

Anesthésie - Réanimation Divers

ID: 420

Création d'une base de données rétrospective monocentrique à partir de données informatisées : apport de l'intelligence artificielle

C. De tymowski*(1), J.Dovale(1), K.Sallah(2), P.Montravers(1), S.Provenchère(1)

(1) Département d'anesthésie-réanimation chirurgicale, Hôpital Bichat, Paris, France, (2) Unité de recherche Clinique, Hôpital Bichat, Paris, France

**Auteur présenté comme orateur*

Position du problème et objectif(s) de l'étude:

Les entrepôts des données de santé (EDS) contiennent de nombreuses informations comme les comptes rendus médicaux (CRM) qui sont difficilement inclusibles directement dans une base de données. L'intelligence artificielle (IA) via des algorithmes de texte non structuré pourrait aider à extraire les données cliniques des CRM. Nous avons évalué la pertinence de l'IA afin de créer une base de données à partir des données d'EDS.

Matériel et méthodes:

Il s'agit d'une étude monocentrique rétrospective (2017-19) incluant tous les patients (pts) ayant eu une assistance extracorporelle veino-artérielle (ECMO-VA) sur cette période, qui compare les données récupérées de manière automatisée à partir d'un algorithme en utilisant des expressions régulières dans Python (IA), à celles récupérées par un clinicien (CL). Pour l'IA, les informations étaient récupérées dans les CRH et dans les codages stockés dans l'EDS ; pour la base clinique, celles-ci étaient récupérées dans ORBIS®. La qualité du recueil de 15 antécédents (sexe, âge, poids, taille, diabète, insuffisance cardiaque, HTA, BPCO, dyslipidémie, insuffisance rénale chronique, angor, tabagisme, immunodépression, cirrhose, cancer) était comparée entre l'IA et le CL. Un examinateur extérieur a revu les différences entre les deux bases pour comprendre les causes de discordance.

Résultats & Discussion:

Sur la période 2017-2019, le recueil des données des 130 patients inclus a été comparé entre l'IA et le CL. Les taux moyens d'exactitude, de sensibilité et de spécificité étaient de 0,82, 0,75 et 0,85 avec une forte variabilité selon les variables ; le sexe ayant la meilleure exactitude à 0,97 et l'insuffisance cardiaque chronique la moins bonne à 0,69. Au total 123 discordances ont été notées. Les causes de discordance étaient par ordre décroissant : absence des données dans l'EDS (30%), mauvais positionnement de l'antécédent dans le CRM (22%), mauvaise retranscription des données dans l'EDS avec des phrases manquantes (20%), erreurs de l'algorithme (17%), erreurs du clinicien (10%). Le poids et la taille étaient respectivement récupérés pour 104 et 90 pts, avec une corrélation de 0,93 et 0,77.

Conclusion:

Les algorithmes de texte non structuré peuvent aider au recueil des données stockées par les EDS, mais les principales limites actuelles sont la qualité de ce stockage et la qualité des données retranscrites dans les comptes rendus. Cette étude ouvre la possibilité de réaliser des bases de données à grande échelle par une lecture automatisée des CRM par IA.

Remerciements:

Les auteurs déclarent ne pas avoir toute relation financière impliquant l'auteur ou ses proches (salaires, honoraires, soutien financier éducationnel) et susceptible d'affecter l'impartialité de la présentation.