

Détection de la sensibilité aux corticostéroïdes chez les patients septiques à l'aide d'une approche personnalisée de fouille de données : une étude clinique

Rahma Hellali¹, Zaineb Chelly Dagdia^{2,3}, Ahmed Ktaish⁴,
Karine Zeitouni¹, Djillali Annane⁵

¹ Université Paris-Saclay, UVSQ, DAVID, Paris, France

² Univ. Lille, CNRS, Inria, Centrale Lille, UMR 9189 CRISTAL, F-59000 Lille, France

³ Univ. Lille, Inserm, CHU Lille, U1286, INFINITE, Institute for Translational Research in Inflammation, F-59000 Lille, France

⁴ GCI, France

⁵ Réanimation medico-chirurgicale, hôpital Raymond-Poincaré, AP-HP, Garches, France

8 mars 2026

Résumé

Le sepsis est une pathologie grave associée à une forte mortalité et morbidité. Les corticostéroïdes (CS) sont recommandés dans sa prise en charge, mais certains patients y répondent défavorablement. Une prédiction précoce de la réponse aux CS pourrait améliorer la prise en charge et réduire la mortalité. Cette étude propose une méthodologie de fouille de données pour prédire la sensibilité aux CS chez les patients septiques. Les données proviennent de l'essai randomisé APROCCHSS incluant 1 241 patients. Plusieurs modèles d'apprentissage automatique ont été comparés afin d'identifier la meilleure « signature » prédictive. Les performances ont été évaluées par la précision, la sensibilité et la spécificité. La régression logistique s'est révélée la plus performante avec une Aire sous la courbe (AUC) de 72%. Des analyses supplémentaires ont examiné l'ajout de variables et la généralisabilité du modèle. La signature a été déployée dans le système d'information de l'AP-HP sous forme de service web afin de prédire la cortico-sensibilité ou la cortico-résistance. Elle est actuellement utilisée par les cliniciens comme outil d'aide à la randomisation des patients. Cet article correspond à un résumé d'un travail publié à Computer Methods And Programs In Biomedicine Journal [1].

Mots-clés

Sepsis, Corticostéroïdes, Sensible ou résistant, Apprentissage automatique, Système d'aide à la décision clinique.

1 Introduction

Le sepsis est une défaillance organique grave due à une réponse immunitaire excessive à une infection, pouvant évoluer vers un choc septique et le décès. Il représente un enjeu majeur, notamment en contexte maternel et néonatal. Malgré les progrès thérapeutiques, la mortalité reste élevée

en raison de la difficulté à administrer le traitement approprié au bon moment. La corticothérapie est largement utilisée, mais la réponse varie selon les patients. Les déterminants biologiques et cliniques de cette sensibilité restent complexes et insuffisamment maîtrisés. Cette étude s'inscrit dans un essai randomisé en vie réelle mené avec l'Assistance Publique – Hôpitaux de Paris (AP-HP). L'objectif est de prédire précocement (dès l'admission ou dans les 48 premières heures) la sensibilité ou la résistance aux corticostéroïdes à l'aide d'une approche robuste de data mining. Les données cliniques, biologiques et démographiques des patients sont exploitées pour développer une « signature » prédictive basée sur des modèles de machine learning. Cette signature, déployée dans le système d'information de l'AP-HP (AI signature 1), est utilisée comme biomarqueur lors du processus de randomisation bayésienne des patients, aux côtés d'autres biomarqueurs. L'étude analyse également l'effet individuel du traitement en comparant les prédictions sous corticothérapie et sous placebo.

Les contributions principales sont : (i) La prédiction très précoce de la réponse aux corticostéroïdes (à l'admission). (ii) Une évaluation progressive de l'impact des variables patients sur la performance prédictive. (iii) Une estimation plus précise de l'effet individuel du traitement basée sur les patients correctement classifiés. (iiii) La validation et l'intégration clinique réelle du modèle dans un protocole hospitalier. Les questions de recherche examinées dans cette étude sont les suivantes : (1) Les modèles prédictifs peuvent-ils distinguer répondeurs et non-répondeurs ? (2) Quel est l'impact des variables cliniques dans les 48 premières heures sur la performance ? (3) Quel est l'effet individuel et global de la corticothérapie ? (4) Le modèle est-il généralisable à d'autres cohortes ? L'objectif final est d'améliorer la médecine personnalisée en réanimation via une aide à la décision clinique fondée sur l'intelligence artificielle, afin d'optimi-

ser l'administration des corticostéroïdes et réduire la mortalité liée au sepsis.

2 Matériels et méthodes

Cette étude s'inscrit dans le projet national RECORDS, coordonné par l'AP-HP, dont l'objectif est d'identifier rapidement la sensibilité ou la résistance des patients septiques à la corticothérapie. La signature prédictive initiale a été développée à partir de la cohorte APROCCHSS, issue d'un essai contrôlé randomisé. La méthodologie repose sur une chaîne de fouille de données en plusieurs étapes : collecte des données cliniques, prétraitement, préparation des données, apprentissage des modèles, évaluation statistique et déploiement clinique. Les données des patients ont été extraites du dossier électronique médical à l'admission. La cohorte APROCCHSS comprend initialement 1 241 patients décrits par 5 645 variables. Après révision médicale et exclusion de certains cas, 1 234 patients ont été retenus, dont 612 traités par corticostéroïdes. Comme l'objectif est de prédire précocement la réponse au traitement, seules les variables disponibles à J0 ont été utilisées pour construire la signature principale. Une sélection a d'abord réduit les 5 645 variables à 238, puis à 24 variables cliniquement pertinentes choisies avec les experts médicaux. Des étapes de nettoyage ont été appliquées, notamment la gestion des valeurs manquantes, l'encodage des variables catégorielles, la transformation des dates, ainsi que la création de nouvelles variables cliniques. Ces données ont ensuite été préparées pour le machine learning : labellisation des patients selon leur réponse à la corticothérapie à J14, séparation en ensembles d'apprentissage (80%) et de test (20%), équilibrage des classes avec SMOTE, puis normalisation des variables sur l'intervalle [0,1]. Plusieurs modèles de machine learning ont été comparés : régression logistique, méthode d'amplification de gradient, forêt aléatoire, machine à vecteurs de support, descente de gradient stochastique et arbre de décision. Chaque modèle a été évalué selon plusieurs configurations : version par défaut, optimisation des hyperparamètres, sélection de variables, ou combinaison des deux. L'expérimentation a été menée sur deux cohortes : APROCCHSS et RECORDS, afin d'évaluer à la fois la performance et la capacité de généralisation du modèle. Le protocole expérimental comprend deux grandes phases. La première vise à identifier le meilleur modèle pour distinguer les patients répondeurs et non-répondeurs à la corticothérapie à partir des données de J0. La seconde étudie l'apport de variables supplémentaires disponibles à J1 et J2, ainsi que la robustesse du modèle lorsqu'il est appliqué à une autre cohorte. L'évaluation repose sur des métriques classiques de classification : accuracy, précision, rappel, F1-score et AUC, complétées par des analyses statistiques sur 30 exécutions. L'effet du traitement a aussi été étudié à l'échelle individuelle (ITE) et populationnelle (ATE). Enfin, le meilleur modèle a été déployé dans le système d'information de l'AP-HP sous forme d'un service web développé avec Flask. Intégré à l'application CleanWeb, il permet de prédire automatiquement si un patient est sensible

ou résistant aux corticostéroïdes à partir de ses données cliniques, et constitue ainsi un biomarqueur d'aide à la décision dans le protocole RECORDS.

3 Résultats et discussions

Les résultats montrent que la régression logistique est le modèle le plus performant pour prédire la sensibilité aux corticostéroïdes chez les patients septiques. Dans la première expérimentation, elle obtient les meilleurs résultats globaux, avec une AUC de 72%, devant les autres modèles testés. Les performances sont améliorées lorsque les données sont équilibrées avec SMOTE et lorsque les variables les plus pertinentes sont sélectionnées. Les analyses statistiques confirment que cette supériorité est significative. Dans la deuxième expérimentation, plusieurs variantes du meilleur modèle ont été comparées selon les jeux de variables utilisés. Les résultats montrent que l'ajout de données cliniques disponibles à J1 et J2 améliore la prédiction. Le meilleur score est obtenu en intégrant les variables de J0, J1, J2 ainsi que la différence entre J1 et J2, avec une AUC de 76%. Cela indique qu'une information clinique plus riche améliore la qualité de la prédiction. En revanche, la généralisation du modèle à une autre cohorte (RECORDS) reste limitée. Lorsque le modèle appris sur APROCCHSS est testé sur RECORDS, les performances chutent fortement, jusqu'à une AUC de 55% dans certains cas. Cette baisse suggère que les signatures prédictives ne sont pas totalement transférables entre cohortes, en particulier lorsque les profils cliniques diffèrent, par exemple avec la présence de patients atteints de COVID-19. L'évaluation de l'effet du traitement montre des résultats plus nuancés. À l'échelle individuelle (ITE), la majorité des patients présentent un effet neutre, ce qui rend l'interprétation clinique difficile. À l'échelle populationnelle (ATE), les modèles suggèrent un effet global plutôt bénéfique de la corticothérapie. Globalement, le bénéfice semble exister, mais reste modéré et dépend fortement du modèle et des variables considérées.

4 Conclusion

Nous avons développé une signature basée sur le machine learning pour prédire précocement la réponse aux corticostéroïdes chez les patients septiques (cohortes APROCCHSS et RECORDS). La régression logistique s'est révélée la plus performante (AUC 72%, jusqu'à 76% avec l'ajout des données J1-J2), mais sa généralisation à la cohorte RECORDS demeure limitée. L'effet du traitement apparaît globalement modéré. Ces résultats restent préliminaires et devront être consolidés par des cohortes plus larges et l'intégration de données multi-omiques.

Références

- [1] Rahma Hellali, Zaineb Chelly Dagdia, Ahmed Ktaish, Karine Zeitouni, and Djillali Annane. Corticosteroid sensitivity detection in sepsis patients using a personalized data mining approach : a clinical investigation. *Computer Methods and Programs in Biomedicine*, 245 :108017, 2024.