

UNIVERSITE DE LILLE

**FACULTE DE MEDECINE HENRI WAREMBOURG**

Année 2025

THESE POUR LE DIPLOME D'ÉTAT  
DE DOCTEUR EN MEDECINE

**Impact de l'intelligence artificielle sur la prise en charge  
des patients admis aux urgences pour traumatismes du  
squelette appendiculaire**

Présentée et soutenue publiquement le 13 octobre 2025  
à 16:00 au pôle formation

**Par Lubin BRABANT**

---

**JURY**

**Président :**

Monsieur le Professeur Éric WIEL

**Assesseurs :**

Monsieur le Docteur Rémy DIESNIS

Monsieur le Docteur Marc-Aurèle RUSU

**Directrice de thèse :**

Madame le Docteur Bérengère VOGEL

---

# Avertissement

La Faculté n'entend donner aucune approbation aux opinions émises dans les thèses : celles-ci sont propres à leurs auteurs.

# Sigles

|              |                                                                            |
|--------------|----------------------------------------------------------------------------|
| <b>IA</b>    | Intelligence artificielle                                                  |
| <b>ML</b>    | Machine learning                                                           |
| <b>DL</b>    | Deep learning                                                              |
| <b>Se</b>    | Sensibilité                                                                |
| <b>Spe</b>   | Spécificité                                                                |
| <b>VPP</b>   | Valeur prédictive positive                                                 |
| <b>VPN</b>   | Valeur prédictive négative                                                 |
| <b>NaN</b>   | Non analysable                                                             |
| <b>STEMI</b> | Infarctus du myocarde avec élévation du segment ST                         |
| <b>CH</b>    | Centre hospitalier                                                         |
| <b>DRESS</b> | Direction de la recherche, des études, de l'évaluation et des statistiques |
| <b>RGPD</b>  | Règlement général sur la protection des données                            |
| <b>SAU</b>   | Service d'accueil des urgences                                             |
| <b>USA</b>   | United States of America                                                   |
| <b>ESI</b>   | Emergency Severity Index                                                   |
| <b>PACS</b>  | Pictures Archieving and Communicating System                               |

# Sommaire

|                                                                                  |    |
|----------------------------------------------------------------------------------|----|
| Avertissement.....                                                               | 2  |
| Sigles.....                                                                      | 3  |
| Sommaire .....                                                                   | 4  |
| Introduction.....                                                                | 6  |
| 1 Qu'est-ce que l'intelligence artificielle (IA) ? .....                         | 6  |
| 1.1 Un peu d'histoire.....                                                       | 6  |
| 1.2 Comprendre le concept d'intelligence artificielle.....                       | 6  |
| 1.3 Différence entre l'intelligence artificielle et l'intelligence humaine ..... | 7  |
| 1.4 Concept de Machine Learning et Deep Learning .....                           | 7  |
| 1.4.1 Machine Learning (ML) .....                                                | 7  |
| 1.4.2 Deep Learning (DL) .....                                                   | 8  |
| 1.5 L'enjeux des données, le BIG DATA .....                                      | 9  |
| 1.6 Les limites de l'IA.....                                                     | 10 |
| 2 Applications de l'IA en médecine d'urgence .....                               | 10 |
| 3 Gleamer.....                                                                   | 11 |
| 4 Contexte et enjeux actuel de l'IA en radiologie aux urgences .....             | 12 |
| Matériel et méthodes .....                                                       | 16 |
| 1 Design de l'étude.....                                                         | 16 |
| 2 Patients et données.....                                                       | 16 |
| 2.1 Critères d'inclusion .....                                                   | 16 |
| 2.2 Critères de non inclusion .....                                              | 17 |
| 2.3 Schéma de l'étude.....                                                       | 17 |
| 2.4 Recueil de données.....                                                      | 19 |
| 3 Objectif de l'étude.....                                                       | 19 |
| 3.1 Objectif principal et critère de jugement principal .....                    | 19 |
| 3.2 Objectif secondaires (OS) et leurs critères de jugement .....                | 19 |
| 3.3 Gold Standard et contrôle des prises en charge .....                         | 20 |
| 4 Analyse statistique .....                                                      | 20 |
| 5 Cadre réglementaire.....                                                       | 21 |
| 5.1 Considération éthique et réglementaire.....                                  | 21 |
| 1.1. Principes généraux .....                                                    | 21 |
| 1.2. Aspect réglementaire .....                                                  | 22 |

|      |                                                                                  |    |
|------|----------------------------------------------------------------------------------|----|
| 1.3. | Information des patients.....                                                    | 22 |
| 1.4. | Protection des données .....                                                     | 22 |
| 5.2  | Gestion des données.....                                                         | 23 |
| 1.1. | Modalité de recueil des données .....                                            | 23 |
| 1.2. | Droit d'accès aux données et documents sources.....                              | 23 |
| 1.3. | Durée de conservation et archivage .....                                         | 23 |
|      | Résultats.....                                                                   | 25 |
| 1    | Flow chart.....                                                                  | 25 |
| 2    | Description des radiographies à l'inclusion.....                                 | 25 |
| 2.1  | Radiographies des patients .....                                                 | 25 |
| 2.2  | Intervenants.....                                                                | 26 |
| 3    | Impact de l'IA sur la prise en charge des patients .....                         | 27 |
| 4    | Impact d'un logiciel d'IA sur le suivi des patients.....                         | 27 |
| 5    | Analyse en sous-groupes de l'impact de l'IA sur la prise en charge.....          | 28 |
| 6    | Impact d'un logiciel d'IA sur la détection des lésions.....                      | 28 |
| 6.1  | Performances diagnostiques selon le type de lésion.....                          | 29 |
| 6.2  | Performances diagnostiques selon localisation du traumatisme et type de lésion30 |    |
|      | Discussion .....                                                                 | 33 |
| 7    | Principaux résultats .....                                                       | 33 |
| 8    | Discussion des résultats.....                                                    | 33 |
| 9    | Forces et limites de l'étude.....                                                | 35 |
|      | Conclusion.....                                                                  | 38 |
|      | Liste des tables.....                                                            | 39 |
|      | Liste des figures .....                                                          | 40 |
|      | Références .....                                                                 | 41 |
|      | Annexe 1 .....                                                                   | 44 |
|      | Annexe 2 .....                                                                   | 46 |

# Introduction

## 1 Qu'est-ce que l'intelligence artificielle (IA) ?

### 1.1 Un peu d'histoire

Développé dans les années 1950 par John McCarty [1], le concept d'IA se définit comme « la science et l'ingénierie de la fabrication de machines intelligentes » dans un contexte de développement des nouvelles technologies, notamment informatique, qui permet de traiter des informations de manière automatique, avec le développement d'algorithmes.

On s'est vu rêver à pouvoir créer des machines aussi intelligentes que les humains.

Le concept d'intelligence artificielle a connu un essor des années 50 jusqu'à la fin des années 70, puis un ralentissement de la recherche dans les années 80 au vu de l'absence de résultats avec un arrêt des fonds financiers. A la fin des années 90 avec le développement de nouvelles technologies, d'ordinateurs plus puissants, l'optimisation des algorithmes et la présence d'une quantité de données beaucoup plus importantes, l'IA connaît un nouvel essor.

### 1.2 Comprendre le concept d'intelligence artificielle

Pour comprendre ce qu'est l'intelligence artificielle, il faut pouvoir la définir. Cela passe par l'étude et la compréhension de l'intelligence humaine dite « biologique », au travers des sciences cognitives. Les sciences cognitives ont permis de réfléchir sur cette intelligence biologique et d'identifier les différents mécanismes à l'œuvre dans le fonctionnement de la pensée. Cette pensée qui permet de percevoir, planifier, bouger, mémoriser, s'adapter.

L'intelligence est notre capacité à s'adapter à une situation particulière, à trouver la réponse la plus adaptée à une problématique en utilisant les ressources à notre disposition : nos fonctions cognitives et notre mémoire. Elle découle de notre capacité à comprendre une situation et à l'interpréter, à apprendre et à s'adapter aux changements.

Le but de l'intelligence artificielle est d'essayer de reproduire cette intelligence « biologique ». Pour cela, deux approches ont été développées : via la transmission de connaissances et via l'expérimentation.

La première approche, dite symbolique, imite le raisonnement humain et sa capacité à exploiter ses connaissances. La création d'un programme capable de reproduire les différentes étapes du raisonnement humain est possible en injectant dans la

machine des données, des règles et des procédures. On parle alors de système expert.

La deuxième approche, dite par l'expérience (faire du vélo, reconnaître un animal), est l'approche numérique qui tente de reproduire le savoir humain acquis au travers de ses expériences, qu'il s'agisse de réussites ou d'échecs. À ce jour, cette approche est privilégiée, notamment avec le développement du Machine Learning et du Deep Learning.

### **1.3 Différence entre l'intelligence artificielle et l'intelligence humaine**

L'intelligence artificielle se limite à un domaine précis et nécessite un apprentissage qui repose sur beaucoup de données. En effet, la machine ne comprend pas la tâche qu'on lui demande, elle ne fait que l'appliquer. L'intelligence humaine, en revanche, permet de tirer des déductions pertinentes à partir de peu d'exemples grâce à des capacités cognitives et émotionnelles.

Il existe deux catégories d'IA :

- L'intelligence artificielle complète ou forte : une intelligence artificielle capable de copier les aptitudes humaines (apprendre, comprendre, appréhender, raisonner, prendre des décisions, avoir une conscience, des émotions, etc.) À ce jour, cette intelligence artificielle n'existe pas ;
- L'intelligence artificielle faible : c'est l'intelligence artificielle que l'on connaît aujourd'hui. Il s'agit d'un algorithme « apprenant » qui adapte ses paramètres selon des données d'apprentissage. Cette intelligence est dépourvue de capacités mentales et cognitives, mais elle peut effectuer une tâche précise très efficacement, parfois plus qu'un être humain.

### **1.4 Concept de Machine Learning et Deep Learning**

#### **1.4.1 Machine Learning (ML)**

L'intelligence artificielle utilisée actuellement est basée sur une approche « numérique » qui correspond à l'apprentissage par l'expérience.[1,2]

Le « Machine Learning » ou « apprentissage automatique » est défini par l'utilisation d'algorithmes qui permettent à la machine, après analyse de nombreux exemples, d'identifier des modèles récurrents et ainsi émettre des prédictions statistiques. Elle peut réaliser des tâches spécifiques (ex : prédire la taille d'un arbre en fonction de son diamètre, reconnaître un animal sur une image, etc.) en s'appuyant sur ces modèles récurrents.

Des données d'entrées et de sorties sont injectées dans la machine, qui s'entraîne à trouver une corrélation entre les deux : il s'agit de la phase d'apprentissage. Par la suite, la machine doit être capable d'appliquer cet apprentissage à de nouvelles données, c'est ce qu'on appelle la phase de généralisation.

L'intelligence artificielle apprend de trois manières :

- L'apprentissage supervisé : les données sont étiquetées par les humains et permettent à la machine de faire des prédictions sur un modèle déjà établi ;
- L'apprentissage non supervisé : les données sont soumises de manière brute sans annotation ni résultats prédéfinis, la machine essaye alors d'identifier les regroupements latents d'elle-même ;
- L'apprentissage par renforcement : l'algorithme apprend et ajuste ses réponses en fonction de la réaction positive ou négative qui suit chacune de ses actions et en déduit des modèles.

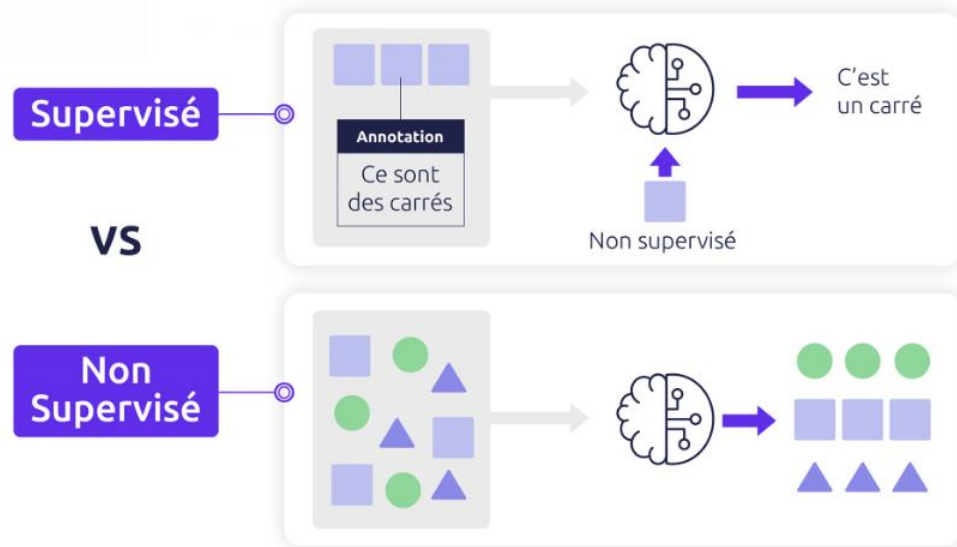


Figure 1. Schémas des deux principaux modes d'apprentissage (source : Diabolocom ; traitement du langage naturel : guide complet)

Une fois l'apprentissage terminé vient la phase de prédiction : la machine produit des données similaires sur lesquelles le programme n'a pas été entraîné.

Le ML permet deux types d'applications :

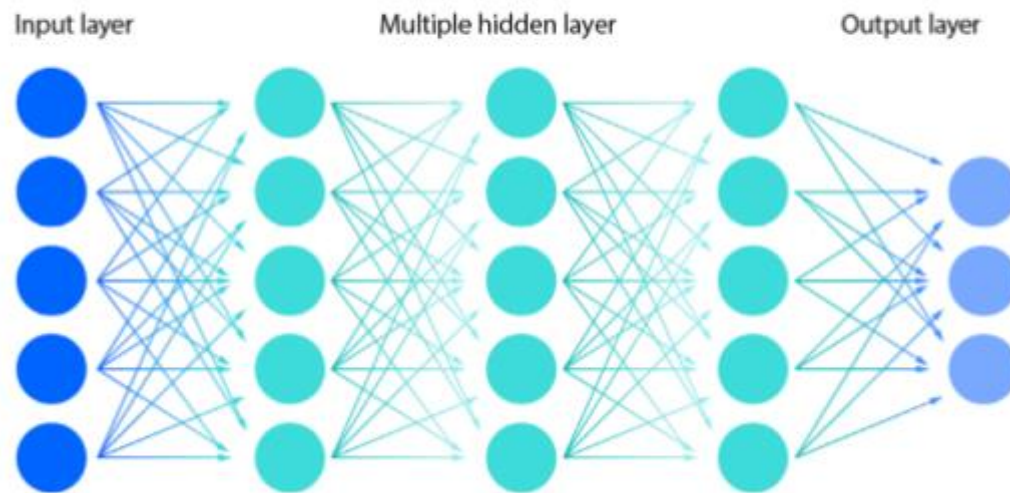
- Système d'analyse : permet d'analyser des données, d'identifier des schémas récurrents et de mettre en évidence des corrélations.
- Système de création : utilise des techniques d'analyse pour créer un modèle pouvant être ensuite utilisé pour produire un texte écrit ou une image.

#### 1.4.2 Deep Learning (DL)

Il s'agit d'un sous-ensemble du « Machine Learning ». Cette technique d'apprentissage supervisé s'appuie sur un réseau de neurones artificiels inspiré du cerveau humain, interconnectés par des fonctions mathématiques et dans lequel chaque neurone combine ses entrées et fournit une sortie. Plusieurs réseaux de neurones peuvent s'imbriquer. Les neurones sont disposés en couches successives : plus il y a de couches, plus le réseau est profond, d'où le terme « Deep Learning ».



## Deep neural network



*Figure 2. illustration du DL avec un réseau de neurones artificiels (source : IBM ;AI vs ML vs DL vs neuralnetworks : what's the difference ?)*

Le DL est utilisé pour des fonctions complexes (plusieurs données de départ doivent être traitées). Par exemple, afin de reconnaître l'image brute d'un mouton, les différentes couches de neurones vont ajuster les paramètres de calcul pour extraire de lui-même le concept de « mouton » et ce qui le définit (yeux, poils, oreilles, mufle, etc.)

C'est ce modèle d'intelligence artificielle qui est utilisé pour la reconnaissance des fractures.

### 1.5 L'enjeux des données, le BIG DATA

Pour que leur apprentissage soit efficace et fiable, les algorithmes nécessitent des données de qualités variées, suffisamment nombreuses et bien annotées.

Ces données peuvent être collectées automatiquement par des télescopes, des caméras de surveillance et des microphones mais elles peuvent également être produites par les internautes : requêtes sur les moteurs de recherche, publications et commentaires sur les réseaux sociaux, objets connectés, achats en ligne, etc.

Se pose alors le problème de la protection de ces données. En effet, nos données personnelles peuvent être exploitées par de grands groupes privés à des fins commerciales. C'est pourquoi le RGPD (règlement général sur la protection des données) est entré en application en 2018 afin de protéger les données personnelles des citoyens et leur exploitation.

## 1.6 Les limites de l'IA

L'avènement de l'IA soulève de nombreuses questions quant à son impact éthique et environnemental.

Nous avons rapidement évoqué ses limites en terme de protection des données.

De nombreux débats agitent la communauté scientifique, notamment sur son impact environnemental, aussi bien positif que néfaste[3], L'IA pourrait avoir un impact environnemental positif grâce à l'amélioration de la production d'énergies renouvelables, à la recherche de matériaux plus durables ou encore au monitoring des émissions énergétiques à grande échelle.

Mais elle pourrait également avoir un impact négatif direct sur l'environnement, causé par les Data center qui consomment beaucoup d'énergie non renouvelable, l'extraction des minéraux utiles à la fabrication du matériel et l'utilisation excessive d'eau ; tout comme un impact négatif indirect sur celui-ci causé par la surexploitation des ressources facilitée par l'optimisation des technologies. Ce concept est appelé « le paradoxe de Jevons » qui énonce qu'« à mesure que les améliorations technologiques augmentent l'efficacité avec laquelle une ressource est employée, la consommation totale de cette ressource peut augmenter au lieu de diminuer ».

Différentes études ont également montré la persistance voire l'aggravation de certaines discriminations par l'IA. Ce phénomène est dû au fait que les données utilisées pour créer les algorithmes reflètent les discriminations et biais présents dans nos sociétés, dans lesquelles certaines minorités ne sont pas ou peu étudiées (femme enceinte, population noire...) Par exemple : sous-diagnostic dans la population féminine, noires, pauvre [4].

L'impact de l'IA sur l'économie et notamment sur l'emploi soulève également des questions. Un rapport de l'organisation internationale du travail datant de 2023 et mis à jour en 2025 s'intéresse aux risques potentiels de l'IA, qui pourrait entraîner une transformation des métiers (10 à 13 % des professions dans le monde) [5].

Le défi actuel est d'accompagner, d'organiser et de réfléchir le déploiement de l'IA pour limiter les conséquences sociales.

## 2 Applications de l'IA en médecine d'urgence

Le développement de l'IA en médecine est en plein essor et celle-ci est utilisée dans la pratique quotidienne des médecins, que ce soit pour la recherche, la prévention ou la détection précoce de lésions cancéreuses pulmonaire [6,7].

Dans le domaine de la médecine d'urgence, l'IA pourrait avoir un réel impact sur la prise en charge des patients grâce à des outils utiles à chaque étape qui la compose.

En pré-hospitalier, elle permettrait d'aider à la prise de décision au cours de la régulation médicale et à la gestion des ressources humaines.

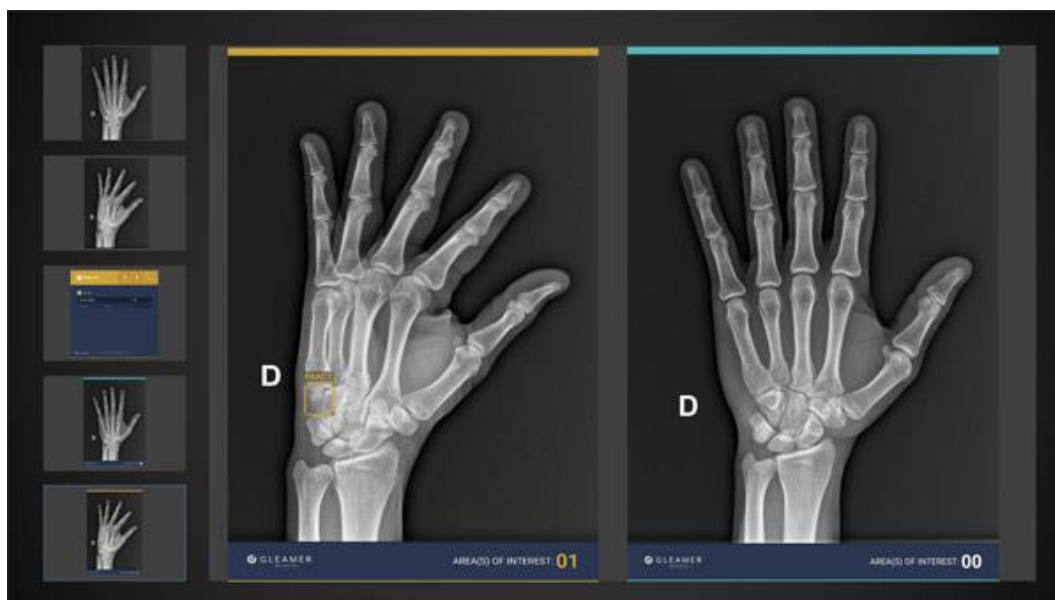
Aux urgences, elle apporterait une aide au triage (des revues de littérature ont d'ailleurs montré des modèles d'IA plus précis que l'ESI pour prédire la mortalité, les besoins de soins critiques ou d'hospitalisation) [8], une aide à la rédaction de l'examen clinique, à la récupération des antécédents et de l'histoire de la maladie, permettant ainsi un véritable gain de temps [9].

L'IA optimiserait également l'interprétation d'examens cliniques comme les électrocardiogrammes avec reconnaissance des STEMI ou de la fibrillation atriale et l'interprétation des radiographies [10].

### 3 Gleamer

Cofondée en 2017 à Paris par Christian Allouche, Alexis Ducarouge et le Dr. Nor-Eddine Regnard, l'entreprise GLEAMER développe des logiciels d'intelligence artificielle destinés à la radiologie et dotés d'outils capables de fournir un diagnostic préliminaire fiable basé sur des images médicales. Ces outils sont utilisés en radiologie ostéo-articulaire, en mammographie, en IRM et scanner. Ils sont utilisés par plus de 15 000 praticiens sur 2 000 sites répartis dans une quarantaine de pays. La plateforme de Gleamer traite plus de 30 millions d'examens médicaux chaque année [11]. Leur logiciel est utilisé au centre hospitalier de Roubaix depuis 2023 [12].

Gleamer utilise un modèle d'approche supervisée forte qui permet de détecter et de délimiter une région spécifique de l'image où la probabilité de présence d'une lésion est la plus élevée [13].



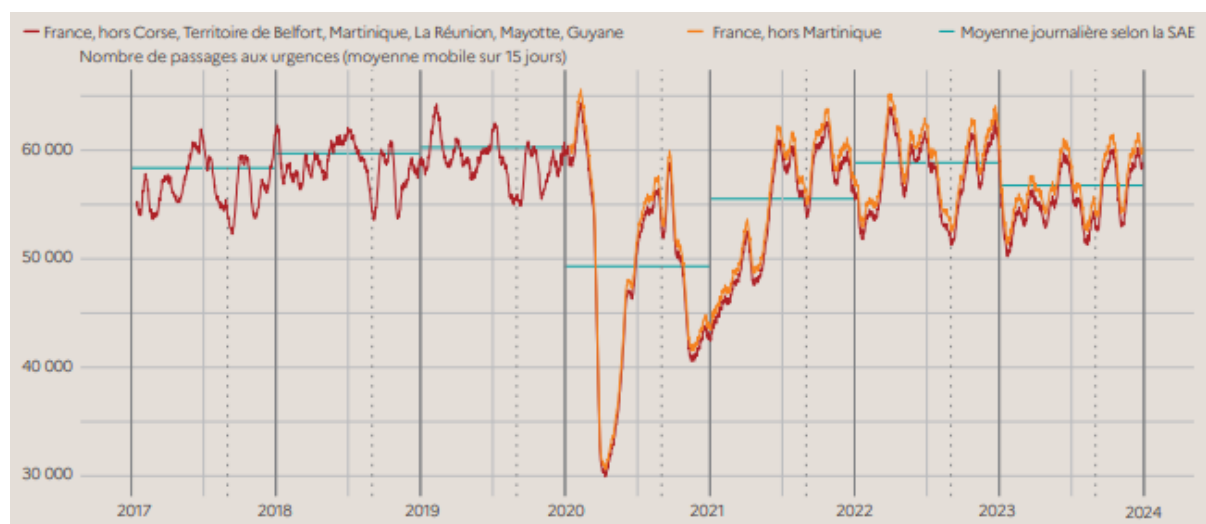
*Figure 3. Exemple d'interprétation d'une radiographie par Gleamer ( source : MED-KI Marktplatz ; BoneView Trauma)*

## 4 Contexte et enjeux actuel de l'IA en radiologie aux urgences

Le nombre de passages aux urgences est en augmentation constante depuis le début des années 2000, augmentant de 3,3% par an en moyenne (il est passé de 10,1 millions en 1996 à 22 millions en 2019).

Celui-ci a brutalement diminué en 2020 dû à l'épidémie de COVID19 pour atteindre 18,1 millions. Les consultations au SAU ont par la suite fortement augmentées en 2021 pour finalement légèrement diminuer en 2023 (environ 20,9 millions de passages selon les dernières statistiques de la DRESS en décembre 2024) [14].

Les urgences adultes et pédiatriques CH de Roubaix suivent également cette tendance avec 90 000 passages en 2024 soit une moyenne journalière de 246 patients.



*Figure 4 Nombre de passages aux urgences par jour de 2017 à 2023, (source : DREES de 2024 ; passage aux urgences entre 2017 et 2023 : des dynamiques contrastées selon les départements).*

Les lésions traumatiques constituent toujours la principale raison de passage aux urgences selon la dernière étude de la DRESS de mars 2025 sur l'année 2023. Elles concernent 1/3 des patients admis aux urgences [15].

La durée de passage s'est allongée depuis 2013 avec la moitié des patients qui passent plus de 3 heures aux urgences, soit 45 minutes de plus qu'il y a 10 ans, notamment pour les patients dont le diagnostic à la sortie fait état de lésions traumatiques. [15]

Un peu moins de la moitié des patients font l'objet d'un acte d'imagerie (45 % en 2023), cependant le type d'imagerie réalisée a évolué. La radiologie standard reste la plus fréquente, mais dans une moindre proportion (32 % des patients en bénéficient, contre 39 % en 2013), tandis que l'utilisation des scanners, IRM ou échographies a augmenté (17 % des patients sont concernés, contre 12 % en 2013) [15].

Cette situation soulève la question du risque d'erreur médicale qui est quotidien et nettement plus important aux urgences pour de multiples raisons : patient inconnu de

l'urgentiste et admis aux urgences pour des situations allant de la situation critique aux soins de base, état de fatigue de l'urgentiste selon le moment de la prise en charge, charge décisionnelle et cognitive très importante (plusieurs patients simultanés) devant un flux de patients important, interruption de tâches multiples, manque de formation, dossier non seniorisé, manque d'effectif, de lit d'aval... [16–18]

Les erreurs de diagnostic sont difficiles à chiffrer en France, mais ont fait l'objet d'une attention particulière par l'HAS [17,19]

Dans un papier du Canadian Journal of Emergency Medicine datant de 2001 [20]. On estime entre 5 et 10 % le taux de diagnostic erroné aux urgences. Aux États-Unis, celui-ci est chiffré à environ 5,7%, soit 1 patient sur 18 [21].

Parmi les erreurs diagnostics, l'absence de mise en évidence de fracture est la plus fréquente [20,22]. Des études rapportent le taux de fractures non diagnostiquées entre 1,4 et 3,7% mais celui-ci reste difficilement évaluable [23]. Ces erreurs diagnostiques entraînent un retard de prise en charge pour le patient avec un impact sur le pronostic fonctionnel du membre atteint pouvant entraîner un handicap. Aux États-Unis, on estime que les incapacités permanentes liées aux erreurs de diagnostic représentent environ 424 000 cas chaque année. Ces erreurs de diagnostics représentent un coût non négligeable pour la société américaine : environ 100 milliards de dollars, toutes erreurs de diagnostics confondues.[19]

Les erreurs d'interprétation radiologique de fractures sont la deuxième cause de poursuites judiciaires à l'encontre des radiologues, juste après les erreurs de diagnostic de cancer du sein [24].

En résumé, l'urgentiste est actuellement confronté à plusieurs problématiques :

- Gérer un flux efficacement devant le nombre de passages aux urgences en augmentation et des temps d'attente toujours plus longs ;
- Poser un diagnostic correct en minimisant le risque d'erreur et de mauvaise interprétation pouvant avoir de lourdes conséquences.

Le domaine de la traumatologie présente un enjeu de taille. S'agissant de la première cause de consultation aux urgences, il convient de pouvoir prendre en charge le patient de manière efficace en lui assurant un diagnostic précis à travers une interprétation correcte de l'imagerie, un traitement et un suivi adaptés. Il est nécessaire de réfléchir à l'utilisation d'outils permettant d'améliorer la prise en charge du patient : l'intelligence artificielle étant l'une des options disponibles pour les praticiens, elle permettrait de faire face à une demande croissante du volume d'imagerie et d'augmenter la qualité du diagnostic. De nombreux outils de détection des fractures par l'IA ont été développés dans le monde [13] et ils pourraient donc améliorer la prise en charge des patients en facilitant cette étape.

Une méta analyse réalisée en 2022 [25], comparant les performances diagnostiques entre IA et praticiens et incluant 42 études, suggère que les performances diagnostiques de l'IA sont non inférieures à celle des cliniciens et l'aide de l'IA dans l'interprétations permettrait une amélioration de la sensibilité et de spécificité. Mais cet

article a de nombreuses limites, regroupant des études hétérogènes, dont 52% d'entre elles comprenaient des biais.

Parmi les logiciels développés, Gleamer (Boneview), utilisé au CH de Roubaix apparaît comme une aide certaine à la détection des fractures, en effet plusieurs études ont démontré son efficacité dans l'aide à l'interprétation.

L'étude de Duron et al [26] a permis de montrer une augmentation de la sensibilité dans la détection des fractures pour de multiples localisations, notamment chez les urgentistes passant d'une sensibilité de 61,3 % à 74,3 % avec l'aide de l'IA ; mais également chez les radiologues avec une amélioration de 4,3 %, passant d'une sensibilité de 80,2 % à 84,6 %. L'aide de l'IA a amélioré la sensibilité des médecins de 8,7 % et la spécificité de 4,1%. Elle a également réduit le nombre de faux positifs par patient de 41,9 % chez les patients sans fracture et semble avoir amélioré le temps de lecture de 15 %. Enfin, les performances en autonomie du système d'IA étaient supérieures à celles de tous les lecteurs (sans aide de l'IA), y compris les radiologues, avec une aire sous la courbe caractéristique de fonctionnement du récepteur de 0,94 (IC à 95 % : 0,92, 0,96).

La deuxième étude américaine de Guermazi et al. [27] a elle aussi montré une amélioration de la sensibilité de détection des fractures grâce à l'IA de 10,4 % (75,2 % contre 64,8 %, supériorité  $P = 0,001$ ) avec des résultats similaires en termes de spécificité, de temps gagné et de performance en autonomie du logiciel d'IA.

Ces deux études ont mis en évidence une réduction du temps de lecture significative pour l'étude américaine avec - 6,3 secondes et l'étude française a montré une baisse non significative de 10 secondes.

Ainsi, Gleamer est un logiciel aidant à l'interprétation des radiographies que l'on sait efficace, fiable et semblant réduire le temps d'interprétation.

Une dernière étude s'est penchée sur les performances de ce logiciel de détection des fractures en pratique courante et en prospectif, au sein des urgences. En effet, cette étude française a comparé 3 logiciels de détection dont Gleamer [28] utilisant des radiographies provenant des passages aux urgences et non sélectionnées en amont contrairement aux études précédentes. Les résultats retrouvés pour Gleamer sont superposables aux autres études.

C'est en cela que notre travail est novateur, il se déroule en prospectif, la plupart des études concernant Gleamer sont réalisées de façon rétrospective sans accès à la clinique du patient. Ce travail se déroule en situation de pratique courante, avec des radiographies incluses à tout moment de la journée, de plus ou moins bonne qualité, et surtout, interprétées par des urgentistes et des internes qui sont les médecins de première ligne dans l'interprétation de ces radiographies de traumatismes.

En effet, bien que décrite comme une étude des performances diagnostiques des logiciels d'IA en vie réelle, les radiographies de l'étude de Bousson et al.[28] étaient interprétés à posteriori par des radiologues.

De plus nous nous intéresserons ici à son impact sur la prise en charge des patients : l'IA nous permet-elle d'améliorer la prise en charge des patients en adaptant le traitement, en limitant nos erreurs et en améliorant le suivi ?

# Matériel et méthodes

## 1 Design de l'étude

C'est une étude :

- Observationnelle
- Prospective
- Mono centrique
- Contrôlée
- Non randomisée
- Ouverte

L'étude s'est effectuée dans le service des urgences de l'hôpital de Roubaix sur le mois d'avril 2025, l'inclusion des patients a duré 1 mois : du 7 avril 2025 au 3 mai 2025.

Cette période a été choisie car elle correspond à la fin du semestre des internes, ils avaient donc une meilleure maîtrise du logiciel des urgences et du logiciel d'interprétations des radiographies.

Chaque patient âgé de 18 ans ou plus, admis aux urgences avec présence d'un traumatisme de membre et bénéficiant d'une radiographie a été inclus.

Après un examen clinique et la prescription d'une radiographie, une feuille de recueil est remplie par la personne prenant en charge le patient (interne ou médecin senior) où est renseignée une première interprétation de l'imagerie, sans l'aide de l'IA avec une conduite à tenir adaptée en fonction des résultats. Puis, une seconde partie où est renseignée une nouvelle interprétation de la même radiographie avec l'aide de l'IA, l'intervenant peut exprimer son accord ou son désaccord avec l'interprétation du logiciel de détection et la nouvelle prise en charge si celle-ci est modifiée.

Enfin, une relecture à distance par les radiologues de l'hôpital de Roubaix a été réalisée. Cette relecture sera utilisée comme Gold standard, en aveugle de l'interprétation des urgentistes.

## 2 Patients et données

### 2.1 Critères d'inclusion

Patient âgé au minimum de 18 ans.

Admis aux urgences du CH de Roubaix avec un traumatisme du squelette appendiculaire.



## 2.2 Critères de non inclusion

Traumatisme autre qu'appendiculaire.

Patient traumatisé grave.

Imagerie réalisée en ville avec résultats déjà disponibles.

Refus du patient.

Patient sous tutelle ou curatelle.

Patient en garde à vue.

Douleur d'un membre appendiculaire sans traumatisme.

Patient mineur.

## 2.3 Schéma de l'étude

Le patient se présente aux urgences avec un traumatisme de membre.

Un médecin sénior ou un interne recevra le patient en consultation et retiendra l'indication d'une radiographie concernant le ou les membre(s) traumatisé(s) après un examen clinique. Ce sera au médecin sénior ou à l'interne d'inclure le patient dans l'étude. Une explication orale de l'étude sera donnée au patient et une note d'information lui sera remise. Son accord de participation à l'étude sera recueilli dans son dossier médical et son opposition sera recueillie via le formulaire joint à la note d'information. Le patient bénéficiera par la suite de son imagerie qui sera interprétée par le médecin sénior ou l'interne conformément à la pratique courante. Si l'inclusion du patient est effective, le médecin sénior ou l'interne qui le prend en charge remplira la feuille concernant la prise en charge adaptée en fonction des résultats de la radiographie.

Le document à remplir sur la prise en charge sera rattaché au bon de radiographie (1 fiche à remplir par localisation). Il se présente sous la forme suivante (cf. annexe 1) :

Partie 1 : Interprétation de la radiographie sans Gleamer où la personne prenant en charge le patient (médecin sénior ou interne) devra renseigner :

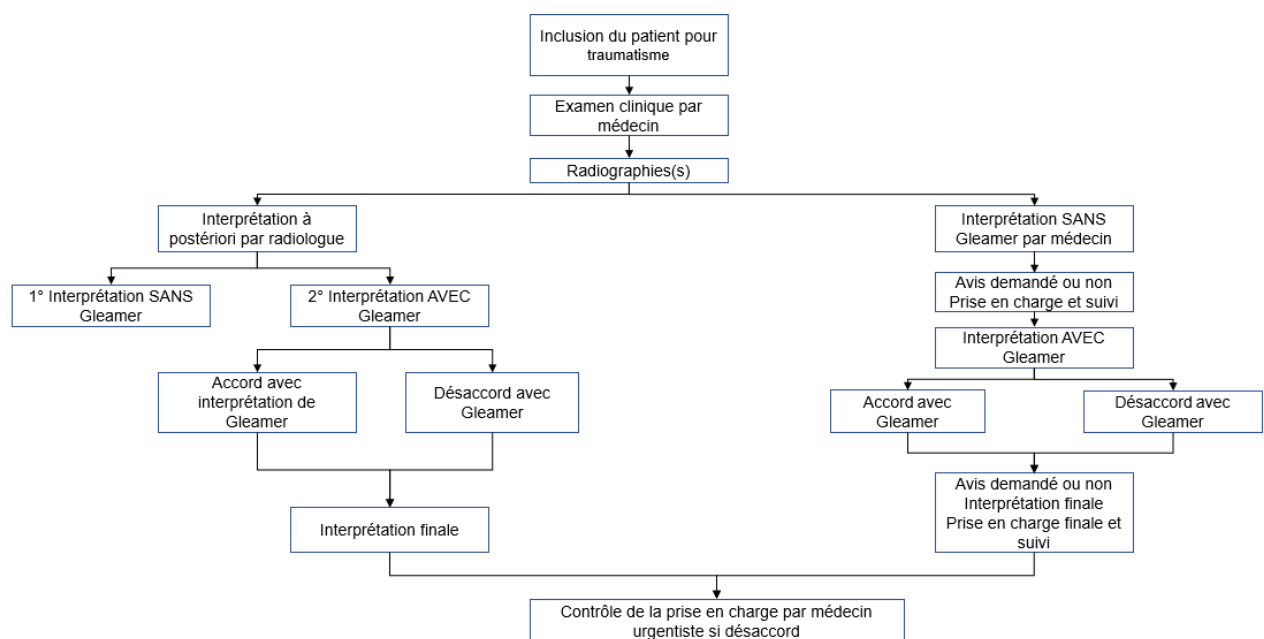
- Statut du praticien
- Moment de la prise en charge (journée de 8h-18h, garde de 18h à 8h)
- Localisation du traumatisme
- Clinique du patient
- Interprétation de la radiographie
- Avis demandé avec modification ou non de l'interprétation radiographique
- Prise en charge
- Suivi

Partie 2 : Seconde interprétation de la même radiographie avec l'aide de Gleamer, remplie par le même interne ou le médecin senior prenant en charge le patient :

- Résultat de l'interprétation par Gleamer

- Accord ou désaccord avec Gleamer (exemple : Gleamer met en évidence une lésion osseuse mais qui finalement se trouve être une lésion ancienne ou ne correspondant pas à l'endroit de la lésion lors de l'examen clinique. Le médecin senior ou l'interne peut donc décider que cette lésion rapportée par l'IA n'est pas en rapport avec le traumatisme actuel)
- Avis demandé (si demandé à cause de l'IA)
- Prise en charge finale après interprétation avec l'IA (modification de la prise en charge ?)
- Suivi (si modification avec nouvelle interprétation)

Le logiciel Gleamer est utilisé par le centre hospitalier de Roubaix depuis 2 ans, après discussion avec les responsables du logiciel, aucune solution n'a pu être trouvée pour séparer de façon claire les deux interprétations. Il a donc été décidé de réaliser l'étude sans modifier le fonctionnement habituel. Les intervenants auront comme consigne d'interpréter sans regarder les clichés où l'IA est présente pour la première étape. À Roubaix, l'interprétation par Gleamer se trouve après les radiographies originales, il est donc possible d'interpréter sans, puis avec l'IA en faisant dérouler les clichés.



**Figure 5. Schéma de l'étude**

La relecture à postériori par notre gold standard (ici le radiologue), se fera de la même manière avec un document à remplir avec une première partie où le médecin interprète sans l'aide de l'IA et une seconde partie où il interprète cette fois-ci avec l'IA et peut également dire s'il est d'accord ou non avec l'interprétation du logiciel (cf. annexe 2).

## **2.4 Recueil de données**

Les données seront recueillies de manière anonyme dans les dossiers papiers ou informatisés des patients sur le logiciel Easily, utilisé au CH de Roubaix. Les patients seront identifiés via un numéro unique dans l'étude et seul l'investigateur du centre détiendra la table des correspondances entre ces numéros et l'identité des participants à la recherche. Cette correspondance sera établie à l'aide du logiciel Excel sur un fichier codé, protégé par un mot de passe, sur l'un des ordinateurs de l'établissement, accessible lui-même par mot de passe, et ne sortira pas de l'établissement.

Le recueil de données en lui-même se fera sur un tableur Excel. Le tableau sera anonymisé en fonction de la liste de correspondance des patients détenue par l'investigateur sans qu'il n'y soit fait mention des noms, prénoms ou dates de naissance des patients. Ce tableur sera stocké sur les serveurs du Centre Hospitalier de Roubaix.

## **3 Objectif de l'étude**

### **3.1 Objectif principal et critère de jugement principal**

Évaluer l'impact d'un logiciel de détection de fracture du squelette appendiculaire basé sur l'IA sur la prise en charge des patients : traitement fonctionnel, orthopédique ou chirurgical.

Le critère de jugement principal sera une variable qualitative sur le type de prise en charge : fonctionnelle, orthopédique, chirurgicale après interprétation avec ou sans Gleamer.

### **3.2 Objectif secondaires (OS) et leurs critères de jugement**

OS 1 : Impact d'un logiciel de détection de fracture sur le suivi des patients.

Critère de jugement OS 1 : Suivi et délai de suivi pour évaluer l'impact du logiciel de détection des fractures sur le suivi des patients

OS 2 : Analyse en sous-groupe de l'impact d'un logiciel de détection de fracture sur la prise en charge des patients : journée/garde, sénior/interne

Critère de jugement OS 2 : Les analyses en sous-groupes regroupent plusieurs variables qualitatives :

- Heure de réalisation de l'examen pour l'analyse entre journée et garde
- Personne ayant demandé l'examen pour l'analyse entre sénior et interne

OS 3 : Évaluation des performances diagnostiques des médecins avec ou sans aide de l'IA en condition réelle.

Critère de jugement OS 3 : Concordance entre l'interprétation avec l'IA et l'interprétation sans l'IA vs gold standard (médecin expert : radiologue).

### 3.3 Gold Standard et contrôle des prises en charge

La majorité des études utilisent comme gold standard le consensus d'interprétation des radiographies par 3 radiologues experts. Dans notre étude, le Gold standard choisi est la relecture des radiographies par des radiologues seniors avec l'aide de l'IA.

La décision de procéder comme tel comprend plusieurs raisons :

- 1) C'est la méthode qui se rapproche le plus de la réalité, le scanner bien qu'accessible n'est pas réalisé à chaque radiographie, l'aide du radiologue ou du chirurgien est dans un premier temps demandé.
- 2) Une étude Lilloise[29] s'intéressant aux performances diagnostiques de Gleamer pour les fractures de poignets et de mains avec comme gold standard le scanner ayant une sensibilité de quasi 100 % mettait en évidence des résultats similaires aux études précédentes concernant l'amélioration de la sensibilité de détection des fractures.

Concernant le contrôle des décisions de prise en charge aux urgences, le gold standard a été la relecture à posteriori par un urgentiste de la prise en charge, si une discordance entre l'interprétation de l'urgentiste et du radiologue existait. Si la concordance était bonne entre l'interprétation de l'urgentiste et celle du radiologue, alors la prise en charge est considérée comme correcte.

## 4 Analyse statistique

Pour l'analyse descriptive, les données qualitatives ont été présentées avec leur effectif et taux et les données quantitatives avec leur moyenne et déviation standard.

S'agissant d'une étude observationnelle, nous n'avons pas calculé de nombre de sujets nécessaires a priori. Les urgences de Roubaix réalisent entre 50 et 100 radiographies standards par jour aux urgences, nous estimons donc notre nombre d'inclusions à 300 radiographies de zones anatomiques différentes.

Afin d'évaluer la concordance entre deux variables qualitatives (avec ou sans IA) et une mesure de référence (gold standard), nous avons mené une analyse de la concordance sur des données appariées comportant trois modalités ordinales (traitement fonctionnel, orthopédique ou chirurgical).

Le choix d'une méthode de calcul par rapport à une autre repose sur le déséquilibre dans la distribution de la variable analysée, nous avons considéré que si l'une des modalités représentait plus de 80 % des cas, une méthode de Gwet devait être appliquée, sinon nous devions utiliser le kappa pondéré.

Le kappa pondéré est utilisé pour estimer l'accord entre chaque test et le gold standard, en tenant compte du degré de désaccord entre les catégories. Le kappa pondéré est particulièrement adapté aux variables ordinales, car il accorde des poids

moindres aux discordances proches et permet une interprétation plus nuancée que le kappa standard de Cohen.

Le choix du kappa pondéré repose sur la nature ordinale des réponses, ainsi que sur la présence de déséquilibres de fréquence entre les modalités, qui peuvent biaiser les mesures de concordance standard comme le kappa de Cohen. L'indice AC1 de Gwet, bien que plus robuste aux prévalences extrêmes, n'a pas été retenu dans l'analyse finale car les déséquilibres observés ne justifiaient pas son utilisation, et l'interprétation du kappa pondéré restait stable et cohérente dans notre jeu de données.

Pour chaque test, nous avons estimé la valeur du kappa pondéré, accompagné de leur intervalle de confiance à 95 %. Une comparaison des deux tests a ensuite été effectuée en examinant les différences absolues et relatives de ces coefficients de concordance avec leur intervalle de confiance à 95 %. Les différences étaient jugées significatives si l'intervalle de confiance excluait 0.

Nous avons ensuite calculé les performances diagnostiques de l'interprétation avec et sans IA en calculant les sensibilité et spécificité, VPP et VPN, par rapport au gold standard pour chaque type de diagnostic (Fracture, Luxation, Epanchement) et par rapport à la présence d'au moins une lésion avec leur intervalle de confiance à 95 %. Les performances diagnostiques avec et sans IA ont été comparées en examinant les différences absolues et relatives de la sensibilité et de la spécificité avec leur intervalle de confiance à 95 %. Les différences étaient jugées significatives si l'intervalle de confiance excluait 0.

Nous avons également réalisé des tests du chi<sup>2</sup> (ou Fisher si les conditions de réalisation d'un test paramétrique n'étaient pas réunies) pour comparer des variables qualitatives.

Le seuil de significativité pour l'ensemble des analyses était de 5 %.

Les analyses ont été réalisées avec le logiciel R 4.5.1.

## 5 Cadre réglementaire

### 5.1 Considération éthique et réglementaire

#### 1.1. Principes généraux

Conformément à la réglementation en vigueur, le promoteur de l'étude ainsi que le principal investigateur attestent que le protocole contient toutes les informations nécessaires à la conduite de l'étude. Ils s'engagent à réaliser l'étude en respectant le protocole et les termes et conditions qui y sont définis.

Le promoteur de l'étude et le principal investigateur s'engagent également à ce que les co-investigateurs et les autres membres qualifiés de l'équipe aient accès aux copies de ce protocole et des documents relatifs à la conduite de l'étude pour leur permettre de travailler dans le respect des dispositions figurant dans ces documents.

### 1.2. Aspect réglementaire

Le projet est déclaré conforme à la méthodologie de référence MR-004.

Le projet a été validé devant le comité d'éthique interne du Centre Hospitalier de Roubaix.

### 1.3. Information des patients

Conformément aux dispositions du code de santé publique, des Bonnes Pratiques Cliniques et du Règlement Européen sur la Protection des Données (RGPD), le patient recevra une information claire et compréhensible de la recherche par l'investigateur.

Une note d'information écrite lui sera remise. Celle-ci sera remise soit en main propre si le patient est présent au sein du centre hospitalier au moment du recueil des données, soit par courrier au dernier domicile connu du patient.

Pour les patients décédés, et pour lesquels l'investigateur en a la connaissance et la certitude, l'information ne pouvant être délivrée, l'investigateur pourra utiliser les données de ces patients conformément au point 2.5.1 du Titre II de la MR- 004.

Cette utilisation peut avoir lieu sous réserve que le patient ne se soit pas opposé, de son vivant, à l'utilisation de ses données à des fins de recherche. Cette vérification doit être réalisée en amont par l'investigateur.

La note d'information reprendra l'objectif de la recherche, son déroulement, les bénéfices et risques liés pour le patient ainsi que les contraintes de la recherche et ses droits. Les modalités de traitement et conservation de ses données personnelles seront également exposées conformément à l'article 13 du RGPD.

À l'issu de l'information, le patient bénéficiera d'un délai de réflexion et pourra poser toutes les questions nécessaires à l'investigateur.

Dans le cas où l'information a été délivrée par courrier. Les patients disposent d'un délai de trois (3) semaines pour s'opposer à l'utilisation de ses données. Passé ce délai, il sera considéré que le patient accepte l'étude et il sera inclus dans celle-ci.

La note d'information contiendra un coupon-réponse que les patients pourront renvoyer afin de notifier leur refus de participer à l'étude.

L'investigateur mentionnera dans le dossier médical du patient sa participation à la recherche.

Le patient peut, à tout moment, décider d'arrêter sa participation à la recherche sans préjudice pour la prise en charge de sa pathologie ni sur sa relation avec son médecin.

### 1.4. Protection des données

Le traitement des données sera réalisé conformément aux dispositions de la loi Informatique et Liberté du 06 janvier 1978 et du Règlement Européen sur la Protection des Données (RGPD). L'étude a été enregistrée sur le dispositif MR-004 concernant les recherches n'impliquant pas la personne humaine de l'établissement (déclaration n° 2213245 en date du 26 avril 2019).

Le projet est enregistré au registre des traitements du CH Roubaix.

Afin de mener à bien cette recherche, le promoteur organisera un traitement des données administratives des centres investigateurs participant.

Les investigateurs et professionnels impliqués dans la recherche peuvent exercer leurs droits d'accès, de rectification, d'effacement, de limitation, de portabilité, de transmission ou d'opposition au traitement de leurs données personnelles. Ces droits s'exercent auprès du Délégué à la Protection des Données du Centre Hospitalier de Roubaix : [dpo@ch-roubaix.fr](mailto:dpo@ch-roubaix.fr)

## **5.2 Gestion des données**

### 1.1. Modalité de recueil des données

Les données seront recueillies de manière anonyme dans les dossiers informatisés des patients sur le logiciel Easily utilisé au CH de Roubaix. Les patients seront identifiés via un numéro unique dans l'étude et seul l'investigateur du centre détiendra la table des correspondances entre ces numéros et l'identité des participants à la recherche. Cette correspondance sera établie à l'aide du logiciel Excel sur un fichier codé, protégé par un mot de passe, sur l'un des ordinateurs de l'établissement, accessible lui-même par mot de passe, et ne sortira pas de l'établissement.

Le recueil de données en lui-même se fera sur un tableur Excel. Le tableau sera pseudonymisé en fonction de la liste de correspondance des patients détenue par l'investigateur sans qu'il n'y soit fait mention des noms, prénoms ou dates de naissance des patients. Ce tableur sera stocké sur les serveurs du Centre Hospitalier de Roubaix.

### 1.2. Droit d'accès aux données et documents sources

Cette étude faisant l'objet de la thèse en médecine du porteur de projet, Dr BRABANT Lubin, les CRF seront remplis par celui-ci sous la responsabilité de l'investigateur coordonnateur.

Cette collecte des données sera réalisée directement dans le centre d'investigation à l'aide du matériel informatique à disposition dans ce centre.

### 1.3. Durée de conservation et archivage

Conformément à la méthodologie de référence MR-004 établie par la Commission Nationale Informatique et Liberté (CNIL) et à laquelle l'étude se déclare conforme, la base de données constituée sera conservée dans les systèmes d'information du Centre Hospitalier de Roubaix pendant la durée de réalisation de l'étude et jusqu'à deux ans après la dernière publication des résultats de l'étude (ou, à défaut, à la date de signature du rapport final de l'étude).

Puis, la base de données sera archivée pendant 5 ans maximum au sein de

l'établissement. Passé ce délai, la base de données ainsi que tous les documents relatifs à la recherche seront définitivement détruits.



# Résultats

## 1 Flow chart

Sur la période d'inclusion du 7 avril au 3 mai 2025, il y a eu 3975 passages aux urgences, parmi lesquels 125 ont été admis avec comme motif d'entrée : traumatisme du membre supérieur, membre inférieur et /ou multiples traumatismes.

Au total, 188 personnes ont été incluses, 1 personne en garde à vue et 15 patients mineurs ont secondairement été exclus, soit 172 patients restants.

Dans l'ensemble des patients inclus, 1 patient a été inclus 2 fois sur 2 passages différents, 3 patients auraient pu être à nouveau inclus étant revenus aux urgences pour un traumatisme.

Après analyse des questionnaires, 7 patients ont été exclus devant l'absence de données suffisantes.

Au final, 165 patients ont été analysés, soit 185 radiographies.

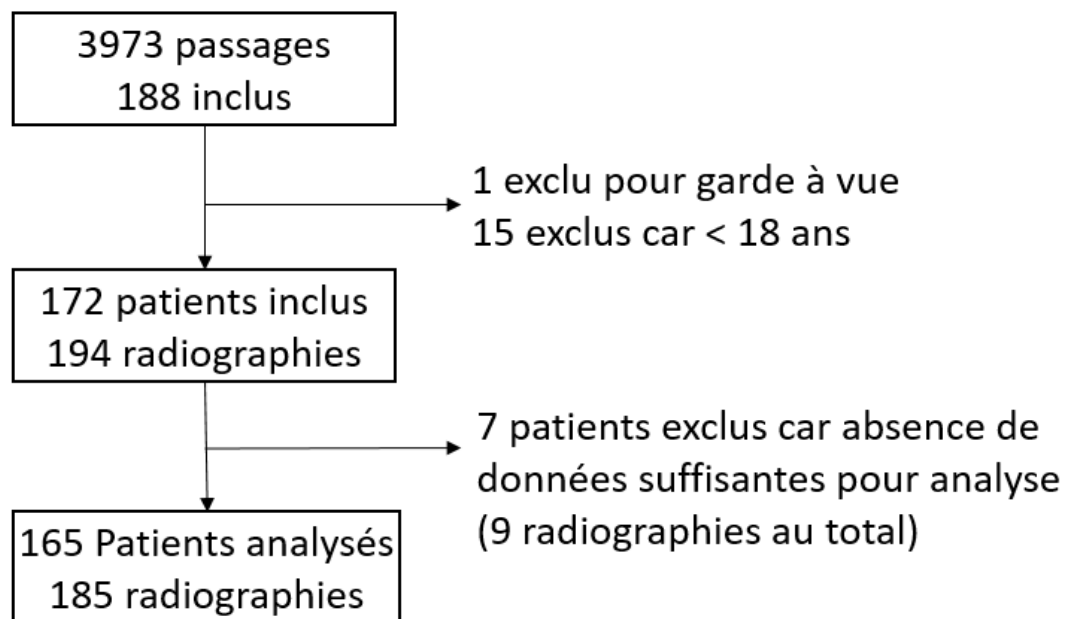


Figure 6. Flow chart de l'étude

## 2 Description des radiographies à l'inclusion

### 2.1 Radiographies des patients

Sur tous les clichés radiographiques, 185 au total, 52,4 % sont des hommes et la moyenne d'âge est de 46,9 ans.

La localisation principale était le membre inférieur avec 56,2 % des clichés contre 43,8 % pour le membre supérieur avec respectivement une présence majoritaire des traumatismes de la cheville et du pied ainsi que du poignet et de la main.

Les principaux éléments cliniques retrouvés sont la douleur pour 78,9 % des cas et l'impotence fonctionnelle pour 56,8 %.

On retrouve 49 fractures, 16 épanchements, 2 luxations et 120 radiographies sans lésions.

**Tableau 1** *Tableau descriptif des clichés analysés par patient*

| Overall (N= 185)                   |             |                                 |             |
|------------------------------------|-------------|---------------------------------|-------------|
| Démographie                        |             | Clinique, n (%)                 |             |
| <b>Sexe, Homme, n (%)</b>          | 97 (52.4%)  | <b>Hématome</b>                 | 41 (22.2%)  |
| <b>Âge, en année, moyenne (DS)</b> | 46.9 (21)   | <b>Déformation</b>              | 20 (10.8%)  |
| <b>Localisation, n (%)</b>         |             | <b>Oedeme</b>                   | 73 (39.5%)  |
| <b>Membre supérieur</b>            | 81 (43.8%)  | <b>Douleur</b>                  | 146 (78.9%) |
| <b>Épaule</b>                      | 16 (8.6%)   | <b>Impotence</b>                | 105 (56.8%) |
| <b>Coude</b>                       | 17 (9.2%)   | <b>Plaie</b>                    | 13 (7.0%)   |
| <b>Poignet/main</b>                | 48 (25.9%)  | <b>Déficit sensitivo-moteur</b> | 2 (1.1%)    |
| <b>Membre inférieur</b>            | 104 (56.2%) |                                 |             |
| <b>Hanche</b>                      | 13 (7.0%)   |                                 |             |
| <b>Genou</b>                       | 31 (16.8%)  |                                 |             |
| <b>Cheville/Pied</b>               | 60 (32.4%)  |                                 |             |

## 2.2 Intervenants

Les intervenants sont en majorité des internes/externes représentant 58,4 % des intervenants contre 41,6 % de sénior.

La plupart des inclusions ont eu lieu en journée, de 9h à 18h, soit 71,4 %.

Dans 47,03 % des cas, un avis a été demandé, 40,23 % à un senior et 59,77 % à un spécialiste (radiologue ou chirurgien). Dans 5 cas, il y a eu deux avis pour la même radiographie, avec un avis auprès d'un senior ainsi qu'auprès du spécialiste.

Sur les 83 avis demandés, seulement 7 ont permis de modifier l'interprétation de la radiographie.

**Tableau 2** *Intervenants, période d'inclusion et avis demandé*

| Overall (N=185)        |             |
|------------------------|-------------|
| <b>Clinicien</b>       |             |
| <b>Seniors</b>         | 77 (41.6%)  |
| <b>Interne/Externe</b> | 108 (58.4%) |
| <b>Période</b>         |             |
| <b>Journée (9-18h)</b> | 132 (71.4%) |

|                                                |                 |
|------------------------------------------------|-----------------|
| <b>Soirée (18h-00h)</b>                        | 45 (24.3%)      |
| <b>Nuit profonde (00h- 9h)</b>                 | 8 (4.3%)        |
| <b>Avis demandé</b>                            | 87 (47.03%)     |
| <b>Avis sénior</b>                             | 35/87 (40.23 %) |
| <b>Avis spécialiste</b>                        | 52/87 (59.77%)  |
| <b>Modification interprétation après avis.</b> | 7/83 (8.4%)     |

### 3 Impact de l'IA sur la prise en charge des patients

L'objectif principal de cette étude est d'étudier l'impact de l'IA sur la prise en charge des patients. On s'intéresse donc aux erreurs de prise en charge avec et sans IA.

*Tableau 3 Nombre d'erreurs avec ou sans IA*

|                           | <b>Nombre d'erreur</b> | <b>Kappa pondéré</b> | <b>AC1 de Gwet</b>  |
|---------------------------|------------------------|----------------------|---------------------|
|                           | n (%)                  | r [IC95]             | r [IC95]            |
| <b>Sans IA</b>            | 13 (7.0)               | 0.89 [0.82 – 0.97]   | 0.91 [0.86 – 0.96]  |
| <b>Avec IA</b>            | 6 (3.2)                | 0.93 [0.87 – 1.00]   | 0.96 [0.92 – 0.99]  |
| <b>Différence absolue</b> | -                      | 0.04 [-0.01, 0.04]   | 0.05 [-0.01 – 0.11] |

Après analyse, on remarque que l'utilisation de l'intelligence artificielle s'accompagne d'une diminution du nombre d'erreurs commises par les cliniciens, passant de 13 (7,0 %) sans IA à 6 (3,2 %) avec IA. Les indices de concordance confirment cette tendance. Le kappa pondéré augmente légèrement avec l'IA : 0,89 [0,82–0,97] sans IA vs 0,93 [0,87–1,00] avec IA, avec une différence absolue de +0,04, bien que l'intervalle de confiance croisé inclue 0.

La concordance avec et sans IA est considérée comme bonne étant supérieure à 0,8, les décisions de prise en charge sont dans l'ensemble correctes.

Ces résultats suggèrent que l'assistance par IA réduit les erreurs et tend à améliorer la fiabilité du diagnostic, même si les intervalles de confiance indiquent que l'ampleur réelle du bénéfice pourrait varier.

### 4 Impact d'un logiciel d'IA sur le suivi des patients

Le premier objectif secondaire était d'étudier l'impact d'un logiciel de détection des fractures sur le suivi des patients. Malheureusement, après analyse des inclusions, de nombreuses données sont manquantes ne permettant pas d'analyser cet objectif secondaire.

## 5 Analyse en sous-groupes de l'impact de l'IA sur la prise en charge

Tableau 4 Nombre d'erreur selon sous-groupes, Test Khi2 (\*), Test de Fisher (\*\*)

| Nombre d'erreur            | Sans IA    |          |         | Avec IA    |         |         |
|----------------------------|------------|----------|---------|------------|---------|---------|
|                            | Non        | Oui      | P-value | Non        | Oui     | P-value |
| <b>Intervenant, n (%)</b>  |            |          |         |            |         |         |
| Senior                     | 71 (41.5)  | 6 (46.2) | 0.97*   | 74 (41.6)  | 3 (50)  | 0.70**  |
| Interne/Externe            | 100 (58.5) | 7 (53.8) |         | 104 (58.4) | 3 (50)  |         |
| <b>Période, n (%)</b>      |            |          |         |            |         |         |
| Journée                    | 123 (71.5) | 9 (69.2) | 1**     | 129 (72.1) | 3 (50)  | 0.36**  |
| Garde                      | 49 (28.5)  | 4 (30.8) |         | 50 (27.9)  | 3 (50)  |         |
| <b>Localisation, n (%)</b> |            |          |         |            |         |         |
| Membre supérieur           | 77 (55.2)  | 4 (30.8) | 0.49*   | 81 (45.3)  | 0 (0)   | 0.036** |
| Membre inférieur           | 95 (44.8)  | 9 (69.2) |         | 98 (54.7)  | 6 (100) |         |

Le deuxième objectif secondaire est l'analyse en sous-groupe de l'impact d'un logiciel de détection de fracture sur la prise en charge des patients en comparant le moment de l'inclusion (Journée/garde), les intervenants (Sénior/interne) et également selon la localisation (membre supérieur, membre inférieur).

On constate que le nombre d'erreurs est équivalent entre les seniors et le groupe interne/externe sans l'IA, avec respectivement 6 et 7 erreurs. Grâce à l'IA, on remarque une diminution du nombre d'erreurs avec 3 erreurs dans les deux groupes.

La différence entre les deux groupes sans IA n'est pas significative avec une p-value à 0,97, la différence n'est également pas significative avec IA avec une p-value à 0,7.

Il n'y pas de différence significative entre la période d'inclusion et la baisse des erreurs, notamment pour la période « garde » où le nombre d'erreur de prise en charge comparé à la période « journée » est similaire avec une p-value à 0,36.

Concernant le nombre d'erreurs selon la localisation, on retrouve 9 erreurs pour le membre inférieur et 4 erreurs pour le membre supérieur sans l'utilisation de l'IA. Avec l'IA, le nombre d'erreur chute à 0 pour le membre supérieur et augmente à 6 pour le membre inférieur. La diminution des erreurs membre supérieur et inférieur avec IA est significative avec une p-value à 0,036.

## 6 Impact d'un logiciel d'IA sur la détection des lésions

Le troisième objectif secondaire est d'évaluer les performances diagnostiques des médecins avec ou sans aide de l'IA en condition réelle. On s'intéresse donc à la concordance entre l'interprétation avec l'IA et l'interprétation sans l'IA par notre gold standard, ici le radiologue.

Les performances diagnostiques ont été analysées selon le type de lésion (fracture, épanchement etc...). Puis selon le type de lésion par membre (fracture sur membre supérieur etc...)

## 6.1 Performances diagnostiques selon le type de lésion

*Tableau 5 Tableau des performances diagnostiques selon le type de lésion. NaN (non analysé)*

| <b>Fracture</b>       |                   |                   |                    |                     |
|-----------------------|-------------------|-------------------|--------------------|---------------------|
| <b>Paramètre</b>      | Clinicien sans IA | Clinicien avec IA | Différence absolue | Différence relative |
| <b>Sensibilité</b>    | 0.84 (0.70, 0.93) | 0.94 (0.83, 0.99) | 0.1 (0, 0.20)      | 11.6 (-26, 68.5)    |
| <b>Spécificité</b>    | 0.97 (0.93, 0.99) | 0.95 (0.90, 0.98) | -0.02 (-0.06, 0.0) | -2.2 (-23.2, 24.4)  |
| <b>VPP</b>            | 0.91 (0.79, 0.98) | 0.87 (0.75, 0.95) | -                  | -                   |
| <b>VPN</b>            | 0.94 (0.89, 0.98) | 0.98 (0.94, 1.00) | -                  | -                   |
| <b>Luxation</b>       |                   |                   |                    |                     |
| <b>Paramètre</b>      | Clinicien sans IA | Clinicien avec IA | Différence absolue | Différence relative |
| <b>Sensibilité</b>    | 1.00 (0.16, 1.00) | 1.00 (0.16, 1.00) | 0 (0, 0)           | 0 (-85.9, 609.9)    |
| <b>Spécificité</b>    | 0.99 (0.97, 1.00) | 1.00 (0.98, 1.00) | 0.01 (-0.01, 0.02) | 0.5 (-18.1, 23.4)   |
| <b>VPP</b>            | 0.67 (0.09, 0.99) | 1.00 (0.16, 1.00) | -                  | -                   |
| <b>VPN</b>            | 1.00 (0.98, 1.00) | 1.00 (0.98, 1.00) | -                  | -                   |
| <b>Epanchement</b>    |                   |                   |                    |                     |
| <b>Paramètre</b>      | Clinicien sans IA | Clinicien avec IA | Différence absolue | Différence relative |
| <b>Sensibilité</b>    | 0.98 (1.00, 0.95) | 0.98 (1.00, 0.95) | NaN                | NaN                 |
| <b>Spécificité</b>    | 0.62 (0.35, 0.85) | 0.38 (0.15, 0.65) | NaN                | NaN                 |
| <b>VPP</b>            | 0.33 (0.07, 0.70) | 0.23 (0.05, 0.54) | -                  | -                   |
| <b>VPN</b>            | 0.94 (0.03, 0.10) | 0.97 (0.01, 0.07) | -                  | -                   |
| <b>Toutes lésions</b> |                   |                   |                    |                     |
| <b>Paramètre</b>      | Clinicien sans IA | Clinicien avec IA | Différence absolue | Différence relative |
| <b>Sensibilité</b>    | 0.82 (0.70, 0.90) | 0.91 (0.81, 0.97) | 0.09 (0, 0.18)     | 10.5 (-22.9, 58.1)  |
| <b>Spécificité</b>    | 0.92 (0.85, 0.96) | 0.94 (0.88, 0.98) | 0.03 (-0.02, 0.07) | 2.6 (-20.6, 32.7)   |
| <b>VPP</b>            | 0.84 (0.73, 0.92) | 0.89 (0.79, 0.96) | -                  | -                   |
| <b>VPN</b>            | 0.90 (0.83, 0.95) | 0.95 (0.89, 0.98) | -                  | -                   |

L'analyse comparative des performances diagnostiques des cliniciens avec et sans assistance par l'IA montre des résultats contrastés selon le type de lésion. Pour les fractures, l'IA semble améliorer la sensibilité (Se) (0,84 à 0,94 ; différence absolue +0,10, différence relative +11,6 %), sans impact défavorable sur la spécificité, qui reste élevée (0,97 vs 0,95). Concernant les luxations, aucune amélioration notable n'est observée avec une sensibilité qui demeure identique (1,00) et la spécificité n'évolue que marginalement (+0,01). En revanche, pour l'épanchement, l'IA n'apporte pas de bénéfice, avec une sensibilité inchangée (0,98) mais une spécificité réduite (0,62 à 0,38), traduisant une perte de performance diagnostique, également reflétée par une diminution du VPP et du VPN.

Enfin, lorsque toutes les lésions sont considérées conjointement, l'IA semble apporter un gain de sensibilité (0,82 à 0,91 ; différence absolue +0,09, différence relative +10,5 %), tandis que la spécificité reste globalement stable (0,94 vs 0,93).

## 6.2 Performances diagnostiques selon localisation du traumatisme et type de lésion

*Tableau 6 Performances diagnostiques des fractures selon la localisation du traumatisme*

| Fracture                | Clinicien sans IA | Clinicien avec IA | Différence absolue  | Différence relative |
|-------------------------|-------------------|-------------------|---------------------|---------------------|
| <b>Membre supérieur</b> |                   |                   |                     |                     |
| Sensibilité             | 0.82 (0.60, 0.95) | 0.95 (0.77, 1.00) | 0.14 (-0.05, 0.33)  | 16.7 (-37.8, 119)   |
| Spécificité             | 0.97 (0.88, 1.00) | 0.97 (0.88, 1.00) | 0 (-0.05, 0.05)     | 0 (-30.5, 43.9)     |
| <b>Membre inférieur</b> |                   |                   |                     |                     |
| Sensibilité             | 0.85 (0.66, 0.96) | 0.93 (0.76, 0.99) | 0.07 (-0.03, 0.18)  | 8 (-37.3, 86.1)     |
| Spécificité             | 0.97 (0.91, 1.00) | 0.94 (0.85, 0.98) | -0.04 (-0.1, 0.02)  | -3.9 (-30.3, 32.4)  |
| <b>Epaule</b>           |                   |                   |                     |                     |
| Sensibilité             | 1.00 (0.4, 1.00)  | 1.00 (0.4, 1.00)  | 0 (0, 0)            | 0 (-75, 299.8)      |
| Spécificité             | 0.92 (0.62, 1.00) | 0.92 (0.62, 1.00) | 0 (0, 0)            | 0 (-55.1, 122.6)    |
| <b>Coude</b>            |                   |                   |                     |                     |
| Sensibilité             | 0.6 (0.15, 0.95)  | 1.00 (0.48, 1.00) | 0.4 (-0.03, 0.83)   | 66.7 (-60.2, 597.4) |
| Spécificité             | 1.00 (0.74, 1.00) | 1.00 (0.74, 1.00) | 0 (0, 0)            | 0 (-55.1, 122.6)    |
| <b>Poignet/Main</b>     |                   |                   |                     |                     |
| Sensibilité             | 0.85 (0.55, 0.98) | 0.92 (0.64, 1.00) | 0.08 (-0.18, 0.33)  | 9.1 (-51.9, 147.2)  |
| Spécificité             | 0.97 (0.85, 1.00) | 0.97 (0.85, 1.00) | 0 (-0.08, 0.08)     | 0 (-37.8, 60.9)     |
| <b>Hanche</b>           |                   |                   |                     |                     |
| Sensibilité             | 1.00 (0.54, 1.00) | 1.00 (0.54, 1.00) | 0 (0, 0)            | 0 (-67.7, 210.1)    |
| Spécificité             | 0.86 (0.42, 1.00) | 0.71 (0.29, 0.96) | -0.14 (-0.4, 0.12)  | -14.3 (-71.2, 155)  |
| <b>Genou</b>            |                   |                   |                     |                     |
| Sensibilité             | 1.00 (0.16, 1.00) | 1.00 (0.16, 1.00) | 0 (0, 0)            | 0 (-85.9, 609.9)    |
| Spécificité             | 1.00 (0.88, 1.00) | 1.00 (0.88, 1.00) | 0 (0, 0)            | 0 (-40.2, 67.3)     |
| <b>Cheville/Pied</b>    |                   |                   |                     |                     |
| Sensibilité             | 0.79 (0.54, 0.94) | 0.89 (0.67, 0.99) | 0.11 (-0.03, 0.24)  | 11.8 (-41.9, 115)   |
| Spécificité             | 0.98 (0.87, 1.00) | 0.93 (0.80, 0.98) | -0.05 (-0.39, 0.29) | -5 (-39.1, 28.9)    |

Pour les fractures, plusieurs localisations montrent une amélioration notable de la sensibilité (Se), comme pour le membre supérieur (0,82 à 0,95) avec des localisations comme le poignet/main (0,85 à 0,92) et surtout le coude (0,60 à 1,00), avec une différence absolue de +0,40, traduisant une réduction des faux négatifs non significative dû à l'intervalle de confiance (-0,03 ; 0, 83).

La détection des fractures du membre inférieur semble impactée de manière positive par l'IA (Se 0,85 à 0,93) comme la cheville/pied (0,79 à 0,89), sans différence pour la hanche ou le genou en terme de Se.

En revanche, les spécificités (Spe) tendent à rester stables ou à diminuer légèrement, parfois de manière marquée comme pour le membre inférieur (0,97 à 0,94) avec par exemple la hanche (0,86 à 0,71), suggérant un risque accru de faux positifs dans certaines situations.

Concernant les luxations, les résultats montrent globalement des sensibilités parfaites (1,00) dans la plupart des localisations, aussi bien sans qu'avec assistance par intelligence artificielle. Cette absence de variation explique les différences absolues nulles et l'absence d'impact relatif. Les spécificités présentent des valeurs légèrement inférieures, mais stables, avec de très faibles variations entre les deux cas (par exemple +0,01 pour le membre supérieur ou +0,06 pour le coude). Ces résultats traduisent une excellente performance diagnostique globale, indépendamment de l'utilisation de l'IA, avec un éventuel gain marginal de spécificité pour certaines articulations. Les valeurs manquantes pour certaines localisations reflètent la rareté des cas dans l'échantillon, empêchant un calcul fiable.

*Tableau 7 Performances diagnostiques des luxations selon la localisation du traumatisme*

| Luxation                | Clinicien sans IA | Clinicien avec IA | Différence absolue | Différence relative |
|-------------------------|-------------------|-------------------|--------------------|---------------------|
| <b>Membre supérieur</b> |                   |                   |                    |                     |
| <b>Sensibilité</b>      | 1.00 (0.02, 1.00) | 1.00 (0.02, 1.00) | 0 (0, 0)           | 0 (-93.7, 1498.8)   |
| <b>Spécificité</b>      | 0.99 (0.93, 1.00) | 1.00 (0.95, 1.00) | 0.01 (-0.01, 0.04) | 1.3 (-25.8, 38.2)   |
| <b>Membre inférieur</b> |                   |                   |                    |                     |
| <b>Sensibilité</b>      | 1.00 (0.02, 1.00) | 1.00 (0.02, 1.00) | 0 (0, 0)           | 0 (-93.7, 1498.8)   |
| <b>Spécificité</b>      | 1.00 (0.96, 1.00) | 1.00 (0.96, 1.00) | 0 (0, 0)           | 0 (-23.9, 31.4)     |
| <b>Epaule</b>           |                   |                   |                    |                     |
| <b>Sensibilité</b>      | 1.00 (0.02, 1.00) | 1.00 (0.02, 1.00) | 0 (0, 0)           | 0 (-93.7, 1498.8)   |
| <b>Spécificité</b>      | 1.00 (0.78, 1.00) | 1.00 (0.78, 1.00) | 0 (0, 0)           | 0 (-51.1, 104.6)    |
| <b>Coude</b>            |                   |                   |                    |                     |
| <b>Sensibilité</b>      | NaN               | NaN               | NaN                | NaN                 |
| <b>Spécificité</b>      | 0.94 (0.71, 1.00) | 1.00 (0.80, 1.00) | 0.06 (-0.05, 0.17) | 6.2 (-46.3, 110.3)  |
| <b>Poignet/Main</b>     |                   |                   |                    |                     |
| <b>Sensibilité</b>      | NaN               | NaN               | NaN                | NaN                 |
| <b>Spécificité</b>      | 1.00 (0.93, 1.00) | 1.00 (0.93, 1.00) | 0 (0, 0)           | 0 (-33, 49.2)       |
| <b>Hanche</b>           |                   |                   |                    |                     |
| <b>Sensibilité</b>      | 1.00 (0.02, 1.00) | 1.00 (0.02, 1.00) | 0 (0, 0)           | 0 (-93.7, 1498.8)   |
| <b>Spécificité</b>      | 1.00 (0.74, 1.00) | 1.00 (0.74, 1.00) | 0 (0, 0)           | 0 (-51.1, 104.6)    |
| <b>Genou</b>            |                   |                   |                    |                     |
| <b>Sensibilité</b>      | NaN               | NaN               | NaN                | NaN                 |
| <b>Spécificité</b>      | 1.00 (0.89, 1.00) | 1.00 (0.89, 1.00) | 0 (0, 0)           | 0 (-39.2, 64.5)     |
| <b>Cheville/Pied</b>    |                   |                   |                    |                     |
| <b>Sensibilité</b>      | NaN               | NaN               | NaN                | NaN                 |
| <b>Spécificité</b>      | 1.00 (0.94, 1.00) | 1.00 (0.94, 1.00) | 0 (0, 0)           | 0 (-30.1, 43)       |

L'analyse comparative des performances diagnostiques sans et avec assistance par l'IA sur toute lésions confondues selon la localisation, montre globalement une amélioration des sensibilités, particulièrement marquée pour le membre inférieur avec le genou et de la cheville/pied. Par exemple, la sensibilité passe de 0,76 à 0,90 pour le membre inférieur et de 0,64 à 0,91 pour le genou, avec des différences absolues respectives de 0,15 et 0,27, traduisant un gain substantiel de détection.

Pour le membre supérieur toute localisation confondue, la sensibilité reste similaire avec ou sans IA.

En revanche, les spécificités apparaissent globalement stables, avec des variations faibles et systématiquement non significatives (par exemple, +0,04 pour le membre supérieur, mais 0 pour l'épaule ou la hanche).

**Tableau 8 Performances diagnostiques toutes lésions confondues selon la localisation du traumatisme**

| Toutes lésions          | Clinicien sans IA | Clinicien avec IA | Différence absolue | Différence relative |
|-------------------------|-------------------|-------------------|--------------------|---------------------|
| <b>Membre supérieur</b> |                   |                   |                    |                     |
| <b>Sensibilité</b>      | 0.92 (0.73, 0.99) | 0.92 (0.73, 0.99) | 0 (-0.12, 0.12)    | 0 (-43.9, 78.2)     |
| <b>Spécificité</b>      | 0.91 (0.81, 0.97) | 0.95 (0.85, 0.99) | 0.04 (-0.03, 0.1)  | 3.7 (-28.6, 50.7)   |
| <b>Membre inférieur</b> |                   |                   |                    |                     |
| <b>Sensibilité</b>      | 0.76 (0.60, 0.88) | 0.90 (0.77, 0.97) | 0.15 (0.02, 0.27)  | 17.6 (-25.5, 85.8)  |
| <b>Spécificité</b>      | 0.92 (0.82, 0.97) | 0.94 (0.85, 0.98) | 0.02 (-0.04, 0.07) | 1.6 (-28.6, 44.7)   |
| <b>Epaule</b>           |                   |                   |                    |                     |
| <b>Sensibilité</b>      | 0.83 (0.36, 1.00) | 0.83 (0.36, 1.00) | 0 (0, 0)           | 0 (-67.7, 210.1)    |
| <b>Spécificité</b>      | 0.90 (0.55, 1.00) | 0.90 (0.55, 1.00) | 0 (0, 0)           | 0 (-58.4, 140.3)    |
| <b>Coude</b>            |                   |                   |                    |                     |
| <b>Sensibilité</b>      | 1.00 (0.48, 1.00) | 1.00 (0.48, 1.00) | 0 (0, 0)           | 0 (-71, 245.4)      |
| <b>Spécificité</b>      | 0.83 (0.52, 0.98) | 0.92 (0.62, 1.00) | 0.08 (-0.2, 0.36)  | 10 (-53.3, 159)     |
| <b>Poignet/Main</b>     |                   |                   |                    |                     |
| <b>Sensibilité</b>      | 0.92 (0.64, 1.00) | 0.92 (0.64, 1.00) | 0 (-0.21, 0.21)    | 0 (-55.1, 122.6)    |
| <b>Spécificité</b>      | 0.94 (0.81, 0.99) | 0.97 (0.85, 1.00) | 0.03 (-0.03, 0.08) | 2.9 (-35.8, 65)     |
| <b>Hanche</b>           |                   |                   |                    |                     |
| <b>Sensibilité</b>      | 1.00 (0.63, 1.00) | 1.00 (0.63, 1.00) | 0 (0, 0)           | 0 (-62.5, 166.4)    |
| <b>Spécificité</b>      | 0.60 (0.15, 0.95) | 0.60 (0.15, 0.95) | 0 (0, 0)           | 0 (-71, 245.4)      |
| <b>Genou</b>            |                   |                   |                    |                     |
| <b>Sensibilité</b>      | 0.64 (0.31, 0.89) | 0.91 (0.59, 1.00) | 0.27 (0.01, 0.54)  | 37.5 (-44.7, 241.8) |
| <b>Spécificité</b>      | 0.95 (0.75, 1.00) | 1.00 (0.83, 1.00) | 0.05 (-0.05, 0.15) | 5.3 (-43.8, 97.2)   |
| <b>Cheville/Pied</b>    |                   |                   |                    |                     |
| <b>Sensibilité</b>      | 0.73 (0.50, 0.89) | 0.86 (0.65, 0.97) | 0.14 (-0.05, 0.33) | 16.7 (-37.8, 119)   |
| <b>Spécificité</b>      | 0.95 (0.82, 0.99) | 0.95 (0.82, 0.99) | 0 (-0.07, 0.07)    | 0 (-36.6, 57.7)     |



# Discussion

## 7 Principaux résultats

Dans cette étude, on a voulu montrer l'impact d'un logiciel d'intelligence artificielle pour la détection des fractures sur la prise en charge. On constate qu'il y a une tendance à la baisse des erreurs de prise en charge grâce à l'IA, bien que cette baisse ne soit pas significative.

Un de nos objectifs était également de s'intéresser à l'impact sur le suivi des patients, malheureusement, au vu du manque de données trop important, nous n'avons pas pu répondre à cette question.

Concernant les analyses en sous-groupes en fonction de la personne prenant en charge le patient, le moment de prise en charge ou la localisation : une baisse des erreurs grâce au logiciel a été mise en évidence pour les sous-groupes senior/interne et journée/garde en proportion, bien que celle-ci ne soit pas significative. En revanche, une baisse significative des erreurs de prise en charge a été constatée selon la localisation (membre supérieur et inférieur).

L'objectif secondaire s'intéressant aux performances diagnostiques du logiciel en fonction du type de lésion suggère que l'apport de l'IA est surtout perceptible pour l'amélioration de la sensibilité dans la détection des fractures et des lésions globales, mais peut ne pas apporter de changement important comme pour les épanchements ou les luxations, ce qui suggère que les cliniciens identifiaient déjà correctement ce type de lésion sans l'aide de l'IA.

Les performances diagnostiques par lésion et localisation du logiciel d'IA nous montrent de manière non significative que l'IA contribue principalement à réduire les faux négatifs en renforçant la sensibilité, mais ce gain de sensibilité s'accompagne parfois d'une perte de spécificité. L'effet semble plus marqué pour les localisations traditionnellement plus difficiles à interpréter, comme le genou et la cheville/pied, ce qui appuie l'intérêt de l'IA dans l'aide au diagnostic dans des contextes anatomiques complexes, au prix d'un léger surdiagnostic potentiel.

## 8 Discussion des résultats

Le résultat en terme de nombre d'erreurs de prise en charge est un résultat que nous ne pouvons comparer aux données de la littérature, car aucune étude ne s'est intéressée à l'impact de ce logiciel sur la prise en charge des patients. L'augmentation du nombre d'erreurs de prise en charge à l'IA pour le membre inférieur s'explique par des radiographies de hanche avec prothèses qui mettaient en évidence une fracture,

infirmée par le scanner ou alors ne relevant finalement pas d'une prise en charge chirurgicale.

Concernant les performances diagnostiques selon les lésions, les études existantes ne se sont pas intéressées aux performances diagnostiques des radiologues aidés par l'IA pour d'autres lésions que les fractures. Nos résultats ne sont pas comparables aux données de la littérature. Les données existantes pour les autres lésions sont des données sur les performances diagnostiques des logiciels d'IA comparées aux radiologues.

Pour les épanchements, on remarque dans notre étude une très bonne sensibilité à 0,98 avec et sans IA, et une spécificité passant de 0,62 à 0,38 avec l'aide de l'IA. Un travail réalisé par Israel Cohen et al [30] s'intéresse aux performances diagnostiques de l'IA sur la détection des épanchements de genou (sur les radiographies et les scanners) et retrouve des sensibilités de détection des épanchements de 74,5 % avec une Spe de 73 %, des résultats qui sont moins bons que dans notre étude.

Une autre étude de Nor-Eddine Regnard et al.[31] a comparé les performances diagnostiques de Gleamer pour des lésions des membres et du pelvis à l'interprétation des radiologues en routine et a retrouvé une Se à 91,5 % pour les épanchements de coude, comparé à une sensibilité de 84,7 % pour le radiologue, ce qui diffère également de notre travail. Dans notre étude, la spécificité diminue entre l'interprétation sans puis avec IA, ce qui peut traduire un surdiagnostic des lésions de la part de Gleamer.

On retrouve également des données sur les luxations, où dans notre étude les performances sont très bonnes avec une sensibilité de 100 % mais pour seulement 2 luxations au total, un résultat qui peut être comparé aux études Regnard N-E et al[31] où la Se passe de 100 % pour la détection des luxations complètes pour le radiologue à 79,3 % pour Gleamer seul. On remarque dans le même article que la sensibilité de Gleamer seul pour d'autre lésions, comme les subluxations ou diastasis, est supérieur à celle du radiologue seul. Il semble donc que l'IA n'a pas d'avantage à être utilisée dans les luxations complètes étant des lésions considérées comme évidentes, mais aurait un impact positif sur la détection des anomalies plus difficiles à diagnostiquer.

Dans notre étude, nous retrouvons une amélioration de la sensibilité de la détection de fracture avec l'aide de l'IA de +11,6 %, bien que notre résultat ne soit pas significatif, il correspond aux données de la littérature. En effet, on retrouve une augmentation de la sensibilité de +12,2 % dans l'étude de Duron et al.[26] grâce à l'IA, +10,4 % dans l'étude de Guermazi et al.[27]

Notre travail diffère en terme de spécificité retrouvée pour la détection des fractures. En effet, nous retrouvons une baisse de la spécificité de -2,2 %, ce qui n'est pas le cas dans la littérature. L'étude française de Duron et al.[26] retrouve une spécificité qui augmente de 5 %, passant de 90,6 % à 95,6 %. L'étude américaine de Guermazi et al. [27] retrouve, elle, un gain de 4,6 %. Cette variation de résultat pourrait s'expliquer par un manque de puissance de l'étude avec un nombre de clichés bien moindre dans

notre cas. L'effectif faible explique ces intervalles de confiance importants, ce qui peut donner l'impression de résultats divergents. Des différences de population ont également pu avoir un impact, notre population était plus masculine et plus jeune que celle de l'étude de Duron et al.[26].

En terme de performance diagnostique de détection des fractures selon la localisation, on remarque également des résultats plutôt similaires en terme de sensibilité, excepté pour certaines localisations comme l'épaule, la hanche ou le genou avec des sensibilités dans notre étude se rapprochant de 1. Dans une revue de la littérature par Robert M. Wee et al. [32], la sensibilité de détection des fractures selon la localisation diffère des résultats de notre étude pour l'épaule, la hanche et le genou retrouvant des sensibilités moins élevées. Ces différences peuvent s'expliquer par des différences de catégorisation de ces localisations dans la littérature, avec des regroupements comme épaule/bras, genou/jambe ou ne faisant pas apparaître certaines localisations dans leurs résultats. Le taux de radiographies pour chaque région dans cette étude semble être moins important que dans la littérature, ce qui peut entraîner un manque de reproductibilité à une autre population. Cette sensibilité parfaite ou quasi parfaite dans notre travail peut aussi s'expliquer par le petit nombre de radiographies de ces localisations comparativement aux études, ou alors par la présence de lésions que l'on pourrait caractériser « d'évidentes » où l'IA n'a pas de plus-value dans l'aide au diagnostic. De plus, dans notre étude, il manquait les cases « épaules » et « hanches », localisations que nous avons rajoutées à posteriori lors de l'analyse des données, ce qui a pu avoir un impact avec des traumatismes potentiellement non inclus.

Il n'y a pas, concernant la spécificité, de différence avec les données de la littérature concernant la détection des fractures selon la localisation. Ce qui semble différer est plutôt la tendance des Spe entre l'interprétation sans puis avec IA qui ne varient pas ou diminuent dans notre étude. On remarque notamment dans l'étude de M. Wee [32] que les différents articles mettent en évidence des spécificités qui, au contraire, ont tendance à augmenter ou à rester stables avec ou sans utilisation de l'IA. Les raisons pouvant expliquer ces différences sont encore une fois le manque de puissance de l'étude, avec un nombre insuffisant de radiographies. La performance des radiologues était probablement bonne dès le départ et l'IA dans ce contexte a pu surdiagnostiquer et ainsi créer plus de faux positifs.

## **9 Forces et limites de l'étude**

Le principal point fort de cette étude est son caractère novateur, en effet, comme décrit dans l'introduction, une seule étude a été réalisée en prospectif sans sélection des radiographies, les deux autres études principales sélectionnaient les radiographies interprétées à posteriori sans s'intéresser à l'impact sur la prise en charge.

Ici, l'étude se déroule en condition de vie réelle avec des radiographies non sélectionnées et donc de qualité variable, interprétées directement par l'urgentiste. De

plus, notre étude s'intéresse à l'impact de l'IA sur la prise en charge des patients, sujet qui n'a pas encore été traité à notre connaissance.

Mais cette étude comporte certaines limites. Premièrement, nous n'avons pas réussi à atteindre le nombre de clichés attendu qui était de 300, avec 185 radiographies analysées ce qui a pu amener à une perte de puissance. Sur les passages aux urgences, 125 ont eu le motif « traumatisme » attribué par l'infirmier(e) d'accueil et d'orientation, des patients se présentant pour un autre motif en plus d'avoir un traumatisme n'ont sûrement pas été étiquetés comme traumatisés (exemple : motif d'entrée malaise, mais patient qui présente un traumatisme de genou causé par sa chute), ce qui a pu entraîner un biais de recueil des données. Il y a également pu avoir des différences de prise en charge entre les urgentistes pour une même lésion mais sans discordance entre leur interprétation et celle du radiologue, n'entraînant pas de contrôle à posteriori, il y a donc pu avoir des erreurs de prise en charge non relevées. Il y a eu un biais de sélection car plus d'internes/externes ont participé que de seniors, ce qui a pu avoir un impact sur l'interprétation des radiographies et la prise en charge des patients. L'interprétation par Gleamer n'a pu être séparée complètement des radiographies dans le PACS, l'interprétation par l'IA a donc pu être visualisée par erreur avant l'interprétation sans le logiciel, impactant le reste de la prise en charge. Il y a pu également avoir des erreurs au moment de remplir la fiche d'information, pendant l'analyse des données certains intervenants ont oubliés de cocher certaines cases ayant pour conséquence une perte de données. C'est également une étude mono centrique, les résultats ne sont donc pas extrapolables aux autres centres hospitaliers.

S'il y avait discordance entre l'interprétation du radiologue et de l'IA, le diagnostic final retenu était celui du radiologue sans relecture par un pair. Contrairement à la majorité des études dans lesquelles une relecture est effectuée dans cette situation, cette absence de contrôle a aussi pu avoir un impact sur les performances diagnostiques du logiciel de détection. La relecture des radiographies a été réalisée par plusieurs radiologues du CH de Roubaix, comprenant 2 radiologues spécialisés en musculo squelettique et 2 autres en imagerie digestive, il y a donc pu avoir un impact sur les performances diagnostiques entre les praticiens.

Nous n'avons pas pu étudier l'impact d'un logiciel d'IA sur le suivi des patients, qui est pourtant une partie importante dans leur prise en charge.

Un des aspects que nous n'avons pas étudiés est l'impact d'un logiciel d'IA sur le temps de lecture des radiographies, il aurait été intéressant de pouvoir l'étudier dans cette situation de pratique courante, malheureusement l'étude de cette variable est compliquée avec la difficulté de savoir comment mesurer un temps fiable d'interprétation. Les données de la littérature suggèrent un gain de temps comme décrit dans l'introduction avec une réduction du temps de lecture significative pour Guerhazi et al[27] avec - 6,3 secondes et l'étude du Duron et al[26] a montré une baisse non significative de 10 secondes. D'autres études comme la revue par M. Wee

et al[32] montrent un temps de lecture réduit dans l'ensemble, de façon significative ou non selon les études.

Un élément intéressant aurait été d'analyser les performances diagnostiques d'un logiciel d'IA en pratique courante par les urgentistes, chose qui n'a pas été encore réalisée. Dans notre étude, ces performances diagnostiques ont été calculées à partir de l'interprétation des radiologues. En effet, les performances diagnostiques par spécialité ont été analysées dans les études de Guermazi et al [27] et Duron et al [26]. Elles ont mis en évidence une amélioration de la sensibilité de détection des fractures pour les urgentistes grâce à l'IA, quel que soit le degré d'expérience des praticiens. Mais ces études, comme expliqué auparavant, sont des études rétrospectives avec des radiographies sélectionnées.

# Conclusion

Notre étude n'a pas permis de mettre en évidence un impact significatif de l'intelligence artificielle sur la prise en charge des patients, bien qu'elle semble faire diminuer le nombre d'erreurs.

Il convient également de remarquer que les performances diagnostiques de Gleamer semblent inférieures dans notre étude à celles retrouvées dans la littérature.

Ces résultats peuvent s'expliquer par un manque de puissance de notre étude et de nombreux biais.

Pour conclure, nos résultats ne doivent pas remettre en cause l'utilisation de l'IA dans notre pratique quotidienne. D'autres études avec des effectifs plus importants et multicentriques seront nécessaires afin de pouvoir répondre à la question de l'impact de l'IA sur la prise en charge des patients.

L'IA reste un atout dans notre exercice quotidien ayant fait ses preuves dans de nombreuses études, avec également un potentiel gain de temps dans la prise en charge. Par ailleurs, toute lésion non diagnostiquée rattrapée par un logiciel de détection de fracture est un bénéfice pour le patient et peut lui éviter des séquelles à long terme.

# Liste des tables

|           |                                                                                                 |    |
|-----------|-------------------------------------------------------------------------------------------------|----|
| Tableau 1 | Tableau descriptif des clichés analysés par patient.....                                        | 26 |
| Tableau 2 | Intervenants, période d'inclusion et avis demandé .....                                         | 26 |
| Tableau 3 | Nombre d'erreurs avec ou sans IA .....                                                          | 27 |
| Tableau 4 | Nombre d'erreur selon sous-groupes, Test Khi2 (*), Test de Fisher (**)..                        | 28 |
| Tableau 5 | Tableau des performances diagnostiques selon le type de lésion. NaN (non analysé) .....         | 29 |
| Tableau 6 | Performances diagnostiques des fractures selon la localisation du traumatisme .....             | 30 |
| Tableau 7 | Performances diagnostiques des luxations selon la localisation du traumatisme .....             | 31 |
| Tableau 8 | Performances diagnostiques toutes lésions confondues selon la localisation du traumatisme ..... | 32 |

# Liste des figures

|                                                                                                                                                                                               |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 1. Schémas des deux principaux modes d'apprentissage (source : Diabolocom ; traitement du langage naturel : guide complet).....                                                        | 8  |
| Figure 2. illustration du DL avec un réseau de neurones artificiels (source : IBM ;AI vs ML vs DL vs neuralnetworks : what's the difference ?).....                                           | 9  |
| Figure 3. Exemple d'interprétation d'une radiographie par Gleamer ( source : MED-KI Marktplatz ; BoneView Trauma).....                                                                        | 11 |
| Figure 4 Nombre de passages aux urgences par jour de 2017 à 2023, (source : DREES de 2024 ; passage aux urgences entre 2017 et 2023 : des dynamiques contrastées selon les départements)..... | 12 |
| Figure 5. Schéma de l'étude .....                                                                                                                                                             | 18 |
| Figure 6. Flow chart de l'étude .....                                                                                                                                                         | 25 |



# Références

- [1] L'Intelligence Artificielle... avec intelligence ! FUN MOOC n.d. <http://www.fun-mooc.fr/fr/cours/lintelligence-artificielle-avec-intelligence/> (accessed August 16, 2025).
- [2] Smith ME, Zalesky CC, Lee S, Gottlieb M, Adhikari S, Goebel M, et al. Artificial Intelligence in Emergency Medicine: A Primer for the Nonexpert. *J Am Coll Emerg Physicians Open* 2025;6:100051. <https://doi.org/10.1016/j.acepjo.2025.100051>.
- [3] Luccioni AS, Strubell E, Crawford K. From Efficiency Gains to Rebound Effects: The Problem of Jevons' Paradox in AI's Polarized Environmental Debate 2025. <https://doi.org/10.48550/arXiv.2501.16548>.
- [4] Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019;366:447–53. <https://doi.org/10.1126/science.aax2342>.
- [5] Gmyrek P, Berg J, Bescond D. Generative AI and Jobs: A Global Analysis of Potential Effects on Job Quantity and Quality 2023. <https://doi.org/10.2139/ssrn.4584219>.
- [6] Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol Divers* 2021;25:1315–60. <https://doi.org/10.1007/s11030-021-10217-3>.
- [7] Chassagnon G, De Margerie-Mellon C, Vakalopoulou M, Marini R, Hoang-Thi T-N, Revel M-P, et al. Artificial intelligence in lung cancer: current applications and perspectives. *Jpn J Radiol* 2023;41:235–44. <https://doi.org/10.1007/s11604-022-01359-x>.
- [8] Sánchez-Salmerón R, Gómez-Urquiza JL, Albendín-García L, Correa-Rodríguez M, Martos-Cabrera MB, Velando-Soriano A, et al. Machine learning methods applied to triage in emergency services: A systematic review. *Int Emerg Nurs* 2022;60:101109. <https://doi.org/10.1016/j.ienj.2021.101109>.
- [9] Chi EA, Chi G, Tsui CT, Jiang Y, Jarr K, Kulkarni CV, et al. Development and Validation of an Artificial Intelligence System to Optimize Clinician Review of Patient Records. *JAMA Netw Open* 2021;4:e2117391. <https://doi.org/10.1001/jamanetworkopen.2021.17391>.
- [10] Martínez-Sellés M, Marina-Breyse M. Current and Future Use of Artificial Intelligence in Electrocardiography. *J Cardiovasc Dev Dis* 2023;10:175. <https://doi.org/10.3390/jcdd10040175>.
- [11] Sommet de l'IA : Gleamer ou l'intelligence artificielle au service du diagnostic radiologique 2025. <https://bigmedia.bpi-france.fr/nos-actualites/sommet-de-lia-gleamer-ou-lintelligence-artificielle-au-service-du-diagnostic-radiologique> (accessed June 1, 2025).

- [12] Demilly PA. Au CH de Roubaix, l'intelligence artificielle au plus près des patients. Nord Éclair 2025;50ROUBAIX8.
- [13] Zech JR, Santomartino SM, Yi PH. Artificial Intelligence (AI) for Fracture Diagnosis: An Overview of Current Products and Considerations for Clinical Adoption, From the AJR Special Series on AI Applications. AJR Am J Roentgenol 2022;219:869–78. <https://doi.org/10.2214/AJR.22.27873>.
- [14] Passages aux urgences entre 2017 et 2023 : des dynamiques contrastées selon les départements | Direction de la recherche, des études, de l'évaluation et des statistiques n.d. [https://drees.solidarites-sante.gouv.fr/publications-communique-de-presse/etudes-et-resultats/241212\\_ER\\_passages-aux-urgences](https://drees.solidarites-sante.gouv.fr/publications-communique-de-presse/etudes-et-resultats/241212_ER_passages-aux-urgences) (accessed June 4, 2025).
- [15] Urgences : la moitié des patients y restent plus de 3 heures en 2023, 45 minutes de plus qu'en 2013 | Direction de la recherche, des études, de l'évaluation et des statistiques n.d. [https://drees.solidarites-sante.gouv.fr/publications-communique-de-presse/etudes-et-resultats/250319\\_ER\\_urgences-la-moitie-des-patients-y-restent-plus-de-trois-heures-en-2023](https://drees.solidarites-sante.gouv.fr/publications-communique-de-presse/etudes-et-resultats/250319_ER_urgences-la-moitie-des-patients-y-restent-plus-de-trois-heures-en-2023) (accessed June 23, 2025).
- [16] Jacques E, Jacques J-M. L'erreur de diagnostic en médecine d'urgence - son incidence et ses causes n.d.
- [17] Flash Sécurité Patient - « Erreurs diagnostiques dans les services des urgences. Diagnostic manqué, urgence aggravée ». Haute Aut Santé n.d. [https://www.has-sante.fr/jcms/p\\_3598581/fr/flash-securite-patient-erreurs-diagnostiques-dans-les-services-des-urgences-diagnostic-manque-urgence-aggravee](https://www.has-sante.fr/jcms/p_3598581/fr/flash-securite-patient-erreurs-diagnostiques-dans-les-services-des-urgences-diagnostic-manque-urgence-aggravee) (accessed June 23, 2025).
- [18] Actualités de l'Urgence - APM / Société Française de Médecine d'Urgence - SFMU n.d. [https://www.sfm.org/fr/actualites/actualites-de-l-urgences/erreurs-diagnostiques-aux-urgences-les-professionnels-repondent-a-la-has-en-appelant-a-des-decisions-politiques/-new\\_id/70406](https://www.sfm.org/fr/actualites/actualites-de-l-urgences/erreurs-diagnostiques-aux-urgences-les-professionnels-repondent-a-la-has-en-appelant-a-des-decisions-politiques/-new_id/70406) (accessed June 23, 2025).
- [19] Les erreurs diagnostiques en médecine. Haute Aut Santé n.d. [https://www.has-sante.fr/jcms/p\\_3555106/fr/les-erreurs-diagnostiques-en-medecine](https://www.has-sante.fr/jcms/p_3555106/fr/les-erreurs-diagnostiques-en-medecine) (accessed June 23, 2025).
- [20] Guly HR. Diagnostic errors in an accident and emergency department. Emerg Med J 2001;18:263–9. <https://doi.org/10.1136/emj.18.4.263>.
- [21] Newman-Toker DE, Peterson SM, Badihian S, Hassoon A, Nassery N, Parizadeh D, et al. Diagnostic Errors in the Emergency Department: A Systematic Review. Rockville (MD): Agency for Healthcare Research and Quality (US); 2022.
- [22] Pinto A, Berritto D, Russo A, Riccitiello F, Caruso M, Paola Belfiore M, et al. Traumatic fractures in adults: missed diagnosis on plain radiographs in the Emergency Department. Acta Bio Medica Atenei Parm 2018;89:111–23. <https://doi.org/10.23750/abm.v89i1-S.7015>.
- [23] Mattijssen-Horstink L, Langeraar JJ, Mauritz GJ, van der Stappen W, Baggelaar M, Tan ECTH. Radiologic discrepancies in diagnosis of fractures in a Dutch teaching emergency department: a retrospective analysis. Scand J Trauma Resusc Emerg Med 2020;28:38. <https://doi.org/10.1186/s13049-020-00727-8>.

- [24] Whang JS, Baker SR, Patel R, Luk L, Castro A. The causes of medical malpractice suits against radiologists in the United States. *Radiology* 2013;266:548–54. <https://doi.org/10.1148/radiol.12111119>.
- [25] Kuo RYL, Harrison C, Curran T-A, Jones B, Freethy A, Cussons D, et al. Artificial Intelligence in Fracture Detection: A Systematic Review and Meta-Analysis. *Radiology* 2022;304:50–62. <https://doi.org/10.1148/radiol.211785>.
- [26] Duron L, Ducarouge A, Gillibert A, Lainé J, Allouche C, Cherel N, et al. Assessment of an AI Aid in Detection of Adult Appendicular Skeletal Fractures by Emergency Physicians and Radiologists: A Multicenter Cross-sectional Diagnostic Study. *Radiology* 2021;300:120–9. <https://doi.org/10.1148/radiol.2021203886>.
- [27] Guermazi A, Tannoury C, Kompel AJ, Murakami AM, Ducarouge A, Gillibert A, et al. Improving Radiographic Fracture Recognition Performance and Efficiency Using Artificial Intelligence. *Radiology* 2022;302:627–36. <https://doi.org/10.1148/radiol.210937>.
- [28] Bousson V, Attané G, Benoist N, Perronne L, Diallo A, Hadid-Beurrier L, et al. Artificial Intelligence for Detecting Acute Fractures in Patients Admitted to an Emergency Department: Real-Life Performance of Three Commercial Algorithms. *Acad Radiol* 2023;30:2118–39. <https://doi.org/10.1016/j.acra.2023.06.016>.
- [29] Jacques T, Cardot N, Ventre J, Demondion X, Cotten A. Commercially-available AI algorithm improves radiologists' sensitivity for wrist and hand fracture detection on X-ray, compared to a CT-based ground truth. *Eur Radiol* 2024;34:2885–94. <https://doi.org/10.1007/s00330-023-10380-1>.
- [30] Cohen I, Sorin V, Lekach R, Raskin D, Segev M, Klang E, et al. Artificial intelligence for detection of effusion and lipo-hemarthrosis in X-rays and CT of the knee. *Eur J Radiol* 2024;175:111460. <https://doi.org/10.1016/j.ejrad.2024.111460>.
- [31] Regnard N-E, Lanseur B, Ventre J, Ducarouge A, Clovis L, Lassalle L, et al. Assessment of performances of a deep learning algorithm for the detection of limbs and pelvic fractures, dislocations, focal bone lesions, and elbow effusions on trauma X-rays. *Eur J Radiol* 2022;154:110447. <https://doi.org/10.1016/j.ejrad.2022.110447>.
- [32] Kwee RM, Kwee TC. Artificial intelligence-assisted detection of fractures on radiographs with BoneView: a systematic review. *Eur J Radiol* 2025;190:112230. <https://doi.org/10.1016/j.ejrad.2025.112230>.

# Annexe 1

ETIQUETTE PATIENT

## FEUILLE DE PRISE EN CHARGE A REMPLIR

### 1 FICHE PAR TRAUMATISME

« Impact de l'IA sur la prise en charge des patients admis aux urgences pour traumatismes appendiculaires »

#### PREMIÈRE PARTIE À REMPLIR : RADIOGRAPHIE SANS GLEAMER

##### Intervenant :

☐ Senior

☐ Interne (Semestre : )

☐ Journée

☐ Garde, heure de prise en charge :

##### Membre traumatisé :

☐ Bras

☐ Coude

☐ Avant-bras

☐ Poignet

☐ Main

☐ Doigts

☐ Cuisse

☐ Genou

☐ Jambe

☐ Cheville

☐ Pied

☐ Orteils

##### Clinique :

☐ Hématome

☐ Déformation

☐ Plaie

☐ Déficit sensitivo-  
moteur

☐ Œdème

☐ Douleur localisée

☐ Impotence fonctionnelle

##### Interprétation (Sans GLEAMER) :

☐ Absence de lésion

☐ Fracture

☐ Autre :

☐ Luxation

☐ Epanchement

##### Avis demandé :

☐ Aucun

☐ Senior

☐ Avis spécialisé (Chirurgien / Radiologue)

☐ Modification  
interprétation si avis

☐ Pas de modification de  
l'interprétation

Préciser si modification :

##### Prise en charge et Suivi :

☐ Traitement fonctionnel

☐ Traitement orthopédique

☐ Traitement chirurgical

☐ Suivi à distance

☐ Pas de suivi à distance

Délai souhaité si suivi :

Traitement fonctionnel : immobilisation incomplète, traitement antalgique seul / Traitement orthopédique : immobilisation complète (plâtre) / Traitement chirurgical : Prise en charge au bloc opératoire

Suivi à distance : consultation spécialisée avec chirurgien orthopédique

DEUXIÈME PARTIE À REMPLIR : RADIOGRAPHIE AVEC GLEAMER

**Nouvelle interprétation :**

☐ Accord avec Gleamer

☐ Désaccord avec Gleamer

☐ Absence de lésion

☐ Fracture

☐ Luxation

☐ Epanchement

Autre :

**Avis demandé sur nouvelle interprétation :**

☐ Oui

☐ Non

☐ Avis Senior

☐ Avis spécialisé (Chirurgien / Radiologue)  
Merci d'entourer le spécialiste

**Prise en charge finale**

☐ Traitement fonctionnel

☐ Traitement orthopédique

☐ Traitement chirurgical

☐ Suivi à distance

☐ Pas de suivi à distance

Délai souhaité si suivi :

Traitement fonctionnel : immobilisation incomplète, traitement antalgique seul / Traitement orthopédique : immobilisation complète (plâtre) / Traitement chirurgical : Prise en charge au bloc opératoire

Suivi à distance : consultation spécialisée avec chirurgien orthopédique.

# Annexe 2

## FEUILLE DE PRISE EN CHARGE A REMPLIR

### PAR RADIOLOGUE

*« Impact de l'IA sur la prise en charge des patients admis aux urgences pour traumatismes appendiculaires »*

*Patient tagué avec le tag « THESE IATU »*

Identité patient :

Date de la radiographie :

Spécialité du radiologue lecteur :

#### PREMIÈRE PARTIE À REMPLIR : RADIOGRAPHIE SANS GLEAMER

#### Interprétation (Sans GLEAMER) :

☐ Absence de lésion

☐ Fracture

☐ Luxation

☐ Epanchement

☐ Autre :

#### DEUXIÈME PARTIE À REMPLIR : RADIOGRAPHIE AVEC GLEAMER

#### Nouvelle interprétation :

☐ Accord avec Gleamer

☐ Désaccord avec Gleamer

☐ Absence de lésion

☐ Fracture

☐ Luxation

☐ Epanchement

☐ Autre :

**AUTEUR : Nom :** BRABANT    **Prénom :** Lubin

**Date de Soutenance :** 13/10/2025

**Titre de la Thèse :** Impact de l'intelligence artificielle sur la prise en charge des patients admis aux urgences pour traumatisme du squelette appendiculaire.

**Thèse - Médecine - Lille 2025**

**Cadre de classement :** Radiologie

**DES + FST ou option :** Médecine d'urgence

**Mots-clés :** intelligence artificielle, fracture, traumatisme, Gleamer, urgences.

### **Résumé :**

**Contexte :** l'augmentation des passages aux urgences entraîne le défi important d'une prise en charge sûre et efficace pour les patients, les traumatismes sont le motif le plus fréquent d'admission, les radiographies faisant partie des examens les plus réalisés. L'IA s'est montrée être une aide efficace au diagnostic des fractures, nous nous intéresserons ici à son impact sur la prise en charge.

**Matériel et Méthodes :** c'est une étude prospective, mono centrique. L'inclusion des patients se présentant pour traumatisme d'un membre appendiculaire s'est déroulée sur un mois. Les médecins et internes devaient remplir une fiche avec une première interprétation des radiographies sans l'aide de l'intelligence artificielle et une première prise en charge puis dans un second temps, interpréter avec l'aide de l'IA et proposer une seconde prise en charge.

**Résultats :** notre étude n'a pas permis de mettre en évidence un impact significatif de l'intelligence artificielle sur la prise en charge des patients bien qu'elle semble faire diminuer le nombre d'erreurs. On retrouve 13 erreurs de prise en charge sans l'aide de l'IA contre 6 erreurs avec l'aide de l'IA, différence absolue 0.04 [-0.01, 0.04]. Il convient également de remarquer que les performances diagnostiques de Gleamer semblent inférieures dans notre étude avec une spécificité qui baisse entre l'interprétation sans puis avec IA contrairement aux données de la littérature.

**Conclusion :** l'IA reste un atout dans notre exercice quotidien ayant fait ses preuves dans de nombreuses études, avec également un potentiel gain de temps dans la prise en charge. Par ailleurs, toute lésion non diagnostiquée rattrapée par un logiciel de détection de fracture est un bénéfice pour le patient et peut lui éviter des séquelles à long terme.

### **Composition du Jury :**

**Président :** Monsieur le Professeur Éric WIEL

**Asseseurs :** Monsieur le Docteur Rémy DIESMIS  
Monsieur le Docteur Marc-Aurèle RUSU

**Directeur :** Madame la Docteur Bérengère VOGEL