

Comparaison de l'Intelligence Artificielle (IA) et de l'Attaché de Recherche Clinique (ARC) dans l'analyse rétrospective de 475 hernies discales lombaires sous endoscopie : Précision, Efficience et Fiabilité.

Marc Szadkowski* ¹, Henri d'Astorg ¹, Aymeric Faure ², Xavier Castel ¹, Hadrien Giorgi ²

¹ Centre Orthopédique Santy, Orthopédie, Lyon, France

² IMED, Orthopédie, Marseille, France

INTRODUCTION

La recherche clinique en chirurgie rachidienne repose sur la qualité du recueil de données. Cependant, le passage au crible de centaines de dossiers patients est une tâche chronophage pour l'Attaché de Recherche Clinique (ARC) et comporte un risque d'erreur humaine non négligeable. L'émergence des modèles d'Intelligence Artificielle (IA) offre une alternative potentielle pour l'extraction automatisée de données de santé. L'objectif de cette étude était de comparer la performance de l'IA à celle d'un ARC dans le cadre d'une étude rétrospective sur la hernie discale lombaire (HDL) opérée sous endoscopie, en utilisant un référentiel d'experts ("Golden Dataset").

MATÉRIEL ET MÉTHODE :

Nous avons mené une étude rétrospective multicentrique incluant 475 patients opérés par 5 chirurgiens d'une HDL par voie endoscopique (transforaminale ou interlaminaire) avec un recul minimum de 1 an.

Les variables extraites comprenaient :

Données administratives et cliniques : Nom, date de chirurgie, type de hernie, durée des symptômes et type de traitement médical pré-opératoire.

Scores fonctionnels : Douleur lombaire et radiculaire (EVA), ainsi que les scores de qualité de vie/ fonctionnels pré-opératoires, à 3 mois et à 1 an.

MÉTHODOLOGIE DE COMPARAISON :

Le Golden Dataset : Quatre chirurgiens experts du rachis ont analysé indépendamment les dossiers pour établir une base de données de référence absolue, résolvant les discordances par consensus.

L'ARC : Un attaché de recherche expérimenté a réalisé l'extraction manuelle des données.

L'IA : Un logiciel utilisant un modèle de LLM a été configuré pour extraire les mêmes variables à partir des comptes-rendus opératoires et des notes de consultation.

Les critères de jugement étaient le temps de travail global (en heures) et le taux d'erreur (écart par rapport au Golden Dataset).

RÉSULTATS :

Sur les 475 dossiers analysés :

Efficience temporelle : L'ARC a nécessité un temps de travail total de 80 heures (soit environ 10 minutes par dossier).

L'ARC a présenté un taux d'erreur de 20%, principalement lié à des erreurs d'interprétation du type de hernie, des oublis sur les scores à court terme, et sur l'inclusion ou non des patients dans l'étude.

Nous avons ensuite demandé à l'IA l'extraction des mêmes paramètres afin de les passer au crible des experts et mesuré le taux d'erreur et le temps nécessaire à l'IA pour extraire ces données.

CONCLUSION :

L'IA semble devenir un outil de rupture pour la recherche clinique en chirurgie du rachis. Elle permet une analyse rétrospective massive, rapide et fiable. Si elle ne remplace pas l'expertise humaine, elle transforme le rôle de l'ARC vers une mission de supervision, accélérant ainsi la production de preuves scientifiques en chirurgie endoscopique.