HAL open science

Approche basée sur les réseaux de neurones récurrents pour la prévision de la demande de soins aux urgences hospitalières

Farid Kadri, Kahina Abdennbi, Sondès Chaabane, Issam Nouaouri

To cite this version:

Farid Kadri, Kahina Abdennbi, Sondès Chaabane, Issam Nouaouri. Approche basée sur les réseaux de neurones récurrents pour la prévision de la demande de soins aux urgences hospitalières. 10ème conférence Francophone en Gestion et Ingénierie des Systèmes Hospitaliers, GISEH2020, Oct 2020, Valenciennes (visioconférence), France. hal-03199566

HAL Id: hal-03199566

https://uphf.hal.science/hal-03199566v1

Submitted on 15 Apr 2021

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

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

Approche basée sur les réseaux de neurones récurrents pour la prévision de la demande de soins aux urgences hospitalières

KADRI Farid[1], ABDENNBI Kahina[2], CHAABANE Sondès[3], NOUAOURI Issam[4]

[1] Agence Aeroline Ccustomer Services, Sopra Steria Group, 31770 Colomiers.

[2] Service Communal d'Hygiène et de Santé, Mairie de Mantauban, 82000 les chaumes.

[3] Univ. Polytechnique Hauts-de-France, CNRS UMR 8201 - LAMIH - Laboratoire d'Automatique, de Mécanique et d'Informatique Industrielles et Humaines, F-59313 Valenciennes, France, (+33) (0) 3 27 51 13 55

[4] Université Artois, EA 3926, Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A), F-62400 Béthune, France.

Résumé. Au cours des dernières années, les besoins en soins d'urgences ont considérablement augmenté dans les services des urgences (SU). La gestion du flux de patients est l'un des principaux défis auxquels sont confrontés de nombreux SU. À cet égard, il est essentiel de prédire les demandes de soins afin de mieux adapter les ressources aux demandes de soins et d'atténuer les problèmes de surpopulation dans ces établissements. Les modèles de régression et de série chronologique existants, sont principalement des modèles linéaires et ne peuvent pas décrire la nature stochastique et non linéaire des données. Les approches basées sur l'apprentissage par réseaux de neurones sont considérées comme une nouvelle alternative pour la prévision dans divers domaines. Cet article présente une approche d'apprentissage par réseaux de neurones profonds basée sur le modèle GRU (*Gated Recurrent Unit*) pour prédire les demandes de soins dans un SU. Les données issues du service des urgences pédiatriques (SUP) du centre hospitalier régional de Lille (CHRU), sont utilisées pour tester et valider l'approche proposée. Les résultats montrent une bonne performance de prévision des admissions au SUP.

Mots clés : Services des urgences, surcharge, prévision, réseaux de neurones profonds, GRU

Introduction

Au cours des dernières décennies, en raison de la demande croissante aux soins d'urgences, la gestion des flux de patients et la gestion du personnel médical deviennent de plus en plus importantes dans les établissements hospitaliers, en particulier les services d'urgences (SU). Ces derniers, sont considérés comme la principale porte d'entrée de l'hôpital, car ils constituent un passage presque obligatoire pour tous les patients avant leur admission dans d'autres services hospitaliers (Boyle et al., 2012; Kadri et al., 2014a).

Les SU présentent une activité fluctuante et sujette à des situations imprévisibles et même exceptionnelles, souvent générées par un afflux important de patients. La gestion efficace des demandes de soins d'urgences est un défi auquel de nombreux SU sont confrontés. Ces établissements doivent donc pouvoir recevoir des flux de patients parfois très importants pour les traitements médicaux et chirurgicaux (Kadri et al., 2017, 2014b). Pour remplir leurs missions; Les SU doivent intégrer dans leur mode de fonctionnement la capacité d'anticiper et de prévoir le nombre d'admissions à l'avance pour pouvoir mieux gérer le flux des patients. La prévision des admissions aux SU fournit des informations utiles qui peuvent être utilisées dans l'allocation et

la planification des ressources (optimiser la répartition des ressources humaines et matérielles). La prévision du flux de charge en soins permet également l'amélioration de l'efficacité des services d'urgences en réduisant le nombre de patients dans les files d'attentes et en augmentant le nombre de passages (Aboagye-Sarfo et al., 2015; Benbelkacem et al., 2019).

La revue de la littérature montre que les approches classiques de prévision de la demande de soins comprennent les modèles de régression et les modèles de séries chronologiques (lissage exponentiel, ARMA et ses variantes), sont limitées en termes de capture de variations aléatoires, de caractéristiques non linéaires, de précision et d'horizon de prévision à long terme (Calegari et al., 2016; Jones et al., 2002; Kadri et al., 2019, 2014b; Steins et al., 2019). Les réseaux de neurones récurrents (RNN) ont été utilisés avec succès dans la résolution des problèmes d'apprentissage par réseaux de neurones profonds (algorithmes de Deep Learning). Ces modèles ont été proposés pour résoudre les problèmes d'apprentissage dépendant du temps (Hochreiter and Schmidhuber, 1997). Le modèle GRU (*Gated Recurrent Unit*) est l'un des modèles RNN les plus utilisés pour la prévision de séries chronologiques. Ce modèle est parfaitement adapté aux problèmes de prévision de la demande de soins d'urgences. Plusieurs études de cas montrent que les modèles RNN présentent de meilleures performances que les modèles de séries chronologiques classiques et les réseaux traditionnels. Dans cet article, nous avons appliqué le RNN-GRU pour prédire les arrivées de patients par heure au service des urgences pédiatriques du centre hospitalier régional de Lille, en France. Cet article est organisé comme suit. La section 1 présente les bases de l'apprentissage en profondeur, ses applications dans le cas de demandes de services d'urgence. Elle présente également une approche basée sur le GRU (*Gated Recurrent Unit*) permettant de prévoir les flux de patients dans les services d'urgence. La section 2 présente notre cas d'étude et les données associées. Les résultats sont exposés et discutés dans la section 3. Enfin, la dernière section conclut cette étude.

1. Apprentissage profond et prévision des admissions aux urgences hospitalières

Au cours des dernières années, de nombreuses études ont été consacrées aux problèmes de prévision dans plusieurs domaines d'application. Dans le cas des systèmes hospitaliers, de nombreux modèles de prévision ont été proposés afin d'étudier et d'établir des prévisions appropriées pour les demandes de services d'urgence. Selon la littérature, la plupart des techniques et modèles développés peuvent être classés en deux catégories: 1) modèles statistiques et séries chronologiques tels que modèles de régression et séries chronologiques: modèle ARIMA, lissage exponentiel de Holt – Winters, et 2) techniques d'intelligence artificielle telles que la « machine learning » et le « deep learning ». Grâce au développement des outils informatiques et aux technologies Big-data (capacité de stockage, vitesse de traitement, etc.,), les approches d'apprentissage par réseaux de neurones profonds sont devenues l'une des recherches les plus populaires dans les domaines de recherche universitaires et industriels.

L'apprentissage par réseaux de neurones profonds une des techniques du « Deep Learning » est le résultat de l'ajout de plusieurs couches dans les mécanismes de réseau neuronal (Kotu and Deshpande, 2019). Les modèles d'apprentissage par réseaux de neurones profonds les plus utilisés incluent les machines de Boltzmann, les réseaux de croyances profondes (DBN) et les réseaux de neurones récurrents (RNN). RNN est un type de réseau de neurones qui exploite la nature séquentielle des données d'entrée. Les RNN servent à modéliser des données en fonction du temps (séries chronologiques). Ces derniers, ont fait déjà leurs preuves dans plusieurs domaines d'application. Le modèle GRU (Gated Recurrent Unit) est un type de RNN capable de gérer la mémorisation des informations pendant des périodes beaucoup plus longues (Cho et al., 2014). Il est également considéré comme l'un des modèles RNN les plus utilisés pour la prévision des séries

chronologiques, parfaitement adapté aux problèmes de prévision de la demande de soins dans les systèmes hospitaliers.

Les deux sous-sections suivantes présentent le modèle de base du GRU et la manière dont il peut être construit et mis en œuvre, puis présentent une approche basée sur RNN-GRU permettant de modéliser et de prédire les arrivées de patients aux SU.

1.1 Modèle GRU

Le modèle GRU (Gated Recurrent Unit) est un mécanisme de déclenchement dans les réseaux de neurones récurrents proposé par (Cho et al., 2014). Le modèle GRU est considéré comme une variante des modèles LSTM (Long short-term memory) (Hochreiter and Schmidhuber, 1997). GRU est similaire au LSTM mais comporte moins de paramètres, sa structure combine la porte d'oubli « forget gate » et la porte d'entrée « input gate » en une seule porte nommée « update gate ». Le modèle résultant est plus simple que les modèles LSTM standard, mais ses performances sont comparables à celles de LSTM dans les modélisations séquentielles avec moins de paramètres et une construction plus facile. Le modèle de base de GRU est présenté dans la figure 1.

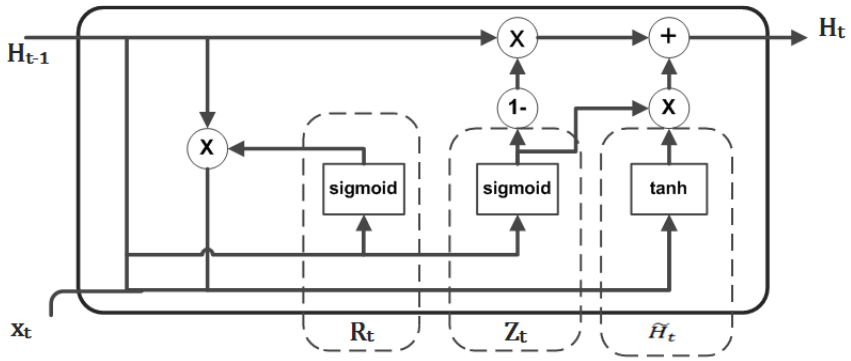

Figure 1. Structure de base d'un modèle GRU.

Le model GRU introduit le concept de la porte de réinitialisation (reset gate) et de la porte de mise à jour (update gate) afin de changer la méthode utilisée pour calculer les états cachés dans les RNNs. Comme le montre la figure 1, les entrées des deux portes (reset et update gates) sont considérées comme l'entrée du pas-de-temps actuel X_t et l'état caché (hidden state) du pas-de-temps précédent H_{t-1}. La sortie est calculée par la couche entièrement connectée à l'aide de la fonction sigmoïde, utilisée comme fonction d'activation. Supposons qu'il y a *h* unités cachées, pour un pas-de-temps donné *t*, l'entrée est X_t, l'état caché du dernier pas-de-temps est H_{t-1}. Par conséquent, la porte de réinitialisation (reset gate R_t), la porte de mise à jour (update gate Z_t) et de l'état caché H_t sont calculées comme suit :

$$R_t = \sigma(X_t W_{xr} + H_{t-1} W_{hr} + b_r) \quad (1)$$

$$Z_t = \sigma(X_t W_{xz} + H_{t-1} W_{hz} + b_z) \quad (2)$$

$$\widetilde{H}_t = \tanh(X_t W_{xh} + (R_t o\ H_{t-1}) W_{hh} + b_h \quad (3)$$

$$H_t = Z_t\ o\ H_{t-1} + (1 - Z_t) o\ \widetilde{H}_t \quad (4)$$

Où W_{xr}, W_{xz} ,W_{hr} W_{hz} W_{xh}, W_{hh} sont des paramètres de pondération et b_r, b_z, b_h est un paramètre de biais. *o* est un opérateur qui exprime la multiplication élément par élément.

1.2 Méthodologie proposée

La méthodologie proposée dans cet article a pour objectif de prédire les arrivées de patients dans un service d'urgences. Cette méthodologie est basée sur l'apprentissage par réseaux de neurones profonds. La figure 2 présente les principales étapes de la méthodologie proposée : i) collection de données à partir d'une base de données d'un SU, ii) prétraitement des données et analyse statistique, iii) standardisation des données pour normaliser les données collectées, iv) la construction et la validation des modèles de Deep learning, v) amélioration des performances du modèle RNN. Enfin, le modèle le plus performant sera utilisé pour la prévision de la demande de soins aux urgences.

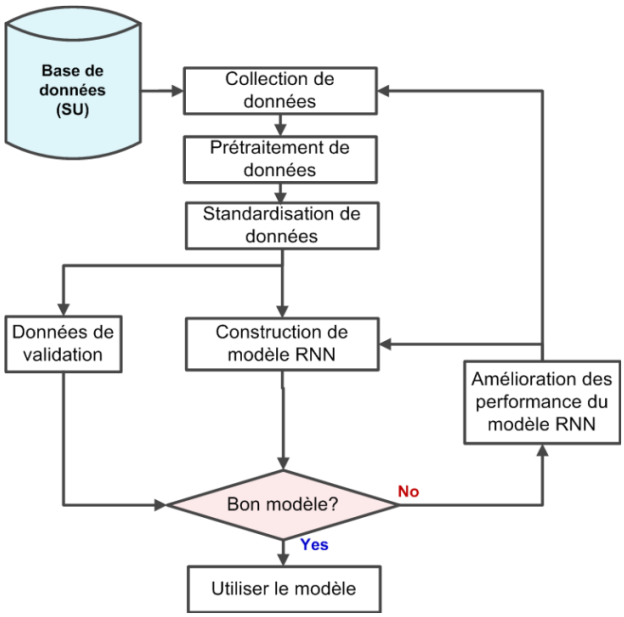

Figure 2. Etapes principales de la méthodologie proposée.

a) **Collection de données :** la première étape de notre méthodologie est la collecte de données. Les données nécessaires à notre étude ont été collectées à partir de la base de données du SU.
b) **Prétraitement de données :** la deuxième étape consiste à traiter les valeurs incohérentes et / ou manquantes. Dans notre étude, cette étape est nécessaire pour générer des données de qualité afin d'améliorer les performances des modèles RNN.
c) **Standardisation de données :** la troisième étape est la standardisation des données collectées. Comme les modèles RNN sont sensibles à l'échelle des données d'entrée, les données d'origine doivent être normalisées entre [0, 1].

d) **Construction et validation des modèles de prévision :** les modèles de prévision sélectionnés devraient offrir des prévisions adéquates. Afin d'évaluer les modèles, les données peuvent être divisées en deux groupes : données d'apprentissage et données de validation. Les données de validation sont utilisées pour évaluer et valider le modèle.

e) **Amélioration des performances du modèle RNN :** les techniques d'apprentissage automatique obtiennent de meilleures performances avec plus de données. La précision et les performances des modèles RNN dépendent de nombreux facteurs et paramètres : i) la quantité et la qualité de données d'apprentissage, ii) la configuration de réseau (architecture), iii) les hyper-paramètres et iv) les optimiseurs.

Plusieurs métriques sont proposées dans la littérature pour évaluer les capacités prédictives des modèles de prévision :

- Temps d'apprentissage (t) : c'est le temps requis pour compléter l'apprentissage des modèles ou pour modéliser un ou plusieurs ensembles de données.
- Erreur absolue moyenne (EAM ou MAE pour Mean Absolute Error) : moyenne arithmétique des valeurs absolues des écarts entre les valeurs prédites et les valeurs réelles.
- Moyenne des carrés des erreurs (MCE ou MSE pour Mean Square Error) : c'est la moyenne arithmétique des carrés des écarts entre chaque valeur prédite et la valeur réelle correspondante. L'erreur quadratique moyenne (EQM ou RMSE MSE pour Root Mean Square Error) est la racine carrée du MCE.
- Coefficient de détermination (R^2) : la qualité de l'ajustement est évaluée à l'aide du coefficient de détermination qui correspond au pourcentage de variabilité expliqué par le modèle.

2. Cas d'études

Le centre hospitalier régional Universitaire de Lille (CHRU) dessert quatre millions d'habitants dans les Hauts-de-France (France). Le service des urgences pédiatriques (SUP) du CHRU accueille environ 24 000 patients par an. En plus de ses capacités internes, le SUP partage de nombreuses ressources avec le service des urgences pour adultes (base de données patients, enregistrement du patient, examens complémentaires, etc.). L'activité au sein de SUP est caractérisée par l'irrégularité de la quantité et de la nature des soins à effectuer. Cette dernière est accentuée par le contexte actuel où les demandes de soins d'urgences augmentent constamment alors que le nombre de personnel stagne ou parfois il diminue.

Dans cette étude, nous avons utilisé un ensemble de données extraites de la base de données de SUP de janvier 2011 à novembre 2013. Les statistiques descriptives des arrivées par jour au SUP sont présentées dans le tableau 1. Selon ce tableau, les arrivées moyennes, médianes et maximales de patients au SUP pour l'ensemble de la période étudiée étaient 65, 66 et 119 respectivement.

Table 1: Statistiques descriptives des arrivées par jour des patients au SUP

Statistics	2011	2012	2013	Toutes les années
Nb de jours	365	366	334	1065
Moyenne	66,2	64,95	64,68	65,3
Médiane	67	65	64	66
Ecart-type	12,95	12,32	11,72	12,36
Minimum	26	34	35	26
1er quartile	58	57	57	57
3em quartile	75	74	72	74
Maximum	119	95	100	119

La figure 3 présente le nombre moyen d'arrivées de patients par année et par heure au SUP. Selon cette figure, le nombre d'arrivées est relativement léger avant 9h du matin, mais reprend après 10 heures. Le nombre de patients est relativement élevé entre [10h et 20h] avec une moyenne de 4 patients par heure.

La figure 4 présente quelques statistiques descriptives (moyenne, médiane, maximale) pour les arrivées par heure des patients au SUP sur l'ensemble des trois années étudiées (2011, 2012 et 2013). Selon cette figure, entre [9h et 19h], nous avons remarqué que la moyenne et la médiane sont presque égales; cela implique que la distribution d'arrivée des patients dans cette plage horaire est symétrique.

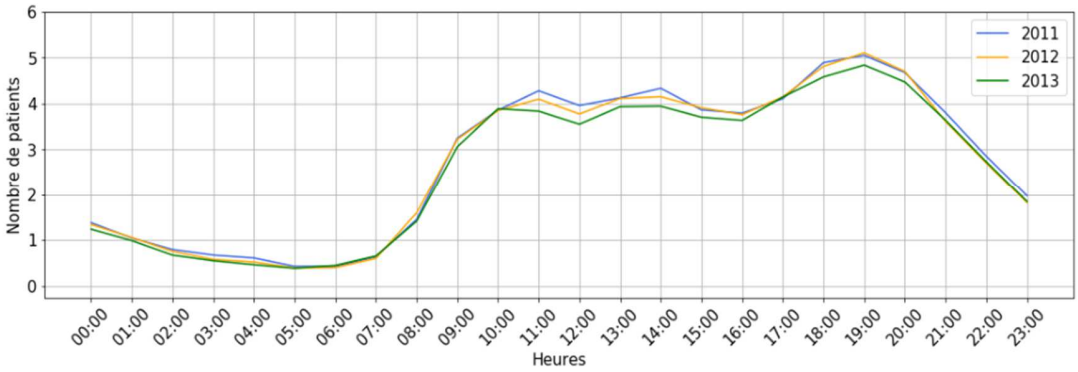

Figure 3. Nombre moyen d'arrivées de patients par année et par heure au SUP.

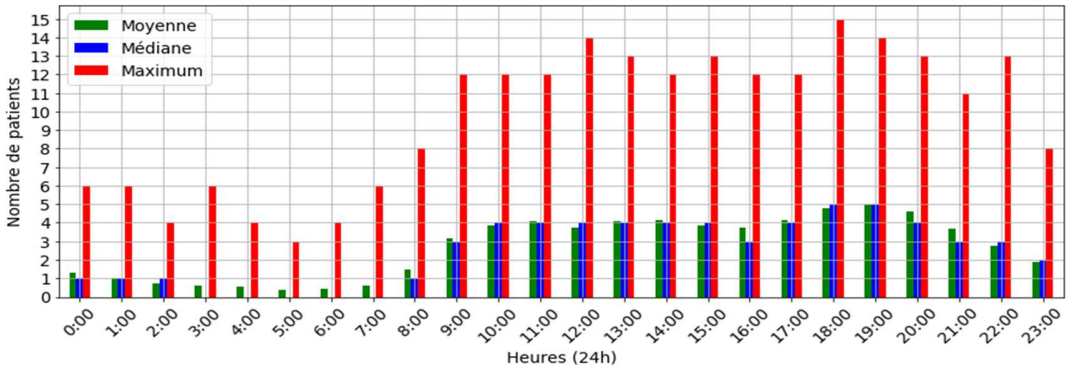

Figure 4. Statistiques descriptives des arrivées par heure des patients au SUP sur l'ensemble des trois années étudiées (2011, 2012 et 2013).

3. Résultats

3.1. Sélection du modèle GRU

Le modèle GRU a été construit sur la base des paramètres présentés ci-dessous (voir tableau 2). Les données ont été divisées en deux ensembles de données : les données d'apprentissage (80% des données) et les données de validation (20%). Le tableau 2 résume les algorithmes et les paramètres utilisés dans la modélisation et l'ajustement du modèle GRU. Dans cette étude, les modèles GRU sélectionnés ont été évalués en utilisant MAE, RMSE, MSE, et R^2. Les résultats sont présentés dans le tableau 3.

Table 2: Paramètres, métriques et algorithmes d'apprentissage utilisés pour la construction du modèle GRU

Paramètres, métriques et algorithmes d'apprentissage	**Valeur/ nom**	**Paramètres, métriques et algorithmes d'apprentissage**	**Valeur/ nom**
« Droupout »	0,2	Métriques	MAE, RMSE et R^2
Taille de lot	64	« Epochs »	100
Nombre de neurones	150	Séquence	24
Algorithmes d'optimisation	ADAM	Nombre de couches	2
Fonction de pertes	MSE	Régularisation, L_1 and L_2	$1x10^{-2}$
Fonction d'activation	ReLu	Taux d'apprentissage	$2x10^{-4}$

Table 3: Métriques de validation du modèle GRU

Paramètres	**Valeur**
R^2	0,972
RMSE	0,117
MAE	0,088
MSE	0,014

La figure 5 présente l'évolution de MSE, RMSE, MAE et R^2 en fonction du nombre d'époques. Comme le montre le tableau 3, les valeurs élevées de R^2 (0,972), de faible RMSE (0,117) et de faible MAE (0,088) montrent que le modèle GRU sélectionné représente bien les séries chronologiques observées.

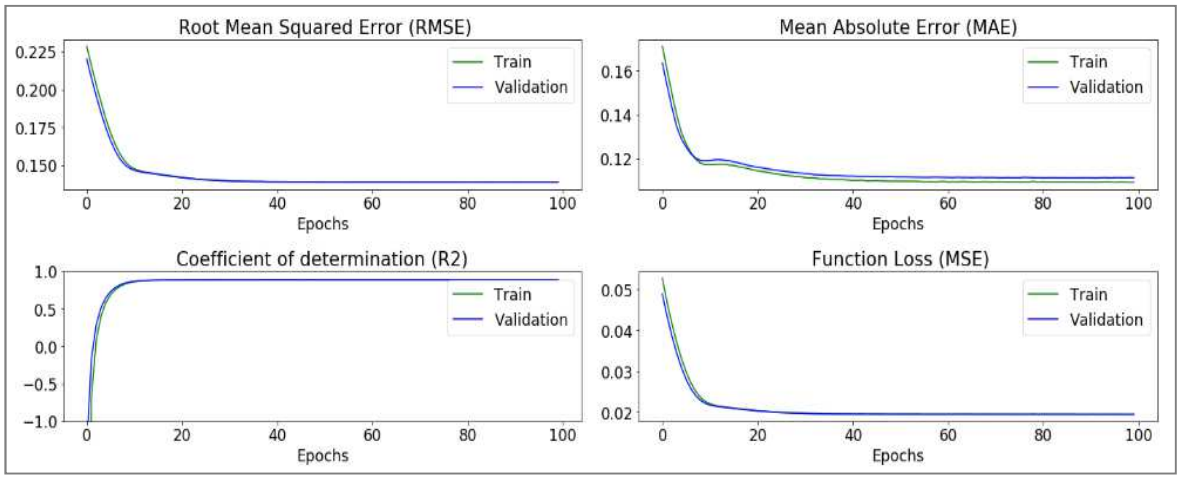

Figure 5. Evolution de la fonction de perte MSE, RMSE, MAE et R^2 en fonction du nombre d'Epochs.

3.2. Prévisions

Une fois que le modèle GRU a été construit et validé par rapport aux données réelles, les valeurs futures des arrivées des patients au SUP peuvent être prédites. Dans cette sous-section, le modèle GRU sélectionné a été examiné pour sa capacité prédictive sur un horizon de 3 jours (72 heures). La sortie du modèle a été comparée aux valeurs de données réelles du SUP qui ne sont pas utilisées dans la construction du modèle GRU. Les résultats des valeurs prédites et réelles sont présentés dans la figure 6, enregistrée pour un horizon H = 72 heures (trois jours).

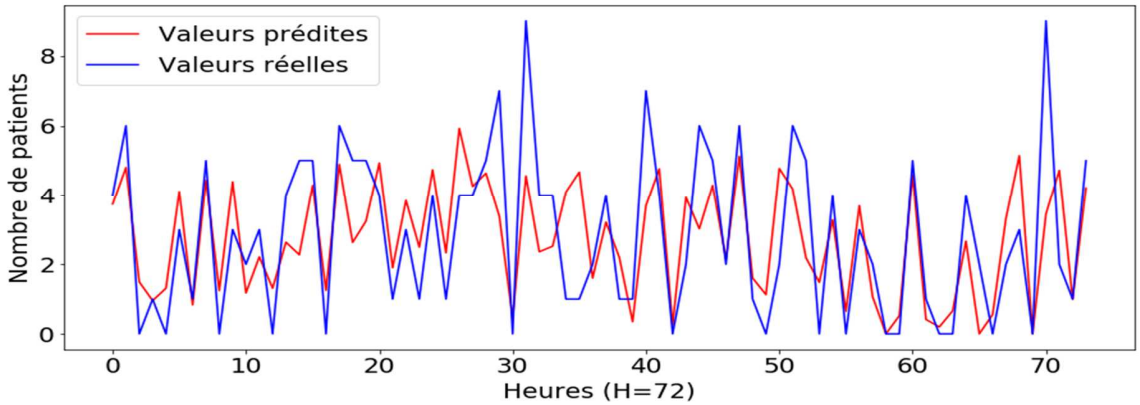

Figure 6. Valeurs prédites Vs valeurs réelles sur un horizon de 72 heures.

Le tableau 4 présente une description statistique des erreurs de prédiction. Nous avons remarqué que l'erreur moyenne de prévision pour les arrivées par heure est de 1,3 patients avec un écart type des erreurs de prévision de 1,08. Selon cette étude, le modèle GRU sélectionné a une bonne capacité de prévision et peut être utilisé pour prédire le nombre d'arrivée par heure des patients au sein du SUP

Table 4: Description statistique des erreurs de prédiction.

Métriques	Valeurs
1er quartile (25%)	0,61
Moyenne	1,30
Médiane	1,01
3em quartile (75%)	1,60
Ecart-type	1,08

Conclusion

Une approche d'apprentissage par réseaux de neurones profonds basée sur le modèle GRU a été proposée pour la prévision des arrivées de patients par heure au service d'urgences pédiatrique SUP au sein de CHRU de Lille. En se basant sur les métriques calculées, les résultats de prévision indiquent que le modèle proposé offre une bonne description et une performance acceptable de la prévision du nombre d'admissions au service d'urgences pédiatriques (SUP).

Cette étude soulève plusieurs questions sur les problèmes de prévision de demandes de soins au sein des SU. La prévision des admissions aux urgences dans un horizon adéquat, associé aux i) événement épidémiques, ii) à l'état situationnel du SU et iii) aux diagnostiques et types de demandes de soins comme les examens complémentaires (la biologie, la radiologie, le scanner), fournit des informations utiles qui peuvent être utilisées dans l'optimisation de la répartition des ressources humaines et matérielles au sein des SU.

Références

Aboagye-Sarfo, P., Mai, Q., Sanfilippo, F.M., Preen, D.B., Stewart, L.M., Fatovich, D.M., 2015. A comparison of multivariate and univariate time series approaches to modelling and forecasting emergency department demand in Western Australia. Journal of Biomedical Informatics 57, 62–73.

Benbelkacem, S., Kadri, F., Atmani, B., Chaabane, S., 2019. Machine Learning for Emergency Department Management. IJISSS 11, 19–36. https://doi.org/10.4018/IJISSS.2019070102

Boyle, A., Beniuk, K., Higginson, I., Atkinson, P., 2012. Emergency Department Crowding: Time for Interventions and Policy Evaluations. Emergency Medicine International 2012, 1–8.

Calegari, R., Fogliatto, F.S., Lucini, F.R., Neyeloff, J., Kuchenbecker, R.S., Schaan, B.D., 2016. Forecasting Daily Volume and Acuity of Patients in the Emergency Department. Comput Math Methods Med 2016.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. arXiv:1406.1078

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural computation 1735–1780.

Jones, S.A., Joy, M.P., Pearson, J., 2002. Forecasting demand of emergency care. Health Care Manag Sci, 297–305.

Kadri, F., Barraoui, M., Nouaouri, I., 2019. An LSTM-based Deep Learning Approach with Application to Predicting Hospital Emergency Department Admissions. Presented at the IEEE-IESM, Shanghai, China.

Kadri, F., Chaabane, S., Tahon, C., 2014a. A simulation-based decision support system to prevent and predict strain situations in emergency department systems. Simulation Modelling Practice and Theory 42, 32–52.

Kadri, F., Harrou, F., Chaabane, S., Tahon, C., 2014b. Time series modelling and forecasting of emergency department overcrowding. Journal of medical systems 38, 1–20.

Kadri, F., Harrou, F., Sun, Y., 2017. A multivariate time series approach to forecasting daily attendances at hospital emergency department, in: 2017 IEEE Symposium Series on Computational Intelligence (SSCI). Presented at the 2017 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1–6.

Kotu, V., Deshpande, B., 2019. Chapter 10 - Deep Learning, in: Kotu, V., Deshpande, B. (Eds.), Data Science (Second Edition). Morgan Kaufmann, pp. 307–342. https://doi.org/10.1016/B978-0-12-814761-0.00010-1

Steins, K., Matinrad, N., Andersson Granberg, T., 2019. Forecasting the Demand for Emergency Medical Services, in: Proceedings of the 52nd Hawaii International Conference on System Sciences, 2019, January 8-11, Grand Wailea, Maui, Hawai'i, University of Hawai'i at Manoa, pp. 1855–1864.