

INTELLIGENCE ARTIFICIELLE CAS PRATIQUES EN SENOLOGIE

Pr Isabelle Thomassin-Naggara

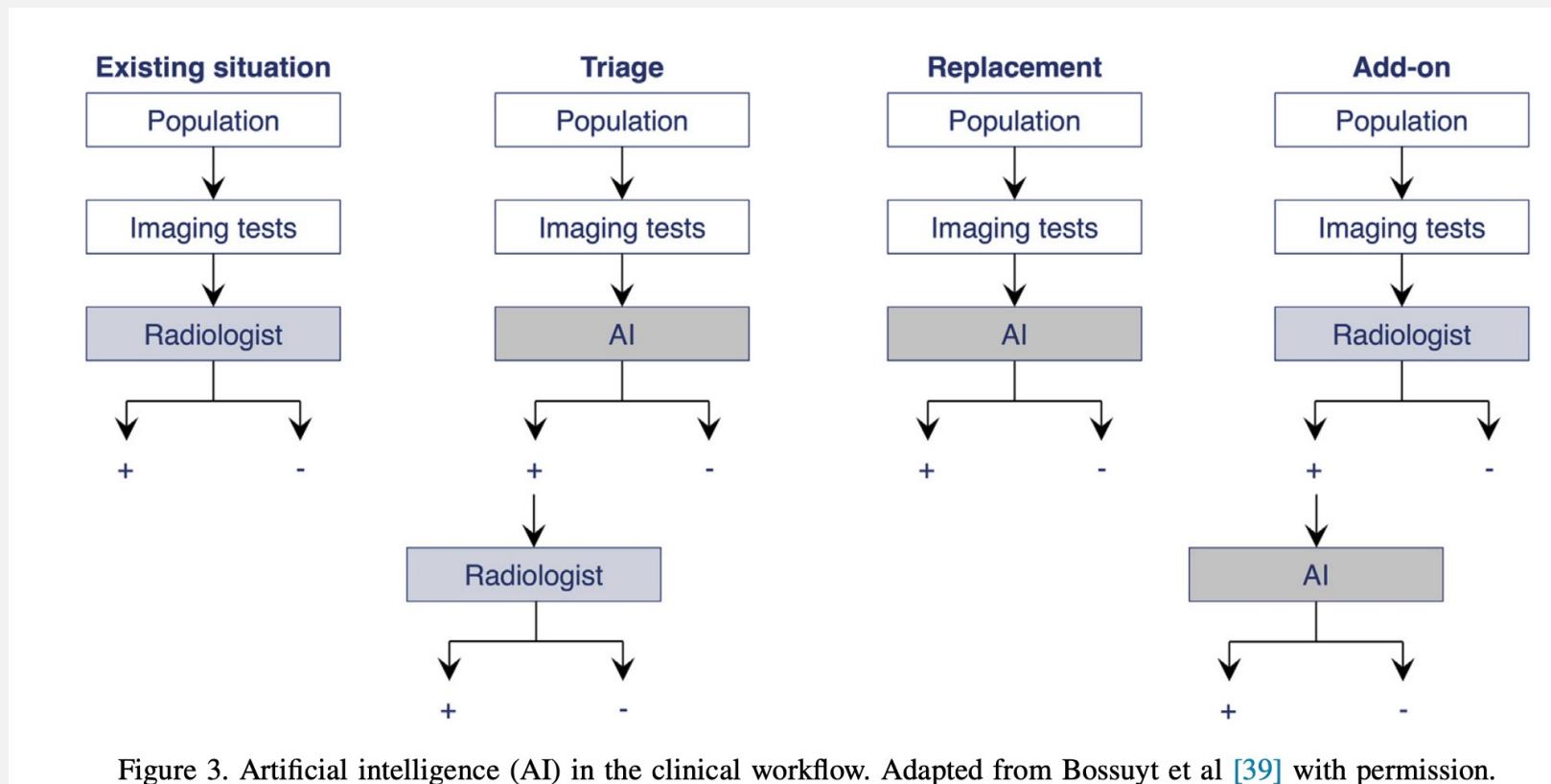
28 Novembre 2022



INTRODUCTION

- Il y a eu une explosion de la littérature sur l'IA au cours des dernières années, alors qu'il y a relativement peu d'études qui évaluent réellement l'expérience clinique réelle des radiologues atteints d'IA (1)
- L'utilisation prospective de l'IA est une étape essentielle avant son utilisation clinique généralisée
- Les leçons tirées du déploiement de la technologie de détection assistée par ordinateur (CAO) au début des années 2000 ont montré que l'expérience clinique réelle n'était pas à la hauteur des attentes
- L'objectif de cet exposé sera de comprendre comment peut on utiliser l'IA aujourd'hui et dans l'avenir

COMMENT UTILISER L'ALGORITHME D'IA DANS LA PRATIQUE CLINIQUE ?



MAMMOGRAPHIE OU TOMOSYNTHÈSE INTERPRÉTATION AUTONOME

- **Modèle d'IA par rapport à la lecture humaine**
 - Revue systématique (12 études) sur 79910 femmes (taux de détection 23/1000) avec 36 systèmes d'IA différents : 94 % des systèmes d'IA sont moins précis que le radiologue d'origine (1)
 - L'IA en tant que modalité autonome a une sensibilité égale à celle des radiologues, mais une spécificité plus faible, en particulier sur la tomosynthèse (2)
 - INSPlRED003 (essai prospectif multicentrique) : Le diagnostic systématique du cancer du sein est une approche d'information multimodale (ASC=0,95) qui surpassé tous les algorithmes (ASC=0,90) (3)

MAMMOGRAPHIE OU TOMOSYNTHÈSE

INTERPRETATION AVEC AI

- **Radiologue augmenté par IA**

- Radiologue AUC avec radiologue AI >AUC sans système AI ou AUCAI (1,2, 3, 4)
- IA + radiologue = sensibilité accrue (86 % contre 83 %) et spécificité (79 % contre 77 %) (2)
- Valeur ajoutée différente du modèle d'IA appliqué sur mammographie par rapport à Tomosynthèse selon le type d'anomalie

(1) DREAM CHALLENGE Jama Network open 2020

(2) Rodriguez Ruiz Radiology 2019

(3) Wu et al. IJET 2018

(4) Jiang et al. Radiology 2020

(5) Le Gullo et al. Eur Radiol 2020

(6) Kim et al. Lancet digit 2020

TOMOSYNTHÈSE MAMMAIRE

INTERPRÉTATION AVEC INTELLIGENCE ARTIFICIELLE

- Résultats concordants**
 - Cohorte enrichie, lecteurs multiples
 - >15 radiologues (principalement de petite taille K invasive)
- Durée de lecture = Sélection de la coupe positive**
 - Comparaison entre le modèle d'IA appliqué sur 2DMG, la coupe optimale sur DBT et la MG synthétique

Mendel et al. Acad Radiol 2019

Profound AI (24 lecteurs) ICAD	Sans IA	Avec IA	
Sensitivity (patient)	77%	85%	P<0,01
Sensitivity (lesion)	77%	85.3%	P<0,01
Specificity	62.7%	69.6%	P<0,01
AUC	0,795	0,852	P<0,01
Reading time	64sec	30sec	P<0,01
Recall rate	38%	31%	P<0,01

Conant et. al. Radiology AI January 2019

Transpara (18 lecteurs) INCEPTO	Sans IA	Avec IA	
Sensitivity (patient)	75%	79%	P<0,01
Specificity	equal	equal	
AUC	0,863	0,833	P<0,01
Reading time	41sec	36 sec	P<0,01

Van Winkel et. al. Eur Radiol 2021

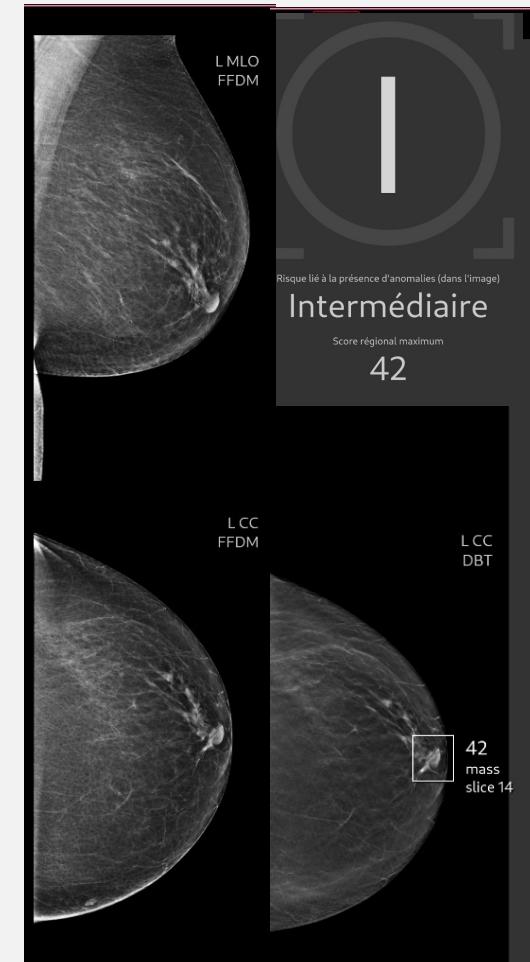
MAMMOGRAPHIE OU DBT

FACTEURS DE VARIATION DE L' AI

- **Facteurs de variation de la performance du modèle d'IA**
 - Population de training différente de la population testée (1)
 - Analyse d'une seule vue par rapport aux deux vues >> Détection d'asymétrie >> Spécificité améliorée (2)
 - Pas d'intégration d'erreur de positionnement (FP)
 - Aucune intégration des mammographies précédentes (FP)

(1) Sasaki, Breast cancer 2020

(2) Kooi et al. Proceedings of SPIE 2017



MAMMOGRAPHIE OU DBT TRIAGE

- **IA = Triage**

- Note 1 ou 2 : 0 cancer omis et peu de faux positifs = 20 % de la population

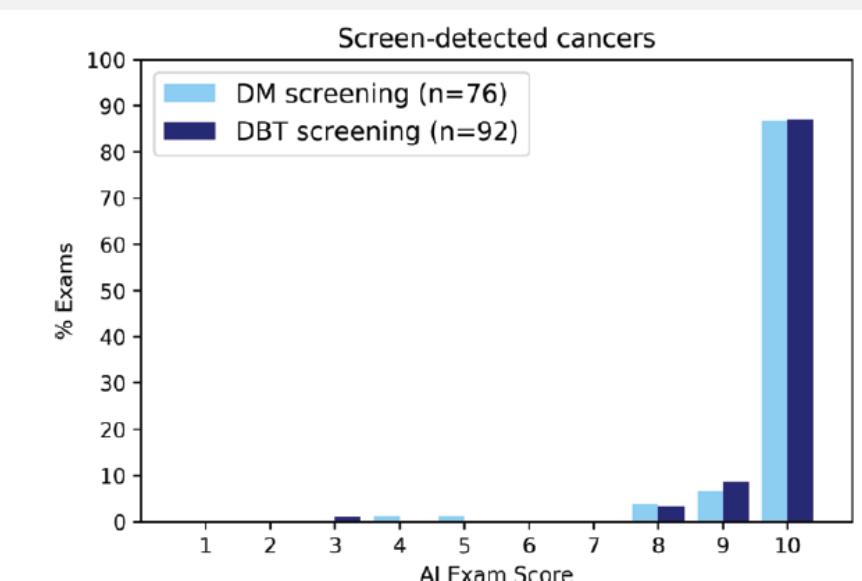
Zaricksson et al. Eur Radiol 2020

- Note ≤ 7 : seulement 5 % de la perte de sensibilité représente 70 % de la population

Povedano et al. Radiology 2021

- Examen systématique : Large éventail : de 7 % à 91 % des mammographies normales identifiées pourraient être lues par un dépistage adapté, avec 0 % à 7 % des cancers manqués.

Hickman and Gilbert et al. Radiology 2022



Cordoba Tomosynthesis trial

15986 women

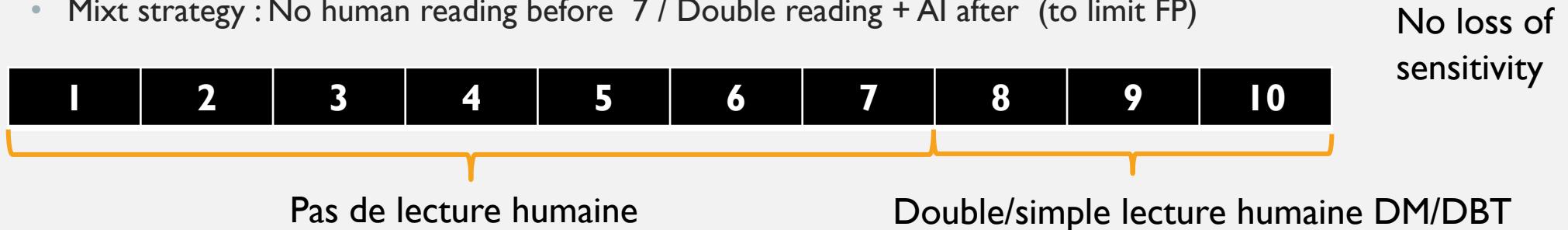
Detection rate 6 / 1000

Interval K Rate 1 / 1000

MAMMOGRAPHIE OU DBT

TRIAGE

- Selection of examination which need second reading (MG)
 - Mixt strategy : No human reading before 7 / Double reading + AI after (to limit FP)



	Classical	With AI triage	P (Nb,H,RR)
DM double human reading	Nb 31974 / H 222 / Se 67.3% / RR 5,1	Nb 9100 / H 63 / Se 69% / RR 4,2	<0.01
DBT double human reading	Nb 31974 / H 568 / Se 81,4% / RR 4,4	Nb 8830 / H 156 / Se 84,1% / RR 3,7	<0.01
DBT simple human reading	Nb 15987 / H 284 / Se 77% / RR 3	Nb 4415 / H 78 / Se 79,5% / RR 3,1	<0.01

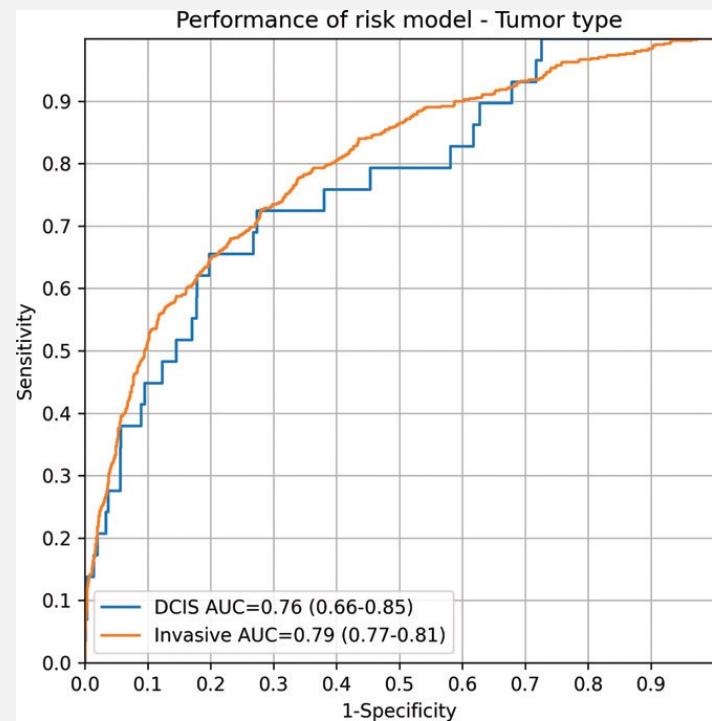
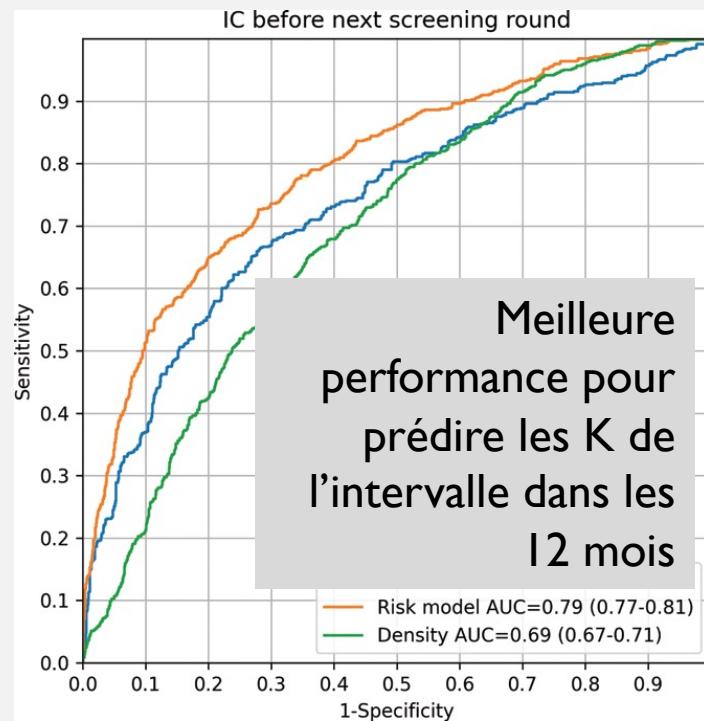
2D MAMMOGRAPHY OR DBT BREAST SCREENING : INTERVAL CANCER RATE

- Taux de réduction: 50%

Lang Eur Radiol. 2021 (Transpara)

- Étude cas-témoin rétrospective
 - 2222 femmes avec CI
 - 4661 femmes dans le groupe témoin
 - Un modèle de réseau neuronal a été entraîné en combinant le score d'IA et de densité mammaire (validation croisée $\times 10$)

Wanders et al. Radiology 2022



L'évaluation combinée d'un système de détection par intelligence artificielle et de mesures de la densité mammaire a permis d'identifier une plus grande proportion de femmes qui développeraient un cancer de l'intervalle comparativement à l'une ou l'autre des méthodes seules.

« URGE FOR CAUTION »

JAMA Health Forum™

Viewpoint

Artificial Intelligence in Medical Imaging—Learning From Past Mistakes in Mammography

Joann G. Elmore, MD, MPH; Christoph I. Lee, MD

LCI

Dépistage du cancer du sein : la petite révolution de l'intelligence artificielle

A.P

Publié le 4 février 2021 à 13h48, mis à jour le 5 février 2021 à 15h06

L'I.A. de Google détecte mieux les cancers du sein que les meilleurs radiologues

Cédric Ingrand

Publié le 3 janvier 2020 à 16h20

Study reference	Risk of bias						Applicability concerns			
	Patient selection	Index test	Comparator test	Reference standard	Flow and timing	Patient selection	Index test	Comparator test	Reference standard	
Standalone AI systems (5 studies)										
Lotter 2021 ²⁸	High	High	High	Unclear	Unclear	High	High	High	High	High
McKinney 2020 ²⁹	High	High	Low	High	High	High	High	High	High	Low
Rodriguez-Ruiz 2019 ³³	High	High	High	Unclear	Unclear	High	High	High	High	Unclear
Salim 2020 ³⁵	High	High	Low	High	High	High	High	Low*	High*	High
Schaffter 2020 ³⁶	Low	High	Low	High	High	Unclear	High	Low*	High*	High
AI as reader aid (3 studies)										
Pacilè 2020 ³⁰	High	High	High	High	Unclear	High	High	High	High	High
Rodriguez-Ruiz 2019 ^{32,34}	High	High	High	High	Unclear	High	High	High	High	High
Watanabe 2019 ³⁷	High	High	High	Unclear	Unclear	High	High	High	High	Low
AI for triage (4 studies)										
Balta 2020 ²⁵	Low	High	None	High	High	Low	High	None	High	High
Dembrower 2020 ²⁶	High	High	None	Low†	High†	High	High	High	None	Low†
Lång 2020 ²⁷	Low	High	None	High	High	Low	High	None	High	High
Raya-Povedano 2021 ³¹	Low	Low	Low	Low	High	Low	High	High	High	Low

Freeman et al. BMJ 2021

« URGE FOR CAUTION »

	Type de jeux de données	Prévalence de cancer	Design
Entrainement + Validation interne	Cohorte enrichie et annotée	100%	Rétrospective sur data (données annotées)
Validation externe	Prévalence supérieure à la normale / Cas subtils	5-10%	Rétrospective sur data (relecture cas discordants)
Validation externe	Prévalence de population	5-7/1000	Rétrospective sur data (données non annotées)
Validation	En condition clinique	5-7/1000	Prospective multicentrique

Steps to avoid overuse and misuse of machine learning in clinical research

Machine learning algorithms are a powerful tool in healthcare, but sometimes perform no better than traditional statistical techniques. Steps should be taken to ensure that algorithms are not overused or misused, in order to provide genuine benefit for patients.

Victor Volovici, Nicholas L. Syn, Ari Ercole, Joseph J. Zhao and Nan Liu

NATURE MEDICINE | VOL 28 | OCTOBER 2022 | 1996-1999 | www.nature.com/naturemedicine

CONCLUSION

Human patients are likely to want human doctors to continue making medical decisions, no matter how well an algorithm can predict outcomes. ML should, therefore, be studied and implemented as an integral part of a complete system of care.

The clinical integration of ML and big data is poised to improve medicine. ML researchers should recognize the limits of their algorithms and models in order to prevent their overuse and misuse, which could otherwise sow distrust and cause patient harm.



EVALUATION DES LOGICIELS IA

Table 2

Overview of proposed Canadian Association of Radiologists AI Working Groups

Working group	Mandate(s)
Research and development	<ul style="list-style-type: none">- Define a research and development framework- Oversee the creation of annotated data- Organize competitions- <u>Develop training, validation, and test datasets for clinically relevant applications</u>
Technique and applications	<ul style="list-style-type: none">- Evaluate and compare datasets, algorithms, and applications that have a high potential of being used in clinical practice.
Education	<ul style="list-style-type: none">- Contribute to outreach and education- Propose content on AI at the annual scientific meeting- Provide guidance on introductory pieces- Suggest objectives and content for radiology residency curriculum
Legal and ethics	<ul style="list-style-type: none">- Define scope of access to health care data for training of AI models- Advise on best practices in data management- Inform on anonymization requirements- Clarify scope of medico-legal responsibilities of AI-supported clinical decisions
Writing	<ul style="list-style-type: none">- Write a white paper updated on a regular basis- Development of practice guidelines- Definition of clinical use scenarios
Industrial partners	<ul style="list-style-type: none">- Liaison with industrial partners in technology and AI applications to radiology- Interoperability between AI systems

CAR staff

AI advisory WG

ACR staff

AI = artificial intelligence.

Tang A. et al JCAR 2018

INITIATIVE DRIM FRANCE IA CNP RADILOGIE

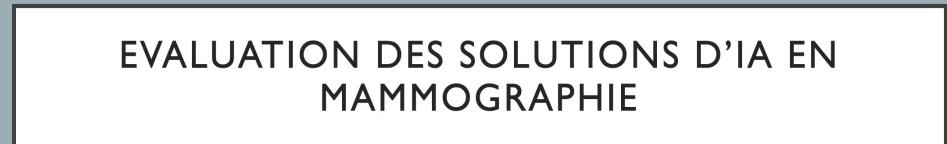
- Création d'une grille descriptive et analytique des solutions d'IA en mammographie
- Travail collaboratif avec les éditeurs de logiciel
- Première sollicitation et retour des différents constructeurs pour les JFR 2022



JFR
JOURNÉES FRANCOPHONES
DE RADILOGIE 2022
DIAGNOSTIQUE & INTERVENTIONNELLE



DRIM FRANCE IA



EVALUATION DES SOLUTIONS D'IA EN
MAMMOGRAPHIE

Pr Isabelle Thomassin-Naggara

9 octobre 2022



iCAD[®]



HERA-MI

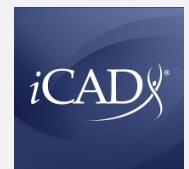


THERAPIXEL



INCEPTO[®]

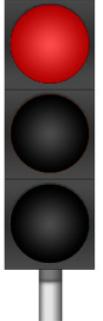
Conseil National professionnel de radiologie et imagerie médicale (G4)
Associe toutes les composantes de la radiologie française



GRILLE DESCRIPTIVE ET ANALYTIQUE DES SOLUTIONS D'IA EN MG



- Objectifs de l'outil d'IA (Standalone, Triage, Prediction du risque)
- Descriptif de fonctionnement (MG/DBT, clicéh antérieur, comparative..)
- Certification – Marquage
- Base de données (Entrainement, validation, test)
- Intégration SI, expression des résultats (serveur local, temps de process, PACS..)
- Aspects médico légaux et cybersécurité (règles RGPD, Serveurs agréés, anonymisation..)
- Performance et validation scientifique



GRILLE DESCRIPTIVE ET ANALYTIQUE DES SOLUTIONS D'IA EN MG 7-PERFORMANCE ET VALIDATION SCIENTIFIQUE

PARTIE A = Publications scientifiques à présenter sous forme de tableau

- Type d'utilisation : Standalone / Triage / Aide au radiologue
- Type de publication : peer review (pubmed) ou en « gray literature », Communications scientifiques dans les congrès
- Caractéristiques de l'étude (Design de l'étude, population, mammographe (marque et %)), test index, comparateur, reference standard) Cf modèle : freeman BMJ 2021 doi: 10.1136/bmj.n1872 | BMJ 2021;374:n1872)
- Résultats de l'étude : Sensibilité / spécificité / ROC curve ...
- Qualité de l'étude : Score DECIDE-AI (Nat Med. 2022 May;28(5):924-933. doi: 10.1038/s41591-022-01772-9

Table 2 | DECIDE-AI checklist

Item number	Theme	Recommendation
1-17	AI-specific reporting items	
1-X	Generic reporting items	
Title and abstract		
I 1	Title	Identify the study as early clinical evaluation of a decision support system based on AI or machine learning, specifying the problem addressed.
I 2	Abstract	Provide a structured summary of the study. Consider including: intended use of the AI system, type of underlying algorithm, study setting, number of patients and users included, primary and secondary outcomes, key safety endpoints, human factors evaluated, main results and conclusions.
Introduction		
II 2	Intended use	a) Describe the targeted medical condition(s) and problem(s), including the current standard practice, and the intended patient population(s). b) Describe the intended users of the AI system, its planned integration in the care pathway, and the potential effect, including patient outcomes, that it is intended to have.
II 3	Objectives	State the study objectives.
Methods		
III 3	Research governance	Provide a reference to any study protocol, study registration number, and ethics approval.
III 4	Participants	a) Describe how patients were recruited, stating the inclusion and exclusion criteria at both patient and data level, and how the number of recruited patients was decided. b) Describe how users were recruited, stating the inclusion and exclusion criteria, and how the intended number of recruited users was decided. c) Describe steps taken to familiarize the users with the AI system, including any training received before the study.
III 5	AI system	a) Briefly describe the AI system, specifying its version and type of underlying algorithm used. Describe, or provide a direct reference to, the characteristics of the patient population on which the algorithm was trained and its performance in preclinical development/validation studies. b) Identify the data used as inputs. Describe how the data were acquired, the process needed to enter the input data, the pre-processing applied, and how missing/low-quality data were handled. c) Describe the AI system outputs and how they were presented to the users (an image may be useful).
III 6	Implementation	a) Describe the settings in which the AI system was evaluated. b) Describe the clinical workflow/care pathway in which the AI system was evaluated, the timing of its use, and how the final supported decision was reached and by whom.
IV 6	Outcomes	Specify the primary and secondary outcomes measured.
IV 7	Safety and errors	a) Provide a description of how significant errors/malfunctions were defined and identified. b) Describe how any risks to patient safety or instances of harm were identified, analyzed, and minimized.
IV 8	Human factors	Describe the human factors tools, methods or frameworks used, the use cases considered, and the users involved.
IV 9	Analysis	Describe the statistical methods by which the primary and secondary outcomes were analyzed, as well as any pre-specified additional analyses, including subgroup analyses and their rationale.
IV 10	Ethics	Describe whether specific methodologies were used to fulfil an ethics-related goal (such as algorithmic fairness) and their rationale.
IV 11	Patient Involvement	State how patients were involved in any aspect of: the development of the research question, the study design, and the conduct of the study.
Results		
V 9	Participants	a) Describe the baseline characteristics of the patients included in the study and report on input data missingness. b) Describe the baseline characteristics of the users included in the study.
V 10	Implementation	a) Report on the user exposure to the AI system, on the number of instances the AI system was used, and on the users' adherence to the intended implementation. b) Report any significant changes to the clinical workflow or care pathway caused by the AI system.
V 11	Main results	Report on the pre-specified outcomes, including outcomes for any comparison group if applicable.
V 12	Subgroups analysis	Report on the differences in the main outcomes according to the pre-specified subgroups.
V 13	Modifications	Report any changes made to the AI system or its hardware platform during the study. Report the timing of these modifications, the rationale for each, and any changes in outcomes observed after each of them.

Continued

Table 2 | DECIDE-AI checklist (continued)

Item number	Theme	Recommendation
12	Human-computer agreement	Report on the user agreement with the AI system. Describe any instances of and reasons for user variation from the AI system's recommendations and, if applicable, users changing their mind based on the AI system's recommendations.
13	Safety and errors	a) List any significant errors/malfunctions related to: AI system recommendations, supporting software/hardware, or users. Include details of: (i) rate of occurrence, (ii) apparent causes, (iii) whether they could be corrected, and (iv) any significant potential effects on patient care. b) Report on any risks to patient safety or observed instances of harm (including indirect harm) identified during the study.
14	Human factors	a) Report on the usability evaluation, according to recognized standards or frameworks. b) Report on the user learning curves evaluation.
Discussion		
15	Support for intended use	Discuss whether the results obtained support the intended use of the AI system in clinical settings.
16	Safety and errors	Discuss what the results indicate about the safety profile of the AI system. Discuss any observed errors/malfunctions and instances of harm, their implications for patient care, and whether/how they can be mitigated.
IX	Strengths and limitations	Discuss the strengths and limitations of the study.
Statements		
17	Data availability	Disclose if and how data and relevant code are available.
X	Conflicts of interest	Disclose any relevant conflicts of interest, including the source of funding for the study, the role of funders, any other roles played by commercial companies, and personal conflicts of interest for each author.

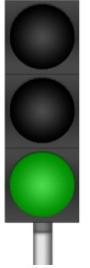
AI-specific items are numbered in Arabic numerals; generic items are numbered in Roman numerals.

CONSENSUS STATEMENT<https://doi.org/10.1038/s41591-022-01772-9>**Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI**

Baptiste Vasey^{1,2,3}✉, Myura Nagendran⁴, Bruce Campbell^{5,6}, David A. Clifton², Gary S. Collins^{1,7}, Spiros Denaxas^{8,9,10,11}, Alastair K. Denniston^{12,13,14}, Livia Faes¹⁴, Bart Geerts¹⁵, Mudathir Ibrahim^{1,16}, Xiaoxuan Liu^{12,13}, Bilal A. Mateen^{17,18}, Piyush Mathur¹⁹, Melissa D. McCradden^{20,21}, Lauren Morgan²², Johan Ordish²³, Campbell Rogers²⁴, Suchi Sarria^{25,26}, Daniel S. W. Ting^{27,28}, Peter Watkinson²⁹, Wim Weber³⁰, Peter Wheatstone³¹, Peter McCulloch¹ and the DECIDE-AI expert group*

GRILLE DESCRIPTIVE ET ANALYTIQUE DES SOLUTIONS D'IA EN MG

8-DESCRIPTIF DE LA SOCIÉTÉ ET MODÈLE ÉCONOMIQUE



- Date de création de la société
- Nombre d'employés
- Nombre de clients : en France / à L'étranger
- Modèle de facturation : Pay per view / Abonnement / Autre
 - Détail
- Des formations à l'usage de l'outil sont-elles proposées ?
- Détails des maintenances et service après-vente
- Le modèle économique inclue-t-il la mise à disposition de nouvelles versions ?

TAKE HOME MESSAGE

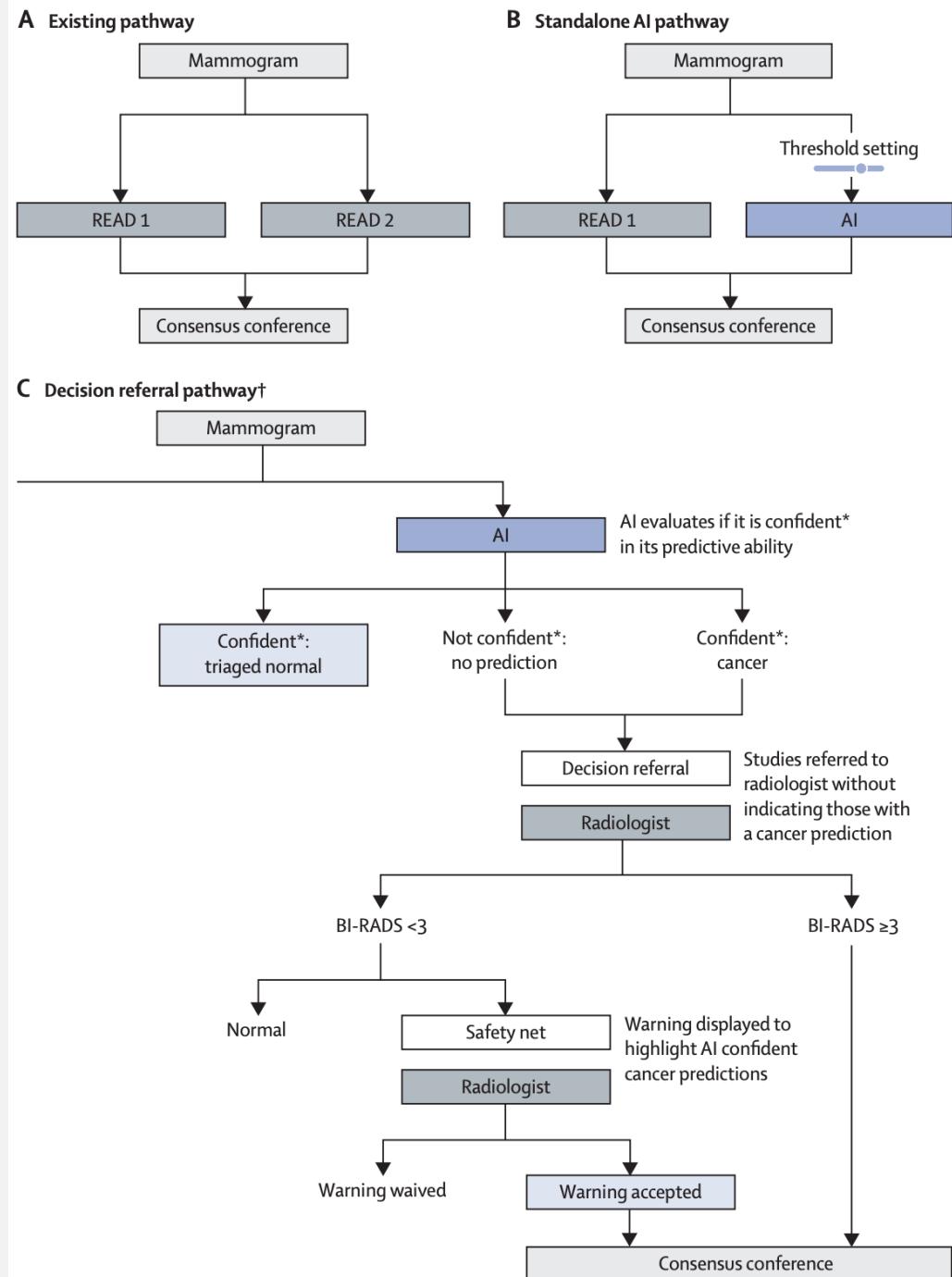
- L'interaction humaine et de la machine est complexe et les lecteurs ne peuvent pas prédire de façon fiable les résultats
 - L'applicabilité doit être testée dans la vraie vie >> Expérience Française
- Quel avenir peut on envisager ?
 - Solutions de détection du cancer du sein

Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis

Christian Leibig*, Moritz Brehmer*, Stefan Bunk, Danalyn Byng, Katja Pinkert†, Lale Umutlut

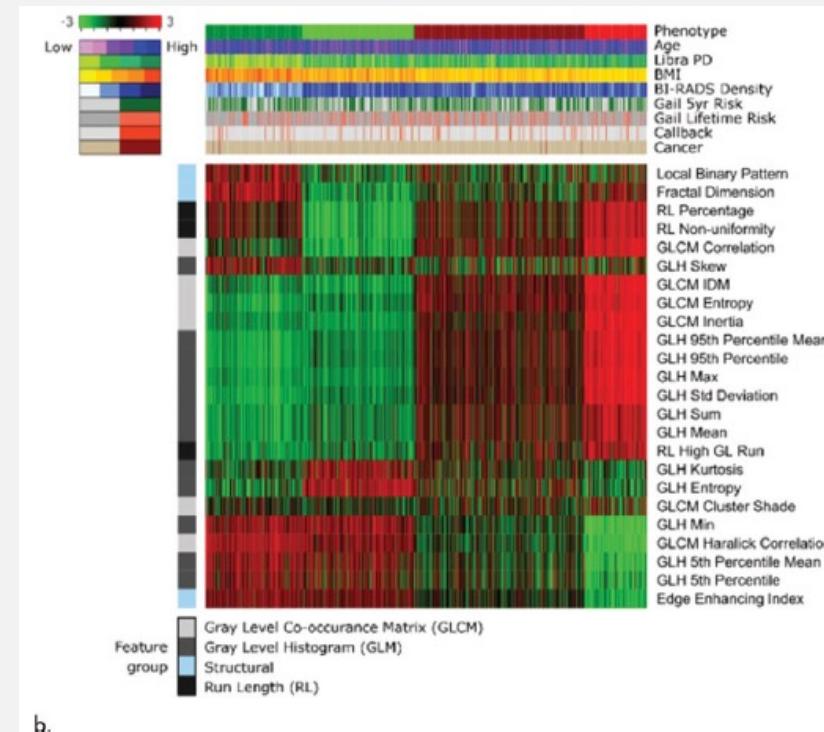
www.thelancet.com/digital-health Vol 4 July 2022

- 1 million de mammographies (Allemagne)
- 8 centres dépistage du cancer du sein
- Prévalence populationnelle
- 3 constructeurs
- Sensibilité et spécificité IA seul < radiologue
- Decision referral pathway (triage) >> meilleur performance que le radiologue sans IA



TAKE HOME MESSAGE

- Quel avenir peut on envisager ?
 - Ne pas tout investir sur le remplacement de l'humain mais imaginer faire bien au-delà
 - >>> Modèles non supervisés (radiomique)
 - Score de complexité du parenchyme mammaire : facteur prédictif indépendant de survenue de cancer du sein



b.