

Intelligence artificielle en médecine interne: développement d'un modèle prédictif des durées de séjour

JÉRÉMIE DESPRAZ^a, Dr ANTOINE GARNIER^b, Dre MARIE MÉAN^b, Dr JULIEN VAUCHER^b, Dre VANESSA KRAEGE^b et Pr PETER VOLLENWEIDER^b

Rev Med Suisse 2021; 17: 2042-8

Une prise en charge efficiente des patients nécessite une planification minutieuse des soins en fonction de la pathologie et des contraintes hospitalières. Dans ce contexte, une estimation de la durée de séjour permet de mieux coordonner les tâches interprofessionnelles, de gérer le flux des patients et d'anticiper la préparation à la sortie. Cet article présente la construction et l'évaluation d'un modèle prédictif de la durée de séjour à l'aide de données cliniques présentes à l'admission dans un service de médecine interne universitaire. Nous démontrons que les approches basées sur le Machine Learning sont capables de prédire des durées de séjour avec une performance similaire à celle des professionnels.

Artificial Intelligence in internal medicine: development of a model predicting length of stay for non-elective admissions

Efficient management of hospitalized patients requires carefully planning each stay by taking into account patients' pathologies and hospital constraints. Therefore, the ability to accurately estimate length of stays allows for better interprofessional tasks coordination, improved patient flow management, and anticipated discharge preparation. This article presents how we built and evaluated a predictive model of length of stay based on clinical data available upon admission to a division of internal medicine. We show that Machine Learning-based approaches can predict lengths of stay with a similar level of accuracy as field experts.

INTRODUCTION

Les services de médecine interne sont confrontés à une patientèle en grande majorité âgée et polymorbide qui requiert une prise en charge souvent complexe et des efforts conséquents en matière d'organisation des soins et de préparation à la sortie. Ainsi, depuis 2017, un grand effort a été déployé dans le Service de médecine interne (SMI) du CHUV à Lausanne (Suisse) pour optimiser la durée moyenne de séjour en mettant en place une réorganisation de la journée de travail, en instaurant un colloque interprofessionnel et en favorisant une meilleure préparation à la sortie. Une estimation quotidienne de la durée de séjour permet de mieux coordonner le

travail interprofessionnel, l'enseignement thérapeutique, et de procéder à la réconciliation médicamenteuse, pour sécuriser le transfert vers l'ambulatoire et potentiellement diminuer les réadmissions. Connaître la date prévisionnelle de sortie permet aussi de mieux informer le patient et ses proches du projet hospitalier.

Actuellement, l'estimation de la durée de séjour se base sur une appréciation de l'équipe médico-soignante qui fixe une date prévisionnelle de sortie lors d'une rencontre d'équipe chaque matin du lundi au vendredi. Cette estimation se fait par les médecins, l'infirmier-ère de liaison et l'infirmier-ère-chef-fe d'unité en fonction notamment de la pathologie du patient, de son contexte social et de son évolution depuis l'entrée. Elle est très fortement basée sur l'expérience. Par ailleurs, le SMI est régulièrement confronté à des cas complexes pour lesquels des diagnostics sont difficiles à établir et dont les complications peuvent modifier grandement la date de retour à domicile. Dans ce contexte, avoir une estimation plus objective et fiable de la durée de séjour améliore significativement l'anticipation et la gestion du travail de l'équipe.

Plusieurs études ont montré que des modèles informatiques sont capables de prédire des durées de séjour à partir de données cliniques.¹⁻⁴ Des sociétés privées proposent des logiciels capables de calculer une date de sortie prévisionnelle.⁵⁻⁷ Dans ce contexte, le SMI et le groupe *Data Science* du CHUV ont collaboré pour mesurer la capacité de tels modèles à produire une estimation fiable de la date de sortie dans un délai de 48 heures après l'admission. Nous présentons ici les méthodes développées et des résultats préliminaires démontrant la faisabilité d'une telle approche.

MATÉRIEL ET MÉTHODES

Cohorte

Les données sources proviennent des dossiers électroniques des patients adultes hospitalisés dans le SMI du CHUV via les urgences entre le 1^{er} novembre 2017 et le 29 février 2020.

Parmi ces séjours, ont été exclus:

- Ceux liés à des patients s'étant opposés à la réutilisation de leurs données pour la recherche.
- Ceux se terminant par un décès.
- Ceux suivis d'une réadmission au CHUV dans un délai de moins de 30 jours.
- Ceux dont la durée passée au SMI était inférieure à 75% de la durée totale.

^aGroupe Data Science, Direction des systèmes d'information, Département infrastructures, Centre hospitalier universitaire vaudois, 1011 Lausanne, ^bService de médecine interne, Département de médecine, Centre hospitalier universitaire vaudois, 1011 Lausanne
jeremie.despraz@chuv.ch | antoine.garnier@chuv.ch
marie.mean@chuv.ch | julien.vaucher@chuv.ch
vanessa.kraege@chuv.ch | peter.vollenweider@chuv.ch

Traiter individuellement. Patient après patient. Patiente après patiente.

Pour la thérapie au magnésium
selon le besoin.¹



Magnesiocard®

Magnesiocard® (formes orales)

C: Magnesii aspartatis hydrochloridum trihydricum. **I:** Carence en magnésium, troubles du rythme cardiaque, besoins accrus liés à la pratique sportive de haut niveau et pendant la grossesse, éclampsie et pré-éclampsie, tétanie, crampes dans les mollets, myoclonies, jambes sans repos (restless legs). **P:** De 4.5 mg (= 0.185 mmol) à 9 mg (= 0.37 mmol) de magnésium par kg de poids corporel / 10 – 20 mmol de magnésium par jour, en 1 – 3 prises orales selon la forme d'administration (granulés, comprimés effervescents, comprimés pelliculés). **CI:** Hypersensibilité à l'un des composants du médicament. **P:** Insuffisance rénale. Il est indispensable de surveiller la concentration sérique de magnésium chez les insuffisants rénaux. Magnesiocard 7.5 mmol: ne pas utiliser en cas de phénylcétonurie. **IA:** Les tétracyclines et Magnesiocard devraient être pris à 3 – 4 heures d'intervalle (inhibition mutuelle au niveau de l'absorption). Tendance à l'hypercalcémie lors de l'administration concomitante de magnésium et de cholécalférol. **G/A:** Peut être administré. **EI:** Occasionnellement: troubles gastro-intestinaux. **E:** Comprimés pelliculés (2.5 mmol) 50, 100; granulés (5 mmol) citron et granulés (5 mmol) orange 20*, 50*; comprimés effervescents (7.5 mmol) 20*, 60*; granulés (10 mmol) grapefruit et granulés (10 mmol) orange 20*, 50*. **Cat. B.** Pour des informations détaillées, voir www.swissmedicinfo.ch. *admis par les caisses-maladie V030820

Références: 1: www.swissmedicinfo.ch, consulté le 15.10.2020.

- Ceux dont le transfert au SMI a eu lieu plus de 48 heures après l'entrée.
- Ceux comportant un séjour dans une unité de soins intensifs.

Durée de séjour et qualité de la prédiction

La durée de séjour est définie comme le temps entre l'admission aux urgences et la sortie: retour à domicile, transfert ou reclassement en lit d'attente vers un centre de traitement et de réadaptation ou un établissement médico-social. Cette durée sert de référence pour l'entraînement du modèle et pour le calcul de l'erreur de prédiction. La date de sortie prévisionnelle estimée par les médecins et soignants est aussi utilisée comme point de comparaison. Celle-ci est calculée à partir des données consignées dans le dossier patient informatisé lors des colloques d'équipe du matin. Lorsqu'elle est disponible, nous utilisons la dernière valeur enregistrée à 48 heures après l'admission.

Préparation des données

Les données sources proviennent du dossier informatique du patient et sont collectées jusqu'à maximum 48 heures après l'admission. En raison de la grande quantité de variables, la prédiction repose sur des paramètres présélectionnés en fonction de quatre critères:

1. L'expertise médicale.
2. La corrélation des variables avec la durée de séjour via une analyse univariée.
3. La complétude des données (au moins 50% de données valides sur l'ensemble des séjours).
4. La robustesse des variables, à savoir peu d'influence externe ou humaine sur les valeurs mesurées.

Les sources de données retenues totalisent 262 variables distinctes.

Les données présentes dans un format complexe ou non structuré (par exemple, dates, textes) sont transformées de manière à être représentées sous une forme numérique. Les informations sous forme textuelle sont exprimées par des vecteurs encapsulant leur signification en utilisant la méthode Word Embeddings.^{8,9} Ainsi, les groupes de mots proches sémantiquement se retrouvent avec une représentation numérique similaire. Cela permet d'effectuer des opérations mathématiques sur ces mots dans le but, par exemple, de mesurer la similitude entre certains concepts. Un exemple est présenté sur la **figure 1** illustrant l'utilité de ce type de représentation.

Les données ayant une composante temporelle sont agrégées par périodes de 24 heures pour n'en conserver que la valeur moyenne et l'écart-type (**figure 2**). Cette approche permet d'obtenir une représentation identique pour tous les séjours, indépendamment du fait que certaines mesures aient pu être effectuées à divers moments de la journée ou avec des fréquences différentes. En outre, la valeur de référence de chaque variable ainsi que les variations éventuelles sont capturées par ces deux statistiques.

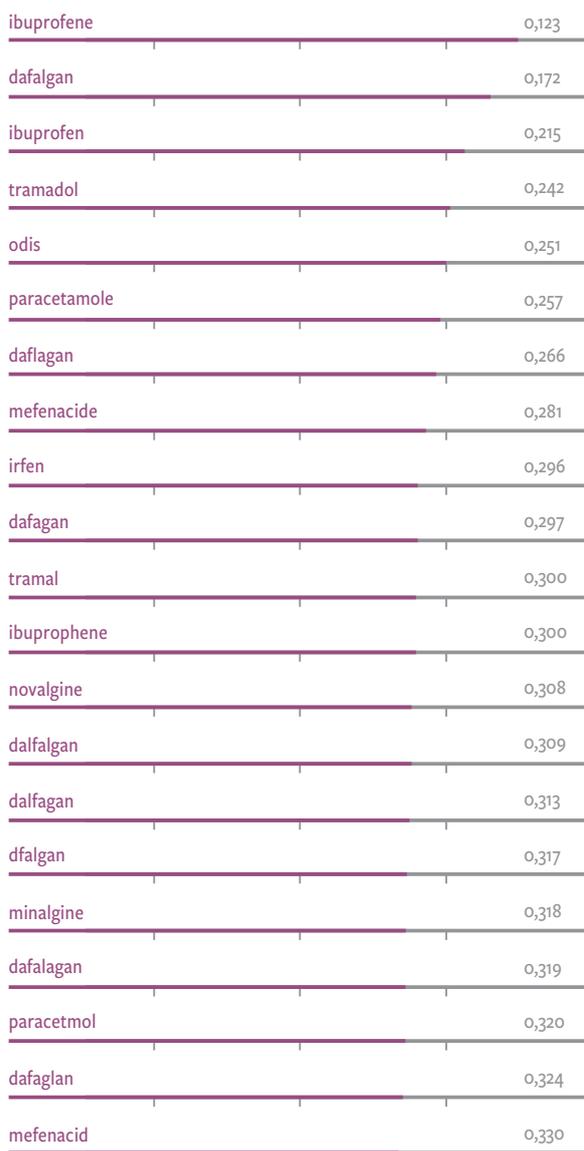
Rôle des facteurs absents du dossier informatique

Il apparaît que de multiples facteurs externes influencent de manière significative la durée de séjour, comme le jour de la

FIG 1 Méthode Word Embeddings

Mots les plus proches sémantiquement de «paracétamol» dans l'espace numérique des «Word Embeddings». Plus la distance est faible, plus la similarité est grande et plus ce mot se retrouve en haut de la liste. La distance est indiquée par le nombre sur la droite de la figure et par la barre au-dessous du mot.

Points les plus proches dans l'espace d'origine



semaine (**figure 3**), les périodes de vacances et les jours fériés. Si disponibles, ces variables ont été incluses dans les données sources utilisées par le modèle.

Création et entraînement du modèle prédictif

Afin de capturer au mieux les dépendances temporelles des différentes variables présentes dans le dossier patient, nous avons choisi le modèle de réseau de neurones récurrent Long Short-Term Memory (LSTM).¹⁰ Trois réseaux de neurones ont été construits en parallèle, l'un permettant de déterminer

FIG 2 Agrégation des données du dossier patient

Représentation schématique de l'agrégation des données du dossier patient sur un intervalle de temps donné. Les abréviations A, T et FC représentent ici les admissions de médicament, les mesures de température et de la fréquence cardiaque, respectivement.

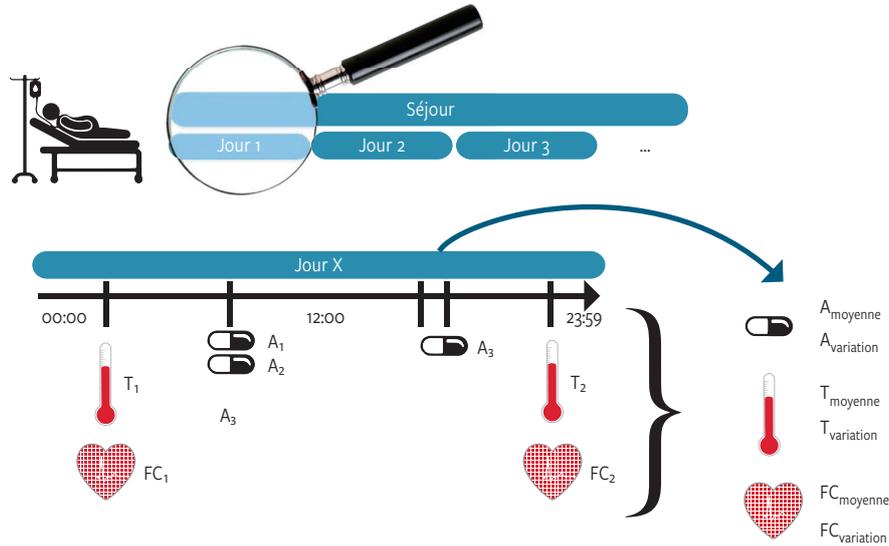
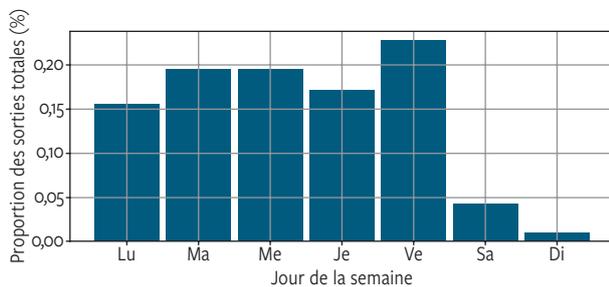


FIG 3 Distribution des sorties en fonction du jour de la semaine



semaine (figure 3). L'architecture complète de la solution adoptée est présentée sur la figure 4.

Pour déterminer la performance du modèle, la cohorte a été divisée en deux sous-ensembles, l'un servant à l'optimisation du modèle et l'autre au calcul des erreurs. Ce processus est répété six fois afin d'obtenir une mesure pour l'ensemble des séjours (méthode de validation croisée). Les prédictions du modèle sont ensuite comparées:

- À celles faites par les médecins et soignants.
- À une référence: la durée de séjour moyenne des patients de la cohorte.

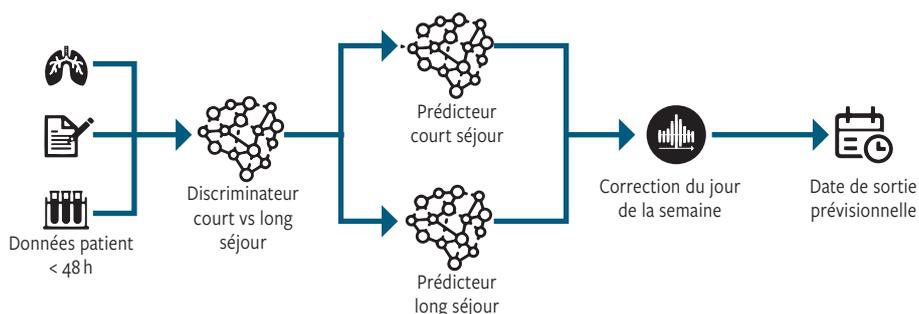
si le séjour serait plus long ou plus court que la moyenne et les deux autres spécialisés dans les prédictions de longs et de courts séjours, respectivement. Finalement, un module appliquant des corrections selon une analyse bayésienne a été ajouté en sortie du modèle afin de s'assurer que les dates de sortie prévisionnelles soient correctement réparties dans la

Post-traitement des résultats

Estimation de l'incertitude

L'agrégation de résultats générés par le modèle peut fournir des indications sur l'incertitude de la prédiction. Pour ce faire, plusieurs prédictions peuvent être calculées simultanément en modifiant légèrement les données d'entrée du modèle pour un séjour donné. Les séjours pour lesquels la sensibilité

FIG 4 Processus de prédiction de la date de sortie prévisionnelle à partir des données patient



aux données d'entrée est grande généreront une variabilité plus importante dans les prédictions. En analysant ensuite la distribution de ces prédictions, il est possible d'estimer l'incertitude du modèle.

Prédictions agrégées

L'agrégation de résultats générés par le modèle peut aussi fournir des indications sur le service dans son ensemble. Ainsi, en générant des dates de sortie prévisionnelles pour tous les patients hospitalisés à un temps donné, il devient possible d'estimer le nombre de lits occupés dans le service ainsi que le nombre de sorties qui vont se produire dans un jour ou une semaine à venir. De plus, l'agrégation des prédictions permet d'obtenir des résultats robustes, les erreurs de prédiction individuelles s'annulant mutuellement.

RÉSULTATS

La cohorte étudiée contient un total de 6053 séjours pour 4823 patients. La durée médiane de séjour se situe à 6,8 jours et l'âge médian à 76 ans. Les figures 5 et 6 représentent les distributions d'âges et de durées de séjour.

Le tableau 1 et la figure 7 représentent les performances obtenues par le modèle en comparaison avec les prévisions

FIG 5 Distribution des âges des patients inclus dans la cohorte

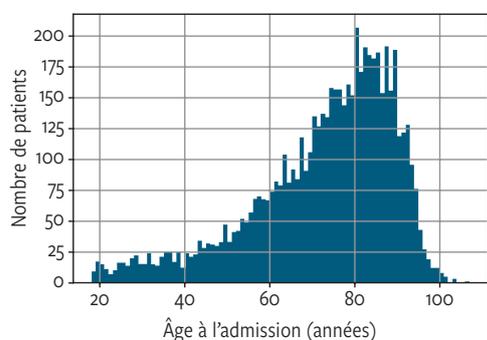


FIG 6 Distribution des durées de séjour des patients inclus dans la cohorte

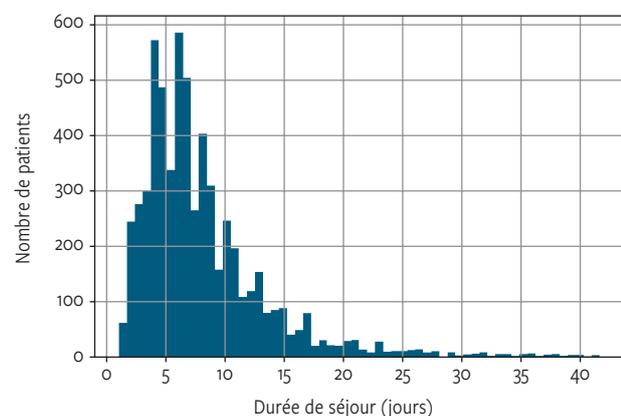
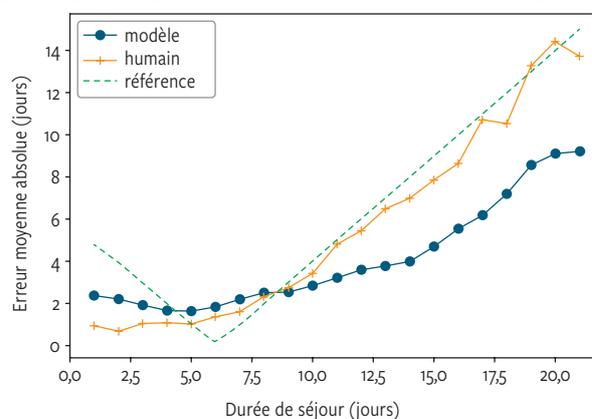


TABLEAU 1 Performances du modèle

	Référence (prédiction = durée de séjour moyenne)	Prédiction par l'humain	Prédiction par le modèle
Erreur moyenne absolue (jours)	3,62	2,77	2,82

FIG 7 Erreur moyenne absolue entre la prédiction et la durée réelle de séjour

Le modèle est représenté en bleu, les médecins et soignants en orange et la valeur de référence (durée moyenne des séjours) en vert. Plus les courbes sont proches de la ligne verticale correspondant à y = 0, plus les prédictions sont précises. Les courbes se séparent pour des durées de séjour de 10 jours et plus où le modèle devient meilleur que les humains.



humaines et celles de la référence. Il est intéressant de noter que seuls 38% des séjours possèdent une date prévisionnelle de sortie renseignée par les médecins et soignants 48 heures après l'admission aux urgences, alors que le modèle est en mesure de générer une prédiction pour tous les séjours de la cohorte.

Mesure de l'incertitude

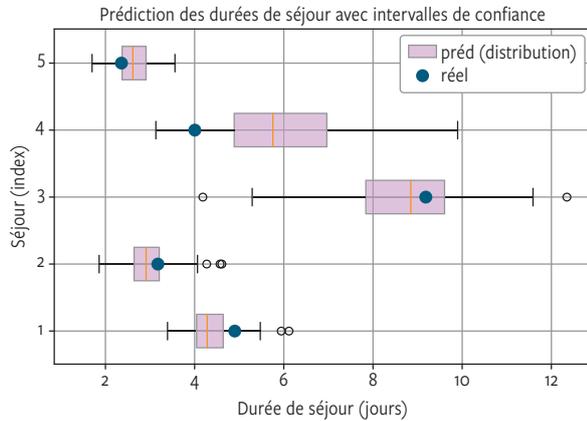
La figure 8 représente la distribution des prédictions pour 5 séjours choisis aléatoirement parmi les patients de la cohorte. Ces résultats rendent possible une estimation de l'incertitude sur la date de sortie pour chacun de ces séjours. En règle générale, plus les séjours sont courts, plus l'incertitude est faible. Ces résultats peuvent permettre de juger de la crédibilité de la prédiction si celle-ci devait s'inscrire dans un processus décisionnel.

Résultats agrégés

Les figures 9 et 10 illustrent la qualité de la prédiction des résultats agrégés par rapport aux données réelles de la période étudiée. La figure 9 représente le nombre de lits occupés par jour dans le service et la figure 10 celui des sorties chaque semaine de la période étudiée. Les plages vertes représentent les périodes de vacances scolaires dans le canton de Vaud.

FIG 8 Durées réelles vs prédictions du modèle

Comparaison entre les durées réelles et les prédictions du modèle incluant les incertitudes. Les données sont celles de 5 séjours correspondant à 5 patients de la cohorte, représentés ici à titre d'exemple. Les lignes continues indiquent les max et min 25% et la zone colorée la région où tombent les 50% des prédictions les plus proches de la médiane.



DISCUSSION

Nous avons développé un modèle qui illustre les possibilités d'application de l'apprentissage automatique pour la prédiction de la durée de séjours hospitaliers dans un service de médecine interne. Nous avons pu observer que les performances du modèle sont comparables à celles des médecins et soignants, avec une erreur plus grande sur les courts séjours (moins d'une semaine) mais s'améliorant progressivement pour les séjours plus longs. De plus, le modèle est en mesure de générer une prédiction pour tous les séjours de la cohorte, alors que 60% des séjours ne possèdent pas de date prévisionnelle de sortie renseignée par les médecins et soignants 48 heures après l'admission. L'agrégation des prédictions individuelles permet d'effectuer des prévisions globales et robustes au niveau du service, notamment en matière d'occupation des lits et d'anticipation du nombre de sorties.

Dans l'optique d'une intégration de la prédiction informatisée dans l'organisation des soins, une sensibilisation du personnel serait nécessaire afin de s'assurer de la bonne compréhension des possibilités et des limites du modèle par les utilisateurs. En effet, cette prédiction ne constitue qu'une information

FIG 9 Comparaison du nombre de lits occupés en fonction des durées de séjour

Nombre de lits occupés par jour dans le Service de médecine interne (prédiction en rouge et valeur réelle en bleu) en fonction des durées de séjour et des périodes de vacances du canton de Vaud (zones vertes).

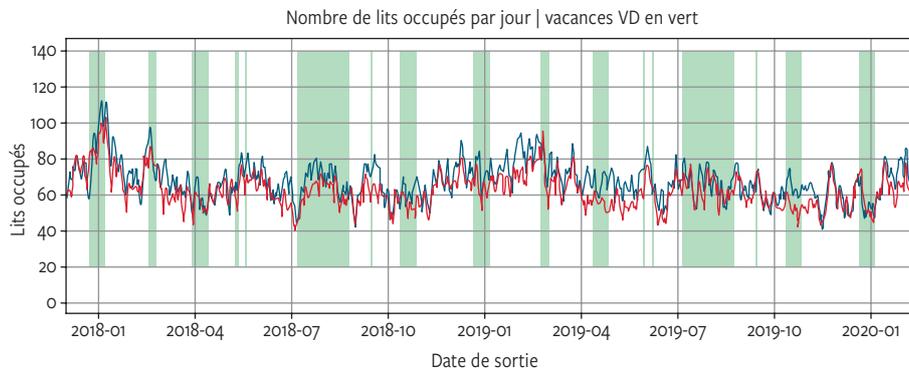
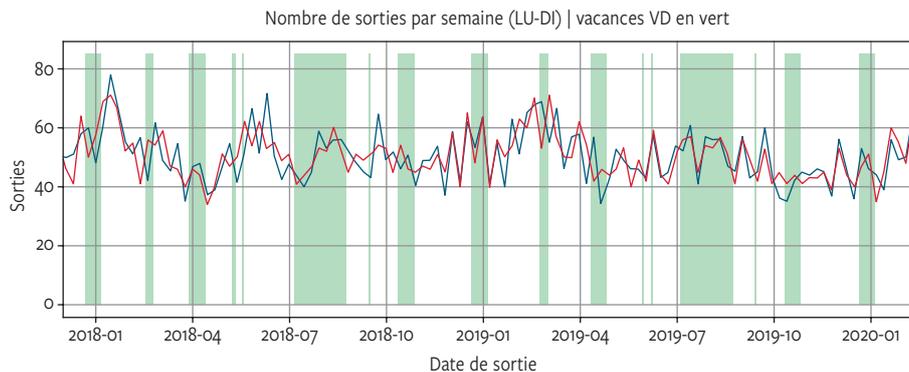


FIG 10 Comparaison du nombre de sorties par semaine en fonction des durées de séjour

Nombre de sorties par semaine dans le Service de médecine interne (prédiction en rouge et valeur réelle en bleu) en fonction des durées de séjour et des périodes de vacances du canton de Vaud (zones vertes).



supplémentaire, dont l'utilité doit être intégrée tant pour des décisions sur un cas individuel que pour la gestion de l'ensemble d'un service.

Du fait de complications imprévisibles pendant l'hospitalisation (chutes, infections, etc.), une limite de performance est inévitable. Le modèle pourrait encore augmenter en précision en incluant les aspects organisationnels du séjour (tourneurs des médecins, planning des soins, réservations et utilisation des ressources matérielles, etc.) qui ne font pas strictement partie du dossier patient informatisé.

CONCLUSION

Au travers de cette étude, nous avons présenté les étapes de conception et d'analyse des résultats d'un modèle d'apprentissage automatique capable de prédire des dates de sortie pour toutes les entrées non électives dans le SMI du CHUV.

Conflit d'intérêts: Les auteurs n'ont déclaré aucun conflit d'intérêts en relation avec cet article

IMPLICATIONS PRATIQUES

- Les outils basés sur l'apprentissage automatique sont à présent capables de prédire la durée de séjour de patients hospitalisés
- Le calcul de la date de sortie peut se faire avec les données collectées dans le dossier électronique du patient
- Les prédictions peuvent être utilisées dans le but de mieux planifier les soins et la préparation à la sortie
- Les prédictions individuelles de la durée de séjour peuvent être agrégées afin de créer des indicateurs robustes pour le service entier

1 Wilk M, Marsh DWR, De Freitas S, Prowle J. Predicting Length of Stay in Hospital Using Electronic Records Available at the Time of Admission. *Stud Health Technol Inform* 2020;270:377-1.

2 *Rajkomar A, Oren E, Chen K, et al. Scalable and Accurate Deep Learning with Electronic Health Records. *NPJ Digit Med* 2018;1:1-10.

3 Bacchi S, Gluck S, Tan Y, et al. Prediction of General Medical Admission Length of Stay with Natural Language Processing and Deep Learning: a Pilot

Study. *Intern Emerg Med* 2020;15:989-95.

4 *Lu M, Sajobi T, Lucyk K, Lorenzetti D, Quan H. Systematic Review of Risk Adjustment Models of Hospital Length of Stay (LOS). *Med Care* 2015;53:355-65.

5 Calyps SA. Calyps SANIIA (En ligne). 2021 (cité le 22 août 2021). Disponible sur : www.calyps.ch/en/calyps-saniiia-artificialintelligence/

6 Qventus. Qventus for Inpatient Units (En ligne). 2021 (cité le 22 août 2021).

7 Microsoft Machine Learning Services.

Predicting Hospital Length of Stay (En ligne). 2021 (cité le 22 août 2021). Disponible sur : microsoft.github.io/r-server-hospital-length-of-stay/

8 Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation of Word Representations in Vector Space. *arXiv:13013781 (cs)* (En ligne). 2013 (cité le 7 septembre 2013). Disponible sur : arxiv.org/abs/1301.3781

9 Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed Representations of Words and Phrases

and their Compositionality. In: *Advances in Neural Information Processing Systems*. Curran Associates, Inc. (En ligne). 2013 (cité le 07 septembre 2013). Available from: proceedings.neurips.cc/paper/2013/hash/9aa42b31882ec039965f-3c4923ce901b-Abstract.html

10 Hochreiter S, Schmidhuber J. Long Short-Term Memory. *Neural Comput* 1997;9:1735-80.

* à lire

** à lire absolument