Détection précoce du cancer du sein : réentraînement et évaluation de modèles d'intelligence artificielle sur des mammographies de dépistage organisé en Occitanie

Emilien Jemelen[1,2], Francisco Orchard[2], William Madie[2], Dr. Bernard Valentin[3], Dr. Josine Belin[3], Guillaume Jeannerod[2], Pr. Pierre Mares[3], Pr. Sandrine Katsahian[1], Pr. Agathe Guilloux[1]

[1] Heka Team (INRIA, INSERM, UPC), [2] Epiconcept, [3] CRCDC-OC

Contexte

- Le cancer du sein (CS) était la **première cause de mortalité par cancer** chez les femmes en France en 2023 (12 100) [1,2].
- Le programme national de dépistage (DOCS, années 2000), repose sur la **mammographie pour la détection** précoce.
- Les **modèles récents de réseaux de neurones montrent de bonnes performances** pour la détection de malignité à partir de mammographies numériques [3–5].
- **Évaluer leur impact sur le DOCS** — en termes d'amélioration du dépistage et de réduction de la charge de travail — **est essentiel** pour l'avenir du programme en France.

Objectif de l'étude

Réentraîner et tester des modèles de détection open source à l'état de l'art sur une base de dépistage rétrospective, afin d'évaluer leurs performances.

1. Données utilisées

- 266 624 **examens de dépistage** (2011–2018) réalisés dans le Gard et la Lozère, chez des femmes de 50 à 74 ans (âge moyen : 61 ans, sans facteur de risque connu), ont été **appariés au SNDS** (Sécurité Sociale) pour identifier les cas de cancer du sein.
- 45 013 examens comportaient ≥ 4 images (mammographies, échographies, agrandissements). **41 884 examens avec 4 clichés** valides (R-CC, R-MLO, L-CC, L-MLO) après application d'un algorithme de sélection (voir section 3 et Fig. 1).

2. Identification des cancers (Fig. 2)

- Les CS sont identifiés par le DOCS ou dans le SNDS, dans un délai de 24 mois (ou jusqu'au dépistage suivant). Dans le SNDS, l'identification passe par:
- **Décès** par CS
- **Chirurgie spécifique** CS
- **Chirurgie suspectée** CS **+ traitement** cancer
- **Inscription aux Affections Longue Durée (ALD)** pour CS **+ traitement** cancer
- **Diagnostic** pour CS **+ traitement** du cancer

➔Ces labels CS sont au niveau de l'examen (pas de distinction sein gauche/droit)

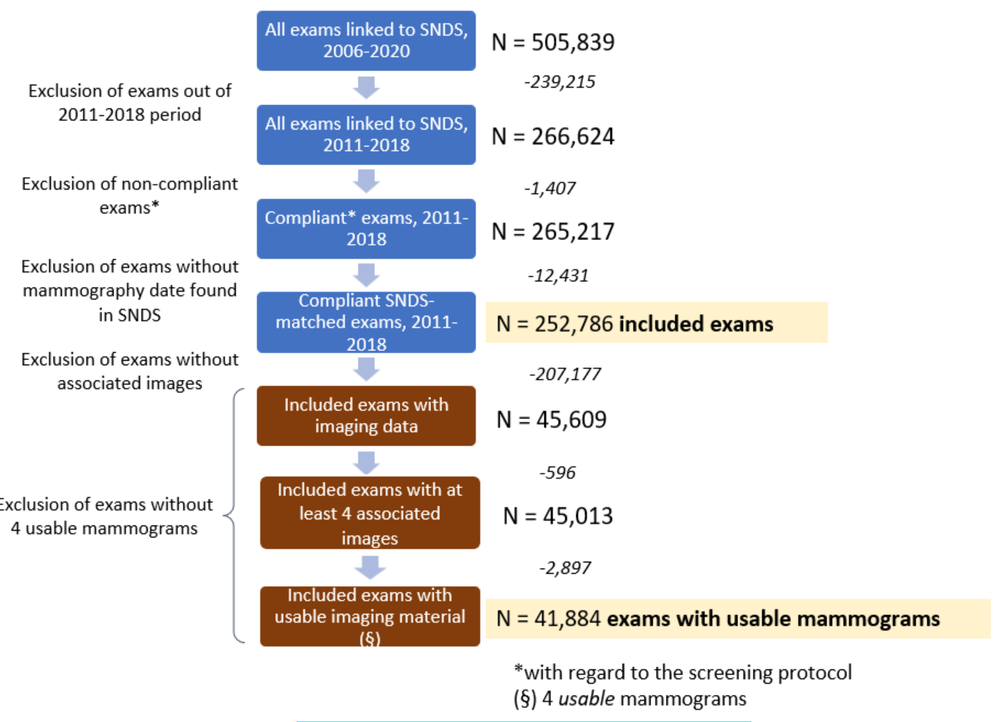


Fig 1. Critères d'inclusion des mammographies

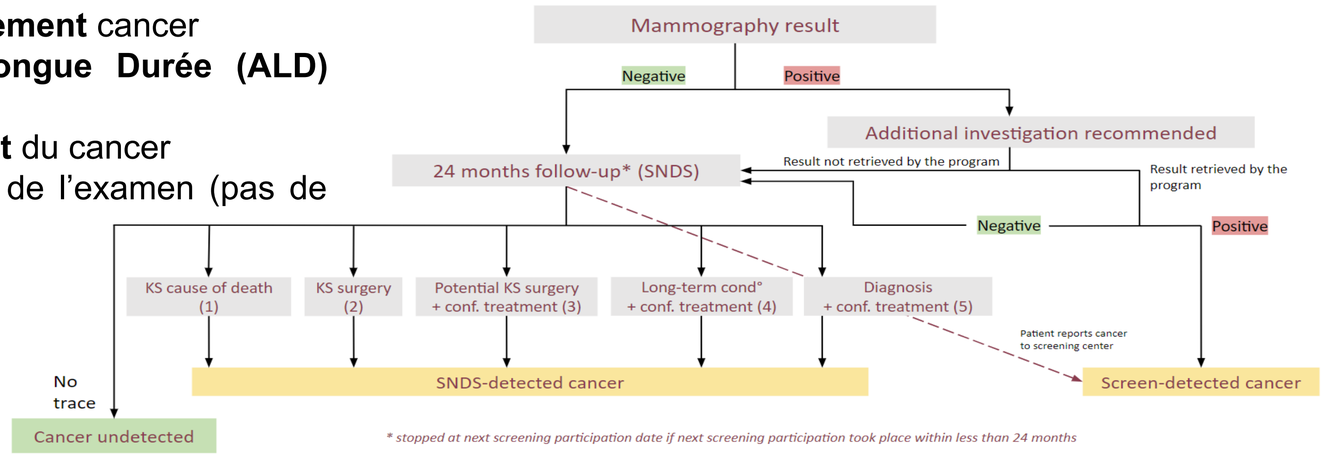


Fig 2. Suivi des femmes dans le SNDS

3. Procédure de filtrage des clichés

- Un **réseau de neurones ResNet-18** a été entraîné pour détecter la latéralité (gauche/droite) et la position (CC/MLO) sur des mammographies avec 4 clichés valides (Fig. 3).
- La *réponse softmax* (en anglais, *Softmax Response*; SR) et l'entropie de prédiction du modèle ont servi d'**indicateurs de qualité**, croisés avec des seuils spécifiques (Fig. 4). **Tous les clichés de chaque examen ont ainsi été filtrés** (Fig. 5).

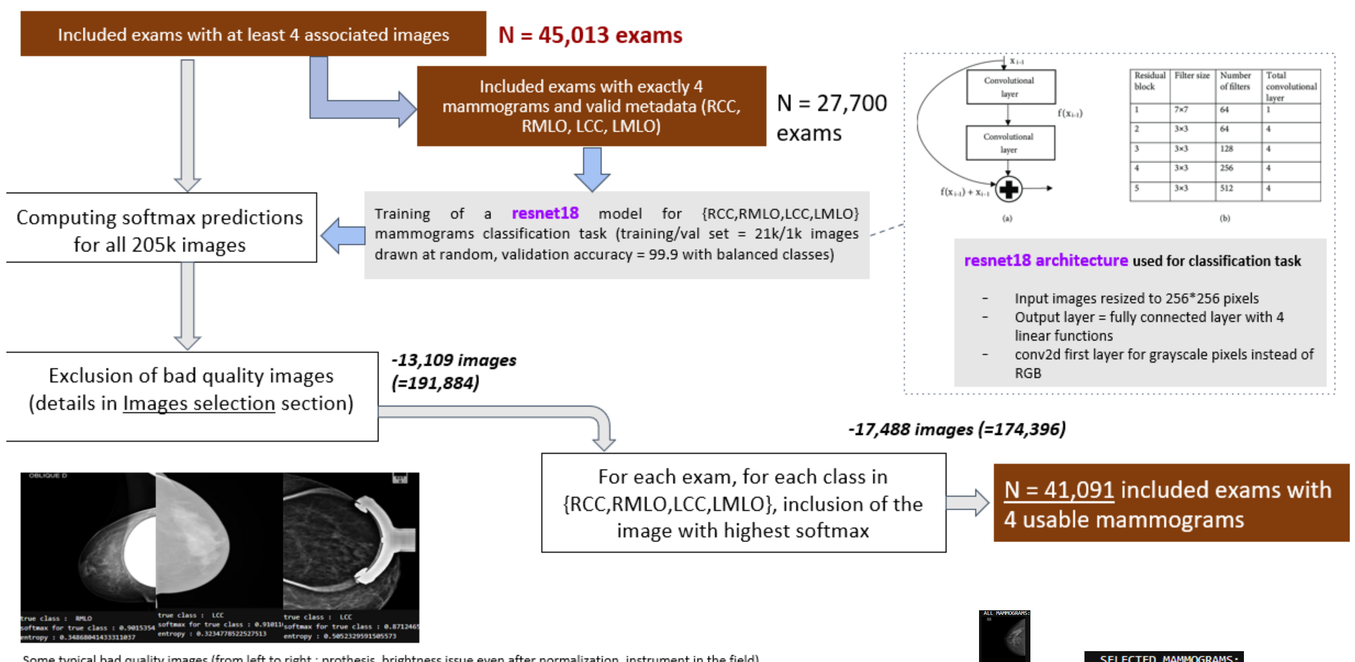


Fig 3. Procédure détaillée de filtrage des clichés inutilisables

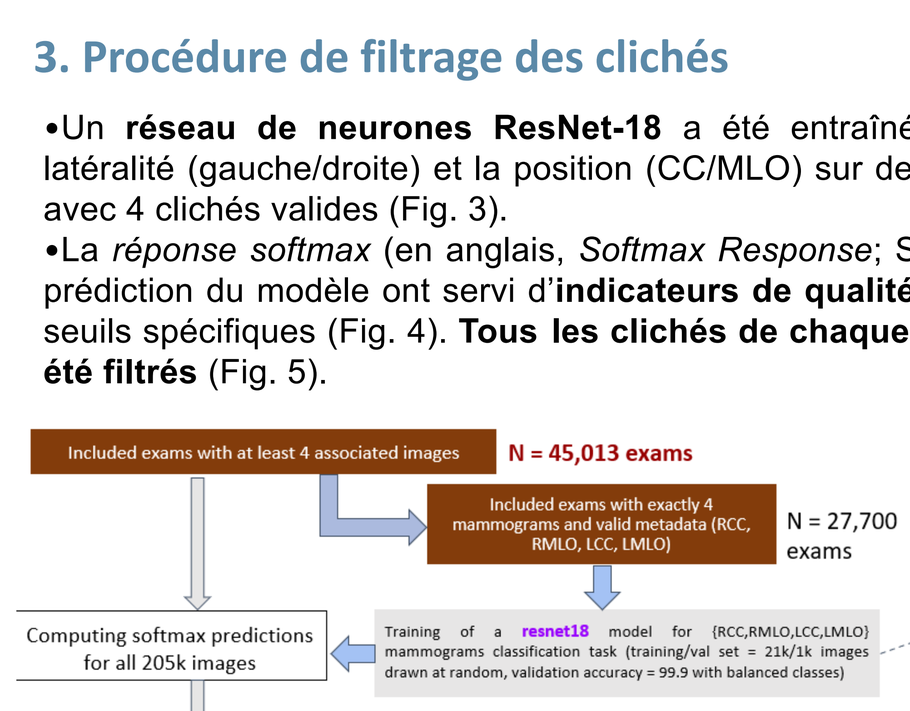


Fig 4. Méthode de choix des seuils de qualité des mammographies (SR et entropie)

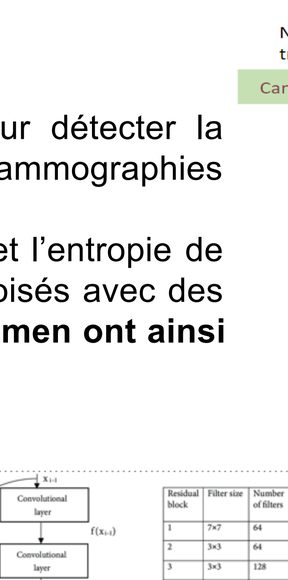


Fig 5. Un examen de dépistage, avant et après filtrage

4. Architectures utilisées pour détecter les CS

1. Réseau de neurones **NYU** (Wu et al., 2020) avec poids d'origine des auteurs.
2. Réseau de neurones **GMIC** (Shen et al., 2021) : poids d'origine + réentraînement sur 1 200 images cancer / 1 200 non-cancer (non-cancer rééchantillonnées à chaque nouvelle séance d'entraînement du modèle).

Des **variantes** sont également entraînées pour des raisons spécifiques:
– GMIC avec étiquettes CS raffinées au niveau du sein (selon ACR BIRADS) pour meilleure adéquation avec l'architecture
– GMIC modifié pour prendre 4 vues en entrée (LCC, LMLO, RCC, RMLO) pour meilleure adéquation avec labels CS au niveau de l'examen

Architecture GMIC ; Shen et al. (2021)

5. Résultats de performances

Jeu de données de test utilisé:

- Approximation du jeu utilisé par les auteur Shen et al. [4]
- 2,426 mammographies:
  - 4.3% cancers (N=100) identifiés sous 120 jours.
  - 95.7% sans cancer (N=2,326) avec 2 ans de suivi SNDS

Model	Training loss	AUC	PRAUC	F1	PPV @ SE 80%	Test time per exam* (4 images)
NYU (images+heatmaps)	-	0.60	0.1	0.12	4%	~5 minutes
GMIC (N_ensemble = 5)	-	0.76	0.29	0.34	9%	2-3 seconds
GMIC 3 classes (N_ensemble = 5)	Custom loss	0.73	0.15	0.25	8%	2-3 seconds
GMIC (N_ensemble = 2)	BCE	0.87	0.375	0.43	12%	2-3 seconds
GMIC with birads refined breast-level labels (N_ensemble = 2)	BCE	0.86	0.36	0.41	11%	2-3 seconds
GMIC EXAM (N_ensemble = 6)	BCE	0.86	0.254	0.37	12%	4-5 seconds
GMIC (N_ensemble = 2)	Custom loss	**0.92**	**0.43**	0.44	19%	2-3 seconds
GMIC EXAM (N_ensemble = 6)	Custom loss	0.92	0.353	0.42	17%	4-5 seconds
GMIC + SNDS risk module	Custom loss					

Retrained

Best results for cancer <120 days detection task

Conclusion

- **Le modèle le plus performant est GMIC réentraîné avec les labels CS au niveau de l'examen** → AUC = 0.92, PRAUC = 0.43 vs. performances publiées par Shen et al.: AUC = 0.89, PRAUC = 0.39.
- La **VPP du modèle est 19% à 80% de sensibilité** (avec une prévalence de 4.3%)

- Intégrer ce modèle rétrospectivement au DOCS permettra d'évaluer son potentiel pour la réduction de charge et l'amélioration de détection des CS.

References

[1] International Agency for Research on Cancer, WHO. Global cancer observatory. https://gco.iarc.fr/. Accessed: 2024-01-29.

[2] *Ligue contre le Cancer* website, Breast cancer section. https://www.ligue-cancer.net/questce-que-le-cancer/les-types-de-cancer/cancer-du-sein, Accessed: 2024/05/07.

[3] Wu N, Phang J, Park J, Shen Y, Huang Z, Zorin M, Jastrzebski S, Fevry T, Katsnelson J, Kim E, Wolfson S, Parikh U, Gaddam S, Lin LLY, Ho K, Weinstein JD, Reig B, Gao Y, Toth H, Pysarenko K, Lewin A, Lee J, Airola K, Mema E, Chung S, Hwang E, Samreen N, Kim SG, Heacock L, Moy L, Cho K, Geras KJ. *Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening*. IEEE Trans Med Imaging. 2020 Apr;39(4):1184-1194. doi: 10.1109/TMI.2019.2945514. Epub 2019 Oct 7. PMID: 31603772; PMCID: PMC7427471.

[4] Y. Shen, N. Wu, J. Phang, J. Park, K. Liu, S. Tyagi, L. Heacock, S.G. Kim, L. Moy, K. Cho and K.J. Geras. *An interpretable classifier for high-resolution breast cancer screening images utilizing weakly supervised localization*. Medical Image Analysis. Volume 68. 2021.

[5] Adam Yala et al. *Toward robust mammography-based models for breast cancer risk*.Science Translational Medicine. 13,eaba4373(2021).DOI:10.1126/scitranslmed.aba4373