradient-boosted classification trees for all the time series. In Figure 5.4, the feature importance of the ten most important features is illustrated for the models using variable-specific lags for a prediction in a minimum of 24 and 48 hours. In both models, the EDWIN score 24 hours ago remains the most crucial feature, with EDWIN scores 12 and 48 hours ago, length of stay in surgical beds, and delay to see an emergency physician being present in both graphs. Compared to the feature importance of the random forest model, the most important features remain globally similar, with the EDWIN scores having an increased importance in GBM models.

Finally, we fitted optimal classification trees. The tree for a prediction in 24 hours using the variable-specific model represented in Figure 4.7 shows again the importance of past overcrowding scores and occupancy of surgical and medical beds. The hour of the day was also a determinant factor, which likely encompasses the reduced resources available after 5 PM in most departments, including radiology, consultation services, and reduced personnel for admissions and ED services.

Tables 4.2 and 4.3 summarize the performance of the five models trained on the five different time series.

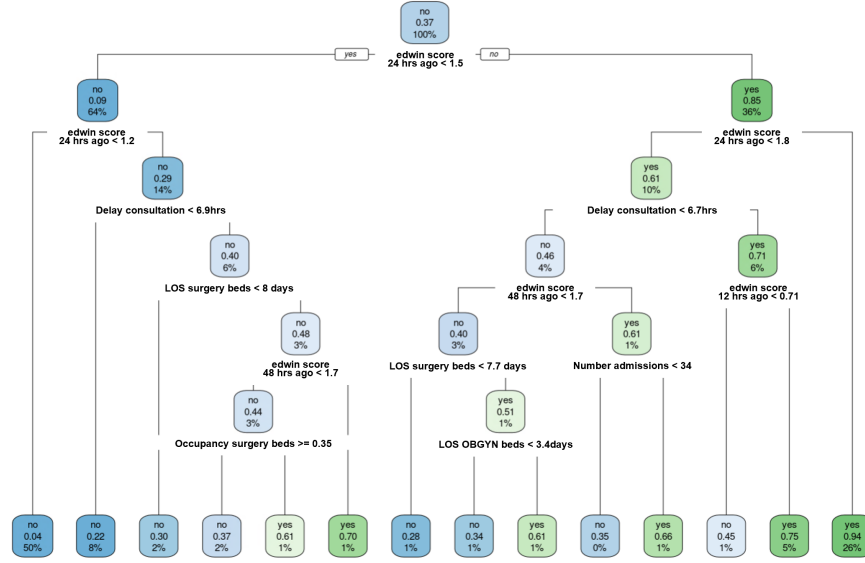


FIGURE 4.5 A classification tree forecasting overcrowding in 24 hours with a variable-specific lag

TABLE 4.2 Comparison of performance between ML models for classification

Forecasting	Accuracy				
	LR	CART	RF	GBM trees	OCT
No lag - T_0	0.943	0.896	0.926	0.930	0.887
24hrs fixed lag - T_{24}^{FIX}	0.881	0.864	0.893	0.898	0.815
48hrs fixed lag - T_{48}^{FIX}	0.848	0.848	0.874	0.869	0.816
24hrs CCF lag - T_{24}^{CCF}	0.881	0.888	0.895	0.901	0.861
48hrs CCF lag - T_{48}^{CCF}	0.884	0.883	0.890	0.899	0.829

Performance comparison The most accurate model was the gradient-boosted trees, followed by the random forests. Gradient-boosted trees use fewer trees than random forests, representing a computational advantage.

Although both these models perform better, they are less interpretable than the classification and optimal classification trees. CART's and OCT's predictive performances are lower, but their interpretability and clear cut-offs for different variables make these models highly interesting for practical purposes.

Crosscorrelation function Using variable-specific delays through a crosscorrelation function has improved the accuracy and F1-score for all models for 24 and 48 hours. At 24 hours, the most significant increase was for the OCT model, with a 4.6% increase in accuracy, and

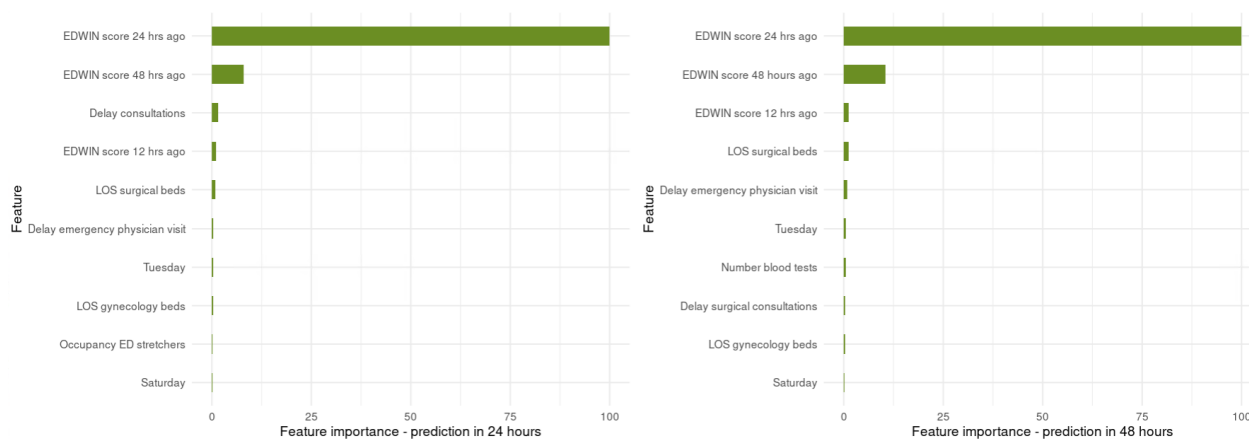


FIGURE 4.6 Feature importance of variable-specific GBM models for predictions in 24 hours and 48 hours

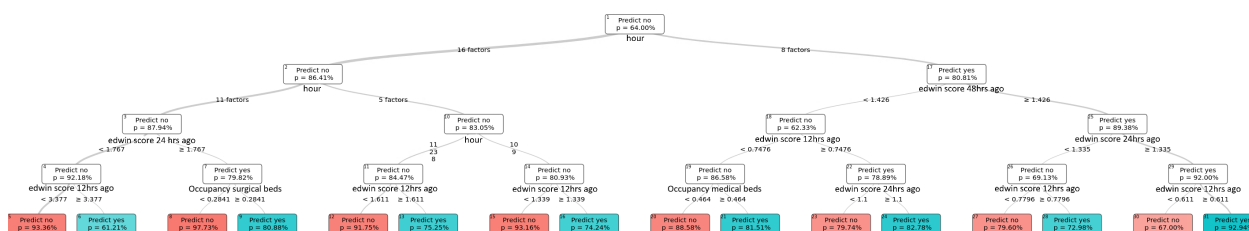


FIGURE 4.7 Optimal classification tree using variable-specific lags to predict overcrowding in 24 hours.

TABLE 4.3 Comparison of performance between ML models for classification

Forecasting	F1-score				
	LR	CART	RF	GBM trees	OCT
No lag - T_0	0.911	0.837	0.879	0.892	0.766
24hrs fixed lag - T_{24}^{FIX}	0.799	0.777	0.817	0.832	0.696
48hrs fixed lag - T_{48}^{FIX}	0.728	0.753	0.783	0.781	0.680
24hrs CCF lag - T_{24}^{CCF}	0.803	0.819	0.824	0.841	0.715
48hrs CCF lag - T_{48}^{CCF}	0.803	0.818	0.817	0.839	0.665

for the CART model, with an F1-score increase of 4.2%. For the 48-hour prediction, the best increase was for the logistic regression model, with a rise in 3.6% in accuracy and an increase of 7.5% in the F1-score.

4.5.3 Local interpretation

The SHAP values were calculated for variable-specific 24-hour and 48-hour lags for the best model, which was the GBM model.⁷ We chose to present the SHAP values of two random overcrowded moments in Figures 4.8 and 4.9. While a few of the same variables are present in both graphs, it is clear that their values are different depending on the time. In Figure 4.9, the delay for general medicine admissions is one of the most positive features, while in Figure 4.8, the SHAP value for this feature is negative.

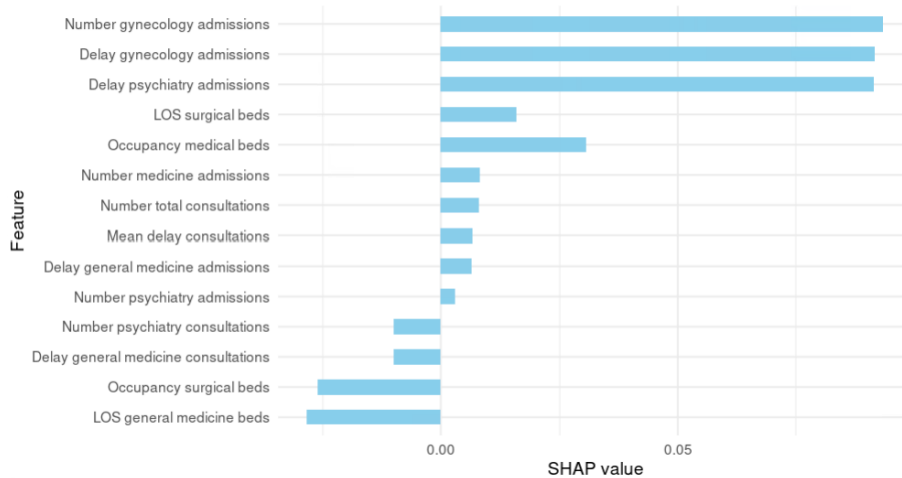


FIGURE 4.8 Highest SHAP values for 24 hrs forecasting using the variable-specific GBM model for a random overcrowded moment (August 13th, 2021, at 1 :15 AM)

7. We used the "Shapley" function of the `iml` library to calculate SHAP values in R.

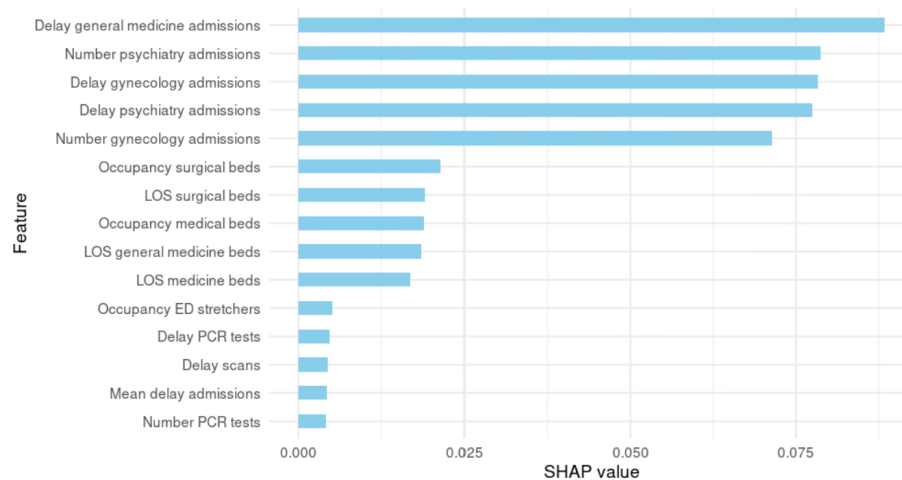


FIGURE 4.9 Highest SHAP values for 24 hrs forecasting using variable-specific GBM model for another random overcrowded moment (September 5th, 2021, at 9 PM)

Furthermore, while trending the SHAP values over 120 hours, there is a lot of variability in time and between the different variables, as seen in Figure 4.10. The SHAP values compared for five different independent variables show not only variability over time for each variable but also variations between variables, with some values increasing, decreasing, or peaking simultaneously for the different variables.

Using the same random overcrowded moment as for the SHAP value graphs, we calculated Counterfactuals, keeping the top 15 features that could switch the prediction if they were changed. The most essential features for the first random moment are the delay in PCR tests, surgical consultations, medical admissions, and the number of scans pending, as seen in Figure 4.11. The delay in medical admissions is still essential for another random moment represented in Figure 4.12. Still, the LOS in surgical beds, the occupancy of surgical beds, and delays in consultations and ultrasounds are new top counterfactuals that were not as important in the first random moment. This shows that counterfactuals vary depending on the moment, and the same modification would likely not have the same impact on different causes. Managers could use the insight on the other models and explanations to prevent overcrowding.

4.6 Managerial Insights

At the source of this model is the aim to better overcome overcrowding and prevent it while simultaneously using as few resources as possible.

We propose consolidating the prediction models and explainable AI methods into an easily

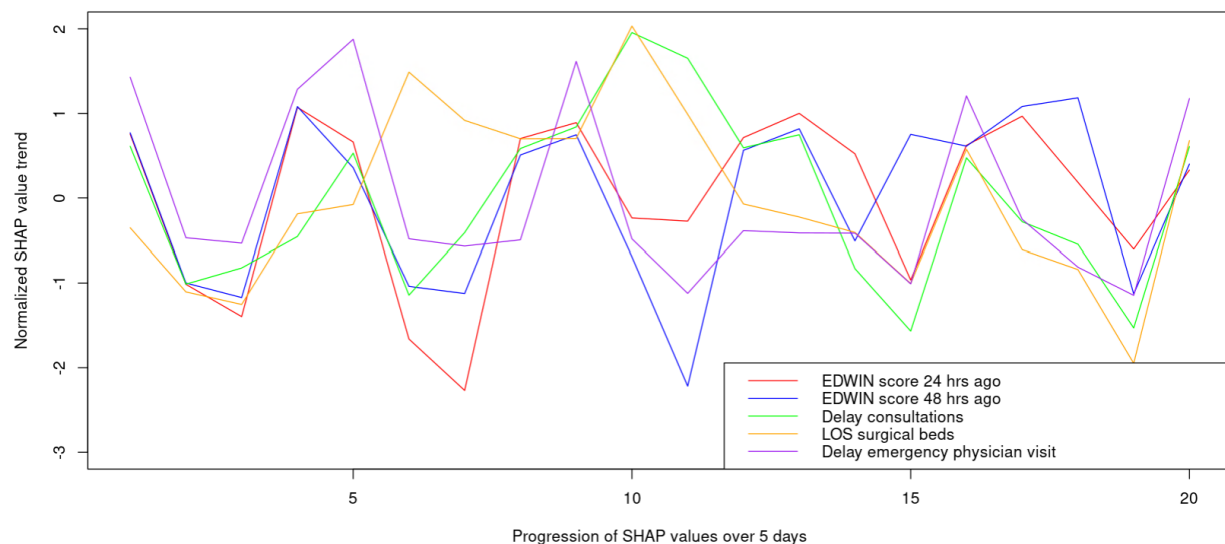


FIGURE 4.10 SHAP values trended every 6 hours over 120 hours for five key features

interpretable virtual dashboard for hospital managers to predict overcrowding, identify potential influential factors, and prevent overcrowding. This virtual dashboard is illustrated in Figure 4.13. This dashboard illustrates real-time forecasting of the likelihood of overcrowding in 24 and 48 hours. It incorporates the SHAP values, representing what variables contribute to the forecasting. Counterfactuals, representing what variables need to be modified to have the opposite prediction, are represented for each forecasting.

With this accurate and dynamic information, managers could implement short-term and long-term changes to prevent overcrowding in the ED. The short-term changes include modifying staffing levels and resource allocations.

As overcrowding can be predicted with high accuracy, the staffing levels could also be adjusted 24 hours or 48 hours in advance to fit the crowding levels. Compared to usual staffing, this could result in understaffing with economic benefits and overstaffing with clinical benefits [68]. In most EDs, nursing overtime is determined on a much shorter notice than 24 and 48 hours in advance. This also represents an opportunity to improve staff satisfaction with scheduling.

With predictions explaining which variable or department is predicted to contribute to overcrowding in 24 and 48 hours, there is an excellent opportunity to allocate resources specifically to these targeted departments. Depending on the predictions, a float team comprising float nurses, personal support workers, and hospital porters could be dedicated to targeted help. Over time, short-term modifications could reveal patterns in the departments targeted.

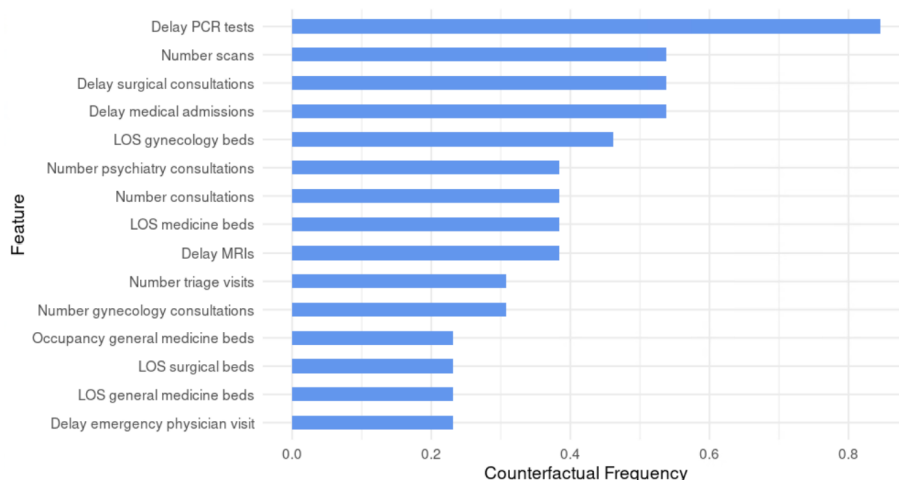


FIGURE 4.11 Highest Counterfactual values for 24 hrs forecasting using GBM model for a random overcrowded moment

Long-term changes could then also be made to prevent overcrowding.

Some features are globally more critical in their impact on overcrowding. These variables represent actions by teams that might need to be better supported in their role or investigated to find bottlenecks and specific problems to resolve. Highlighting the key issues would guide the following steps to tackle the chronic obstacles.

Overall, the tool could implement a culture shift in a few different ways. Using an objective measure specifically for overcrowding could help identify overcrowding through continuous monitoring and assess the results of implemented solutions. Using targeted interventions instead of hospital-wide overcrowding signals could increase the effectiveness of interventions. Finally, it could influence a mindset change where clinicians could position themselves as preventing overcrowding instead of reacting to overcrowding.

4.7 Conclusion

In conclusion, our retrospective study, conducted on Centre Hospitalier de l'Université de Montréal data, aims to predict ED overcrowding 24 and 48 hours in advance. Leveraging logistical delays within the ED to obtain services and utilizing the validated EDWIN score as an objective measure of overcrowding, we employ time series analysis with variable-specific delays through a crosscorrelation function. Our study employs various classification models, including logistic regression, classification trees, random forests, gradient-boosted trees, and optimal classification trees, with gradient-boosted trees emerging as the most accurate, achieving a good F1 score and accuracy. Specifically, the best accuracy rates obtained are 90.1% at

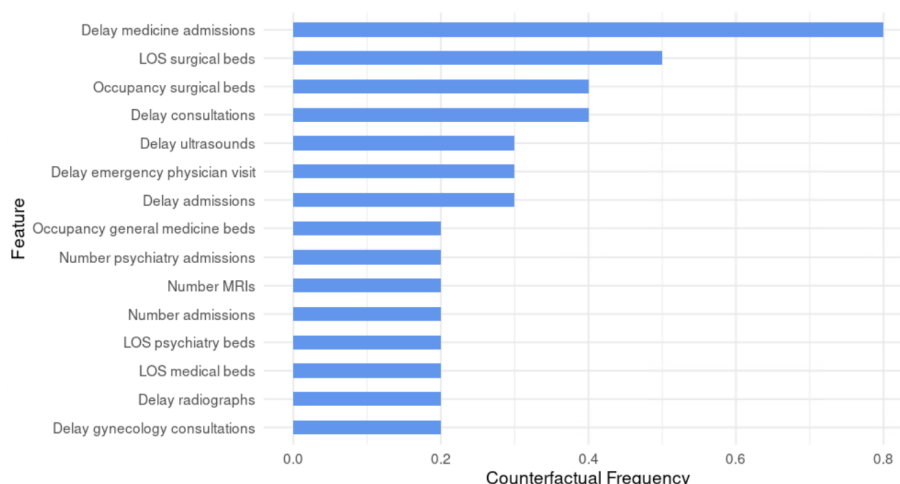


FIGURE 4.12 Highest Counterfactual values for 24 hrs forecasting using variable-specific GBM model for another random overcrowded moment (September 5th, 2021, at 9 PM)

24 hours and 89.9% at 48 hours. Past overcrowding scores represent the most critical factor in predicting emergency overcrowding in all the models trained. This leads to the hypothesis that there is an inherent inertia in the system where, despite efforts being made, the situation will remain the same in terms of overcrowding. Other important features, including delays in consultations and length of stay in surgical and gynecological beds, recur in most models, increasing the likelihood that they genuinely have an impact. SHAP values and counterfactuals for local moments show the punctual importance of certain variables, which have no effect on the overall model but have a punctual value that varies daily. Furthermore, by integrating SHAP values and counterfactuals, we developed a virtual dashboard enabling managers to allocate and plan resources based on the likelihood of overcrowding and provided explanations. There are a few limitations related to the data used for this study. The timeline includes the COVID period, during which significant changes have shifted features in the ED. Regarding our outcome, the EDWIN scores have significantly dropped during this period as the number of patients visiting the ED had been reduced.

Another limitation of this study is that only logistical factors have been included as predictors. Clinical features like age, past medical history, or chief concern, and out-of-hospital features, including ambulances, availability of walk-in clinics, or weather, would likely improve the models' accuracy. To address this, future studies should incorporate clinical data and out-of-hospital variables to enhance the accuracy of predicting ED overcrowding.

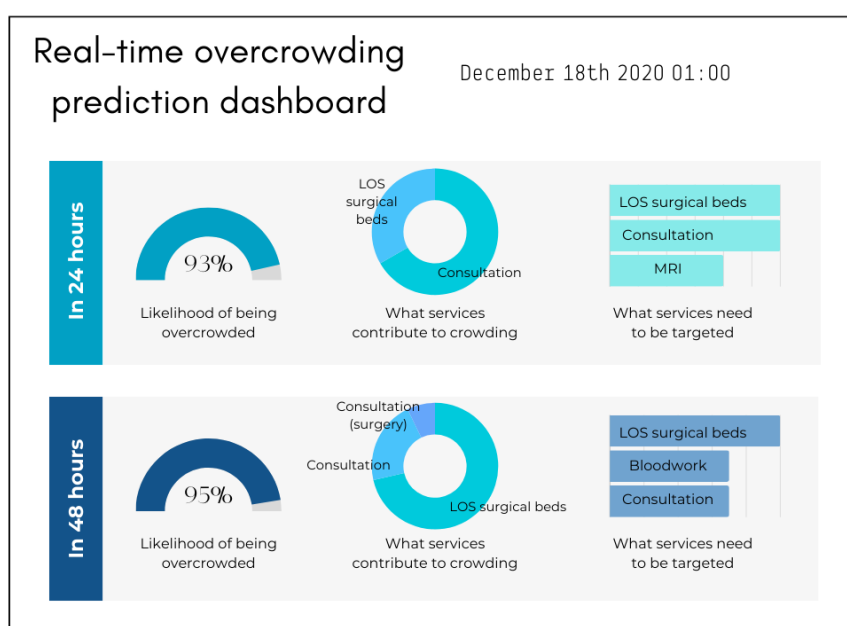


FIGURE 4.13 A virtual dashboard summarizing the results of the study.

CHAPITRE 5 ANALYSES COMPLÉMENTAIRES

5.1 Portrait de l'urgence

Avant de se lancer vers des analyses de prédiction, un portrait de la démographie de l'urgence du CHUM a été adressée afin d'identifier l'acuité des patients et leur cheminement à l'urgence. Tel qu'illustré à la Figure 5.1, la majorité des patients sont traités à partir de la salle d'attente, avec une minorité de patients nécessitant d'être installés sur civière (30%) ou d'être installés en salle de réanimation (5%). Accompagné de la Figure 5.2, ceci met en lumière qu'une très petite partie des patients représentent des conditions de santé mettant en danger leur survie ou nécessitant une attention médicale immédiate. Ceci soulève également l'hypothèse qu'une proportion des patients se présentant à l'urgence n'ont simplement pas accès à un médecin de famille pour une consultation et doivent recourir aux services d'urgence.

Tel qu'expliqué dans l'article, le modèle entrée-débit-sortie est souvent peu linéaire, avec une durée de séjour à l'urgence qui augmente par l'impossibilité d'accéder à la phase de sortie. En regardant la distribution des prises en charge (PEC) des patients illustré à la Figure 5.3, on identifie une proportion de patients qui ne suivent pas le schéma classique. Incluant les patients réorientés vers une clinique sans rendez-vous (2%) et les patients qui ont quitté l'urgence après le triage sans avoir vu le médecin d'urgence (11%), 13 % des patients quittent à la phase d'entrée. Les patients ayant un parcours classique incluent les patients qui voient le médecin d'urgence et sont ensuite congédiés (52%), les patients qui sont consultés en spécialité et ensuite congédiés (11%) et les patients admis dans une unité d'hospitalisation (13%), soit 76% des patients. Le 11% restant est expliqué par 10% des patients qui sont vu par un spécialiste et activement traité à l'urgence sans penser que le patient nécessite une admission (de plusieurs jours) à l'hôpital et 1% des patients qui ont été vus par un spécialiste qui pensait qu'ils nécessitaient une admission de plusieurs jours mais qui ont attendu une admission et ont été congédiés avant d'être admis. Ces deux catégories de patients n'ont pas un processus hospitalier normal ou optimal.

5.2 Gestion des variables

5.2.1 Sélection des variables

Une matrice de corrélation avec toutes les variables indépendantes a été générée. Parmi les variables initiales, il y a une corrélation entre les variables qui étaient produites à partir

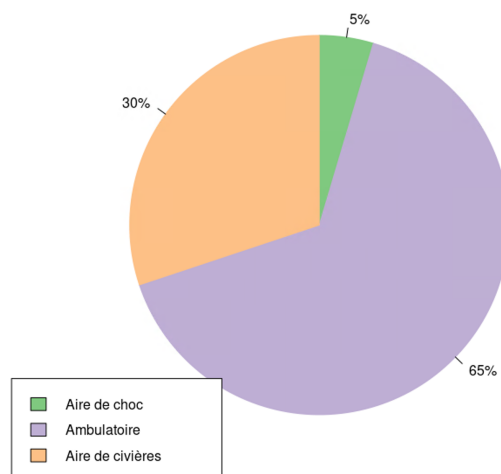


FIGURE 5.1 Orientation des patients après le triage

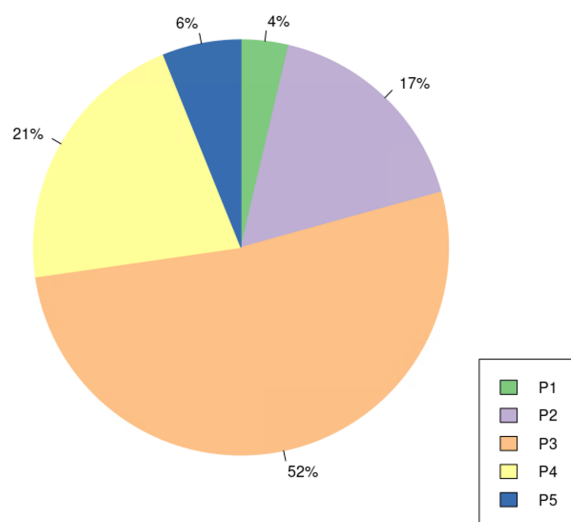


FIGURE 5.2 Distribution des catégories de triage des patients, où P1 représente la catégorie la plus urgente et P5 la moins urgente

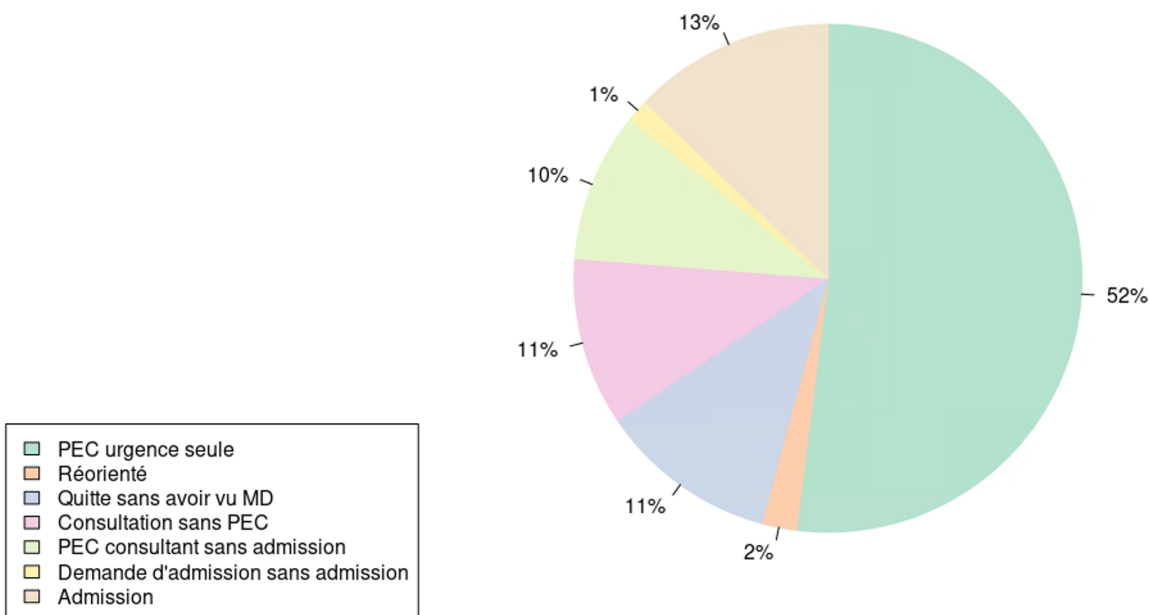


FIGURE 5.3 Prise en charge (PEC) des patients à partir du département d'urgence

d'un facteur d'une autre. Par exemple, la variable de nombre de patients hospitalisés dans l'unité de psychiatrie et la variable du taux d'occupation des lits de l'unité de psychiatrie sont corrélées entre elles, car le taux d'occupation est égal au nombre de patients hospitalisés à une division près. Ainsi, seulement le taux d'occupation a été conservé comme variable.

Tel qu'illustré à la Figure 5.4, le score EDWIN a nettement augmenté au fil des années, une conséquence de la fermeture de civières licenciées. La variable "année" initialement intégrée au modèle a été retirée dans un deuxième temps, dans l'optique d'augmenter l'exportabilité du modèle et de l'utiliser en temps réel. Une variabilité liée au jour est également perçue dans cette figure et a été conservée comme il s'agit de variables cycliques.

Au final, 66 variables ont donc été utilisées pour les analyses.

5.2.2 Relations non-linéaires

Pour chacune des variables conservées, des graphiques de nuages de points mettant en relation la variable indépendante et la variable dépendante ont été produits, similaire à la Figure 5.5. Parmi les graphiques produits, il n'y a pas de relation autre que linéaire qui a été découverte. Pour la majorité des variables, une relation linéaire n'explique pas tout à fait les nuages de points.

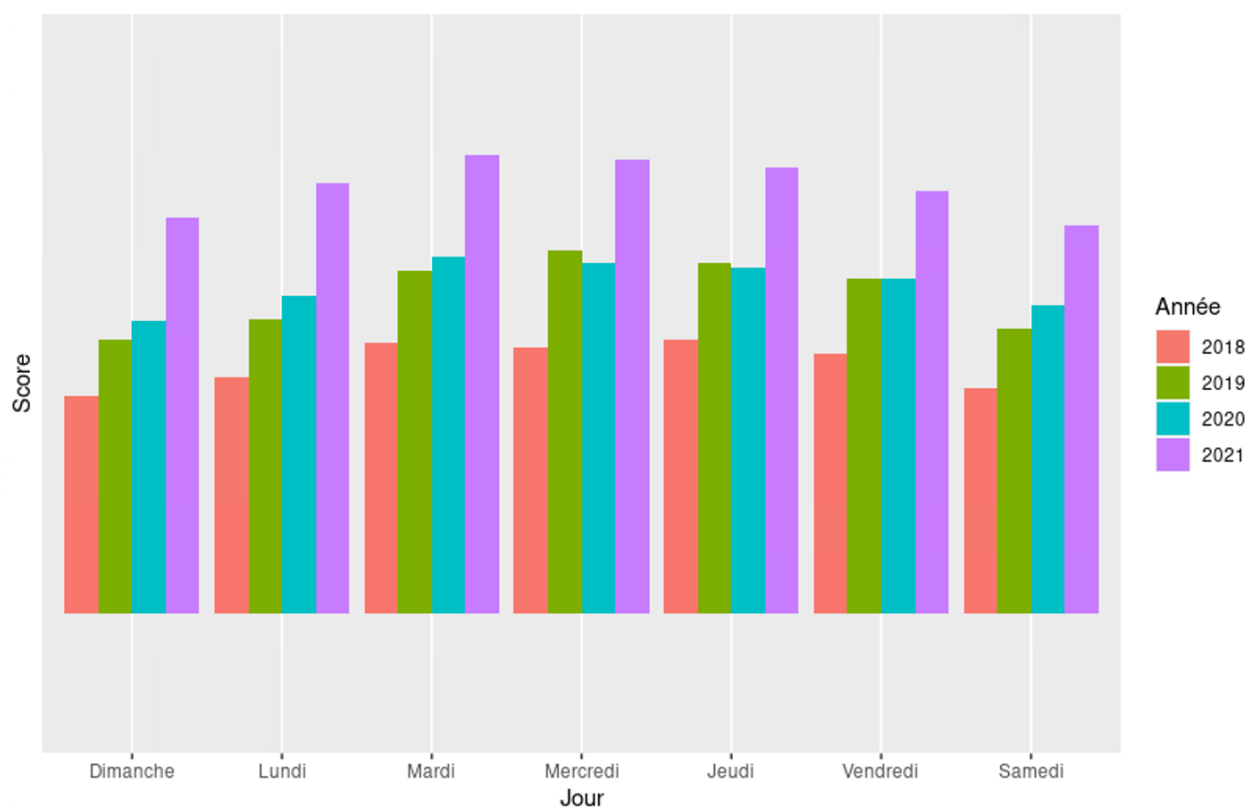


FIGURE 5.4 Score EDWIN normalisé selon le jour de la semaine et l'année

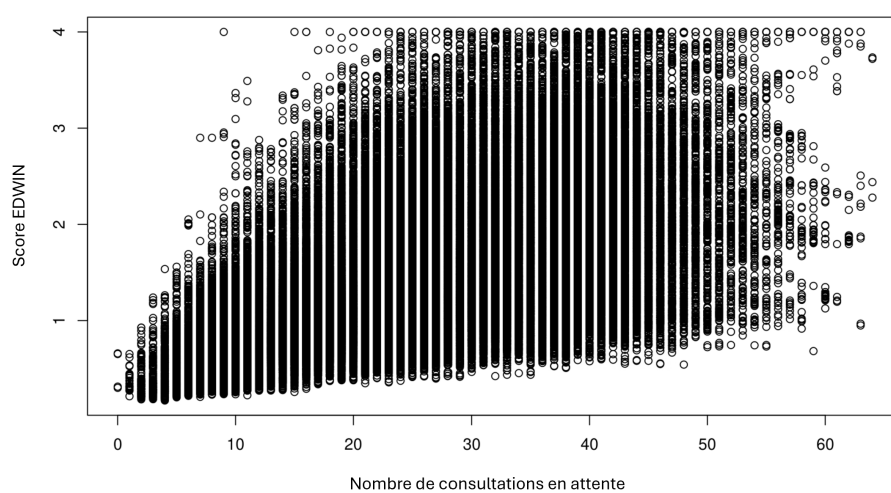


FIGURE 5.5 Relation entre le score EDWIN et le nombre de consultations en attente

5.2.3 Variable dépendante - Score EDWIN

Pour le score EDWIN, la distribution des valeurs numériques dans le temps, tel que visualisé à la Figure 5.6, montre une grande diminution vers mars 2020. Ceci représente le début de la pandémie à COVID-19, où une grande diminution des visites à l'urgence a résulté des manoeuvres d'isolement et de distanciation sociales.

En regardant la distribution des quartiles, à la Figure 5.7, on conclut que la médiane est située à 1.24, avec un premier quartile à 0.167 et un 3e quartile à 1.84.

On utilise la formule suivante pour déterminer quelles sont les données aberrantes :

$$1.5 \times IQR + Q3$$

On obtient une marge supérieure à 4. Il y a 1.1% des données qui sont situées au dessus de cette limite et qui représentent donc des données aberrantes.

En analysant où se situent temporellement ces données, la majorité sont situées entre les mois d'août à novembre 2021. En questionnant les experts sur place ainsi qu'en regardant les données à notre disposition, il est très clair que la fermeture de civières avec personnel (15 civières/70 ont été fermées) pour l'ouverture de civières sans personnel, a un grand impact dans le calcul du score EDWIN puisqu'il s'agit du dénominateur (B_T).

Pour ne pas totalement éliminer ces données de nos ensembles, nous avons décidé d'imposer la valeur numérique de 4 aux valeurs aberrantes.

5.3 Analyses temporelles

Des séries temporelles ont été ajustées avec des méthodes multiplicatives et additives. Des graphiques de décomposition ont été générés, dont le graphique de décomposition multiplicative à la Figure 5.8. En analysant ce graphique, on dénote que la tendance augmente au fil des années, dénotant une augmentation de la congestion à l'urgence. Le facteur aléatoire est instable sur la période étudiée, avec des pics et des creux nécessaires pour ajuster la saisonnalité et la tendance. La saisonnalité ne semble pas avoir une grande contribution aux données observées. Ces analyses n'ont donc pas été intégrées aux modèles de prédiction.

La librairie `stats` a été utilisée dans le langage de programmation R. La fonction `"ts"` et les spécificateurs `"multiplicative"` et `"additive"` ont été utilisés pour établir des graphiques de séries temporelles.

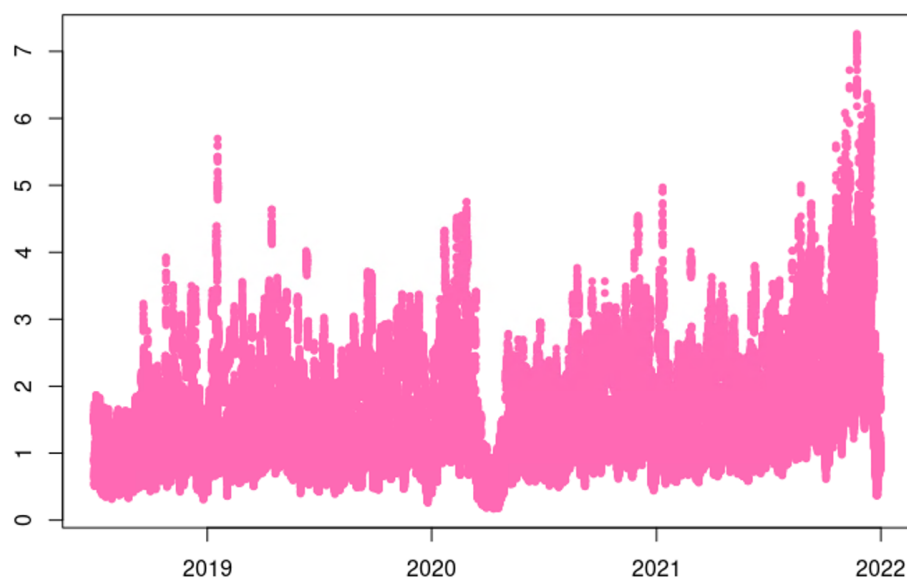


FIGURE 5.6 Valeur numérique du score EDWIN sur la plage de temps étudiée

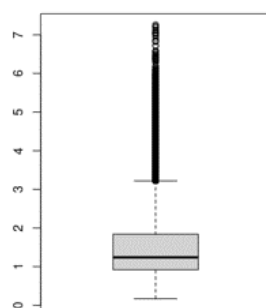


FIGURE 5.7 Distribution des quartiles du score EDWIN

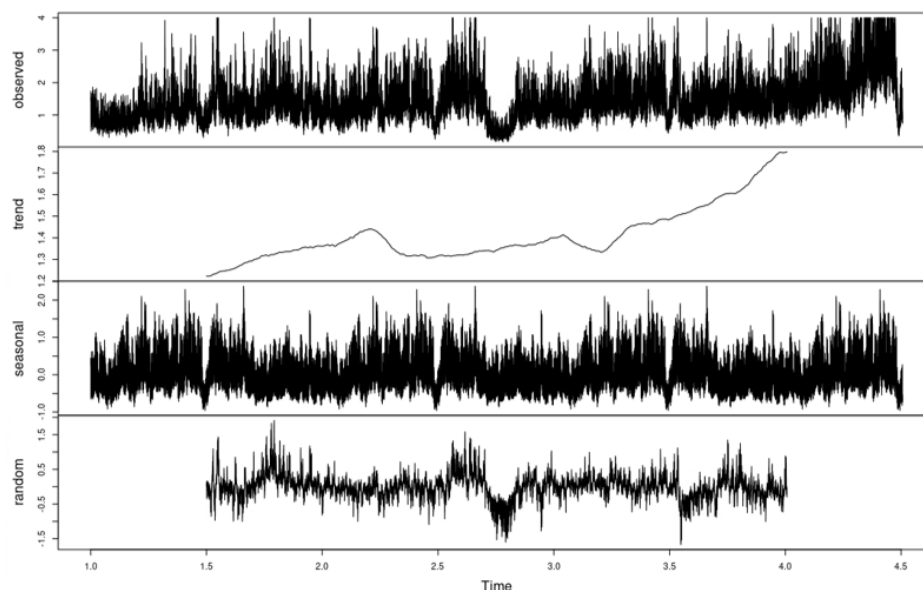


FIGURE 5.8 Décomposition multiplicative des séries temporelles

Validation croisée

En conséquence d'une approche avec des séries temporelles, une validation croisée avec des ensembles aléatoires K-fold provoquerait trop de contamination entre les ensembles d'entraînement et les ensembles tests. En effet, comme les données sont encodées à chaque 15 minutes, plusieurs lignes de données sont très similaires, voire identiques. Pour palier à ce problème et comme nos ensembles de données sont relativement balancés, nous avons choisi d'utiliser la méthode de la fenêtre coulissante comme modèle de validation croisée, tel qu'illustrée à la Figure 5.9 [65]. Ceci empêche que des données dans l'ensemble d'entraînement se retrouvent temporellement après des données dans l'ensemble test. Comme nos données étaient balancées dans les ensembles, nous avons pu procéder à cette méthode.

5.4 Analyses comparatives - Méthodes de régression

Les méthodes de classification discutées dans l'article ont également été entraînées en prenant la valeur numérique du score EDWIN comme donnée de régression.

5.4.1 Régression linéaire

La librairie `caret` a été utilisée dans le langage de programmation R. La méthode `"lm"` a été utilisée pour entraîner un modèle de régression linéaire.

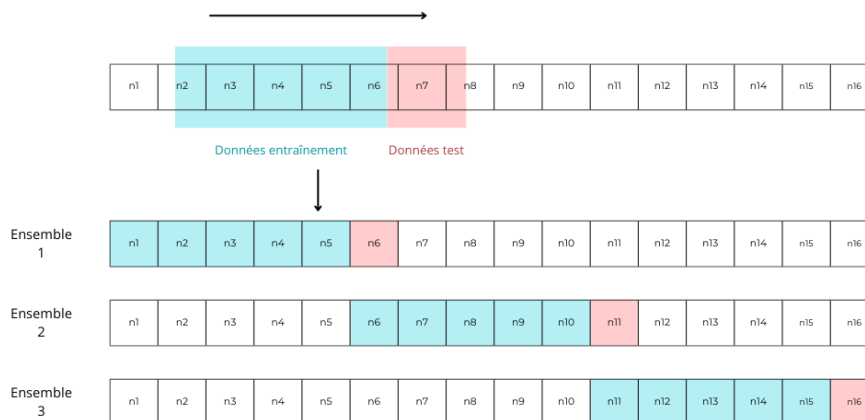


FIGURE 5.9 Méthode de fenêtre coulissante pour éviter une contamination entre les ensembles d'entraînement et les ensembles tests pour des données temporellement liées

Le Tableau 5.1 montre encore une fois une augmentation des R^2 et une diminution des RMSE en passant des séries temporelles à délai fixe aux séries temporelles à délai spécifique aux variables.

TABLEAU 5.1 Modèles de régression linéaires pour différents délais de prédiction

Type de délai	R^2 ajusté	RMSE
Aucun - T_0	0.9187	0.2298
24h fixe - T_{24}^{FIX}	0.7639	0.3563
48h fixe - T_{48}^{FIX}	0.7154	0.4190
CCF 24h - T_{24}^{CCF}	0.7646	0.3230
CCF 48h - T_{48}^{CCF}	0.7642	0.3158

On se rend compte qu'il y a une diminution de la précision des modèles par rapport aux modèles de classification. Ceci évoque que les variables n'ont effectivement pas une relation linéaire avec la variable dépendante, tel qu'il avait également été observé dans l'analyse des graphiques de points de nuages.

5.4.2 Arbres de régression

La librairie `caret` a été utilisée dans le langage de programmation R. La méthode "rpart2" a été utilisée pour entraîner des modèles de classification et de régression. Des valeurs de "cp" entre 0.0001 et 0.1 ont été testées pour sélectionner le meilleur modèle. Le Tableau 5.2 montre de très faibles résultats en comparaison au modèle de régression linéaire. Ceci

est généralement attendu car les arbres de régression séparent tout de même en classes les différents points estimés, ce qui augmente naturellement les valeurs R^2 . Une amélioration entre les délais fixes et les délais spécifiques aux variables est encore notée.

TABLEAU 5.2 Modèles d'arbres de régression pour différents délais de prédiction

Type de délai	R^2 ajusté	RMSE
Aucun - T_0	0.7362	0.3177
24h fixe - T_{24}^{FIX}	0.5956	0.4222
48h fixe - T_{48}^{FIX}	0.5234	0.4906
CCF 24h - T_{24}^{CCF}	0.7131	0.3456
CCF 48h - T_{48}^{CCF}	0.7060	0.3405

L'arbre de régression illustré à la Figure 5.10 démontre une très grande importance des différents scores de congestion antérieurs, qui représente la majorité des noeuds. Ce modèle n'a toutefois pas une grande précision.

5.4.3 Forêts aléatoires

La librairie `caret` a été utilisée dans le langage de programmation R. La méthode "ranger" a été utilisée pour entraîner des modèles de classification et de régression. Des valeurs de "num.tree" entre 200 et 1000 ont été testées pour sélectionner le meilleur modèle. Le Tableau 5.3 résume la performance des différents modèles de forêts aléatoires. La performance est similaire aux modèles de régression linéaire, avec une amélioration marquée entre les modèles avec délais variables par rapport aux modèles avec délai fixe.

TABLEAU 5.3 Modèles de forêts aléatoires de régression pour différents délais de prédiction

Type de délai	R^2 ajusté	RMSE
Aucun - T_0	0.8842	0.2151
24h fixe - T_{24}^{FIX}	0.7099	0.3610
48h fixe - T_{48}^{FIX}	0.6651	0.4057
CCF 24h - T_{24}^{CCF}	0.7788	0.3037
CCF 48h - T_{48}^{CCF}	0.7576	0.3073

5.4.4 Arbres boostés par le gradient

La librairie `caret` a été utilisée dans le langage de programmation R. La méthode "gbm" a été utilisée pour entraîner des modèles de classification et de régression. Des valeurs de "interaction.depth" entre 1 et 9 et des valeurs de "n.tree" entre 400 et 1000 ont été testés pour

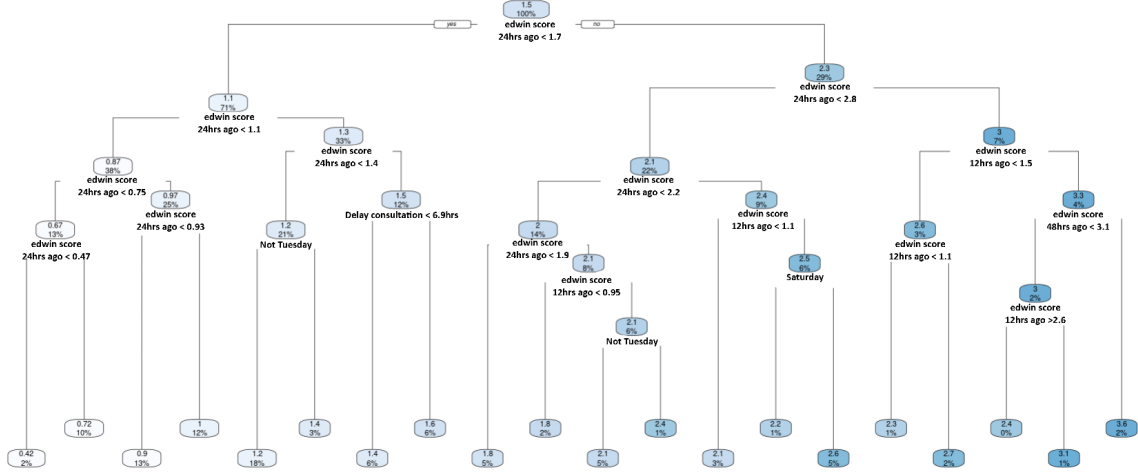


FIGURE 5.10 Arbre de régression avec délais spécifiques aux variables pour prédire la congestion dans 24 heures

sélectionner le meilleur modèle. Le Tableau 5.4 montre une grande amélioration des R^2 et RMSE par rapport aux modèles précédents.

TABLEAU 5.4 Modèles de d'arbres de régression boostés par le gradient pour différents délais de prédiction

Type de délai	R^2 ajusté	RMSE
Aucun - T_0	0.901	0.2401
24h fixe - T_{24}^{FIX}	0.805	0.331
48h fixe - T_{48}^{FIX}	0.7648	0.3638
CCF 24h - T_{24}^{CCF}	0.8486	0.2915
CCF 48h - T_{48}^{CCF}	0.8493	0.2908

En comparant les variables les plus importantes dans les modèles de régression d'arbres boostés par le gradient avec délai spécifique aux variables à 24h et 48h, tels qu'illustré à la Figure 5.11, à ceux produits dans l'article pour la classification, la grande majorité des variables sont présentes dans les deux et souvent dans le même ordre.

5.5 Arbres de classification optimaux

5.5.1 Idée générale

Lorsque Bertismas et Dunn ont décrit ce modèle en 2017, l'idée était de proposer un modèle d'arbre de classification qui, au lieu de sélectionner des embranchements de manière itérative,

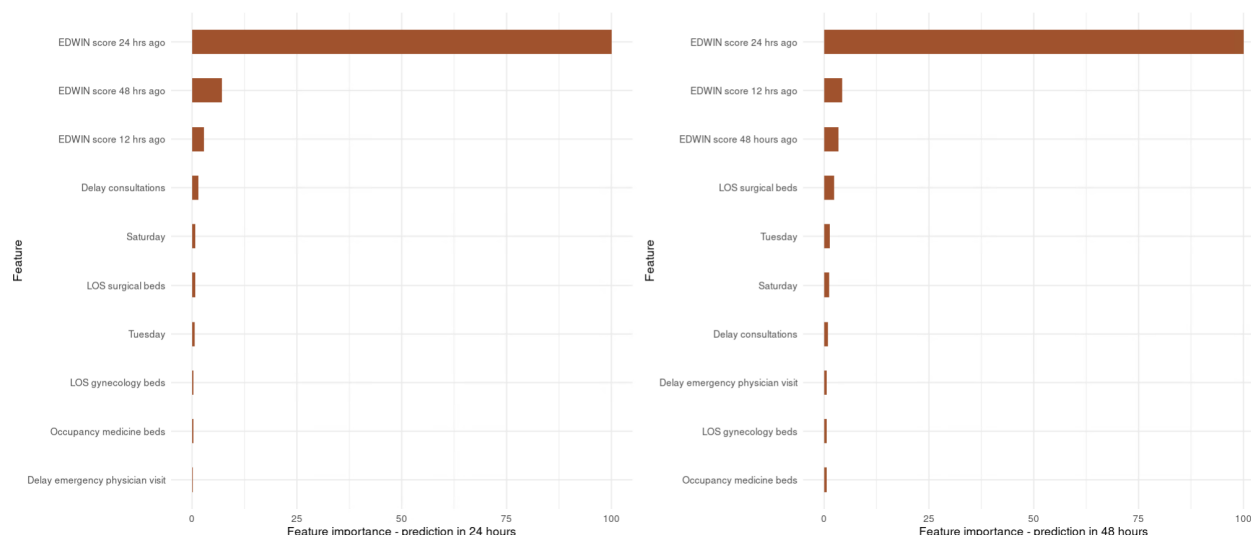


FIGURE 5.11 Variables les plus importantes dans les modèles d'arbres boostés par le gradient avec délais spécifiques aux variables pour des prédictions à 24h et 48h

allait considérer l'entièreté de l'arbre comme un problème d'optimisation. Ceci permettrait d'éviter de "manquer" des embranchements significatifs qui ne seraient pas retenus à cause des embranchements précédents. Dans un arbre de décision traditionnel, les noeuds sont générés top-down, c'est-à-dire que le noeud initial ne tient pas compte des noeuds subséquents. L'étape d'émondage, qui survient après que l'arbre soit construit, représentait également un potentiel d'éliminer un embranchement significatif.

En comparaison au modèle d'arbre de décision traditionnel qui utilise des mesures d'impureté comme le score de Gini pour chaque embranchement, le modèle d'arbre de classification optimal tente plutôt de minimiser l'erreur de classification. Ceci implique qu'au lieu de tenter de trouver des minimums locaux de manière itérative comme les arbres de décisions usuels, l'arbre de classification optimal tente de trouver un minimum global.

Pour accélérer la vitesse de résolution, le modèle prend comme point de départ (*warm start*), une solution trouvée à partir d'un arbre CART.

5.5.2 Modèle théorique

À partir du modèle de Bertsimas et al. décrit en 2017 suivant :

$$\min \frac{1}{L} \sum_{t \in \mathcal{T}_L} L_t + \alpha \sum_{t \in \mathcal{T}_B} d_t \quad (5.1)$$

$$\text{s.c. : } L_t \geq N_t - N_{kt} - n(1 - c_{kt}) \quad k = 1, \dots, K, \forall t \in \mathcal{T}_L \quad (5.2)$$

$$L_t \leq N_t - N_{kt} + nc_{kt} \quad k = 1, \dots, K, \forall t \in \mathcal{T}_L \quad (5.3)$$

$$L_t \geq 0 \quad \forall t \in \mathcal{T}_L \quad (5.4)$$

$$N_{kt} = \frac{1}{2} \sum_{i=1}^n (1 + Y_{ik}) z_{it} \quad k = 1, \dots, K, \forall t \in \mathcal{T}_L \quad (5.5)$$

$$N_t = \sum_{i=1}^n z_{it} \quad \forall t \in \mathcal{T}_L \quad (5.6)$$

$$\sum_{k=1}^K c_{kt} = l_t \quad \forall t \in \mathcal{T}_L \quad (5.7)$$

$$\mathbf{a}_m^T \mathbf{x}_i \geq b_t - (1 - z_{it}) \quad i = 1, \dots, n, \forall t \in \mathcal{T}_B, \forall m \in A_R(t) \quad (5.8)$$

$$\mathbf{a}_m^T (\mathbf{x}_i + \boldsymbol{\epsilon}) \leq b_t + (1 + \epsilon_{\max})(1 - z_{it}) \quad i = 1, \dots, n, \forall t \in \mathcal{T}_B, \forall m \in A_R(t) \quad (5.9)$$

$$\sum_{t \in \mathcal{T}_L} z_{it} = 1 \quad i = 1, \dots, n \quad (5.10)$$

$$z_{it} \leq l_t \quad \forall t \in \mathcal{T}_L \quad (5.11)$$

$$\sum_{i=1}^n z_{it} \geq N_{\min} l_t \quad \forall t \in \mathcal{T}_L \quad (5.12)$$

$$\sum_{j=1}^p a_{jt} = d_t \quad \forall t \in \mathcal{T}_B \quad (5.13)$$

$$0 \leq b_t \leq d_t \quad \forall t \in \mathcal{T}_B \quad (5.14)$$

$$d_t \leq d_{p(t)} \quad \forall t \in \mathcal{T}_B \setminus \{1\} \quad (5.15)$$

$$z_{it}, l_t \in 0, 1 \quad i = 1, \dots, n, \forall t \in \mathcal{T}_L \quad (5.16)$$

$$a_{jt}, d_t \in 0, 1 \quad j = 1, \dots, p, \forall t \in \mathcal{T}_B \quad (5.17)$$

- L'ensemble de données d'entraînement (x_i, y_i) , $i = 1, \dots, n$, ont donc n observations, chacune de ces observations a p variables indépendantes (features), donc $x_i \in \mathbb{R}^p$ et une classe $y_i \in 1, \dots, K$.
- Soient $\mathbf{a}$ et b des paramètres utilisés dans un noeud donné pour $\mathbf{a}^T x_i < b$.
- $N_{\min}$ est le nombre minimal de noeuds inclus.
- α est le paramètre de complexité (complexity parameter).
- T correspond au nombre de noeuds pour une profondeur maximale D , où $T = 2^{(D+1)} - 1$. Les noeuds sont divisés en noeuds de branches $t \in \mathcal{T}_B = \{1, \dots, \lfloor T/2 \rfloor\}$ et les noeuds de feuilles $t \in \mathcal{T}_L = \{\lfloor T/2 \rfloor + 1, \dots, T\}$.

- $A_R(t)$ correspond aux noeuds ancêtres du côté droit du noeud t .
- d_t sont des variables indicatrices telles que $d_t = \mathbb{1}\{\text{noeud } t \text{ applique la séparation}\}$.
- z_{it} sont des variables indicatrices telles que $z_{it} = \mathbb{1}\{\mathbf{x}_i \text{ est dans le noeud } t\}$.
- l_t sont des variables indicatrices telles que $l_t = \mathbb{1}\{\text{feuille } t \text{ contient un point quelconque}\}$.
- L_t est l'optimal misclassification loss dans le noeud t .
- N est le nombre de points, où N_{kt} est le nombre de points de classification k dans le noeud t .
- c_{kt} est une variable binaire pour garder compte de la prédiction k à chaque noeud t telle que $c_{kt} = \mathbb{1}\{c_t = k\}$.
- Soit la matrice $\mathbf{Y}$ où $Y_{ik} = 1$ si $y_i = k$ et -1 sinon.
- $d_{p(t)}$ est le noeud parent de d_t (structure hiérarchique).
- ϵ_j est une petite constante telle que $\epsilon_j = \min\{x_j^{(i+1)} - x_j^{(i)} | x_j^{(i+1)} \neq x_j^{(i)}, i = 1, \dots, n-1\}$
- $\hat{L}$ représente la précision de base, soit en prédisant la classe la plus populaire dans tout l'ensemble de données.

5.5.3 Modèle pratique

Variables connues

Avant de procéder à la résolution du problème mixte, notons que nous connaissons la valeur des variables suivantes :

- $\hat{L} = 0.66$, soit 64% du temps, l'urgence est non-congestionnée.
- $n = 122\,880$, soit 122 880 lignes de données.
- $p = 104$, soit 66 variables indépendantes dont 3 de type catégorique encodées en 43 variables
- $K = 2$, et donc $k = 1, 2$, soit les classes congestionné et non-congestionné pour la classification et 8 classes pour la régression.
- Des profondeurs de $D = \{3, 4, 5, 6\}$ choisies pour l'interprétabilité, la profondeur donnant les meilleurs résultats a été $D = 3$. Ceci implique donc que $T = 2^{(D+1)} - 1 = 2^4 - 1 = 15$.
- Nous savons également que $t \in \mathcal{T}_B = \{1, \dots, \lfloor T/2 \rfloor\} = \{1, \dots, \lfloor 15/2 \rfloor\} = \{1, \dots, 7\}$
- et nous avons $t \in \mathcal{T}_L = \{\lfloor T/2 \rfloor + 1, \dots, T\} = \{\lfloor 15/2 \rfloor + 1, \dots, T\} = \{8, 15\}$

Le modèle d'optimisation a été codé dans le langage de programmation Python. Le solveur utilisé a été Gurobi. L'arbre CART initial à partir duquel l'arbre de classification optimal s'est construit a été programmé avec la librairie sklearn et la fonction "tree".

Modification des hyperparamètres

Il y a 3 hyperparamètres qui peuvent être modifiés pour tenter d'améliorer la performance des arbres de classification optimaux : le paramètre de complexité α , la profondeur D et le nombre minimal de feuilles N_{min} .

Pour sélectionner la meilleur valeur pour chacun de ces hyperparamètres, nous avons débuté avec les valeurs suivantes : $\alpha = 0$ et $D_{max} = 6$ et avons testé des arbres CART afin de trouver la solution qui optimisait la précision :

Pour des valeurs de $D \in \{1, ..., D_{max}\}$:

- et pour des valeurs de nombre d'embranchements $C \in \{1, ..., 2^D - 1\}$

Paramètre de complexité α

Le paramètre de complexité α entre en jeu dans la portion suivante du modèle :

$$\min \frac{1}{\hat{L}} \sum_{t \in \mathcal{T}_L} L_t + \alpha \sum_{t \in \mathcal{T}_B} d_t$$

Des valeurs "alpha" entre 0 et 10 ont été testées pour sélectionner le meilleur modèle.

La valeur α retenue est 7.

Profondeur de l'arbre

La profondeur maximale de l'arbre D entre en jeu via la variable T correspondant au nombre de noeuds. Les noeuds sont divisés en noeuds de branches $t \in \mathcal{T}_B = \{1, ..., \lfloor T/2 \rfloor\}$ et les noeuds de feuilles $t \in \mathcal{T}_L = \{\lfloor T/2 \rfloor + 1, ..., T\}$. Ces variables apparaissent dans chaque contrainte du modèle.

Des valeurs "depth" entre 2 et 6 ont été testées. Ce choix a été fait en fonction du besoin d'interprétabilité. Une profondeur de 6 étages implique qu'il y aurait potentiellement $2^6 = 64$ feuilles, ce qui est la limite de l'interprétable.

La profondeur optimale retenue a été de 4.

Nombre minimal de feuilles

Le nombre minimale de feuilles N_{min} apparaît dans la contrainte :

$$\sum_{i=1}^n z_{it} \geq N_{min} l_t \quad \forall t \in \mathcal{T}_L$$

Le nombre minimal de feuille sélectionné est 5.

En plus de ces trois hyperparamètres, notons également que la limite de temps a été ajustée pour que chaque sous-ensemble aie une solution ainsi qu'une borne trouvée.

À partir du modèle d'optimisation, un arbre de classification a été généré.

Le Tableau 5.5 résume la performance de l'arbre de classification optimal codés manuellement. En comparaison aux modèles de librairie présenté dans l'article, les modèles codés à partir du modèle d'optimisation ont des performances inférieures aux modèles de librairie. En comparaison aux arbres de classification usuels, les modèles de librairie et codé manuellement ont des performances inférieures.

TABLEAU 5.5 Modèles d'arbres de classification optimaux pour différents délais de prédiction

Type de délai	Précision	Score F_1
Aucun - T_0	0.778	0.544
24h fixe - T_{24}^{FIX}	0.6938	0.435
48h fixe - T_{48}^{FIX}	0.6446	0.380
CCF 24h - T_{24}^{CCF}	0.7519	0.514
CCF 48h - T_{48}^{CCF}	0.7272	0.481

Comme notre jeu de données était de taille importante ($n = 122880$) et que la précision des arbres de décision n'était pas très faible, nous avons convenu de ne pas modifier notre algorithme pour inclure des hyperplans. Ceci a été conseillé par les créateurs des arbres de classification optimaux.

5.6 Modèle de librairie

Des arbres de classification optimaux ont également été entraînés à partir de la librairie IAI en R, qui provient de Bertsimas et Dunn. Les modèles de classification sont inclus dans l'article et les modèles de régression sont présentés dans le Tableau 5.6.

TABLEAU 5.6 Modèles d'arbre de classification optimaux ajustés pour la régression pour différents délais de prédiction

Type de délai	R^2 ajusté	RMSE
Aucun - T_0	0.877	0.936
24h fixe - T_{24}^{FIX}	0.777	0.881
48h fixe - T_{48}^{FIX}	0.761	0.872
Minimum 24h - T_{24}^{CCF}	0.794	0.891
Minimum 48h - T_{48}^{CCF}	0.793	0.890

Un arbre de régression a été extrait pour une prédiction dans 24h en utilisation des délais

spécifiques pour chaque variable. Tel qu'illustré à la Figure 5.12, les scores de congestions antérieurs et les variables de dates sont très importants. Les variables liées aux admissions, consultations et hospitalisations complètent l'arbre.

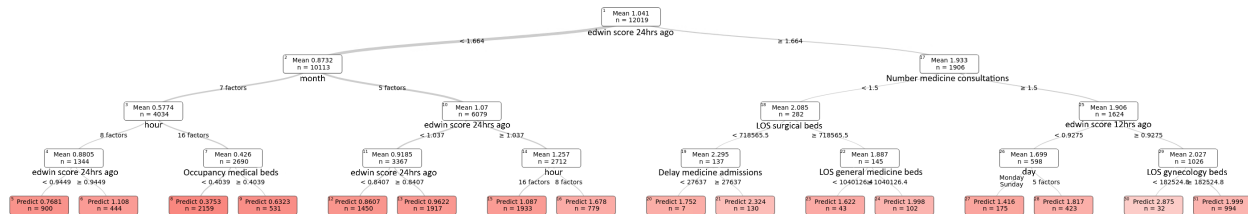


FIGURE 5.12 Arbre de régression optimal pour une prédiction dans 24h en utilisant des délais spécifiques aux variables

CHAPITRE 6 DISCUSSION

Notre objectif de recherche principal, c'est-à-dire développer des modèles d'algorithmes d'apprentissage automatique qui pourraient prédire la congestion de l'urgence, a été accompli à travers l'article présenté pour la classification et les analyses complémentaires pour la régression. Bien que l'objectif principal était de se concentrer sur des modèles hautement interprétables, tels les arbres de classification et les arbres de classification optimaux, ces modèles étant moins performants, ce sous-objectif a été accompli par l'intégration de méthodes d'intelligence artificielle explicable, soit les valeurs SHAP et les explications contrefactuelles. Notre objectif secondaire, soit de déterminer la nature des causes de la congestion, a également été atteint à partir des méthodes d'interprétation globale, tel les courbes de Pareto, l'importance des variables et les arbres, ainsi que les méthodes d'intelligence artificielle explicables. En plus des éléments de discussion présentés dans l'article, soulevons également des éléments liés aux analyses complémentaires présentées au chapitre précédent.

6.1 Analyses temporelles

Pareillement aux modèles de classification discutés dans l'article, tous les modèles de régression ont également vus une nette amélioration des R^2 ajustés et des RMSE avec l'introduction des délais spécifiques aux variables à minimum 24 heures et minimum 48 heures. La meilleure amélioration du score R^2 -ajusté notée est de 11.8% pour les arbres de régression à 24 heures et 18.3% pour les arbres de régression à 48 heures. Ceci représente une amélioration très significative de la performance des modèles et une contribution majeure dans notre compréhension de ce problème.

L'amélioration de la performance à travers l'ensemble des modèles montre que chaque variable n'est pas corrélée à la congestion à l'urgence avec le même délai. Ceci est une conclusion logique du fait que tous les services n'ont pas le même impact sur le séjour d'un patient à l'urgence. Un délai d'admission va directement empêcher un patient de quitter l'urgence, augmentant donc directement la charge sur l'urgence. Un délai pour des investigations, comme une prise de sang, peut signifier qu'un patient reste à l'urgence au lieu d'être congédié, mais peut également ne pas avoir le même impact si le patient doit être hospitalisé de toute façon. Comme plusieurs services au sein de l'hôpital fonctionnent sur des cycles de jour-soir-nuit, il est logique que les graphiques de corrélation croisée obtenus ont également des pics aux 24 heures.

Les analyses de décomposition des séries temporelles ont montré une importance marquée du facteur aléatoire et de la tendance. Tel que vu dans les figures descriptives de la démographie de l'urgence, une augmentation de la congestion selon l'année a été notée. Les analyses de décomposition ARIMA ne semblent pas appropriées pour l'utilisation dans ce contexte qui a représenté des changements majeurs et des instabilités.

6.2 Performance des analyses comparatives

Pour la régression, le meilleur modèle est encore une fois celui des arbres boostés par le gradient, tel que pour la classification, avec des valeurs de R^2 ajustés de 0.8486 à 24h et 0.8493 à 48h. Ce modèle représente un avantage computationnel par rapport aux forêts aléatoires, en plus d'être également plus performant.

Les arbres optimaux ont été décrits comme pouvant améliorer la prédiction des CART surtout lorsque la précision des CART est déjà faible, ce qui dans notre cas n'était pas vrai pour la classification, mais l'était pour la régression. La précision des arbres de régression optimaux est effectivement supérieure à celle des arbres de régression usuels et très similaire à la précision des forêts aléatoires, ce qui correspond au raisonnement derrière le développement des arbres de classification optimaux.

Parmi tous les modèles, les arbres boostés par le gradient couplés des méthodes de valeurs SHAP et des explications contrefactuelles, telles que présentées dans l'article, représentent un modèle avec une haute précision et un bon score F1. Sous un format de tableau de bord virtuel, ces méthodes sont facilement interprétables par des décideurs du système de santé. Leur utilisation dans une optique de prévention de la congestion pourrait améliorer les méthodes de gestion au sein du système hospitalier et changer les comportements face à la congestion, d'une position réactive à une position préventive.

6.3 Score EDWIN

Il demeure plus intéressant pour l'utilisation pratique du score EDWIN de se fier sur la classification plutôt que sur la régression, car la régression est plus complexe d'interprétation.

Il semble que les scores de congestion antérieurs représentent le facteur le plus important dans la capacité de prédire la congestion à l'urgence. Ceci permet d'établir les hypothèses suivantes :

1. La congestion est un phénomène qui n'est pas reconnu dans le système. Ceci implique que lorsque la congestion débute, il n'y a pas nécessairement de méthodes préventives

qui seront mises en place pour rectifier la situation

2. Il y a une inertie inhérente au système où malgré que des efforts soient fait, la situation va demeurer identique au point de vue de la congestion.
3. La congestion est un phénomène qui s'entretient. L'achalandage crée un ralentissement des services qui entretient une durée de séjour plus longue et un achalandage plus long.

Pour les analyses de régression comme pour les analyses de classification, les variables clés étaient les scores EDWIN antérieurs. En décomposant la variation des scores EDWIN sur 24 heures, on constate que le score EDWIN varie très peu (moins de 0.5 dans la grande majorité de nos données), tel qu'illustré à la Figure 6.1, ce qui explique partiellement pourquoi les scores de congestion antérieurs peuvent prédire les scores de congestion futurs.

Certaines variables ont un impact plus important à court terme, comme le délai avant d'avoir une consultation. Certaines variables ont un impact plus important à moyen terme, comme la durée d'hospitalisation moyenne. Ceci est intéressant car il est logique ; dans un ordre temporel une consultation doit être faite avant de pouvoir hospitaliser le patient. Un patient va donc rester à l'urgence plus longtemps si la durée avant qu'il soit consulté par un spécialiste est longue, mais dans 48h ce qui aura un impact sur son départ de l'urgence est le fait qu'il n'ait pas de lit à l'étage.

Les autres variables les plus importantes reviennent dans tous les modèles, incluant les durées de séjour en médecine, chirurgie, gynécologie, les délais de consultation et délais pour voir un médecin d'urgence. Ceci dénote une cohérence entre les modèles de prédiction et augmente la probabilité que ces variables aient réellement un impact sur la congestion.

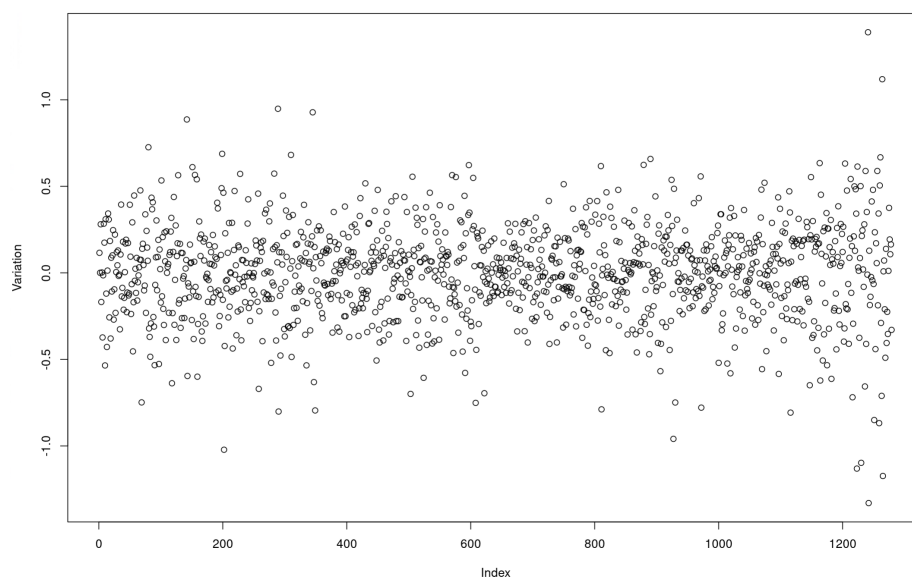


FIGURE 6.1 Variation du score EDWIN sur 24 heures

CHAPITRE 7 CONCLUSION

7.1 Synthèse des travaux

En conclusion, notre étude rétrospective menée sur les données du Centre hospitalier de l'Université de Montréal visait à prédire la congestion de l'urgence 24 et 48 heures à l'avance. En nous appuyant sur les délais logistiques pour obtenir des services à l'urgence et en utilisant le score EDWIN comme mesure objective de la congestion, nous avons utilisé des analyses de séries temporelles avec des délais spécifiques aux variables par le biais d'une fonction de corrélation croisée. Notre étude a utilisé divers modèles de classification et de régression, notamment la régression logistique et linéaire, les arbres de classification et de régression, les forêts aléatoires, les arbres boostés par le gradient et les arbres de classification optimaux. Les arbres boostés par le gradient se sont avérés les plus précis avec des taux de précision de 90,1 % à 24 heures et de 89.9 % à 48 heures. Ils ont également obtenus les meilleurs scores F1. Pour la régression, les arbres boostés par le gradient étaient également les plus performants, avec des valeurs de R^2 ajusté de 84.9% à 24 heures et à 48 heures. Pour l'ensemble des modèles, l'utilisation de délais spécifiques pour chaque variable à l'aide d'une fonction de corrélation croisée a augmenté la performance des modèles. Tous les modèles développés ont montré une grande importance de la congestion passée, dans les 48 dernières heures, sur la congestion future, jusqu'à 48 heures plus tard. Il s'agit de l'élément prédicteur le plus important globalement. L'utilisation de valeurs SHAP et de contrefactuelles pour détailler des moments précis a montré une divergence entre les variables locales les plus importantes à un moment par rapport à un autre ainsi que par rapport aux variables globalement importantes dans le modèle. En outre, grâce à l'intégration des valeurs SHAP et des contrefactuelles, nous avons développé un tableau de bord virtuel permettant aux gestionnaires d'allouer et de planifier les ressources en fonction de la probabilité de congestion et des explications fournies.

7.2 Limitations de la solution proposée

Les données choisies pour cette étude présentent quelques limites. Le calendrier inclut la période COVID, plus précisément de mars 2020 à décembre 2020, où des changements significatifs ont modifié les caractéristiques du département d'urgence. Le personnel disponible, le volume général de patients, les délais d'investigation et les lits variant selon le statut COVID des patients, pour n'en citer que quelques-uns, ont grandement influencé la fluidité des pa-

tients. En ce qui concerne nos résultats, les scores EDWIN ont chuté de manière significative au cours de cette période, car le nombre de patients visitant les urgences a été réduit.

Suite à la période COVID, une augmentation significative des scores EDWIN a résulté de la fermeture de civières à l'urgence. La variation dans les scores EDWIN due à l'instabilité du système hospitalier durant cette période a certainement eut un impact sur la performance des modèles.

Une autre limite de cette étude est que seuls des facteurs logistiques ont été inclus comme prédicteurs. Des variables telles que l'âge, les antécédents médicaux ou la préoccupation principale se sont avérées augmenter la durée de séjour d'un patient. L'inclusion de ces variables pourrait probablement améliorer la capacité à prédire la congestion. De la même manière, les données extrahospitalières, y compris les ambulances, la disponibilité des cliniques sans rendez-vous ou les conditions météorologiques, pourraient aider à prédire une partie du score EDWIN lié au nombre de visites aux urgences.

Le choix du score EDWIN repose sur l'acuité au triage des patients, le nombre de médecins disponibles et le nombre de civières disponibles. L'acuité au triage des patients a représenté une instabilité due à la période COVID, le nombre de civières disponibles a également représenté une instabilité due à la fermeture des civières. Le nombre de médecins disponibles est stable au courant de cette période, mais est réduit pendant la nuit, ce qui met un accent disproportionné sur le poids de la congestion la nuit par rapport au matin.

7.3 Améliorations futures

Dans les recherches futures visant à améliorer la prédiction de la congestion des services d'urgence, plusieurs améliorations clés peuvent être apportées afin d'intégrer un plus large éventail de facteurs influençant le flux de patients et l'utilisation des ressources. Ces améliorations potentielles englobent divers éléments à intra-hospitaliers et extra-hospitaliers, y compris des facteurs cliniques, infirmiers et liés à l'hôpital.

Parmi les améliorations potentielles, l'intégration d'éléments extrahospitaliers revêt une importance particulière, car ces facteurs externes peuvent avoir un impact significatif sur le volume et l'acuité des patients. Premièrement, intégrer les conditions météorologiques locales et les événements spécifiques à la région, tels que les catastrophes naturelles ou les événements sociaux, qui peuvent entraîner une augmentation soudaine des demandes de soins d'urgence, pour améliorer les modèles de prédiction. Deuxièmement, l'intégration des données relatives à la gestion des ambulances pourrait augmenter la précision des modèles. Cela inclut la collecte d'informations sur les délais d'intervention des ambulances, les schémas de détournement vers

d'autres établissements de soins de santé et la disponibilité des ambulances pour répondre aux besoins des patients. Il pourrait s'agir non seulement d'une autre méthode de prédiction de l'achalandage mais procurerait potentiellement de l'information sur les hôpitaux régionaux en cas de diversion. Troisièmement, il serait intéressant d'évaluer la présence et l'accessibilité des cliniques sans rendez-vous et des hôpitaux périphériques. En examinant leurs scores respectifs de congestion et leur capacité à absorber une partie de la demande de soins d'urgence, il est possible d'évaluer leur impact sur les schémas d'utilisation des services d'urgence.

Par ailleurs, en intégrant des éléments cliniques dans les modèles prédictifs, les futures études peuvent obtenir une compréhension plus complète des facteurs contribuant à la congestion et développer des interventions ciblées pour améliorer le flux des patients et l'utilisation des ressources au sein des services de soins d'urgence. Tout d'abord, l'utilisation de données cliniques, notamment des indicateurs tels que le nombre de patients en salle d'attente, les données démographiques des patients (âge, sexe) et les motifs de consultation, offre des informations précieuses sur la diversité et la complexité des besoins médicaux des patients, ce qui permet de comprendre les dynamiques de flux de patients et les points potentiels de congestion au sein du service d'urgence. Ensuite, l'incorporation de données de triage, comprenant des facteurs tels que l'administration de médicaments, les comorbidités des patients et les signes vitaux, fournit une évaluation initiale de l'acuité et de la gravité des patients, aidant à l'identification des cas nécessitant une intervention médicale urgente. Ces informations de triage jouent un rôle crucial dans la priorisation des soins aux patients et l'allocation des ressources, influençant ainsi le débit du service d'urgence et les niveaux de congestion. Enfin, l'intégration de données granulaires sur la gravité des patients, comprenant des indicateurs tels que le besoin d'intubation, de réanimation ou l'activation d'un "code blanc", offre des informations détaillées sur l'acuité et la complexité des cas traités dans le service d'urgence. Ces points de données permettent une évaluation plus nuancée de la charge de travail et des besoins en ressources, facilitant ainsi la mise en œuvre de stratégies de gestion proactive pour atténuer la congestion et optimiser la prestation des soins aux patients.

De plus, en intégrant des éléments de disponibilité du personnel soignant dans nos analyses de la congestion des services d'urgence, il serait potentiellement possible de mieux appréhender les aspects organisationnels et humains qui influent sur les flux de patients et de développer des stratégies de gestion des ressources plus efficaces pour répondre aux besoins des patients de manière opportune et sécuritaire. Il serait judicieux d'évaluer la présence du personnel infirmier aux urgences et à l'hôpital, en examinant le nombre d'infirmières en service aux urgences ainsi que le personnel infirmier disponible dans l'ensemble de l'établissement. Ces informations pourraient fournir des indications précieuses sur la capacité de prise en charge des patients et sur la disponibilité des ressources humaines nécessaires. De même, il serait per-

tiennent de prendre en compte les délais d'acheminement des brancardiers et d'autres facteurs logistiques qui influent sur le flux des patients au sein du service des urgences. Cela pourrait inclure l'évaluation des temps d'attente pour l'admission, le transfert ou le déchargement des patients, ainsi que la gestion des ressources matérielles nécessaires à ces opérations. En outre, il serait intéressant d'évaluer la disponibilité des préposés aux soins et du personnel de soutien pour aider aux tâches de soins et au transport des patients. Ces membres du personnel jouent un rôle essentiel dans le fonctionnement quotidien du service des urgences en fournissant un soutien précieux aux infirmières et aux médecins dans la prestation des soins aux patients.

Finalement, intégrer des éléments hospitaliers dans les modèles de prédiction de la congestion permettrait de mieux prévoir les facteurs internes à l'établissement qui influent sur les flux de patients et de développer des stratégies de gestion plus efficaces pour optimiser l'utilisation des ressources et améliorer la qualité des soins. Il serait utile d'intégrer des données sur les pratiques de gestion des salles d'opération, notamment la planification des interventions chirurgicales, les temps de rotation et les volumes d'interventions. Ces informations pourraient avoir un impact sur le flux des patients aux urgences et aider à anticiper les périodes de pic d'activité ou de congestion. Ensuite, il serait pertinent d'analyser la disponibilité et la répartition des lits dans les unités de soins intensifs afin d'évaluer les contraintes de capacité et d'identifier les éventuels goulets d'étranglement. La gestion efficace des lits est essentielle pour garantir une prise en charge appropriée des patients aux urgences et éviter les retards dans les admissions ou les transferts. De plus, il serait intéressant d'examiner les processus et les délais de congé afin d'identifier les possibilités d'améliorer le débit des patients et la rotation des lits. Une meilleure gestion des sorties des patients pourrait permettre une utilisation plus efficace des ressources et contribuer à réduire la congestion dans les services d'urgence.

RÉFÉRENCES

- [1] A. Affleck, “Emergency department overcrowding and access block,” *Canadian Journal of Emergency Medicine*, vol. 15, n°. 6, p. 359–370, 2013.
- [2] C. HTERP, “CADTH Health Technology Review recommendation : Emergency department overcrowding in canada,” *Canadian Journal of Health Technologies*, vol. 3, n°. 11, 2023.
- [3] B. R. Asplin, “A conceptual model of emergency department crowding,” *Annals of emergency medicine*, vol. 42, n°. 2, p. 173–180, 2003.
- [4] K. Bond, “Frequency, determinants and impact of overcrowding in emergency departments in canada,” *Healthcare Quarterly*, vol. 10, n°. 4, p. 32–40, 2007.
- [5] D. B. Richardson, “Increase in patient mortality at 10 days associated with emergency department overcrowding,” *Medical journal of Australia*, vol. 184, n°. 5, p. 213–216, 2006.
- [6] J. McCusker, “Increases in emergency department occupancy are associated with adverse 30-day outcomes,” *Academic Emergency Medicine*, vol. 21, n°. 10, p. 1092–1100, 2014.
- [7] H. Wang, “The role of patient perception of crowding in the determination of real-time patient satisfaction at emergency department,” *International Journal for Quality in Health Care*, vol. 29, n°. 5, p. 722–727, 2017.
- [8] D. B. Diercks, “Prolonged emergency department stays of non-ST-segment-elevation myocardial infarction patients are associated with worse adherence to the American College of Cardiology/American Heart Association guidelines for management and increased adverse events,” *Annals of emergency medicine*, vol. 50, n°. 5, p. 489–496, 2007.
- [9] T. G. Shin, “The adverse effect of emergency department crowding on compliance with the resuscitation bundle in the management of severe sepsis and septic shock,” *Critical Care*, vol. 17, p. 1–11, 2013.
- [10] G. De Luca, “Time delay to treatment and mortality in primary angioplasty for acute myocardial infarction : every minute of delay counts,” *Circulation*, vol. 109, n°. 10, p. 1223–1225, 2004.
- [11] S. L. Hsieh, “Association between the time to definitive care and trauma patient outcomes : every minute in the golden hour matters,” *European Journal of Trauma and Emergency Surgery*, p. 1–8, 2022.
- [12] M. A. Hajbagheri, “Factors influencing survival after in-hospital cardiopulmonary resuscitation,” *European Journal of Trauma and Emergency Surgery*, vol. 6, n°. 3, p. 317–321, 2022.

2005.

- [13] L. Cardoso, “Impact of delayed admission to intensive care units on mortality of critically ill patients : a cohort study,” *Critical care*, vol. 15, p. 1–8, 2011.
- [14] S. D. Cline, “Expedited admission of patients decreases duration of mechanical ventilation and shortens icu stay,” *The American journal of emergency medicine*, vol. 27, n°. 7, p. 843–846, 2009.
- [15] M. L. Durie, “A “code icu” expedited review of critically ill patients is associated with reduced emergency department length of stay and duration of mechanical ventilation,” *Journal of Critical Care*, vol. 42, p. 123–128, 2017.
- [16] R. García-Gigorro, “Impact on patient outcome of emergency department length of stay prior to icu admission,” *Medicina Intensiva*, vol. 41, n°. 4, p. 201–208, 2017.
- [17] J. Pines, “The effect of emergency department crowding on patient satisfaction for admitted patients,” *Academic Emergency Medicine*, vol. 15, p. 825–831, 2008.
- [18] S. Healy, “Stress in emergency departments : experiences of nurses and doctors,” *Emergency nurse*, vol. 19, n°. 4, 2011.
- [19] J. A. V. Sawatsky, “Exploring the key predictors of retention in emergency nurses,” *Journal of nursing management*, vol. 20, n°. 5, p. 696–707, 2012.
- [20] S. Mahler, “The impact of emergency department overcrowding on resident education,” *The Journal of emergency medicine*, vol. 42, n°. 1, p. 69–73, 2012.
- [21] R. J. Salway, “Emergency department (ed) overcrowding : evidence-based answers to frequently asked questions,” *Revista Médica Clínica Las Condes*, vol. 28, n°. 2, p. 213–219, 2017.
- [22] M. Foley, “Financial impact of emergency department crowding,” *Western Journal of Emergency Medicine*, vol. 12, n°. 2, p. 192, 2011.
- [23] D. M. Fatovich, “Access block causes emergency department overcrowding and ambulance diversion in perth, western australia,” *Emergency medicine journal*, vol. 22, n°. 5, p. 351–354, 2005.
- [24] M. Knapman, “Overcrowding in medium-volume emergency departments : effects of aged patients in emergency departments on wait times for non-emergent triage-level patients,” *International Journal of Nursing Practice*, vol. 16, n°. 3, p. 310–317, 2010.
- [25] M. van der Linden, “Two emergency departments, 6000 km apart : Differences in patient flow and staff perceptions about crowding,” *International Emergency Nursing*, vol. 35, p. 30–36, 2017.
- [26] A. J. Forster, “The effect of hospital occupancy on emergency department length of stay and patient disposition,” *Academic Emergency Medicine*, vol. 10, n°. 2, p. 127–133, 2003.

- [27] D. Fatovich, "Access block causes emergency department overcrowding and ambulance diversion in perth, western australia," *Emergency Medicine Journal*, vol. 22, n^o. 5, p. 351–354, 2005.
- [28] M. B. Ospina, "Key indicators of overcrowding in canadian emergency departments : a delphi study," *Canadian Journal of Emergency Medicine*, vol. 9, n^o. 5, p. 339–346, 2007.
- [29] F. Harrou, "Forecasting emergency department overcrowding : A deep learning framework," *Chaos, Solitons Fractals*, vol. 139, p. 110247, 2020.
- [30] X. Zhao, "Predicting hospital emergency department visits with deep learning approaches," *Biocybernetics and Biomedical Engineering*, vol. 42, n^o. 3, p. 1051–1065, 2022.
- [31] J. Tuominen, "Forecasting daily emergency department arrivals using high-dimensional multivariate data : a feature selection approach." *BMC Medical Informatics and Decision Making*, vol. 22, n^o. 134, 2022.
- [32] O. Araz, D. Olson et A. Ramirez-Nafarrate, "Predictive analytics for hospital admissions from the emergency department using triage information," *International Journal of Production Economics*, vol. 208, p. 199–207, 2019.
- [33] H. Seo, "Prediction of hospitalization and waiting time within 24 hours of emergency department patients with unstructured text data." *Health Care Management Science*, vol. 27, p. 114–129, 2024.
- [34] Z. King, "Machine learning for real-time aggregated prediction of hospital admission for emergency patients." *npj Digital Medicine*, vol. 5, n^o. 104, 2022.
- [35] F. Kadri, "Towards accurate prediction of patient length of stay at emergency department : a gan-driven deep learning framework," *Journal of Ambient Intelligence Humanized Computing*, vol. 14, p. 11 481—11 495, 2023.
- [36] C. Ricciardi, "Evaluation of different machine learning algorithms for predicting the length of stay in the emergency departments : a single-centre study." *Frontiers in Digital Health*, vol. 5, p. 1323849, 2024.
- [37] B. Medina, "Multi-step forecasting of waiting time on emergency department overcrowding using multilayer perceptron neural network algorithm," 2020, iEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC).
- [38] Y. Kuo, "An integrated approach of machine learning and systems thinking for waiting time prediction in an emergency department," *International Journal of Medical Informatics*, vol. 139, p. 104143, 2020.
- [39] S. J. Weiss, "Estimating the degree of emergency department overcrowding in academic medical centers : results of the national ed overcrowding study (nedocs)," *Academic emergency medicine*, vol. 11, n^o. 1, p. 38–50, 2004.

- [40] —, “Development of a novel measure of overcrowding in a pediatric emergency department,” *Pediatric emergency care*, vol. 23, n^o. 9, p. 641–645, 2007.
- [41] —, “Evaluating community ed crowding : the community ed overcrowding scale study,” *The American journal of emergency medicine*, vol. 32, n^o. 11, p. 1357–1363, 2014.
- [42] S. Bernstein, “Development and validation of a new index to measure emergency department crowding,” *Academic Emergency Medicine*, vol. 10, n^o. 9, p. 938–942, 2003.
- [43] S. Weiss, “Comparison of the national emergency department overcrowding scale and the emergency department work index for quantifying emergency department crowding,” *Academic Emergency Medicine*, vol. 13, n^o. 5, p. 513–518, 2006.
- [44] V. Ahalt, “Comparison of emergency department crowding scores : a discrete-event simulation approach,” *Health care management science*, vol. 21, p. 144–155, 2018.
- [45] K. L. Tekwani, “Emergency department crowding is associated with reduced satisfaction scores in patients discharged from the emergency department,” *Western Journal of Emergency Medicine*, vol. 14, n^o. 1, p. 11, 2013.
- [46] S. M. Schneider, “The effect of crowding on the workload and the perception of workload in a community emergency department,” *Annals of Emergency Medicine*, vol. 58, n^o. 4, p. 187, 2011.
- [47] E. B. Kulstad, “Ed overcrowding is associated with an increased frequency of medication errors,” *The American journal of emergency medicine*, vol. 28, n^o. 3, p. 304–309, 2010.
- [48] E. Kulstad, “Occupancy rate and edwin score both correlate significantly with number of patients leaving the emergency department without being seen,” *Annals of Emergency Medicine*, vol. 52, n^o. 4, p. S85, 2008.
- [49] L. Breiman, *Classification and regression trees*. Routledge, 2017.
- [50] H. Li, “Decision tree model for predicting in-hospital cardiac arrest among patients admitted with acute coronary syndrome,” *Clinical Cardiology*, vol. 42, n^o. 11, p. 1087–1093, 2019.
- [51] W. Y. Loh, “Classification and regression trees,” *Wiley interdisciplinary reviews : data mining and knowledge discovery*, vol. 1, n^o. 1, p. 14–23, 2011.
- [52] N. Speybroeck, “Classification and regression trees,” *International journal of public health*, vol. 57, p. 243–246, 2012.
- [53] Y. Zhao, “Comparison of decision tree methods for finding active objects,” *Advances in Space Research*, vol. 41, n^o. 12, p. 1955–1959, 2008.
- [54] L. Breiman, “Random forests,” *Machine learning*, vol. 45, p. 5–32, 2001.
- [55] J. Ali, “Random forests and decision trees,” *International Journal of Computer Science Issues*, vol. 9, n^o. 5, p. 272, 2012.

- [56] K. Lin, “Predicting in-hospital mortality of patients with acute kidney injury in the icu using random forest model,” *International journal of medical informatics*, vol. 125, p. 55–61, 2019.
- [57] T. Zhu, “Analysis on the applicability of the random forest,” *In Journal of Physics : Conference Series*, vol. 1607, n°. 1, p. 012123, 2020.
- [58] G. Biau et E. Scornet, “A random forest guided tour,” *Test*, vol. 25, p. 197–227, 2016.
- [59] J. H. Friedman, “Greedy function approximation : a gradient boosting machine,” *Annals of statistics*, p. 1189–1232, 2001.
- [60] A. Natekin et A. Knoll, “Gradient boosting machines, a tutorial,” *Frontiers in neurorobotics*, vol. 7, p. 21, 2013.
- [61] M. Klug, “A gradient boosting machine learning model for predicting early mortality in the emergency department triage : devising a nine-point triage score,” *Journal of general internal medicine*, vol. 35, p. 220–227, 2020.
- [62] D. Bertsimas et J. Dunn, “Optimal classification trees,” *Machine Learning*, vol. 106, p. 1039–1082, 2017.
- [63] D. Bertsimas, “Applied informatics decision support tool for mortality predictions in patients with cancer,” *JCO clinical cancer informatics*, vol. 2, p. 1–11, 2018.
- [64] M. Sartini, “Overcrowding in emergency department : causes, consequences, and solutions—a narrative review,” *Healthcare*, vol. 10, n°. 9, p. 1625, 2022.
- [65] A. Dehghani, “A quantitative comparison of overlapping and non-overlapping sliding windows for human activity recognition using inertial sensors,” *Sensors*, vol. 19, n°. 22, p. 5026, 2019.
- [66] S. M. Lundberg et S.-I. Lee, “A unified approach to interpreting model predictions,” *Advances in neural information processing systems*, vol. 30, 2017.
- [67] S. Dandl, “Multi-objective counterfactual explanations,” *International Conference on Parallel Problem Solving from Nature. Cham : Springer International Publishing*, p. 448–469, 2020.
- [68] M. Otegbeye, “Designing a data-driven decision support tool for nurse scheduling in the emergency department : a case study of a southern new jersey emergency department,” *Journal of emergency nursing*, vol. 41, n°. 1, p. 30–35, 2015.

ANNEXE A VARIABLES INDÉPENDANTES INTÉGRÉES AUX MODÈLES

TABLEAU A.1 Variables indépendantes intégrées aux modèles

Catégorie	Sous-catégorie	Types	Mesures
Examens diagnostiques	Imagerie	Radiographies Échographies CT-scan IRM	Nombre de tests en attente
	Prises de sang	Groupées en 1 minute Groupées en 1 heure	Délai moyen pour les tests en attente
	Analyses d'urine		
	PCR		
Services à l'urgence	Triage		Nombre de patients en attente Délai moyen pour les patients en attente
	Visite médicale	Médecin d'urgence	
	Consultation spécialiste	Médecine Chirurgie Psychiatrie Obstétrique-Gynécologie Médecine générale	
	Civières		Taux d'occupation
Admissions	Patients sur civère	Médecine Chirurgie Psychiatrie Obstétrique-Gynécologie Médecine générale	Nombre de patients en attente Délai moyen pour les patients en attente
Hospitalisations	Patients sur lits	Médecine Chirurgie Psychiatrie Obstétrique-Gynécologie Médecine générale	Taux d'occupation Durée moyenne de séjour
Dates	Heure Jour Mois		
Scores EDWIN passés	- 12 h - 24 h - 48 h		Valeur numérique Classe

ANNEXE B RAISONS DE CONSULTATION À L'URGENCE

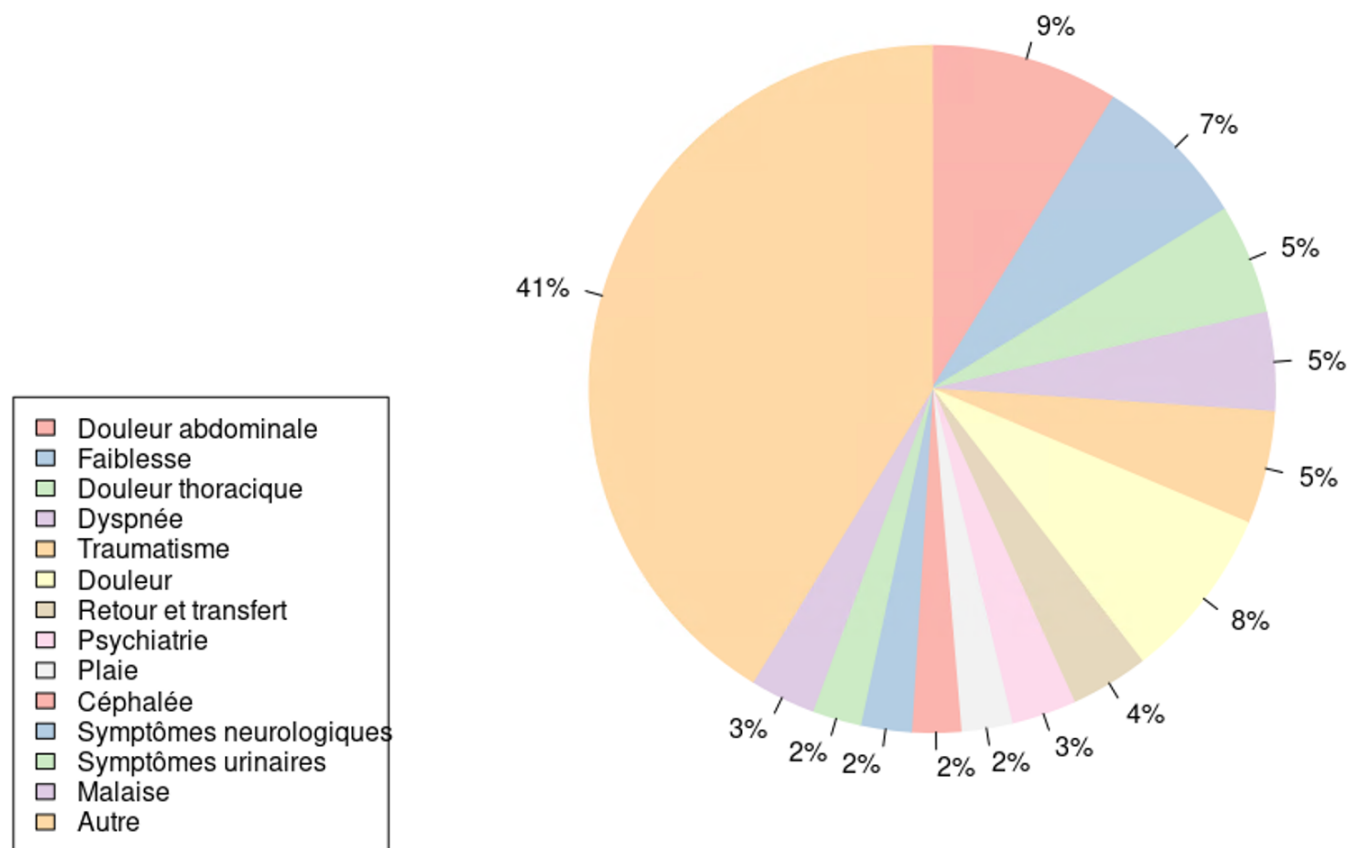


FIGURE B.1 Raisons de consultations des patients ayant visité l'urgence pendant la période étudiée