

# Etat des lieux de l'usage des big data et de l'intelligence artificielle en rhumatologie : une revue systématique de la littérature

## Langage médical

**Auteurs :** Joanna KEDRA, Timothy RADSTAKE, Aridaman PANDIT, Xenofon BARALIAKOS, Francis BERENBAUM, Axel FINCKH, Bruno FAUTREL, Tanja STAMM, David GOMEZ-CABRERO, Christian PRISTIPINO, Rémy CHOQUET, Hervé SERVY, Simon STONES, Gerd BURMESTER, Laure GOSSEC

### Contexte :

Les big data sont définis comme des jeux de données trop volumineux ou complexes pour être traités par des programmes informatiques traditionnels. L'intelligence artificielle (IA) comprend différentes méthodes d'analyses qui peuvent être appliquées aux big data. Le recours à ces concepts dans les publications rhumatologiques est méconnu.

### Objectif :

L'objectif de ce travail était de faire un état des lieux de l'usage des big data et de l'IA dans les études en rapport avec la rhumatologie.

### Matériel et méthodes :

Une revue systématique de la littérature (RSL) a été réalisée sur PubMed MEDLINE en novembre 2018, à l'aide de mots-clés faisant référence aux big data, à l'IA et à la rhumatologie. Toutes les publications originales écrites en anglais ont été analysées. Une revue de la littérature "en miroir" a également été réalisée en dehors de la rhumatologie avec le même nombre d'articles que la RSL, afin de comparer l'usage des big data en rhumatologie par rapport aux autres spécialités médicales. Le nombre de données analysées, les sources de données et les méthodes statistiques utilisées (IA, méthodes traditionnelles ou les deux) ont été collectés.

### Résultats :

Parmi les 567 articles trouvés dans la RSL, 55 remplissaient les critères d'inclusions et ont été analysés, de même que 55 articles dans d'autres spécialités médicales. L'année moyenne de publication était 2014 dans la RSL en rhumatologie, avec 72% d'articles publiés entre 2013 et 2018, tandis que les articles de la revue en dehors de la rhumatologie étaient tous publiés en 2018 ou 2019. En rhumatologie, les 3 pathologies les plus représentées étaient la gonarthrose (N=13, 24%), la polyarthrite rhumatoïde (N=12, 22%) et l'ostéoporose post-ménopausique (N=6, 11%). En dehors de la rhumatologie, les spécialités les plus représentées étaient l'oncologie (N=14, 25%), la neurologie (N=8, 15%), les maladies infectieuses (N=6, 11%), l'ophtalmologie (N=5, 9%) et la psychiatrie (N=5, 9%).

Seuls 2 articles en rhumatologie (4%) et 7 articles en dehors de la rhumatologie (13%) mentionnaient clairement une définition des big data. Le nombre moyen de données était de 746 millions [minimum 2000, maximum 5 milliards] en rhumatologie, et de 9,1 milliards [minimum 100000, maximum 200 milliards] en dehors de la rhumatologie. Les sources de données étaient diversifiées : en rhumatologie, 26 articles (47%) se basaient sur des données cliniques, 8 (15%) sur des données biologiques et 16 (29%) sur des données radiologiques, tandis que dans les autres spécialités médicales, la distribution était plus homogène entre les sources de données cliniques, biologiques ou radiologiques (respectivement 31%, 31% et 29%). Les méthodes statistiques traditionnelles et l'IA étaient toutes deux utilisées pour analyser les big data (respectivement 10 (18%) et 45 (82%) en rhumatologie et 8 (15%) et 47 (85%) en dehors de la rhumatologie). Le machine learning était utilisé dans la quasi-totalité des études recourant à l'IA (44/45 en rhumatologie et 47/47 en dehors de la rhumatologie), et parmi les techniques de machine learning, la plus représentée était les réseaux de neurones artificiels (20/44 en rhumatologie, et 24/47 en dehors de la rhumatologie).

### Conclusions :

Que ce soit en rhumatologie ou dans les autres spécialités médicales, il n'y a pas de définition unique et consensuelle des big data. Les sources de big data sont diversifiées, et les méthodes d'analyses sont variées et hétérogènes. Ces résultats ont permis d'informer le groupe de travail de l'EULAR en charge de l'élaboration des recommandations pour l'usage des big data en rhumatologie.