



HAL
open science

L'intelligence artificielle permettrait-elle de réduire la prescription de scanner pour la détection des fractures du bassin ?

Marc Antoine Durand

► To cite this version:

Marc Antoine Durand. L'intelligence artificielle permettrait-elle de réduire la prescription de scanner pour la détection des fractures du bassin ?. Médecine humaine et pathologie. 2024. <dumas-04910032>

HAL Id: dumas-04910032

<https://dumas.ccsd.cnrs.fr/dumas-04910032v1>

Submitted on 24 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

AVERTISSEMENT

Cette thèse d'exercice est le fruit d'un travail approuvé par le jury de soutenance et réalisé dans le but d'obtenir le diplôme d'Etat de docteur en médecine. Ce document est mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt toute poursuite pénale.

Code de la Propriété Intellectuelle. Articles L 122.4

Code de la Propriété Intellectuelle. Articles L 335.2 - L 335.1

UNIVERSITÉ PARIS CITÉ

Faculté de Santé

UFR de Médecine

Année 2024

THÈSE POUR LE DIPLÔME D'ÉTAT DE DOCTEUR EN MÉDECINE

Présentée et soutenue publiquement le : 14/10/2024

Par

Marc Antoine DURAND

**L'intelligence artificielle permettrait-elle de réduire la prescription
de scanner pour la détection des fractures du bassin ?**

Dirigée par M. le Docteur Valentin Hot

JURY

M. le Professeur Philippe Juvin

Président

M. le Docteur Valentin Hot

Directeur

M. le Docteur Richard Chocron

Membre du jury

M. le Docteur Anthony Mézière

Membre du jury

Remerciements

À M. le Professeur Philippe JUVIN, PU-PH, chef de service d'accueil des urgences de l'HEGP

Merci de m'avoir fait l'honneur de présider le jury de thèse. Merci de m'avoir reçu dans votre service pour mon stage de super externe et mon premier semestre. C'est à cette occasion que j'ai découvert les urgences et l'envie d'en faire mon métier.

À M. le Docteur Valentin HOT, PHC, service d'accueil des urgences de l'HEGP

Merci de m'avoir fait l'honneur de m'encadrer pendant cette thèse. Merci pour tous tes conseils et ta patience, c'est grâce à toi que j'ai autant appris pendant mon semestre à l'HEGP. C'était un véritable plaisir de travailler avec toi !

À M. le Docteur Richard CHOCRON, MCU-PH, service d'accueil des urgences de l'HEGP

Merci de m'avoir si bien accueilli dans le service pendant le semestre : tu m'as appris qu'on pouvait plaisanter et rester efficace. Merci de ton aide et de tes conseils pour ma thèse.

À M. Le Docteur Anthony MEZIERE, PH, chef de service de gériatrie de l'hôpital Corentin Celton

Merci de me faire l'honneur de participer à ce jury, merci de votre gentillesse et de votre disponibilité.

Merci également....

À toute l'équipe du SAU de l'HEGP : merci pour votre accueil et votre bonne humeur. Merci de m'avoir fait découvrir et aimer cette belle spécialité qu'est les urgences. Merci d'avoir été là pour discuter quand le moral n'était pas au rendez-vous. Merci à chacun d'entre vous pour ce que vous m'avez appris et pour l'envie d'en découvrir toujours plus.

À Mme le Docteur Céline TRUONG : Merci pour tes conseils et ta rigueur lorsque tu étais en gériatrie à Bichat, tu as été pour moi un modèle d'efficacité. C'est grâce à toi que j'ai pu découvrir la réanimation d'Antony et m'épanouir là-bas.

À toute l'équipe du SAMU de Garches : Merci de m'avoir fait découvrir le préhospitalier et ses surprises, ça reste un des meilleurs stages de ma vie ! Merci à toi Guillaume de m'avoir encadré pour mes premiers appels et pour ces fous rires en écoutant les enregistrements.

Merci aux Docteurs Valérie Soussan, Apolline Mezaize, Noémie Barthement, Melanie Laurent : Merci pour cette initiation aux joies de la pédiatrie pendant le semestre d'hiver.

Merci aux docteurs Olivier Andremont, Michel Arnaout, Julien Do Vale, Abolfazl Mohebbi-Amoli et Jean Pierre Deme ainsi qu'à toute l'équipe de la réanimation d'Antony : Merci de m'avoir fait découvrir ce monde étrange qu'est la réanimation, merci pour votre rigueur et pour tout ce que vous m'avez appris.

Merci aux Docteurs Jean-Herlé Raphalen et Damien Vimpère : Merci pour vos enseignements

A mes amis :

À Pierre Louis, Philippe, Etienne, Maxime : Merci d'avoir été à mes côtés pendant toutes ces années. Pouvoir compter sur votre présence et vos précieux conseils était et reste inestimable. C'est grâce à vous que j'ai pu traverser l'internat presque sereinement. Je ne pourrais rêver meilleurs amis et témoins.

À Astrid : Ce fut un plaisir de te découvrir, merci pour ces fous rires et à ces balles perdues que tu sèmes de façon aussi innocente. On vous souhaite à tous les deux tout le bonheur du monde.

À ma famille...

À mes parents bien sûr sans qui je ne serais pas là aujourd'hui : merci pour votre soutien sans faille et votre art de toujours réussir à me motiver. Merci à toi maman pour m'avoir fait découvrir et donné le goût de la médecine. Merci à toi papa de supporter nos discussions tellement dépourvues de radar et de tube hyperfréquence.

À mes frères François Xavier et Jean Baptiste : je serai là pour vous, même si vous avez choisi une spécialité pour le moins étrange...

À mes beaux-parents et Capucine, Martin et Jules : merci de m'avoir accepté dans votre famille, pour tous ces beaux moments partagés ensemble et à venir.

Enfin à toi Prune : merci de m'avoir soutenu depuis la D1 dans tous ces projets y compris mon choix de faire urgence. Merci pour ta patience, tes précieux conseils et ton amour. La vie est tellement plus douce dans tes bras.

Résumé et indexation en français

L'intelligence artificielle permettrait-elle de réduire la prescription de scanner pour la détection des fractures du bassin ?

Résumé :

Introduction : Le bilan de première intention d'un traumatisme du bassin à faible cinétique comprend une radiographie de bassin de face (1) qui est relue par les médecins urgentistes (2). Hélas cette interprétation n'est pas sans erreur diagnostique (3,4). Les logiciels d'aide au diagnostic fondés sur l'intelligence artificielle (IA) représentent un nouvel outil (5). Leurs performances sont soit similaires, soit dépassent celles des urgentistes (6). L'objectif de cette étude est d'évaluer ce nouveau test diagnostique qu'est la radiographie de bassin de face relue par IA et de voir si ses performances sont suffisantes pour se passer du scanner de bassin qui est prescrit en cas de doute sur la présence d'une fracture.

Méthode : Etude rétrospective et monocentrique quantitative sur la population des patients ayant réalisé un scanner du bassin à l'HEGP entre le 1^{er} janvier et le 31 décembre 2022. Nous avons utilisé le scanner comme référence pour établir les performances de la radiographie de bassin relue par l'IA et analyser des multiples facteurs de risque et signes cliniques.

Résultats : L'interprétation de la radiographie de bassin par l'IA par rapport au scanner qui est le Gold standard, avait une sensibilité (Se) à 0.69, une spécificité (Sp) à 0.93. Les performances de la radiographie étaient satisfaisantes pour l'anneau pelvien (Se 0.80, Sp 0.94), moyennes pour l'acetabulum et les ailes iliaques mais manquaient systématiquement les fractures du sacrum. L'ostéoporose et ses facteurs de risque n'étaient pas associés à un surrisque de fracture. L'arthrose n'était pas associée à surrisque d'erreur de l'IA. Les radiographies de bassin étaient dans l'ensemble de mauvaise qualité et l'asymétrie des trous obturateurs était associé de façon statistiquement significative à un surrisque de faux positif. Aucun des critères cliniques n'étaient prédictifs d'une fracture.

Conclusion : Notre étude a montré que la relecture d'une radiographie de bassin par l'IA ne possède pas les performances pour remplacer le scanner qui reste indispensable en cas de doute diagnostique en particulier pour les fractures de la partie postérieure du bassin.

Spécialité :

DESMU

Mots clés français :

Urgences, radiologie, intelligence artificielle, traumatologie, fractures osseuses, qualitative, quantitative

Forme ou Genre :

fMeSH : Dissertation universitaire

Rameau : Thèses et écrits académiques

Résumé et indexation en anglais

Can an artificial intelligence software decrease CT-scan prescription to detect pelvic fracture?

Abstract:

Introduction : After a pelvic trauma, the first imagery done to detect pelvic fracture is a pelvic X-ray (1), which is interpreted by the emergency physicians (2). However this interpretation can lead to misdiagnosis (3,4). Artificial intelligence software that detect bone fractures in other locations are a new prospect (5). Their performances are similar if not superior to those of emergency physicians (6). The purpose of this study is to measure the performances of AI interpreted pelvic X-ray and to see if they are enough to avoid prescribing a CT scan.

Method: We made a retrospective, monocentric, quantitative study, on patients who have had a pelvic CT-scan in the HEGP between January 1st and December 31st, 2022. CT-scan was used as gold standard to evaluate performances of AI interpreted pelvic X-ray and analyze risk factors and the relevance of clinical presentation.

Results: AI interpreted pelvic X-ray compared to the CT-scan, that is the gold standard, showed a sensitivity of 0.69 and a specificity of 0.93. Pelvic rings were better diagnosed (Se 0.80, Sp 0.94), acetabulum and iliac wings were average, however sacrum fractures were never detected. Osteoporosis and its risk factors were not associated with a higher chance of pelvic fractures. Arthrosis was not associated with a higher chance of misdiagnosis. Pelvic X-ray were mostly of poor quality and our study showed that the lack of symmetry of the obturator foramens was statistically associated with a higher chance of false positive. None of the clinical criteria were associated with a higher rate of fracture.

Conclusion: Our study showed that AI interpreted pelvic X-ray does not have the required performances to avoid prescribing a CT-scan in case of suspicion of pelvic fracture, especially for the posterior half of the pelvis.

English keywords:

Keywords: [Emergency](#), [radiology](#), [artificial intelligence](#), [traumatology](#), [fractures](#), [qualitative](#), [quantitative](#)

Publication type:

MeSH : Academic Dissertation

Liste des abréviations

- AI : Artificial Intelligence
- DMO : Densité Minérale Osseuse
- DREES : Direction de la Recherche, des Etudes, de l'Évaluation et des Statistiques
- ECG : Electrocardiogramme
- ERLM : Ecrans Radio Luminescents à Mémoire
- FN : Faux Négatif
- FP : Faux Positif
- HAS : Haute Autorité de Santé
- HEGP : Hôpital Européen Georges Pompidou
- IA : Intelligence Artificielle
- IRM : Imagerie par Résonance Magnétique
- NB : Nota Bene
- PACS : Pictures Archiving and Communicating System
- SAE : Statistiques Annuelles des Etablissements de santé
- SAU : Service d'Accueil des Urgences
- SD : Standard Deviation
- Se : Sensibilité
- SMUR : Service Mobile d'Urgence et de Réanimation
- Sp : Spécificité
- SSA : Service de Santé des Armées
- TDM : Tomodensitométrie
- CPP : Comité de Protection des Personnes
- ANSM : Agence Nationale de Sécurité des Médicaments et des produits de santé
- RGPD : Règlement Général sur la Protection des Données
- VN : Vrai Négatif
- VP : Vrai Positif

Table des matières

REMERCIEMENTS	2
RESUME ET INDEXATION EN FRANÇAIS	5
RESUME ET INDEXATION EN ANGLAIS	6
LISTE DES ABREVIATIONS	7
INTRODUCTION	10
1. CONTEXTE	13
1.1 IMAGERIE MEDICALE ET CARACTERISATION DES FRACTURES DE BASSIN	13
1.1.1 Traumatologie fracturaire au SAU	13
1.1.2 Imagerie médicale	16
1.2 L'INTELLIGENCE ARTIFICIELLE (IA) : UN NOUVEL OUTIL	17
1.2.1 Intelligence artificielle et interprétation d'un électrocardiogramme	18
1.2.2 Histoire et définitions des concepts	19
1.2.3 Fonctionnement d'un modèle d'apprentissage	21
1.2.4 Problèmes de traitement d'image pouvant être résolus par un modèle d'apprentissage profond	24
1.2.5 Application en imagerie médicale	25
1.2.6 Fonctionnement de l'IA pour la détection de fracture	26
1.2.7 Performance de l'IA pour la détection de fracture	26
1.2.8 Performance diagnostique sur le bassin	27
2. MATERIEL ET METHODE	29
2.1 OBJECTIF DE L'ETUDE	29
2.2 TYPE D'ETUDE	29
2.3 CARACTERISTIQUES DE LA POPULATION CIBLE	29
2.3.1 Sélection de l'échantillon	29
2.3.2 Critères d'inclusion	29
2.3.3 Critères d'exclusion	30
2.4 CRITERES DE JUGEMENT	30
2.4.1 Critère de jugement principal	30
2.4.2 Critères de jugement secondaires	30
2.5 RECUEIL DE DONNEES	30
2.6 MODELE DE L'IA UTILISEE AU SAU	31
2.7 ANALYSE STATISTIQUE	32

2.8 AVIS ETHIQUE ET PROTECTION DES PERSONNES	32
3 : RESULTATS	34
3.1 DESCRIPTION	34
3.1.1 <i>Population</i>	34
3.1.2 <i>Radiographies</i>	36
3.1.3 <i>Scanners</i>	39
3.2 INTERPRETATION : PRESENCE DE FRACTURE	41
3.2.1 <i>Concordance : performance de l'IA</i>	41
3.2.2 <i>Analyse des résultats et de leur concordance</i>	42
3.3 INTERPRETATION : LOCALISATION DE LA FRACTURE.....	45
4 : DISCUSSION.....	48
4.1 ANALYSE DES RESULTATS	48
4.1.1 <i>Critère de jugement principal</i>	48
4.1.2 <i>Critères de jugement secondaires</i>	48
4.2 LIMITES.....	49
CONCLUSION.....	53
BIBLIOGRAPHIE	54
TABLE DES FIGURES.....	60
TABLE DES TABLEAUX.....	61

Introduction

Depuis la fin des années 90, la fréquentation des Services d'Accueil des Urgences (SAU) n'a cessé d'augmenter. D'après une analyse de la DREES rapportant les Statistiques Annuelles des Etablissements de santé (SAE), on comptait 10,1 millions de passages dans les structures d'urgence en 1996 en France métropolitaine. Après une augmentation moyenne de 3,3% par an, ce chiffre est monté à 21,9 millions en 2019. (7)

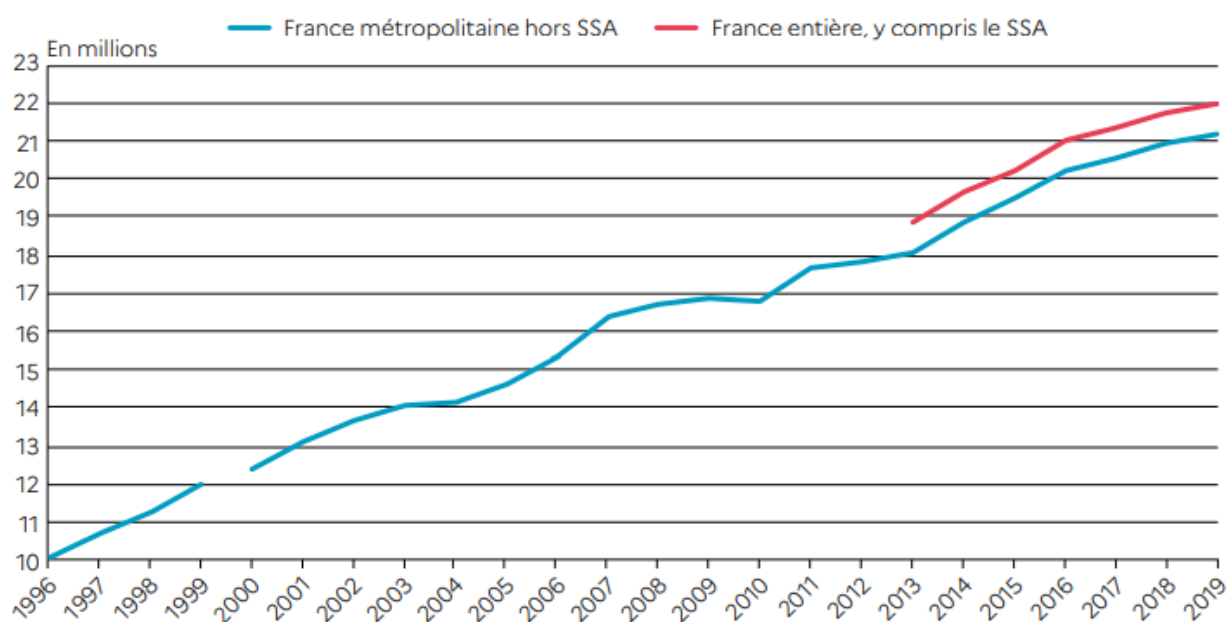


Figure 1 Évolution du nombre de passages annuels aux urgences de 1996 à 2021. SSA : Service de santé des armées (7)

Parallèlement à cette augmentation, le nombre de structures d'urgence a diminué (811 en 2010 contre 698 en 2019) (8), concentrant le flux de patients vers les établissements restants.

Parmi ce flux grandissant de patients, la traumatologie générale constitue entre 20 et 70 % des motifs de consultation aux urgences en fonction de l'âge (9,10). Elle regroupe plusieurs pathologies : fracture osseuse, entorse, luxation, plaie, morsure... Il est difficile de déterminer précisément la proportion de patients venus pour un motif traumatologique qui bénéficient d'une imagerie au SAU du fait de la diversité des pathologies incluses ; cependant une étude nationale faite sur tous les services d'urgences de France nous rapporte que 40%

des patients consultant au SAU ont une imagerie et que la radiographie standard reste l'examen de référence en traumatologie en particulier pour la recherche de fractures osseuses (2,11).

Le résultat de ces radiographies est interprété par les médecins urgentistes dans la majorité des cas (2). Malgré leur expertise, on estime que le diagnostic de fracture est manqué dans 3,7 % à 4,1 % des cas (3,4). On citera en particulier les fractures touchant certaines zones anatomiques telles que le poignet ou le bassin, qui sont connues pour être d'interprétation difficile et source d'erreur diagnostique. Cela reste la première cause d'erreur médicale aux urgences, représentant entre 70 et 80% des diagnostics manqués (12,13). Il a été montré que le diagnostic de fracture est plus souvent raté la nuit du fait de la fatigue et du manque d'avis spécialisé disponible en cas de doute (chirurgien orthopédiste, radiologue). Malheureusement dans le cas où une fracture n'est pas diagnostiquée et traitée correctement, le patient est exposé à un surrisque de pseudarthrose ou d'arthrose post traumatique, entraînant une perte de fonction du membre atteint voire une perte d'autonomie (6).

Se pose alors la question de l'utilisation d'outils d'aide au diagnostic, en théorie disponibles en permanence, pour diminuer ces erreurs. Récemment, des logiciels de relecture d'imagerie se basant sur des techniques d'intelligence artificielle (IA) ont été développés et semblent constituer une option intéressante pour diminuer ce type d'erreur.

En effet ces logiciels interprètent les radiographies et émettent un diagnostic sur la présence de fracture sur les clichés analysés. Les études publiées sont prometteuses et montrent des performances diagnostiques qui, même si elles restent généralement inférieures à celles des spécialistes (orthopédistes ou radiologues) (14) semblent comparables à celles de médecin urgentistes. De plus il a été montré que lorsque qu'un médecin urgentiste utilise ces logiciels, il améliore ses performances (6).

C'est dans ce contexte que nous nous sommes intéressés à l'évaluation de ce nouvel outil pour l'analyse des radiographies sur une zone connue pour être difficile d'interprétation : le bassin. Du fait de sa facilité de réalisation et de son faible coût, la radiographie de bassin de face est l'examen de première intention dans le bilan des traumatismes de cette région. En cas de doute sur la présence de fracture, un scanner est alors réalisé, mais c'est un examen plus irradiant et moins disponible, augmentant ainsi le temps de passage du patient au SAU.

Nous formulons l'hypothèse selon laquelle, dans le cadre du bilan d'un traumatisme du bassin, la relecture d'une radiographie du bassin par un logiciel d'intelligence artificielle aurait les performances diagnostiques nécessaires pour ne pas avoir besoin d'effectuer un scanner.

Ainsi l'objectif de notre étude est d'étudier les performances du logiciel et de les comparer à celles du scanner qui est l'examen de référence. Celle-ci permettra de savoir si ce logiciel d'aide au diagnostic permet de se passer de scanner en cas d'absence de fracture détectée à la radiographie.

1. Contexte

1.1 Imagerie médicale et caractérisation des fractures de bassin

1.1.1 Traumatologie fracturaire au SAU

Une étude faite en 2016 en croisant les bases de données de la sécurité sociale et de l'assurance maladie dénombrait 562 000 fractures sur l'année dont 81% était vues à l'hôpital et 50% nécessitaient une prise en charge chirurgicale (15).

Les fractures rapportées sont essentiellement périphériques à 68% : 36% concernent le membre supérieur avec une forte prédominance de fractures de l'extrémité distale (avant-bras, poignet, doigts), et 32% le membre inférieur avec une majorité de fractures du fémur proximal et de l'extrémité distale (pied et jambe) (15).

Plusieurs types de fractures existent. En fonction de la forme du trait de fracture et de leur localisation, celles-ci sont plus ou moins instables et donc à risque de mauvaise consolidation.

En fonction de la localisation et du type de fracture, plusieurs traitements sont possibles allant d'une immobilisation après réduction (traitement orthopédique) à une intervention chirurgicale avec pose de matériel (plaque ou vis pour assurer la stabilité de la réduction ou retrait de l'os pathologique et pose d'une prothèse).

Le bassin, quant à lui, représente seulement 2% des fractures rapportées ; néanmoins ces fractures ont un profil particulier. On en distingue principalement deux types selon le mécanisme. Chez les personnes jeunes les fractures du bassin sont dues à un traumatisme à haute cinétique : accident de la voie publique, chute d'une hauteur importante, lésions sportives ; les lésions du bassin s'intègrent alors dans le tableau d'un patient polytraumatisé. Les personnes âgées sont plutôt victimes d'accidents à basse cinétique, tels qu'une chute de leur hauteur ; leurs os étant déjà fragilisés par l'ostéoporose (16), le traumatisme du bassin y est alors isolé.

La part de fractures ostéoporotiques est en augmentation. Sur l'ensemble des fractures du bassin survenues en France, on passait de 10 806 cas en 2006 à 18 902 cas en 2016, avec plus de 65% d'entre elles survenant chez des patients de plus de 75 ans (17).

Dans le cas des traumatismes à haute cinétique, les lésions associées au traumatisme du bassin sont fréquentes et le risque de mortalité est élevé (18,19). La mortalité des patients avec une fracture traumatique du bassin est de 8 à 10%, elle atteint 25% en cas de fracture instable et va jusqu'à 40% en cas de fracture ouverte. La recherche et le traitement des traumatismes du bassin sont bien intégrés dans les algorithmes de prise en charge du patient polytraumatisé (20), le risque étant surtout lié à la survenue d'un choc hémorragique. Ce sont souvent des patients pris en charge par le SMUR et amenés directement en unité de surveillance continue dans des centres spécialisés (trauma center).

En cas de traumatisme à faible cinétique la prise en charge est moins codifiée et repose surtout sur des protocoles de service. Une fracture pelvienne doit être suspectée à l'examen clinique en cas de douleur pelvienne se majorant à la palpation et à la mobilisation des ailes iliaques, avec impossibilité de marcher (19). Il n'existe pas de recommandation à fort niveau de preuve concernant les imageries recommandées. En 2008 la HAS recommandait avec un faible niveau de preuve (grade C) de réaliser une radiographie de bassin en cas de chute suivie d'une incapacité de se relever (1). La radiographie de bassin reste, du fait de sa facilité de réalisation, l'examen fait en première intention (21) ; un scanner abdominopelvien peut être réalisé en cas de doute diagnostique ou pour une meilleure caractérisation de la fracture détectée. Ce deuxième examen est plus irradiant et, même s'il est disponible en urgence, reste moins accessible et augmente le temps de passage du patient au SAU.

Malheureusement, du fait de son anatomie, le bassin peut représenter un défi diagnostique. Le bassin est composé du sacrum en arrière et des 2 os coxaux latéralement, qui sont unis par des ligaments très puissants. Du fait de cette architecture, il possède plusieurs zones de fragilité qui sont la symphyse pubienne, qui ferme en avant les 2 os coxaux, les branches ilio- et ischio-pubiennes et l'articulation sacro-iliaque. La figure ci-dessous est une reconstitution en trois dimensions, faite à partir d'un scanner.

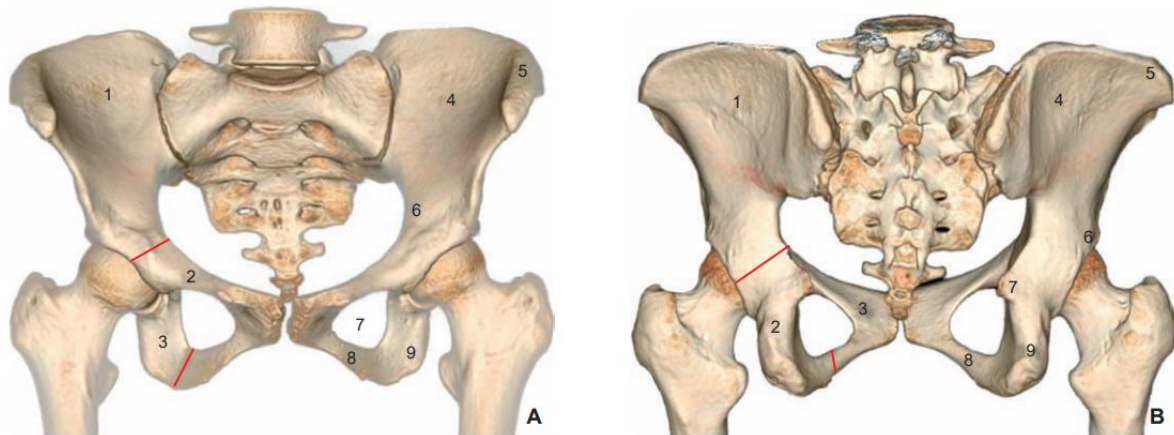


Figure 2 Anatomie du bassin(21)

On retrouve :

« A. **De face.** 1. ilion ; 2. pubis ; 3. ischion ; 4. fosse iliaque ; 5. crête iliaque ; 6. ligne arquée ; 7. foramen obturé ; 8. branche descendante du pubis ; 9. tubérosité ischiatique. B. **Vue postérieure.** 1. ilion ; 2. ischion ; 3. pubis ; 4. fosse glutéale ; 5. crête iliaque ; 6. toit du cotyle ; 7. épine ischiatique ; 8. branche ischiopubienne ; 9. tubérosité ischiatique. » (21)

Le cadre osseux est en rapport avec de nombreux vaisseaux (artères iliaques internes et externes) ainsi que des plexus nerveux, les racines sortant par les trous sacrés antérieurs. Il contient la vessie, l'urètre, le rectum, et l'utérus et le vagin chez la femme. Toutes ces structures peuvent être impactées et endommagées lors d'un traumatisme, soit directement par un fragment osseux soit par un étirement ou une compression brutale de la zone. Ces rapports anatomiques expliquent le risque important d'hémorragie en cas de traumatisme à haute cinétique.

Il existe plusieurs types de fractures du bassin, la figure ci-dessous représente l'ensemble des fractures élémentaires en reprenant la classification de Letournel et Judet. On retrouve donc : « 1. disjonction-fracture de la sacro-iliaque ; 2. fracture verticale de l'aile ; 3. fracture horizontale de l'aile ; 4. disjonction de la symphyse pubienne ; 5. fractures verticales antérieures ; 6. fracture du cotyle ; 7. disjonction de la sacro-iliaque ; 8. fractures verticales du sacrum. » (22) Par mesure de simplicité nous avons classé les fractures en sous-groupes : sacrum (regroupant les fractures 1, 7 et 8), ailes iliaque (fractures 2 et 3), anneau pelvien (fractures 4 et 5) et acetabulum (fracture 6) pour la suite de ce travail.

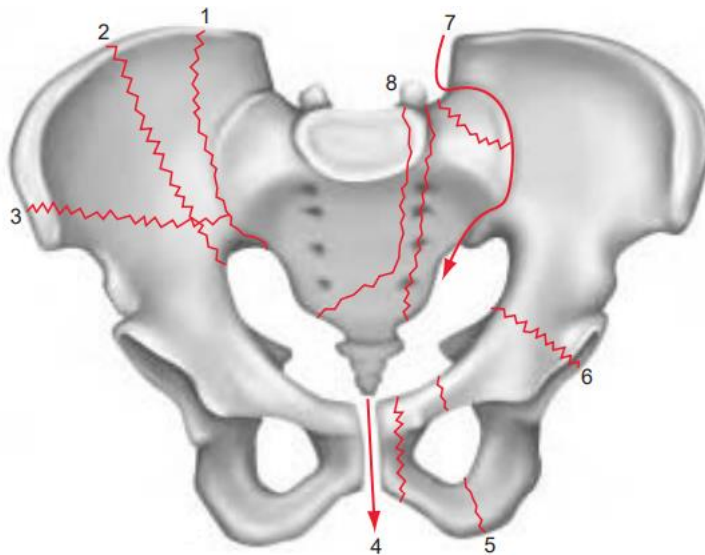


Figure 3 Fractures élémentaire du bassin (22)

L'imagerie est capitale pour la prise en charge des fractures. En plus de confirmer leur présence et de préciser leur localisation, elle permet d'avoir une idée de leur retentissement et de la nécessité d'intervention chirurgicale en fonction de leur localisation et de leur stabilité (23,24).

1.1.2 Imagerie médicale

Après avoir été utilisée pour la première fois par Wilhelm Röntgen en 1895 (25), l'imagerie par rayonnements ionisants n'a cessé de se développer pour arriver à la place incontournable qu'elle connaît aujourd'hui dans le monde médical. L'atténuation des rayons X étant différente d'un tissu à l'autre, il est alors possible de créer un contraste en fonction de l'anatomie du membre ou de la partie du patient irradiée.

On y distingue deux grands types d'imagerie : l'imagerie planaire et celle tomographique. Lorsque l'on projette l'atténuation des rayons X par le patient (objet en trois dimensions) sur une image en deux dimensions on parle d'imagerie planaire (26) : il s'agit de toutes les radiographies effectuées en traumatologie courante. L'imagerie tomographique (scanner) restitue quant à elle une représentation en trois dimensions. Nous nous intéresserons dans la suite majoritairement à l'analyse des radiographies.

Comme nous l'avons vu en introduction, cette interprétation des radiographies pour y rechercher des fractures n'est pas sans erreur. En moyenne on estime qu'il y a une mauvaise

interprétation et que le diagnostic de fracture est manqué dans environ 5% des cas (27). L'architecture du bassin rend l'interprétation des fractures de la partie postérieure de l'anneau pelvien et du sacrum difficile et est connue pour être un piège diagnostique sur les radiographies de bassin de face. Se pose alors la question d'outils d'aide au diagnostic pour augmenter les performances des cliniciens et diminuer le taux de fractures non diagnostiquées.

1.2 L'Intelligence Artificielle (IA) : un nouvel outil

En 1959, le pionnier de l'informatique Arthur Samuel donnait la définition suivante de l'intelligence artificielle : «un champ d'étude qui donne à l'ordinateur la capacité d'apprendre sans besoin d'être programmé » (28). Cette technique s'est peu à peu frayé un chemin hors du monde de la recherche pour trouver de plus en plus d'applications et devenir incontournable. En effet, elle rencontre un succès rapide depuis une dizaine d'années grâce à l'émergence du Big Data. L'accès à des données de plus en plus nombreuses et de plus en plus variées a permis le développement de l'intelligence artificielle dans de nombreux secteurs, allant de l'industrie au service du grand public.

Elle fait désormais partie de notre quotidien : elle est utilisée pour traiter d'immenses ensembles de données des utilisateurs et en dégager des profils et des tendances, afin de personnaliser les suggestions en particulier dans le développement de la publicité ciblée.

Historiquement l'intelligence artificielle a été utilisée pour deux applications en médecine : la relecture de fond d'œil à la recherche de rétinopathie (29,30), et la recherche de lésions mammaires sur les mammographies (31,32). Du fait de son adaptabilité, elle s'est peu à peu développée pour être aujourd'hui présente dans de très nombreuses spécialités. Un des exemples les plus intéressants d'application d'intelligence artificielle au traitement d'image en médecine est l'interprétation d'un électrocardiogramme. A travers celle-ci, nous allons essayer de donner au lecteur un aperçu des possibilités et de la versatilité des logiciels d'intelligence artificielle, avant de nous recentrer sur l'imagerie médicale.

1.2.1 Intelligence artificielle et interprétation d'un électrocardiogramme

L'électrocardiogramme fait partie des outils du quotidien des praticiens. Son interprétation a de nombreuses implications en termes de diagnostic et de thérapeutique.

Depuis 1911, date à laquelle furent commercialisés les premiers électrocardiographes, les machines n'ont cessé de se perfectionner. Au-delà des améliorations des composants (hardware), l'intégration de logiciels de calcul des différents paramètres électrophysiologiques de la contraction cardiaque, tels que la fréquence cardiaque, la distance PR ou même l'analyse du ST (33), fut une avancée majeure. Le diagnostic suggéré par ces algorithmes reste néanmoins souvent erroné (34) : la puissance limitée de ces logiciels, le faible nombre d'études publiées sur leurs performances et l'absence de mises à jour après la date d'achat de l'appareil en font un outil non fiable dans les situations d'urgence (35).

Avec l'apparition des logiciels d'apprentissage profond, il est possible d'avoir une interprétation pertinente de l'ECG lorsqu'on recherche des figures « classiques » (signes d'hyperkaliémie (36), signes d'ischémie myocardique (37)...), mais aussi de dépasser ce qui est habituellement visible par une interprétation humaine (38) et de faire des hypothèses sur la présence d'une dysfonction ventriculaire gauche (39,40), d'une anémie (41). On voit même se développer des applications pour extrapoler le sexe du patient et dans une moindre mesure son âge (42), pour ne citer que quelques exemples.

Parmi les autres examens désormais indispensables à la pratique, on compte l'imagerie médicale, au sein de laquelle les radiographies restent l'examen de premier recours pour de nombreuses indications. Ce domaine a lui aussi été transformé par l'arrivée de logiciels d'intelligence artificielle pour détecter et classifier la présence d'anomalies telles que des fractures osseuses.

Nous allons d'abord aborder quelques notions essentielles d'IA et définir les termes qui seront utilisés par la suite avant de détailler l'apport de l'intelligence artificielle à l'imagerie médicale.

1.2.2 Histoire et définitions des concepts

Dans cette partie les termes anglais seront systématiquement présentés étant donné qu'ils sont plus connus et utilisés, que ce soit dans les publications scientifiques ou par le grand public, mais les termes français seront utilisés de façon préférentielle.

1.2.2.1 Intelligence artificielle

La fiction a, comme dans bien des domaines, précédé la science pour le concept d'intelligence artificielle, avec l'apparition des premières machines pensantes dans les œuvres de science-fiction dès les années 1920 (43).

Il faut attendre 1950 pour que ce soit au tour des scientifiques d'aborder le sujet. C'est sous la plume du mathématicien Alan Turing qu'apparaît pour la première fois le terme d'intelligence artificielle, dans un article intitulé « l'ordinateur et l'intelligence » (44). Dans cet article il examine la question "Les machines peuvent-elles penser ?" et propose d'y répondre par un test, le " jeu de l'imitation". Il y évoque le concept d'intelligence artificielle pour définir une machine capable d'imiter un acteur humain en reproduisant les mécanismes de raisonnement humain à partir d'outils logiques.

Un autre mathématicien, Marvin Lee Minsky, affine le concept en proposant cette définition en 1956 : « la construction de programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des êtres humains, car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique » (45). Ce sera lui qui, accompagné d'une dizaine de chercheurs et ingénieurs en mathématique, fera naître l'intelligence artificielle en tant que discipline scientifique.

Comme on peut le constater, cela reste une définition très générale, probablement du fait des limites techniques de l'époque. On doit citer 2 projets qui firent avancer la mise en forme de l'intelligence artificielle : le *Logic Theorist* créé par Allen Newell, John C. Shaw et Herbert A. Simon en 1955-1956, qui avait pour objectif de démontrer des théorèmes de logique, puis le *General Problem Solver* en 1959, qui reprend et développe les ambitions du premier.

1.2.2.2 Apprentissage automatique

C'est à partir des années 1980 que naît l'apprentissage automatique (*Machine Learning* en anglais).

L'apprentissage automatique est « une manière de développer une intelligence artificielle qui repose sur l'utilisation d'un modèle statistique dont les paramètres sont définis grâce à des données d'entraînement ». (46)

Pour l'illustrer, Daphné Wallach (46) choisit l'exemple suivant : « si notre but est de reconnaître si une image représente une forêt ou la mer il est possible de développer un modèle statistique basé sur la couleur prédominante :

- Si plus de $n_1\%$ des pixels sont verts, et moins de $n_2\%$ sont bleus, alors cette image représente la forêt.
- Si plus de $n_3\%$ des pixels sont bleus, et moins de $n_4\%$ sont verts, alors cette image représente la mer.
- Dans tous les autres cas, il n'est pas possible de conclure. »

Comme on peut le voir, la partie la plus exigeante est le choix des valeurs à choisir pour n_1 , n_2 , n_3 , n_4 . Il faudrait réunir un grand nombre de photographies de mer et de forêt et en extraire les valeurs optimales. De plus, cette action nous questionne sur le choix des meilleures caractéristiques pour répondre au problème posé : une quantité relative de vert ou de bleu est-elle un critère pertinent ou faut-il en chercher un autre, tel que la détection et la reconnaissance d'animaux spécifiques d'un milieu par exemple ?

Plusieurs avancées vont faire progresser la discipline. La première est le développement d'un algorithme d'entraînement des réseaux d'intelligence artificielle appelé rétropropagation qui fut théorisé en 1970, mais qui dut attendre 1986 pour que les chercheurs David Rumelhart, Geoffrey Hinton et Ronald Williams la mettent en pratique. La deuxième est le théorème d'approximation universelle démontré par George Cybenko en 1989 puis complété en 1991 par Kurt Hornik. En pratique celui-ci garantit que sous certaines conditions des architectures de réseaux neuronaux relativement simples ont le potentiel de se rapprocher de fonctions complexes.

C'est le début d'une période d'optimisme, néanmoins la concrétisation des innovations théoriques est freinée par la faible puissance de calcul du matériel informatique de l'époque.

1.2.2.3 Apprentissage profond

Ce sont trois grandes avancées simultanées en 2011 qui vont permettre l'émergence de l'apprentissage profond (ou Deep Learning en anglais) : l'apparition des réseaux de neurones convolutifs, la démocratisation de processeurs performants et à bas coût, et enfin la mise à disposition d'immenses bases de données annotées (autrement dit chaque image possède une étiquette avec la donnée à rechercher). Grâce à ces éléments, les équipes vont avoir à disposition des algorithmes plus sophistiqués, la puissance demandée pour l'incroyable quantité de calculs nécessaires, ainsi qu'un grand nombre d'exemples pour entraîner et tester les fonctions.

La différence majeure avec ce qui s'était fait jusqu'alors est qu'il n'est plus nécessaire de choisir les caractéristiques d'un problème et leurs valeurs pour construire le modèle statistique. Pour reprendre notre exemple d'algorithme de classification mer/forêt : il n'y a alors plus besoin de déterminer les valeurs optimales de quantité seuils de pixels verts ou bleus ou même de choisir les critères de classification. Le système apprendra lui-même à partir de multiples exemples les caractéristiques utiles et pertinentes pour résoudre le problème posé.

C'est cette performance et surtout son côté généralisable qui ont permis la déclinaison de cette technique à tant de champs d'étude différents.

1.2.3 Fonctionnement d'un modèle d'apprentissage

Que ce soit en apprentissage automatique ou profond, un modèle d'apprentissage est une fonction dont le but est de résoudre un problème donné.

Entrées et sorties sont spécifiques d'un problème. Ainsi un modèle dont le but est de reconnaître si une photographie représente la mer ou la forêt aura pour entrée une image et pour sortie un nombre binaire (0 pour une forêt et 1 pour la mer). Pour prendre un autre exemple plus ancré dans notre quotidien, le modèle d'apprentissage intégré à nos messageries électroniques dont le but est de détecter les spams a pour entrée le texte du mail et pour sortie un nombre binaire (0 pour un message légitime et 1 pour les spam).

La première étape est le choix de la fonction la plus adaptée à la résolution du problème.

Commençons par aborder les réseaux de neurones profonds. Chaque neurone, appelé neurone formel, correspond à un calcul. Ils sont regroupés par couche. Au sein d'une couche, chaque neurone est connecté à tous les neurones de la couche précédente et de la couche suivante. A chaque couche le neurone va effectuer 2 opérations : une somme pondérée des variables d'entrée de sa couche puis l'application d'une fonction non linéaire appelée fonction d'activation à cette somme. Augmenter le nombre de couche permet d'augmenter la finesse du calcul et la précision de la réponse. Les couches qui ne sont pas la couche d'entrée (*Input Layer*) ou de sortie (*Output Layer*) sont appelées couches cachées (*Hidden Layer*).

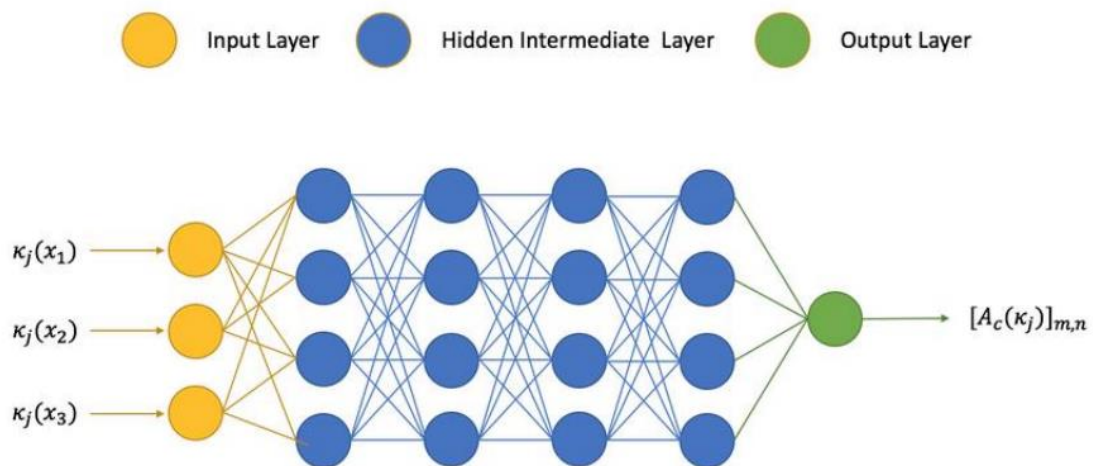


Figure 4 Modèle d'un réseau de neurones profonds(47)

Dans le cas de l'interprétation de radiographies et donc du traitement d'image, ce sont des réseaux de neurones à convolution qui sont utilisés le plus généralement.

Pour comprendre l'intérêt des réseaux de neurones à convolution, il faut prendre conscience du problème que pose l'application d'un réseau de neurones profonds pour classifier des images. Chaque neurone composant les couches cachées doit avoir un poids par pixel à optimiser. Ainsi, même pour une image de petite taille et de faible qualité, avec peu de couches intermédiaires et un petit nombre de classes possible, il faudrait calculer une somme faramineuse de poids lors de l'entraînement. A cela s'ajoute le fait que chaque neurone recherche un motif précis à un endroit précis de l'image. Rechercher ces motifs sur l'ensemble de l'image requiert donc un grand nombre de neurones. Les réseaux à convolution apportent la solution à ce problème en recherchant des motifs indépendamment de leur position dans l'image et en réduisant le nombre de paramètres d'entraînement.

Comme on peut le voir, le réseau présenté ci-dessus peut être parcouru dans deux sens, qui correspondent soit à l'entraînement soit à son utilisation pour prédire des données.

- Lorsque le modèle est utilisé de la gauche vers la droite c'est la propagation (*feed-forward propagation*) : chaque neurone applique ses opérations aux données d'entrée pour obtenir en bout de chaîne une valeur en sortie. En fonction de la nature de la fonction d'activation des neurones de la couche de sortie, la valeur de sortie du réseau pourra être une probabilité, un nombre réel, ou une valeur booléenne (vrai ou faux).

NB : Dans le cas qui nous intéresse d'un problème de classification il y aura autant de neurones dans la couche de sortie que de classes possibles et chaque neurone de couche de sortie renverra une probabilité que l'entrée fasse partie de sa classe.

- Lorsqu'on va dans le sens inverse, de la droite vers la gauche, c'est la rétropropagation (*backpropagation*) qui correspond à la phase d'entraînement : il est alors possible d'identifier les poids optimaux pour chaque couche. Les données d'entraînement sont injectées dans le modèle et la valeur de sortie est comparée à celle connue pour celle-ci. L'erreur entre la valeur prédite par le modèle et la valeur réelle connue est calculée par une fonction appelée fonction coût. L'opération est répétée jusqu'à ce que l'erreur obtenue soit suffisamment faible pour être acceptable.

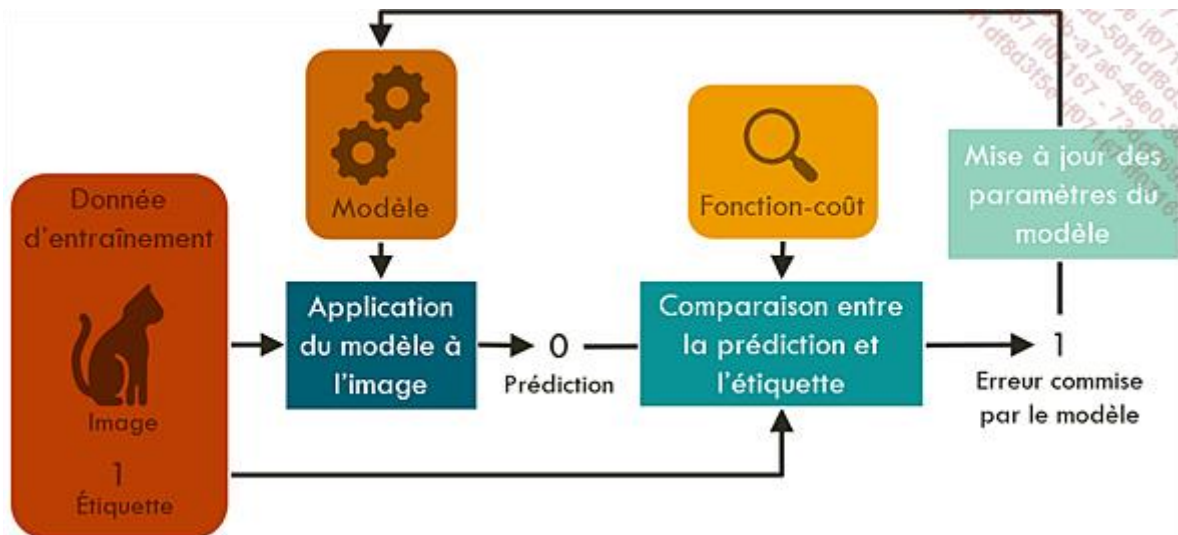


Figure 5 Entraînement d'un réseau de neurone (46)

Ce processus correspond à un entraînement supervisé car les données sont étiquetées et qu'on utilise les « vraies » valeurs pour guider l'apprentissage. Il existe d'autres catégories parmi lesquelles on peut citer l'apprentissage non supervisé, où les données ne possèdent pas d'étiquettes. Le modèle est alors moins performant et surtout nécessite une plus grande quantité d'exemples.

Les données sont séparées en trois sous-ensembles :

- Les données d'entraînement à proprement parler : utilisées pour optimiser le modèle.
- Les données de validation : elles servent à approuver les paramètres sélectionnés pendant la phase d'entraînement sur des données différentes de celles ayant permis d'optimiser le modèle. Le but étant d'éviter un surapprentissage, c'est-à-dire un modèle trop spécifique des données d'entraînement et incapable d'être généralisé à de nouvelles données.
- Les données de test : elles ont pour but d'évaluer les performances du modèle.

1.2.4 Problèmes de traitement d'image pouvant être résolus par un modèle d'apprentissage profond

Nous allons ici principalement citer trois problèmes qui nous intéresseront dans l'interprétation d'une imagerie médicale par l'intelligence artificielle en allant du plus simple au plus complexe. En effet, interpréter une imagerie médicale est avant tout du traitement d'image.

- En premier, la classification : sa résolution permet de détecter la figure présente dans une image, c'est la tâche historique de classement des images de chien et de chat.
- Ensuite vient la détection : cette tâche est plus complexe car il faut déterminer s'ils sont présents sur l'image et où ils se trouvent.
- Enfin on parle de segmentation : il s'agit de délimiter précisément l'objet recherché dans une image.

Bien d'autres fonctions répondant à d'autre problème de traitement d'image existent. Cela va de la détermination de la position de l'animal sur la photo (assis, couché...) à l'amélioration de la netteté de l'image ou la transformation d'une photo prise de nuit en photo prise de jour.

1.2.5 Application en imagerie médicale

Le film photographique à l'argent utilisé à l'origine pour les radiographies a maintenant été remplacé par les écrans radio luminescents à mémoire (ERLM) qui permettent une réutilisation de l'écran et une numérisation des données.

Les radiographies obtenues à partir de ces écrans sont des images de 2000 * 2500 pixels en moyenne dont le contraste est représenté en niveau de gris, le plus fréquemment en 12 bits (48). Il s'agit donc d'analyser une image de 5 000 000 pixels pour en dégager des motifs d'intérêt : fracture, luxation, épanchement...

L'intelligence artificielle s'est également développée dans le domaine de l'imagerie médicale, en particulier dans l'imagerie ostéoarticulaire, sous l'effet des mêmes facteurs que dans l'industrie : apparition d'ordinateurs à bas coût pourvus d'une grande puissance de calcul et de stockage, existence de bases de données conséquentes et d'algorithmes de deep learning haute performance (49).

Ses applications pour le diagnostic sont déjà nombreuses : détections d'anomalies de toutes sortes (fracture, anomalie ligamentaire ou méniscale, lésions dégénératives du rachis ou métastases osseuses), classification de ces lésions (évaluation de l'âge osseux (50) ou de la dégénérescence d'un disque intervertébral (51) par exemple) et parfois segmentation (identifier les pixels ou voxels composant une structure d'intérêt permettant par exemple de mesurer l'épaisseur du ménisque sur les IRM de genou (52,53)). Ces actions ne sont pas exclusives et peuvent être combinées.

Au-delà de l'analyse d'image à but diagnostique, on doit aussi citer les logiciels d'amélioration de la qualité d'image des scanner (54) et ceux permettant d'améliorer la résolution spatiale et même le temps d'acquisition en IRM (55).

Comme pour l'application à l'électrocardiogramme l'intelligence artificielle permet d'aller au-delà des limites de l'interprétation humaine avec la création de jumeau numérique (modèle virtuel reprenant les données des différents examens du patient) permettant de prédire dans une certaine mesure l'efficacité d'un traitement ou le risque de récurrence tumorale (56).

Ce sont les fonctions de détection et plus précisément celle des fractures qui nous intéressent pour la suite de ce travail.

1.2.6 Fonctionnement de l'IA pour la détection de fracture.

Les radiographies acquises sont enregistrées sur le PACS (système d'archivage et de transmission des images médicales) elles sont envoyées automatiquement vers le logiciel qui les analyse et transmet un rapport qui sera intégré sur le PACS dans le fichier du patient. Les fractures authentifiées sont encadrées par un cadre blanc, en cas de doute, le cadre est pointillé. L'ensemble de ce processus prend généralement une minute (49).

1.2.7 Performance de l'IA pour la détection de fracture

Dans le cadre de la détection de fractures, plusieurs logiciels d'aide au diagnostic ont vu le jour. Initialement développés pour détecter les fractures dans des zones anatomiques précises, telles que l'extrémité supérieure du fémur (14) et l'extrémité distale du radius (6), ils sont maintenant applicables sur de multiples sites (57,58) voire à l'ensemble des fractures périphériques et pour certains au rachis, au thorax et au bassin (6).

Ils utilisent plusieurs techniques allant de détection de perte de continuité osseuse (59) à l'intégration de signes plus subtils tels que la présence d'un épanchement articulaire pour les fractures de la tête radiale (49).

Leur précision dépasse celle des médecins urgentistes (6) et même parfois des chirurgiens orthopédistes expérimentés (14) mais reste tout de même en deçà des performances diagnostiques des radiologues experts.

D'autres études montrent que l'usage de logiciels d'apprentissage profond pour la détection de fracture permet une augmentation significative de la sensibilité sans diminution de la spécificité statistiquement significative (60).

Sans dépasser les performances des experts, ces logiciels ont prouvé qu'ils étaient une aide fiable pour les médecins non spécialistes, augmentant leur performance diagnostique (5) et réduisant les erreurs de diagnostic (6). C'est également un outil performant pour les spécialistes car il leur permet d'optimiser leur travail en réalisant un tri en amont de leur analyse.

Une méta analyse récente (5) analysant 42 études publiées entre 2018 et 2021, dont 37 portaient sur l'analyse de radiographies, nous rapportait une absence de différence significative de sensibilité ou de spécificité entre les logiciels et les praticiens. Dans les études

étudiant ces paramètres pour des tests de validation externes les performances poolées étaient les suivantes. Une sensibilité de 91% (95% CI [84, 95]) pour l'IA et de 94% (95% CI [90, 96]) pour les médecins et une spécificité de 91% (95% CI [81, 95]) pour l'IA contre 94% (95% CI [91, 95]) pour les médecins (cf. figure ci-dessous). De plus, quand les médecins avaient accès aux logiciels d'aide au diagnostic, la sensibilité et la spécificité étaient respectivement de 97% (95% CI [83, 99]) et 92% (95% CI [88, 95]).

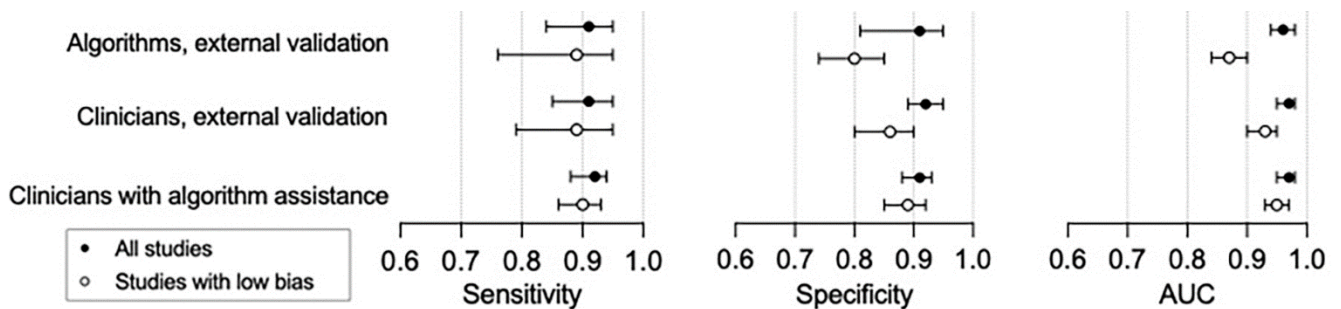


Figure 6 Représentation de la sensibilité, de la spécificité et de l'aire sous la courbe poolées pour les études de validité externes et celles évaluant les performances des médecins aidés des logiciels d'IA (5)

Restent néanmoins des zones anatomiques connues pour être des pièges diagnostiques, telles que le coude et le poignet chez l'enfant, le thorax et le pied chez l'adulte ainsi que le pelvis chez les patients de plus de 65 ans.

C'est justement le bassin, une de ces zones pour lesquelles il reste beaucoup à accomplir, qui nous intéressera pour la suite de ce travail.

1.2.8 Performance diagnostique sur le bassin

Comme nous l'avons vu plus haut, l'une des zones anatomiques étudiée en premier par les logiciels d'aide au diagnostic fut l'extrémité proximale du fémur. Les performances y sont actuellement impressionnantes avec une sensibilité de 96.2% et une spécificité de 94.6% (61) approchant les performances des spécialistes.

Cependant, dans le cadre d'un traumatisme pelvien, si les fractures du fémur sont bien dépistées, les fractures du bassin (acetabulum, anneau pelvien, ailes iliaques et sacrum) sont quant à elles plus difficiles à détecter par radiographie simple. Peu d'outils sont disponibles

pour l'analyse de cette zone et la plupart sont des logiciels de traumatologie périphérique qui ont été progressivement optimisés pour détecter tous types de fractures.

Nous avons répertorié 2 études présentant les performances de logiciels d'intelligence artificielle sur cette zone. La première (62) retrouve une sensibilité de 92% et une spécificité de 97% sans différence statistiquement significative en comparaison des performances des radiologues et des chirurgiens. La deuxième (63) évaluait l'aide apporté au diagnostic par un autre logiciel d'intelligence artificielle : on passait de 64,8% de sensibilité pour les médecins à 69,8% avec l'aide de l'IA et de 86,8% à 91,8% de spécificité sans que la différence soit significative.

Les performances varient donc d'un logiciel à l'autre. De plus les études publiées ont toutes comme référentiel de vérité un consensus d'analyse de radiologues, aucune n'utilise le scanner qui est pourtant l'examen de référence.

Ainsi, même si les performances des logiciels d'intelligence artificielle dans cette zone sont prometteuses, elles possèdent encore une marge de progression. Notre travail aura donc pour but d'évaluer les performances de l'intelligence artificielle sur la zone du bassin, en excluant les fractures de l'extrémité supérieure du fémur, et en se servant du scanner comme référence.

2. Matériel et Méthode

2.1 Objectif de l'étude

Dans le cadre du bilan de chute d'un patient âgé et de la recherche de fracture du bassin, nous formulons l'hypothèse qu'un logiciel d'intelligence artificielle proposant une interprétation des radiographies de bassin permettrait d'éviter la prescription du TDM.

L'objectif de notre étude est d'analyser l'interprétation de l'intelligence artificielle des radiographies de bassin chez les patients ayant bénéficié d'un TDM du bassin.

De plus cette étude permettra d'étudier les performances de l'intelligence artificielle sur un site connu pour être à fort risque d'erreur diagnostique en se servant du TDM comme référence là où les études test se basent sur un consensus d'expert relisant des radiographies.

2.2 Type d'étude

Notre étude a été réalisée dans un SAU parisien : l'Hôpital Européen Georges Pompidou (HEGP). Il s'agit d'une étude rétrospective et monocentrique quantitative concernant les patients ayant réalisé un TDM à l'HEGP.

2.3 Caractéristiques de la population cible

2.3.1 Sélection de l'échantillon

Le recueil est réalisé sur la base de données de l'Hôpital Européen Georges Pompidou (HEGP). Le logiciel d'imagerie est Carestream®, et les données cliniques sont cherchées dans DxCare®. Carestream® comprend toutes les imageries réalisées dans l'hôpital, qu'elles soient faites au SAU ou dans les étages. De plus certains hôpitaux limitrophes comme l'hôpital Corentin Celton réalisent leurs imageries à l'HEGP, dans ce cas les images sont disponibles sur Carestream® mais il n'y a aucun dossier patient accessible dans DxCare®.

2.3.2 Critères d'inclusion

Ont été inclus tous les patients ayant réalisé un scanner du bassin entre le premier janvier 2022 et 31 décembre 2022 à l'HEGP.

2.3.3 Critères d'exclusion

Les critères d'exclusion étaient les suivants : absence de radiographie de bassin de face réalisée dans un délai de 7 jours avant ou après le scanner du bassin, absence d'interprétation de la radiographie par l'AI ou absence d'interprétation du scanner par un radiologue. Les imageries de hanche n'étaient pas incluses dans le recueil.

Les indications étaient recueillies directement dans le logiciel, telles que remplies sur la demande d'imagerie.

2.4 Critères de jugement

2.4.1 Critère de jugement principal

Le critère de jugement principal est la concordance entre le résultat du TDM et l'interprétation de l'intelligence artificielle de radiographie de bassin.

2.4.2 Critères de jugement secondaires

Les critères de jugement secondaires sont :

- Localisation de la fracture au TDM et performances du logiciel pour chaque région (répartition en anneau pelvien, acetabulum, aile iliaque, sacrum).
- Etude du motif de réalisation du scanner.
- Etude des critères de qualité des radios selon 2 types de données : la présence d'arthrose et le respect des critères de qualité de réalisation des radiographies.
- Etude de la pertinence de critères cliniques tels que la possibilité de marcher en cas de fracture ou le recours à un antalgique de palier trois.
- Etude des facteurs de risque de fracture et de leur importance dans la démarche de prescription d'examen complémentaire.

2.5 Recueil de données

Nous avons collecté les paramètres suivants :

- Des paramètres démographiques : âge, sexe.

- Les dates de réalisation du scanner et de la radiographie
- Le motif de réalisation du TDM et identité du prescripteur (SAU ou autre).
- L'interprétation du TDM et la relecture des radiographies par l'intelligence artificielle.
- La localisation des fractures sur le TDM et les radiographies de bassin.
- Les antécédents de fractures du bassin et d'arthrose.
- Les antécédents d'ostéoporose ou la présence de facteurs de risque justifiant la réalisation de DMO.
- La capacité du patient à marcher définie par plusieurs stades :
 - Marche normale
 - Marche possible mais rupture avec l'état antérieur
 - Marche impossible
- La présence de critères cliniques tels que :
 - La présence d'une autre articulation atteinte nécessitant une imagerie
 - La notion de chute à répétition
 - La présence d'une douleur de bassin post traumatique nécessitant le recours à un antalgique de pallier 3.

2.6 Modèle de l'IA utilisée au SAU

Le logiciel d'intelligence artificielle d'aide au diagnostic utilisé dans l'Hôpital Européen Georges Pompidou est BoneView Trauma®. Il s'agit d'un dispositif médical avec un certificat européen de classe IIA développé par Gleamer®. Il utilise un algorithme d'apprentissage supervisé qui a été construit sur la base de Detectron2, qui est le modèle de reconnaissance d'images développé par Facebook AI Research (11). Il a été entraîné sur une base de données de 42 119 radiographies issues de 22 établissements. Ses performances ont été publiées en 2021 (63). De plus le logiciel est mis à jour tous les 6 mois en moyenne avec de nouveaux cas. En janvier 2024 la base de données de développement de l'algorithme comprenait un training de 500 000 radiographies, une validation sur 20 000 images et un test sur 60 000 radiographies.

2.7 Analyse statistique

L'analyse statistique a été réalisée avec le logiciel Jasp® dans sa version 0.19. Une première partie de l'analyse est descriptive.

Les caractéristiques des patients ont été décrites par la moyenne et l'écart type pour les variables quantitatives, et les effectifs par leur nombre et leur pourcentage pour les variables qualitatives. Le dénominateur des effectifs est le nombre de données disponibles pour la variable décrite.

Les variables qualitatives ont été exprimées selon leur nombre et leur fréquence et les variables quantitatives selon leur moyenne ou selon leur médiane et leur écart type. Les moyennes ont été comparées à l'aide d'un test de Student et les médianes à l'aide d'un Mann et Whitney et les variables qualitatives ont été comparées à l'aide d'un test exact de Fisher. L'ensemble des analyses statistiques ont été effectuées à l'aide du logiciel R studio (version R 3.2.2).

Les comparaisons ont été faites selon les variables choisies ci-dessous :

- présence d'une fracture à la radio,
- présence d'une fracture au scanner,
- présence des critères de qualité des radiographies de bassin,
- présence d'antécédents ou facteurs de risque pertinents,
- présence de critères clinique pertinents.

Nous avons choisi un seuil de significativité $\alpha = 0.05$ et une puissance $\beta = 0.80$.

2.8 Avis éthique et protection des personnes

Notre étude est une étude rétrospective. Elle utilise des données récupérées en pratique courante et qui sont déjà disponibles dans les dossiers médicaux. Nous sommes donc en dehors du cadre de la Loi Jardé car elle porte sur données récupérées dans le cadre du soin et qui n'impliquent donc pas directement la personne humaine (64).

Nous n'avons donc pas dû demander un avis de l'ANSM ou d'un Comité de Protection des personnes (CPP). Nous avons cependant respecté le Règlement Général sur la Protection des

Données (RGPD). Lors du recueil, les données de patients ont été anonymisées pour garantir la confidentialité et la protection des données des patients.

3 : Résultats

3.1 Description

3.1.1 Population

Quatre cent cinquante-trois scanners du bassin ont été réalisés entre le 01/01/2022 et le 31/12/2022. Parmi les patients concernés, 205 remplissaient tous les critères d'inclusion, et 248 ont été éliminés : 146 en raison de l'absence de radio de bassin réalisée, 28 car la radio de bassin a été réalisée dans un délai supérieur à 7 jours par rapport au scanner du bassin étudié, 71 en raison de l'absence d'interprétation de la radio de bassin par l'intelligence artificielle, et 3 en raison de l'absence d'interprétation du scanner du bassin par un radiologue.

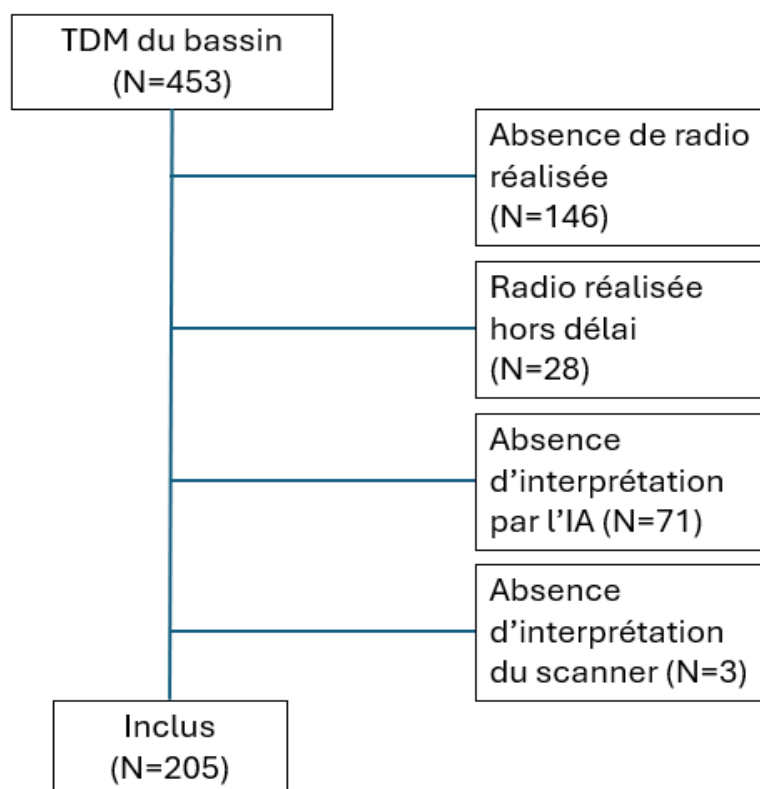


Figure 7 Flow chart

Parmi les 205 patients inclus, 148 (72,2%) étaient des femmes. L'âge moyen (SD) était de 79,22 ans (17,43).

Nous avons effectué une recherche des antécédents et des facteurs de risque de fracture tracés dans les dossiers DxCare, ainsi que de l'état clinique des patients à leur consultation au SAU. 7 patients inclus n'avaient pas de dossier disponible dans DxCare, l'analyse des dossiers se fait donc sur les 198 dossiers restants.

Antécédents (N=198)

	Oui (%)	Non (%)
Antécédent de fracture du bassin	14 (7,1)	184 (92,9)
Antécédent d'arthrose	23 (11,6)	175 (88,4)
Antécédent d'ostéoporose	23 (11,6)	175 (88,4)

Tableau 1 : Antécédents des patients inclus, nombre (%)

Facteurs de risque d'ostéoporose (N=198)

	Oui (%)	Non (%)	Hommes (%)
Fracture vertébrale	16 (8,1)	182 (91,9)	
Fracture périphérique	44 (22,2)	154 (77,8)	
Corticothérapie prolongée	13 (6,6)	185 (93,4)	
Hyperthyroïdie non traitée	0 (0)	198 (100)	
Hypercorticisme	0 (0)	198 (100)	
Hyperparathyroïdie	1 (0,5)	197 (99,5)	
Hypogonadisme prolongé	0 (0)	198 (100)	
Ménopause et antécédent familial de fracture du col	1 (0,5)	144 (72,4)	54 (27,1)
Ménopause et IMC < 19 kg/m ²	0 (0)	145 (72,9)	54 (27,1)
Ménopause avant 40 ans	0 (0)	145 (72,9)	54 (27,1)

Tableau 2 : Facteurs de risque d'ostéoporose des patients inclus, nombre (%)

Nous avons également recherché la présentation clinique des patients analysés, telle que décrite dans l'observation du SAU :

Présentation clinique (N=198)

Notion de chute à répétition	25 (12,5%)
Nécessité d'antalgiques de palier 3	22 (11%)
Marche : impossible	109 (55%)
Marche : rupture avec l'état antérieur	33 (16,5%)
Marche : normale	7 (3,5%)
Ne marche pas habituellement	9 (4,5%)

Tableau 3 : Présentation clinique des patients inclus, nombre (%)

3.1.2 Radiographies

Les 205 radiographies de bassin réalisées ont été analysées pour chercher les critères de qualité requis : vue d'ensemble du bassin, coccyx dégagé, fémurs symétriques, trous obturateurs symétriques.

Critères qualité (N=205)

	oui (%)	non (%)
Vue d'ensemble	153 (74,6)	52 (25,4)
Fémurs symétriques	108 (52,7)	97 (47,3)
Coccyx dégagé	110 (53,7)	95 (46,3)
Trous obturateurs symétriques	56 (27,3)	149 (72,7)

Tableau 4 : Critères qualité des radiographies incluses, nombre (%)

Au total, 33 radios (16,1%) présentaient les 4 critères de qualité, 19 (9,3%) n'en présentaient aucun. Par ailleurs, 164 radios (80% des radiographies) présentaient des signes d'arthrose.

Nombre de critères de qualité remplis (N=205)

0	19 (9,3 %)
1	61 (29,8 %)
2	42 (20,5 %)
3	50 (24,4 %)
4	33 (16,1 %)

Tableau 5 : Nombre de critères qualité des radiographies incluses, nombre (%)

Il est intéressant de noter que peu de radios du bassin remplissaient les 4 critères de qualité. On peut comprendre la difficulté d'avoir une exposition parfaite en situation d'urgence chez un patient douloureux, mais il faut quand même noter le fait que plus de 25% des radios ne montraient pas une vue d'ensemble du bassin.

L'analyse des radiographies par l'intelligence artificielle montrait pour 89 patients (43,4%) l'atteinte d'une autre zone que le bassin, notamment la hanche ; pour 58 patients (27,8%) l'analyse par l'IA montrait une absence de fracture, pour 20 patients (9,8%) un doute sur une fracture, et pour 37 patients (18,5%) la présence d'une fracture du bassin. 1 patient avait une luxation de hanche.

Parmi les fractures du bassin détectées par l'intelligence artificielle, 45 (77%) concernaient l'anneau pelvien, 2 concernaient l'acetabulum, 3 concernaient l'aile iliaque, et 1 concernait le sacrum. 5 fractures associaient des lésions de l'anneau pelvien et de l'acetabulum. 1 concernait l'anneau pelvien et l'aile iliaque.

La population présentant une fracture du bassin était comparable à la population n'ayant pas de fracture du bassin détectée par l'IA, avec un âge moyen à 80 ans, et une répartition des sexes équivalente avec 72% de femmes. Le délai moyen entre la réalisation du scanner et de la radio était légèrement plus long pour les patients ayant une fracture décelée à la radio, sans différence statistiquement significative.

		Fracture vue à la radio (N=57)		Pas de fracture vue à la radio (N=148)	
Âge moyen (SD)		80,11	(19,77)	78,88	(16,51)
Sexe	F	41	72%	107	72%
	M	16	28%	41	28%
Délai moyen (SD)		1,26	(1,92)	0,71	(1,40)

Tableau 6 : Analyse de la population âge (année), sexe et délais de réalisation du scanner (jour) en fonction de la présence de fracture détectée à la radiographie, nombre (%)

		Fracture vue à la radio (N=57)		Pas de fracture vue à la radio (N=127)		p-value
Arthrose	Oui	37	79%	100	79%	0,535
	Non	10	21%	27	21%	
Vue d'ensemble	Oui	36	77%	92	72%	0,646
	Non	11	23%	35	28%	
Fémurs symétriques	Oui	22	47%	70	55%	0,269
	Non	25	53%	57	45%	
Coccyx dégagé	Oui	27	57%	67	53%	0,785
	Non	20	43%	60	47%	
Trous obturateurs symétriques	Oui	8	17%	39	31%	0,006
	Non	39	83%	88	69%	
Nombre de critères qualité remplis	0	4	9%	11	9%	0,638
	1	15	32%	35	28%	
	2	10	21%	30	24%	
	3	12	26%	32	25%	
	4	6	13%	19	15%	

Tableau 7 : Analyse des critères de qualité de réalisation de la radiographie et de la présence d'arthrose en fonction de la présence de fracture détectée à la radiographie, nombre (%)

L'analyse des antécédents et des facteurs de risque de fracture montrait deux populations comparables sur tous les points. L'analyse comparative des critères de qualité de la radio montre une différence statistiquement significative sur la symétrie des trous obturateurs. Cependant, il n'y a pas de différence statistiquement significative sur le nombre total de critère de qualité remplis.

3.1.3 Scanners

Les scanners du bassin étaient réalisés dans un délai moyen (SD) de 0,86 jours (1,58 jours). La demande était faite par un médecin des urgences pour 150 patients (74,3%), par un médecin d'un autre service pour 52 patients, et n'était pas retrouvé pour 3 patients.

Les motifs de demande de scanner du bassin étaient répartis comme indiqué dans le tableau suivant :

Motif du scanner du bassin (N=205)		
Avis spécialisé		
Demande de l'orthopédiste	3	1,46%
Bilan pré-opératoire	18	8,78%
Diagnostic incertain		
Impotence fonctionnelle	65	31,70%
Radio ininterprétable	1	0,48%
Radio douteuse	38	18,53%
Bilan autre localisation		
Présence de matériel	11	5,36%
Fracture de hanche	14	6,82%
Recherche d'autre fracture	29	14,14%
Polytraumatisme	3	1,46%
Autre motif	23	11,21%

Tableau 8 : Motif de réalisation des scanner du bassin, nombre (%)

Parmi les scanners du bassin réalisés, 138 (67,3%) ne montraient pas de fracture du bassin et 67 scanners (32,7%) montraient une fracture. Les zones de fractures sont réparties comme suit :

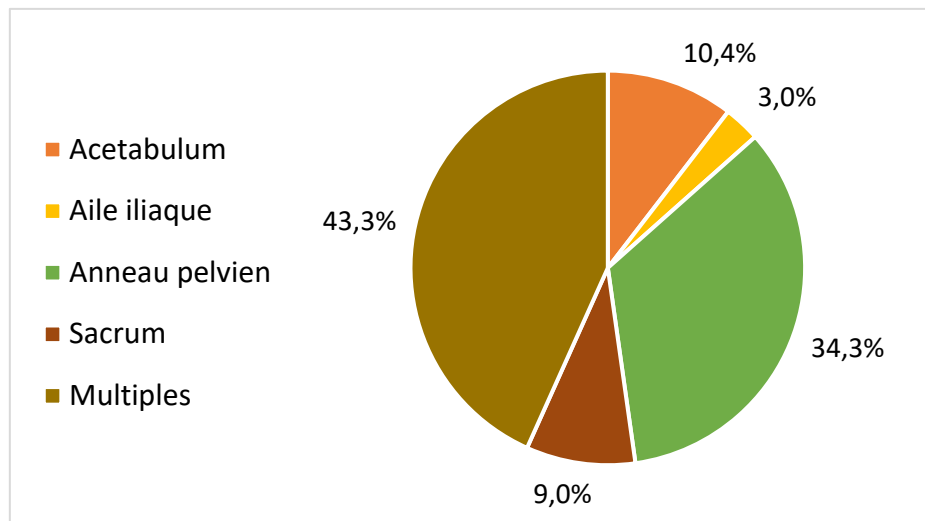


Figure 8 répartition de la localisation des fractures détectées au scanner (%)

Parmi les scanners qui trouvaient de multiples fractures du bassin, la répartition des fractures était la suivante :

Localisation des traits de fracture en cas de fracture multiple (N=29)

Acetabulum - sacrum	1 (3,4%)
Anneau pelvien - acetabulum	4 (13,8%)
Anneau pelvien - acetabulum - aile iliaque	2 (6,9%)
Anneau pelvien - acetabulum - aile iliaque - sacrum	1 (3,4%)
Anneau pelvien - acetabulum - sacrum	5 (17,2%)
Anneau pelvien - sacrum	16 (55,2%)

Tableau 9 : Localisation du trait de fracture en cas de fracture multiple, nombre (%)

La population chez qui une fracture était retrouvée était comparable à la population chez qui il n’y avait pas de fracture vue au scanner : l’âge moyen était de 78,1 ans (VS 79,7 ans), avec 75% de femmes (VS 71%). De même que pour les patients qui présentaient une fracture à la radio, le délai entre la réalisation de la radio et du scanner était légèrement plus long pour ceux présentant une fracture.

		Fracture vue au TDM (N=67)		Pas de fracture vue au TDM (N=138)	
Âge moyen (SD)		78,11	(19,17)	79,75	(16,57)
Sexe	F	50	75%	98	71%
	M	17	25%	40	29%
Délai moyen (SD)		1,15	(1,82)	0,72	(1,42)

Tableau 10 : Analyse de la population âge (année), sexe et délais de réalisation du scanner (jour) en fonction de la présence de fracture détectée au scanner, nombre (%)

L’analyse des antécédents et des facteurs de risque de fracture ne montrait pas de différence statistiquement significative entre les deux populations.

3.2 Interprétation : présence de fracture

3.2.1 Concordance : performance de l’IA

Les performances de l’intelligence artificielle ont été étudiées par rapport au scanner dans la détection des fractures du bassin. Les résultats ont été classés en :

- Vrai positif (VP) : l’IA et le scanner détectent une fracture
- Vrai négatif (VN) : l’IA et le scanner ne détectent pas de fracture
- Faux positif (FP) : l’IA détecte une fracture non vue au scanner
- Faux négatif (FN) : l’IA ne détecte pas une fracture vue au scanner.

		Fracture au scanner	
		OUI	NON
Fracture à la radio	OUI	47	10
	NON	21	127

Tableau 11 : Représentation des fractures en fonction de leur détection par le scanner et la radiographie relue par l'IA, nombre

Ces données nous permettent de calculer une Sensibilité (Se) et une Spécificité (Sp) pour l'interprétation de la radio par l'IA.

On obtient ainsi :

$$Se = \frac{VP}{VP+FN} = 0,69$$

Ce qui signifie que pour 100 fractures, l'IA en détecte 69.

De même :

$$Sp = \frac{VN}{VN+FP} = 0,93$$

Ce qui signifie que pour 100 radios sans fractures, l'IA en interprète 93 correctement.

On peut également calculer la Valeur Prédictive Positive (VPP) et la Valeur Prédictive Négative (VPN) :

$$VPP = \frac{VP}{VP+FP} = 0,82$$

$$VPN = \frac{VN}{VN+FN} = 0,86$$

Autrement dit pour la VPP, pour 100 fractures détectées par l'IA, 82 seront retrouvées au scanner. De même pour la VPN, pour 100 radios ne détectant pas de fracture, 86 verront leur diagnostic confirmé par le scanner.

3.2.2. Analyse des résultats et de leur concordance

Nous avons analysé les critères pouvant influencer la qualité de la radiographie et donc son interprétation, en fonction de leur répartition vrai positif, vrai négatif, faux positif, faux négatif.

		Vrai positif (N=47)		Vrai négatif (N=127)		Faux positif (N=10)		Faux négatif (N=21)		p-value
Arthrose	Oui	37	79%	100	79%	10	100%	17	81%	0,443
	Non	10	21%	27	21%	0	0%	4	19%	
Vue d'ensemble	Oui	36	77%	92	72%	6	60%	19	90%	0,228
	Non	11	23%	35	28%	4	40%	2	10%	
Fémurs symétriques	Oui	22	47%	70	55%	4	40%	12	57%	0,622
	Non	25	53%	57	45%	6	60%	9	43%	
Coccyx dégage	Oui	27	57%	67	53%	4	40%	12	57%	0,761
	Non	20	43%	60	47%	6	60%	9	43%	
Trous obturateurs symétriques	Oui	8	17%	39	31%	0	0%	9	43%	0,023
	Non	39	83%	88	69%	10	100%	12	57%	
Nombre de critères qualité remplis	0	4	9%	11	9%	2	20%	2	10%	0,177
	1	15	32%	35	28%	5	50%	6	29%	
	2	10	21%	30	24%	0	0%	2	10%	
	3	12	26%	32	25%	3	30%	3	14%	
	4	6	13%	19	15%	0	0%	8	38%	

Tableau 12 : Analyse des critères qualité de la radiographie en fonction de leur répartition chez les vrais positifs, vrais négatifs, faux positifs, faux négatifs. Nombre, %

L'absence du critère « trous obturateurs symétriques » à la radiographie est associé de façon statistiquement significative à un plus haut niveau d'erreur de lecture qui se traduit principalement par une proportion plus importante de faux positifs. L'ensemble des autres critères qualité, y compris la présence d'arthrose qui est connue pour être un facteur confondant classique, n'est pas associé à une différence statistiquement significative.

Par ailleurs, l'analyse des antécédents montre qu'un antécédent de fracture du bassin est un facteur associé avec un risque augmenté d'erreur à l'analyse de l'IA, largement associé avec un surrisque de faux positif. Les autres antécédents trouvés dans les dossiers DxCare ne sont pas associés à un risque d'erreur augmenté de façon statistiquement significative, en dehors d'un antécédent familial de fracture du col du fémur chez les femmes ménopausées.

Antécédents :		Vrai positif (N=45)		Vrai négatif (N=122)		Faux positif (N=10)		Faux négatif (N=21)		p-value
Fracture du bassin	Oui	5	11%	5	4%	3	30%	1	5%	0,012
	Non	40	85%	117	92%	7	70%	20	95%	
Arthrose	Oui	6	13%	13	10%	1	10%	3	14%	0,938
	Non	39	83%	109	86%	9	90%	18	86%	
Ostéoporose	Oui	8	17%	12	9%	2	20%	1	5%	0,297
	Non	37	79%	110	87%	8	80%	20	95%	
Fracture vertébrale	Oui	4	9%	8	6%	1	10%	3	14%	0,669
	Non	41	87%	114	90%	9	90%	18	86%	
Fracture périphérique	Oui	9	19%	29	23%	4	40%	2	10%	0,252
	Non	36	77%	93	73%	6	60%	19	90%	
Corticothérapie	Oui	5	11%	6	5%	1	10%	1	5%	0,501
	Non	40	85%	116	91%	9	90%	20	95%	
Hyperthyroïdie non traitée	Oui	0	0%	0	0%	0	0%	0	0%	NC
	Non	45	96%	122	96%	10	100%	21	100%	
Hypercorticisme	Oui	0	0%	0	0%	0	0%	0	0%	NC
	Non	45	96%	122	96%	10	100%	21	100%	
Hyperparathyroïdie	Oui	1	2%	0	0%	0	0%	0	0%	0,332
	Non	44	94%	122	96%	10	100%	21	100%	
Hypogonadisme prolongé	Oui	0	0%	0	0%	0	0%	0	0%	NC
	Non	45	96%	122	96%	10	100%	21	100%	
Ménopause et ATDC familial de fracture du col	Oui	0	0%	0	0%	1	10%	0	0%	0,002
	Non	34	72%	87	69%	6	60%	17	81%	
	Homme	11	23%	36	28%	3	30%	4	19%	
Ménopause et IMC < 19	Oui	0	0%	0	0%	0	0%	0	0%	0,757
	Non	34	72%	87	69%	7	70%	17	81%	
	Homme	11	23%	36	28%	3	30%	4	19%	
Ménopause avant 40 ans	Oui	0	0%	0	0%	0	0%	0	0%	0,757
	Non	34	72%	87	69%	7	70%	17	81%	
	Homme	11	23%	36	28%	3	30%	4	19%	

Tableau 13 : Analyse des antécédents et facteurs de risque d'ostéoporose trouvés dans les dossiers en fonction de leur répartition en vrai positif, vrai négatif, faux positif, faux négatif. Nombre, %. 7 patients n'avaient pas de dossiers (2 VP et 5 FN)

Enfin, nous avons analysé la présentation clinique des patients à leur arrivé pour voir si l'une des présentations cliniques était associée avec un surrisque d'erreur d'interprétation de l'IA. Aucun des critères cliniques recherchés ne présentait d'association statistiquement significative.

		Vrai positif (N=47)		Vrai négatif (N=127)		Faux positif (N=10)		Faux négatif (N=21)		p-value
Clinique	Nécessité d'antalgiques de palier 3	6	13%	12	9%	0	0%	4	19%	0,619
	Notion de chute à répétition	3	6%	18	14%	1	10%	3	14%	
	Impossible	24	51%	65	51%	6	60%	14	67%	
Marche	Rupture avec l'état antérieur	10	21%	18	14%	1	10%	4	19%	0,672
	Normale	1	2%	5	4%	1	10%	0	0%	
	Ne marche pas habituellement	0	0%	7	6%	1	10%	1	5%	

Tableau 14 : Analyse de la présentation clinique des patients en fonction de leur répartition en vrai positif, vrai négatif, faux positif, faux négatif. Nombre, %.

3.3 Interprétation : localisation de la fracture

Une fois que la radio détecte une fracture, la localisation de cette fracture a été analysée. Ainsi pour chaque localisation possible de fracture détectée à la radio (Acétabulum, Aile iliaque, Anneau pelvien et Sacrum), nous avons analysé si le scanner trouvait un trait de fracture sur la même zone anatomique. Pour chaque localisation, nous avons calculé une Sensibilité, une Spécificité, une Valeur Prédictive Positive et une Valeur Prédictive Négative pour évaluer les performances de l'IA.

		Acétabulum			
		Fracture au scanner			
		OUI	NON		
Fracture à la radio	OUI	6	1	Sensibilité :	0,30
	NON	14	184	Spécificité :	0,99
				VPP :	0,86
				VPN :	0,93

Tableau 15 : Performance de L'IA pour la relecture de la radiographie de bassin en fonction de la localisation : acetabulum

Ainsi pour l'acetabulum, on trouve une faible sensibilité, qui reflète que l'IA n'arrive pas à détecter une part non négligeable des fractures de cette zone. Cependant, la spécificité, VPP et VPN sont très satisfaisantes, ce qui montre l'absence de surdiagnostic de fracture par l'IA.

		Aile iliaque			
		Fracture au scanner		Sensibilité :	0,40
		OUI	NON	Spécificité :	0,99
Fracture à la radio	OUI	2	2	VPP :	0,50
	NON	3	198	VPN :	0,99

Tableau 16 : Performance de L'IA pour la relecture de la radiographie de bassin en fonction de la localisation : aile iliaque

De même pour la détection de fracture de l'aile iliaque, l'IA a une mauvaise sensibilité du fait de nombreux faux négatifs, et une mauvaise VPP du fait d'une part importante de faux positifs. Elle a cependant une bonne spécificité et une bonne VPN du fait d'un grand nombre de vrais négatifs.

		Anneau pelvien			
		Fracture au scanner		Sensibilité :	0,80
		OUI	NON	Spécificité :	0,94
Fracture à la radio	OUI	41	10	VPP :	0,80
	NON	10	144	VPN :	0,94

Tableau 17 : Performance de L'IA pour la relecture de la radiographie de bassin en fonction de la localisation : anneau pelvien

L'analyse des fractures de l'anneau pelvien est intéressante car elle fait partie des fractures les plus représentées dans notre étude. Les performances de l'intelligence artificielle en font un bon test pour la recherche de fracture de l'anneau pelvien.

		Sacrum			
		Fracture au scanner		Sensibilité :	0,00
		OUI	NON	Spécificité :	0,99
Fracture à la radio	OUI	0	1	VPP :	0,00
	NON	29	175	VPN :	0,86

Tableau 18 : Performance de L'IA pour la relecture de la radiographie de bassin en fonction de la localisation : sacrum

Enfin l'analyse des performances de l'intelligence artificielle pour la recherche de fracture du sacrum nous montre que la radiographie est un très mauvais test diagnostique pour cette localisation. Aucune des 29 fractures du sacrum ne sont vues par l'IA, la seule fracture détectée est un faux positif. La spécificité reste élevée en raison de la rareté de cette fracture.

4 : Discussion

4.1 Analyse des résultats

Dans cette étude nous sommes partis du Gold Standard qui est le scanner pour rechercher les performances d'une radiographie de bassin relue par un logiciel d'aide au diagnostic.

Les patients que nous avons inclus étaient majoritairement des femmes (70%) âgées (79 ans en moyenne) ce qui correspond à la population classique des traumatismes du bassin à faible cinétique ayant une fracture de l'anneau pelvien (17).

4.1.1 Critère de jugement principal

Nous avons pu calculer les performances de ce nouveau test diagnostic et nous retrouvons des statistiques globalement semblables aux performances déjà publiées pour ce logiciel sur le bassin (63). Notre conclusion, au vu des résultats, est qu'une relecture par l'intelligence artificielle d'une radiographie de bassin n'est pas suffisamment sensible pour se passer d'un scanner du bassin en cas de suspicion de fracture. Elle présente cependant une spécificité élevée et ne crée que peu de surdiagnostic.

4.1.2 Critères de jugement secondaires

Nous avons mis en évidence que l'intelligence artificielle est performante pour la détection des fractures de l'anneau pelvien, dans une moindre mesure de l'acetabulum et de l'aile iliaque, et n'a détecté aucune des fractures du sacrum vues au scanner. Nous n'avons pas trouvé de référence publiée faisant part de ce résultat ; les résultats étant poolés sur l'ensemble du bassin. Si une fracture de la partie postérieure du bassin, en particulier du sacrum, est suspectée, alors la radiographie de bassin n'est pas le test adapté pour la rechercher.

L'analyse du motif de réalisation du scanner montre que celui-ci est demandé dans 50% des cas pour confirmer une suspicion diagnostique, dans 10% des cas à la demande du chirurgien, et dans les 40% restants pour rechercher des fractures autres que le bassin en particulier en cas de présence de prothèse de hanche.

L'étude des critères qualité de réalisation de la radiographie de bassin de face met en évidence que la majorité de ces clichés sont de qualité très médiocre avec 60% ayant entre 0 et 2 critères sur 4. De plus, 25% n'ont pas un bassin vu dans son ensemble ! A notre connaissance aucune étude de qualité de réalisation des radiographies de bassin faites en pratique courante n'existe. Nous ne pouvons savoir s'il s'agit d'un effet dépendant de notre centre ou si les radiographies de bassin sont de mauvaise qualité ailleurs.

Sur plus de 70% des radiographies, les trous obturateurs n'étaient pas symétriques, ce qui est un signe de mauvais positionnement du patient lors de la prise des radiographies. Ce facteur est d'ailleurs associé avec un fort surrisque de fracture détectée par l'intelligence artificielle. Ce qui laisse supposer deux causes potentielles : soit un problème d'analyse de l'intelligence artificielle lorsque le bassin n'est pas bien positionné, auquel cas ce serait un faux positif ; soit que la mobilisation nécessaire au bon positionnement pour la radiographie soit trop douloureuse du fait de la présence de fracture, ce serait alors un vrai positif.

L'analyse de ces critères en fonction de la performance de l'intelligence artificielle montre que 100% des faux positifs ont des trous obturateurs qui ne sont pas symétriques et confirme donc que le mauvais positionnement gêne l'interprétation des clichés par l'intelligence artificielle.

Parmi les faux positifs, on retrouve également une surreprésentation statistiquement significative des patients ayant un antécédent de fracture du bassin. A l'inverse, l'arthrose n'est pas un facteur confondant statistiquement significatif.

Aucun des critères cliniques (intensité de la douleur, impotence fonctionnelle) étudiés n'a montré de corrélation significative avec la présence d'une fracture. Il ne semble donc pas pertinent de se baser seulement sur ceux-ci pour prescrire un scanner ; or il s'agit du seul critère reconnu justifiant une imagerie après un traumatisme à basse cinétique(1).

De même la présence d'ostéoporose ou de ses facteurs de risque n'est pas associée de façon statistiquement significative avec la détection de fracture.

4.2 Limites

Nous nous sommes basés sur la grille STARD version 2015 pour évaluer la validité interne de notre étude. On constate pour la partie méthode que les critères concernant les participants et le test diagnostique sont remplis, mais nous n'avons cependant pas fait

d'analyse de reproductibilité. Quant aux résultats, la part de données manquantes a été évoquée mais nous n'avons pas pu la prendre en compte pour l'analyse des facteurs de risque.

La question que nous nous sommes posée lors de cette étude est pertinente. Les radiographies de bassin sont faites presque systématiquement en pratique courante devant un traumatisme à basse cinétique avec une impotence totale ou partielle. Savoir si un logiciel de relecture des radiographies peut augmenter les performances diagnostiques de celles-ci au point de se dispenser de réaliser un scanner, pourrait changer notre pratique et la démarche diagnostique.

Une objection que l'on pourrait formuler est que même si le diagnostic est posé à la radiographie, un scanner serait nécessairement demandé par le spécialiste pour mieux caractériser la fracture et adapter sa prise en charge. La classification la plus adaptée dans notre cas est celle de Rommens (65). Elle propose une classification des fractures du pelvis sur un os pathologique et une prise en charge tenant compte de spécificités évolutives du sujet âgé. Son utilisation ne nécessite pas de scanner !

Une autre critique pourrait être que les chirurgiens orthopédistes demandent systématiquement un scanner en préopératoire en repérage pour comprendre les lésions. Ainsi il est possible de mesurer le trajet des futures broches qui seront posées. Mais une part non négligeable de ces fractures sont stables et ne nécessitent pas d'être opérées : une étude de registre en Allemagne montrait que ces fractures stables (type 1 de la classification de Tile), même si elles semblent en diminution, représentent 40% des fractures du bassin et sont la forme majoritaire chez les femmes qui composaient alors 75% de l'effectif de l'étude (66).

Notre étude pose cependant plusieurs limites. L'effectif est restreint : sur les 453 scanner trouvés nous n'avons pu en inclure que 205, et seulement 67 présentaient une fracture du bassin. Nous avons été limités dans notre recueil de données par la période de validité du logiciel. En effet avant 2022 il ne prenait pas en compte de nombreux sites et ne rendait pas d'analyse si une radiographie d'une zone non prise en compte (thorax par exemple) était demandée en même temps qu'un bassin ou qu'une radiographie de membre. En 2023 il y a eu un changement de logiciel d'aide au diagnostic, remplaçant BoneView Trauma de Gleamer par le logiciel d'AZmed, raison pour laquelle notre recueil s'arrête dans un souci de validité interne.

Cet effectif réduit se fait sentir dans l'analyse des données sur les fractures peu fréquentes, notamment les fractures de l'aile iliaque (5 au total), qui souffrent de ce manque de puissance. On doit signaler également un biais d'information car les dossiers étaient parfois incomplets et il est probable que nous ayons sous-estimé la prévalence de certains antécédents non demandés en pratique courante tels que les antécédents familiaux, une ménopause précoce, un IMC inférieur à 19 kg/m², etc.

Comme nous pouvons le voir, les patients inclus sont en majorité des femmes âgées avec de l'ostéoporose dans 20% des cas. Cela est probablement révélateur d'un biais de recrutement car notre étude est monocentrique et que l'Hôpital Européen George Pompidou reçoit des patients habitant dans le Sud-Ouest de Paris (14^e, 15^e et une partie du 16^e arrondissement principalement). Il serait intéressant de faire la même analyse avec les données de plusieurs CHU et dans d'autres villes pour se défaire de ce biais, analyser des populations plus jeunes et augmenter la puissance de notre étude.

De plus comme nous l'avons signalé dans la partie méthodologie le logiciel est mis à jour en moyenne tous les six mois. Notre outil d'analyse n'est donc pas le même au début et à la fin de l'inclusion. Nous avons posé la question à Gleamer par l'intermédiaire de Mme Jeanne Ventre qui nous a répondu avec gentillesse et nous a communiqué les dernières données de développement de l'algorithme datant de janvier 2024. Il n'a cependant pas été possible d'obtenir les performances par région anatomique, ni l'évolution de celles-ci au fil des mises à jour.

Nous avons considéré pour nos analyses des facteurs de risque que la présence d'un doute sur la présence de fracture proposée par l'Intelligence artificielle était équivalente au diagnostic de fracture. Or le logiciel ne propose le diagnostic de fracture que s'il y a une concordance à plus de 90% entre la radiographie analysée et le motif « fracture » recherché, et « doute fracture » en cas de concordance comprise entre 50 et 90%. Cela représente un biais de classement même si c'est plus représentatif de la réalité du terrain où le doute doit toujours bénéficier au patient.

Nous devons également signaler que notre étude est rétrospective et part des patients ayant un scanner pour comparer le diagnostic de certitude posé avec l'examen de référence à un nouveau test diagnostique. Nous n'avons donc que les cas où, après une radiographie de bassin, le diagnostic de fracture était suffisamment probable pour qu'un examen supplémentaire, le scanner, soit prescrit pour trancher. En utilisant notre recueil nous n'avons

aucun moyen de savoir le nombre et les caractéristiques des patients qui n'ont eu qu'une radiographie de bassin qui a été jugée suffisante pour poser ou écarter le diagnostic de fracture.

Ce biais n'existerait pas si le scanner était demandé de façon systématique devant toute fracture de bassin. L'analyse des motifs de demande de scanner montre que le motif majoritaire est la confirmation d'un diagnostic incertain (50%) et non une demande du spécialiste (10%). Cela vient donc confirmer l'existence de ce biais de sélection.

Conclusion

Notre étude a montré que la relecture d'une radiographie de bassin par l'intelligence artificielle ne pouvait pas remplacer le scanner comme outil de diagnostic de certitude. Cependant, elle présente une très bonne spécificité (0,93), même si sa sensibilité reste moyenne (0,69) et sa valeur prédictive négative n'est qu'à 0,86. Avec ces résultats il paraît difficile de conclure que la radiographie est un bon test pour dépister les fractures de bassin. En comparaison avec les résultats des études antérieures il apparaît que les performances de l'intelligence artificielle sont meilleures que celles des médecins non radiologues sur cette zone, et qu'elles semblent s'améliorer par rapport à celles publiées en 2021 pour le même logiciel (63).

Les logiciels d'intelligence artificielle sont en développement permanent. Certains arrivent dès aujourd'hui à atteindre les performances de relecture des médecins radiologues (14). Il paraît donc nécessaire de réévaluer sans cesse ce nouvel outil en fonction des évolutions technologiques.

La question de l'intégrer dans les stratégies diagnostiques au SAU reste ouverte. Si ses résultats semblent contrastés dans cette étude, elle reste une piste prometteuse. Parmi les questions qui se poseront par la suite, la valeur légale de cette interprétation sera un des sujets centraux.

Bibliographie

1. Cédric Carbonneil. INDICATIONS DE LA RADIOGRAPHIE DU BASSIN, HAS [Internet]. 2008 [cité 24 août 2024]. Disponible sur: https://www.sfmu.org/upload/consensus/synthese_radiographie_bassin2008.pdf
2. Dominique Pateron JYG. Caducee.net. 2014 [cité 17 août 2024]. L'accès à l'imagerie au cours des urgences : analyse de l'enquête nationale. Disponible sur: <https://www.caducee.net/actualite-medicale/12669/l-acces-a-l-imagerie-au-cours-des-urgences-analyse-de-l-enquete-nationale.html>
3. Hallas P, Ellingsen T. Errors in fracture diagnoses in the emergency department - Characteristics of patients and diurnal variation. *BMC Emerg Med.* 2006;6.
4. Wei CJ, Tsai WC, Tiu CM, Wu HT, Chiou HJ, Chang CY. Systematic Analysis of Missed Extremity Fractures in Emergency Radiology. *Acta Radiol* [Internet]. 1 janv 2006 [cité 17 août 2024];47(7):710-7. Disponible sur: <https://www.tandfonline.com/doi/abs/10.1080/02841850600806340>
5. Kuo RYL, Harrison C, Curran TA, Jones B, Freethy A, Cussons D, et al. Artificial Intelligence in Fracture Detection: A Systematic Review and Meta-Analysis. *Radiology.* juill 2022;304(1):50-62.
6. Lindsey R, Daluiski A, Chopra S, Lachapelle A, Mozer M, Sicular S, et al. Deep neural network improves fracture detection by clinicians. *Proc Natl Acad Sci U S A.* 6 nov 2018;115(45):11591-6.
7. <https://drees.solidarites-sante.gouv.fr/> [Internet]. 2021 [cité 15 août 2024]. Les établissements de santé en 2021; Fiche 25 - La médecine d'urgence. Disponible sur: <https://drees.solidarites-sante.gouv.fr/sites/default/files/2021-07/Fiche%2025%20-%20La%20m%C3%A9decine%20d%E2%80%99urgence.pdf>
8. Actualités de l'Urgence - APM / Société Française de Médecine d'Urgence - SFMU [Internet]. 2020 [cité 15 août 2024]. Disponible sur: https://www.sfmu.org/fr/actualites/actualites-de-l-urgences/la-hausse-des-passages-aux-urgences-a-ralenti-en-2019-analyse-apmnews/-new_id/66734
9. Carrasco B. Les usagers des urgences : premiers résultats d'une enquête nationale. 2003; Disponible sur: https://sante.gouv.fr/IMG/pdf/usagers_urgences.pdf
10. Boisguérin B, Toutlemonde F, Vuagnat A. Caractéristiques de la prise en charge des patients aux urgences. *Soins* [Internet]. mai 2018 [cité 15 août 2024];63(825):16-9. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/S0038081418300811>
11. Louis BENISRI. ARFIAU : Évaluation de l'Acceptabilité d'un logiciel de Reconnaissance de Fracture par Intelligence Artificielle auprès des Urgentistes-thèse d'exercice. 2023.

12. Guly HR. Diagnostic errors in an accident and emergency department. *Emerg Med J*. 2001;18(4):263-9.
13. Leeper WR, Leeper TJ, Vogt KN, Charyk-Stewart T, Gray DK, Parry NG. The role of trauma team leaders in missed injuries: does specialty matter? *J Trauma Acute Care Surg*. 2013;75(3):387-90.
14. Urakawa T, Tanaka Y, Goto S, Matsuzawa H, Watanabe K, Endo N. Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network. *Skeletal Radiol*. févr 2019;48(2):239-44.
15. Bouyer B, Leroy F, Rudant J, Weill A, Coste J. Burden of fractures in France: incidence and severity by age, gender, and site in 2016. *Int Orthop* [Internet]. mai 2020 [cité 19 août 2024];44(5):947-55. Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7190681/>
16. Cortet B, Chauvin P, Feron JM, Grange L, Coulomb A, Launois R, et al. Fragility fractures in France: epidemiology, characteristics and quality of life (the EPIFRACT study). *Arch Osteoporos*. 13 mars 2020;15(1):46.
17. Melhem E, Riouallon G, Habboubi K, Gabbas M, Jouffroy P. Épidémiologie nationale des fractures du bassin et de l'acetabulum en France. *Rev Chir Orthopédique Traumatol* [Internet]. sept 2020 [cité 20 juill 2023];106(5):488-96. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/S1877051720300046>
18. Pohlemann T, Bosch U, Gänsslen A, Tscherné H. The Hannover experience in management of pelvic fractures. *Clin Orthop*. août 1994;(305):69-80.
19. Geeraerts T, Court C, Ozanne A, Cheisson G, Duranteau J. Traumatismes du bassin. *EMC - Médecine Urgence* [Internet]. janv 2010 [cité 13 juin 2024];5(4):1-11. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/S1959518210514879>
20. Incagnoli P, Puidupin A, Ausset S, Beregi J, Bessereau J, Bobbia X, et al. MANAGEMENT OF UNSTABLE PATIENT WITH PELVIC FRACTURE (FIRST 24 HOURS).
21. Amoretti N. Imagerie des fractures du bassin.
22. Jouffroy P, Raoult A. Techniques chirurgicales dans les fractures du bassin. *EMC - Tech Chir - Orthopédie - Traumatol* [Internet]. janv 2006 [cité 13 juin 2024];1(1):1-10. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/S0246046702000491>
23. Mouhsine DE, Garofalo R, Theumann N, Borens O. E. Mouhsine R. Garofalo N. Theumann O. Borens F. Chevalley M. Wettstein. *Rev Médicale Suisse*. 2008;
24. Rommens PM, Wagner D, Hofmann A. Minimal Invasive Surgical Treatment of Fragility Fractures of the Pelvis. *Chirurgia (Bucur)* [Internet]. 2017 [cité 26 août 2024];112(5):524. Disponible sur: <http://www.revistachirurgia.ro/pdfs/2017-5-524.pdf>
25. Tubiana M. [Wilhelm Conrad Röntgen and the discovery of X-rays]. *Bull Acad Natl Med*. 1 janv 1996;180(1):97-108.

26. Bonniaud G. Chapitre 2. Radiologie. In: Chapitre 2 Radiologie [Internet]. EDP Sciences; 2016 [cité 3 juin 2024]. p. 71-146. Disponible sur: <https://www.degruyter.com/document/doi/10.1051/978-2-7598-1992-8.c005/pdf?licenseType=restricted>
27. Pinto A, Berritto D, Russo A, Riccitiello F, Caruso M, Belfiore MP, et al. Traumatic fractures in adults: missed diagnosis on plain radiographs in the Emergency Department. *Acta Bio-Medica Atenei Parm.* 19 janv 2018;89(1-S):111-23.
28. Samuel AL. Some Studies in Machine Learning Using the Game of Checkers. *IBM J Res Dev.* juill 1959;3(3):210-29.
29. Vujosevic S, Aldington SJ, Silva P, Hernández C, Scanlon P, Peto T, et al. Screening for diabetic retinopathy: new perspectives and challenges. *Lancet Diabetes Endocrinol.* avr 2020;8(4):337-47.
30. Ege BM, Hejlesen OK, Larsen OV, Møller K, Jennings B, Kerr D, et al. Screening for diabetic retinopathy using computer based image analysis and statistical classification. *Comput Methods Programs Biomed.* juill 2000;62(3):165-75.
31. Tahmoush D, Samet H. Image Similarity and Asymmetry to Improve Computer-Aided Detection of Breast Cancer. In: Astley SM, Brady M, Rose C, Zwiggelaar R, éditeurs. *Digital Mammography.* Berlin, Heidelberg: Springer; 2006. p. 221-8.
32. Sechopoulos I, Teuwen J, Mann R. Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art. *Semin Cancer Biol* [Internet]. 1 juill 2021 [cité 26 août 2024];72:214-25. Disponible sur: <https://www.sciencedirect.com/science/article/pii/S1044579X20301358>
33. Garvey JL, Zegre-Hemsey J, Gregg R, Studnek JR. Electrocardiographic diagnosis of ST segment elevation myocardial infarction: An evaluation of three automated interpretation algorithms. *J Electrocardiol* [Internet]. 1 sept 2016 [cité 12 juin 2024];49(5):728-32. Disponible sur: <https://www.sciencedirect.com/science/article/pii/S0022073616300255>
34. Lindow T, Kron J, Thulesius H, Ljungström E, Pahlm O. Erroneous computer-based interpretations of atrial fibrillation and atrial flutter in a Swedish primary health care setting. *Scand J Prim Health Care* [Internet]. 2 oct 2019 [cité 12 juin 2024];37(4):426-33. Disponible sur: <https://doi.org/10.1080/02813432.2019.1684429>
35. Taboulet P. ECG. Intelligence artificielle [Internet]. e-cardiogram. 2019 [cité 12 juin 2024]. Disponible sur: <https://www.e-cardiogram.com/intelligence-artificielle/>
36. Galloway CD, Valys AV, Shreibati JB, Treiman DL, Petterson FL, Gundotra VP, et al. Development and Validation of a Deep-Learning Model to Screen for Hyperkalemia From the Electrocardiogram. *JAMA Cardiol* [Internet]. 1 mai 2019 [cité 12 juin 2024];4(5):428-36. Disponible sur: <https://doi.org/10.1001/jamacardio.2019.0640>
37. Al-Zaiti SS, Martin-Gill C, Zègre-Hemsey JK, Bouzid Z, Faramand Z, Alrawashdeh MO, et al. Machine learning for ECG diagnosis and risk stratification of occlusion myocardial

- infarction. *Nat Med* [Internet]. juill 2023 [cité 12 juin 2024];29(7):1804-13. Disponible sur: <https://www.nature.com/articles/s41591-023-02396-3>
38. Attia ZI, Harmon DM, Behr ER, Friedman PA. Application of artificial intelligence to the electrocardiogram. *Eur Heart J* [Internet]. 7 déc 2021 [cité 12 juin 2024];42(46):4717-30. Disponible sur: <https://doi.org/10.1093/eurheartj/ehab649>
 39. Attia ZI, Kapa S, Yao X, Lopez-Jimenez F, Mohan TL, Pellikka PA, et al. Prospective validation of a deep learning electrocardiogram algorithm for the detection of left ventricular systolic dysfunction. *J Cardiovasc Electrophysiol* [Internet]. 2019 [cité 12 juin 2024];30(5):668-74. Disponible sur: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jce.13889>
 40. Adedinsewo D, Carter RE, Attia Z, Johnson P, Kashou AH, Dugan JL, et al. Artificial Intelligence-Enabled ECG Algorithm to Identify Patients With Left Ventricular Systolic Dysfunction Presenting to the Emergency Department With Dyspnea. *Circ Arrhythm Electrophysiol* [Internet]. août 2020 [cité 12 juin 2024];13(8):e008437. Disponible sur: <https://www.ahajournals.org/doi/full/10.1161/CIRCEP.120.008437>
 41. Kwon J myoung, Cho Y, Jeon KH, Cho S, Kim KH, Baek SD, et al. A deep learning algorithm to detect anaemia with ECGs: a retrospective, multicentre study. *Lancet Digit Health* [Internet]. 1 juill 2020 [cité 12 juin 2024];2(7):e358-67. Disponible sur: [https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(20\)30108-4/fulltext#seccestitle130](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30108-4/fulltext#seccestitle130)
 42. Attia ZI, Friedman PA, Noseworthy PA, Lopez-Jimenez F, Ladewig DJ, Satam G, et al. Age and Sex Estimation Using Artificial Intelligence From Standard 12-Lead ECGs. *Circ Arrhythm Electrophysiol* [Internet]. sept 2019 [cité 12 juin 2024];12(9):e007284. Disponible sur: <https://www.ahajournals.org/doi/full/10.1161/CIRCEP.119.007284>
 43. Rossum's universal robots. 1921.
 44. Turing AM. Computing Machinery and Intelligence. In: Epstein R, Roberts G, Beber G, éditeurs. *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer* [Internet]. Dordrecht: Springer Netherlands; 2009 [cité 30 mai 2024]. p. 23-65. Disponible sur: https://doi.org/10.1007/978-1-4020-6710-5_3
 45. Larousse É. LAROUSSE. 2024 [cité 30 mai 2024]. intelligence artificielle - LAROUSSE. Disponible sur: https://www.larousse.fr/encyclopedie/divers/intelligence_artificielle/187257
 46. WALLACH D. Le Deep Learning pour le traitement d'images Classification, détection et segmentation avec Python et TensorFlow [Internet]. Eni Editions. 2024. (Expert It). Disponible sur: <https://www.editions-eni.fr/livre/le-deep-learning-pour-le-traitement-d-images-classification-detection-et-segmentation-avec-python-et-tensorflow-9782409043208>
 47. Wang M, Cheung SW, Chung ET, Efendiev Y, Leung WT, Wang Y. Prediction of Discretization of GMSFEM Using Deep Learning. *Mathematics* [Internet]. mai 2019 [cité 31 mai 2024];7(5):412. Disponible sur: <https://www.mdpi.com/2227-7390/7/5/412>

48. Gruner SM, Eikenberry EF, Tate MW. Comparison of X-ray detectors. urn:isbn: [Internet]. 2006 [cité 13 juin 2024];F:143-7. Disponible sur: http://xrpp.iucr.org/cgi-bin/itr?url_ver=Z39.88-2003&rft_dat=what%3Dchapter%26volid%3DFa%26chnumero%3D7o1%26chvers%3Dv0001
49. Bousson V, Benoist N, Guetat P, Attané G, Salvat C, Perronne L. L'intelligence artificielle appliquée à l'interprétation de l'imagerie ostéoarticulaire : quel est le bilan ? Quelles sont les perspectives ? Rev Rhum [Internet]. oct 2023 [cité 11 oct 2023];90(5):575-84. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/S1169833023003150>
50. Eng DK, Khandwala NB, Long J, Fefferman NR, Lala SV, Strubel NA, et al. Artificial Intelligence Algorithm Improves Radiologist Performance in Skeletal Age Assessment: A Prospective Multicenter Randomized Controlled Trial. Radiology [Internet]. déc 2021 [cité 1 juill 2024];301(3):692-9. Disponible sur: <https://pubs-rsna-org.ezproxy.u-paris.fr/doi/10.1148/radiol.2021204021>
51. Guinebert S, Petit E, Bousson V, Bodard S, Amoretti N, Kastler B. Automatic semantic segmentation and detection of vertebrae and intervertebral discs by neural networks. Comput Methods Programs Biomed Update [Internet]. 1 janv 2022 [cité 1 juill 2024];2:100055. Disponible sur: <https://www.sciencedirect.com/science/article/pii/S2666990022000076>
52. Liu F, Zhou Z, Jang H, Samsonov A, Zhao G, Kijowski R. Deep convolutional neural network and 3D deformable approach for tissue segmentation in musculoskeletal magnetic resonance imaging. Magn Reson Med [Internet]. 2018 [cité 1 juill 2024];79(4):2379-91. Disponible sur: <https://onlinelibrary-wiley-com.ezproxy.u-paris.fr/doi/abs/10.1002/mrm.26841>
53. Zhou Z, Zhao G, Kijowski R, Liu F. Deep convolutional neural network for segmentation of knee joint anatomy. Magn Reson Med [Internet]. 2018 [cité 1 juill 2024];80(6):2759-70. Disponible sur: <https://onlinelibrary-wiley-com.ezproxy.u-paris.fr/doi/abs/10.1002/mrm.27229>
54. Chen H, Zhang Y, Zhang W, Liao P, Li K, Zhou J, et al. Low-dose CT via convolutional neural network. Biomed Opt Express. 2017;8(2):679-94.
55. Chea P, Mandell JC. Current applications and future directions of deep learning in musculoskeletal radiology. Skeletal Radiol [Internet]. 1 févr 2020 [cité 1 juill 2024];49(2):183-97. Disponible sur: <https://doi.org/10.1007/s00256-019-03284-z>
56. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. Nat Med [Internet]. sept 2022 [cité 1 juill 2024];28(9):1773-84. Disponible sur: <https://www-nature-com.ezproxy.u-paris.fr/articles/s41591-022-01981-2>
57. Olczak J, Fahlberg N, Maki A, Razavian AS, Jilert A, Stark A, et al. Artificial intelligence for analyzing orthopedic trauma radiographs. Acta Orthop [Internet]. 2 nov 2017 [cité 1 juill 2024];581-6. Disponible sur: <https://actaorthop.org/actao/article/view/9748>

58. Jones RM, Sharma A, Hotchkiss R, Sperling JW, Hamburger J, Ledig C, et al. Assessment of a deep-learning system for fracture detection in musculoskeletal radiographs. *NPJ Digit Med* [Internet]. 30 oct 2020 [cité 26 août 2024];3:144. Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7599208/>
59. Guy S, Jacquet C, Tsenkoff D, Argenson JN, Ollivier M. Le Deep Learning pour l'aide au diagnostic radiographique des fractures de l'extrémité proximale du fémur : limites et difficultés de programmation. *Rev Chir Orthopédique Traumatol* [Internet]. 1 avr 2021 [cité 11 oct 2023];107(2):204-9. Disponible sur: <https://www.sciencedirect.com/science/article/pii/S1877051721000447>
60. 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*. mars 2022;302(3):627-36.
61. Twinprai N, Boonrod A, Boonrod A, Chindapasirt J, Sirithanaphol W, Chindapasirt P, et al. Artificial intelligence (AI) vs. human in hip fracture detection. *Heliyon*. nov 2022;8(11):e11266.
62. Cheng CT, Wang Y, Chen HW, Hsiao PM, Yeh CN, Hsieh CH, et al. A scalable physician-level deep learning algorithm detects universal trauma on pelvic radiographs. *Nat Commun*. 16 févr 2021;12(1):1066.
63. 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*. juill 2021;300(1):120-9.
64. Noret M. Recherches « Hors Loi Jardé ». 6 oct 2021; Disponible sur: <https://www.infectiologie.com/UserFiles/File/formation/desc/2021/seminaire-octobre-2021/merc-610-t.27/conf-4-hors-loi-jarde-mnoret.pdf>
65. Rommens PM, Wagner D, Hofmann A. Minimal Invasive Surgical Treatment of Fragility Fractures of the Pelvis. *Chir Buchar Rom* 1990. 2017;112(5):524-37.
66. Rollmann MF, Herath SC, Kirchhoff F, Braun BJ, Holstein JH, Pohlemann T, et al. Pelvic ring fractures in the elderly now and then - a pelvic registry study. *Arch Gerontol Geriatr*. juill 2017;71:83-8.

Table des figures

Figure 1 Évolution du nombre de passages annuels aux urgences de 1996 à 2021. SSA : Service de santé des armées (7)	10
Figure 2 Anatomie du bassin(21)	15
Figure 3 Fractures élémentaire du bassin (22).....	16
Figure 4 Modèle d'un réseau de neurones profonds(47)	22
Figure 5 Entraînement d'un réseau de neurone (46)	23
Figure 6 Représentation de la sensibilité, de la spécificité et de l'aire sous la courbe poolées pour les études de validité externes et celles évaluant les performances des médecins aidés des logiciels d'IA (5).....	27
Figure 7 Flow chart	34
Figure 8 répartition de la localisation des fractures détectées au scanner (%).....	40

Table des tableaux

Tableau 1 : Antécédents des patients inclus	35
Tableau 2 : Facteurs de risque d'ostéoporose des patients inclus	35
Tableau 3 : Présentation clinique des patients inclus.....	36
Tableau 4 : Critères qualité des radiographies incluses.....	36
Tableau 5 : Nombre de critères qualité des radiographies incluses	37
Tableau 6 : Analyse de la population âge (année), sexe et délais de réalisation du scanner (jour) en fonction de la présence de fracture détectée à la radiographie	38
Tableau 7 : Analyse des critères de qualité de réalisation de la radiographie et de la présence d'arthrose en fonction de la présence de fracture détectée à la radiographie.....	38
Tableau 8 : Motif de réalisation des scanner du bassin.....	39
Tableau 9 : Localisation du trait de fracture en cas de fracture multiple	40
Tableau 10 : Analyse de la population âge (année), sexe et délais de réalisation du scanner (jour) en fonction de la présence de fracture détectée au scanner	41
Tableau 11 : Représentation des fractures en fonction de leur détection par le scanner et la radiographie relue par l'IA	42
Tableau 12 : Analyse des critères qualité de la radiographie en fonction de leur répartition chez les vrais positifs, vrais négatifs, faux positifs, faux négatif.....	43
Tableau 13 : Analyse des antécédents et facteurs de risque d'ostéoporose trouvés dans les dossiers en fonction de leur répartition en vrai positif, vrai négatif, faux positif, faux négatif	44
Tableau 14 : Analyse de la présentation clinique des patients en fonction de leur répartition en vrai positif, vrai négatif, faux positif, faux négatif	45
Tableau 15 : Performance de L'IA pour la relecture de la radiographie de bassin en fonction de la localisation : acetabulum.....	45
Tableau 16 : Performance de L'IA pour la relecture de la radiographie de bassin en fonction de la localisation : ailes iliaques.....	46
Tableau 17 : Performance de L'IA pour la relecture de la radiographie de bassin en fonction de la localisation : anneau pelvien.....	46

Tableau 18 : Performance de L'IA pour la relecture de la radiographie de bassin en fonction de la localisation : sacrum.....46

L'intelligence artificielle permettait-elle de réduire la prescription de scanner pour la détection des fractures du bassin ?

Résumé :

Introduction : Le bilan de première intention d'un traumatisme du bassin à faible cinétique comprend une radiographie de bassin de face (1) qui est relue par les médecins urgentistes (2). Hélas cette interprétation n'est pas sans erreur diagnostique (3,4). Les logiciels d'aide au diagnostic fondés sur l'intelligence artificielle (IA) représente un nouvel outil (5). Leurs performances sont, soit similaires, soit dépassent celles des urgentistes (6). L'objectif de cette étude est d'évaluer ce nouveau test diagnostique qu'est la radiographie de bassin de face relue par IA et de voir si ses performances sont suffisantes pour se passer du scanner de bassin qui est prescrit en cas de doute sur la présence d'une fracture.

Méthode : Etude rétrospective et monocentrique quantitative sur la population des patients ayant réalisé un TDM à l'HEGP entre le 1^{er} janvier et le 31 décembre 2022. Nous avons utilisé le scanner comme référence pour établir les performances de la radiographie de bassin relue par l'IA et analyser des multiples facteurs de risque et signes cliniques.

Résultats : L'interprétation de la radiographie de bassin par l'IA par rapport au scanner qui est le Gold standard avait une sensibilité (Se) à 0.69, une spécificité (Sp) à 0.93. Les performances de la radiographie étaient satisfaisantes pour l'anneau pelvien (Se 0.80, Sp 0.94), moyennes pour l'acetabulum et les ailes iliaques mais manquaient systématiquement les fractures du sacrum. L'ostéoporose et ses facteurs de risque n'étaient pas associés à un surrisque de fracture. L'arthrose n'était pas associée à surrisque d'erreur de l'IA. Les radiographies de bassin étaient dans l'ensemble de mauvaise qualité et la symétrie des trous obturateurs était associé de façon statistiquement significative à un surrisque de faux positif. Aucun critère clinique n'étaient prédictifs d'une fracture.

Conclusion : Notre étude a montré que la relecture d'une radiographie de bassin par l'IA ne possède pas les performances pour remplacer le scanner qui reste indispensable en cas de doute diagnostique en particulier pour les fractures de la partie postérieure du bassin.

Spécialité :

[DESMU](#)

Mots clés français :

[Urgences](#), [radiologie](#), [intelligence artificielle](#), [traumatologie](#), [fractures osseuses](#), [qualitative](#), [quantitative](#)

Forme ou Genre :

[fMeSH](#) : [Dissertation universitaire](#)

[Rameau](#) : [Thèses et écrits académiques](#)

**Université Paris Cité
UFR de médecine
15 Rue de l'École de Médecine
75006 Paris**