

PROJET TECHNIQUE DE FIN DE E3

2^e SEMESTRE 2024/2025

TITRE DU PROJET : Prédiction de la survenue du cancer du sein par combinaison de méthodes d'IA adaptées à la fusion de données multimodales

MOTS CLES : Traitement d'images, données multimodales, Intelligence Artificielle, machine learning, deep-learning, cancer du sein, application web

DESCRIPTION DU PROJET :

Contexte :

Le cancer du sein est aujourd'hui l'un des cancers les plus courants chez les femmes à travers le monde, représentant une cause majeure de morbidité et de mortalité. Malgré les avancées en matière de dépistage et de traitement, la complexité biologique de la maladie et la variabilité des réponses aux thérapies continuent de poser des défis significatifs. Les approches modernes, incluant l'intelligence artificielle et l'analyse de données multimodales, offrent de nouvelles opportunités pour améliorer la prédiction de la survenue de cette maladie.

Description :

L'utilisation de l'IA dans la prédiction du risque de cancer du sein a connu une croissance exponentielle ces dernières années, grâce à la disponibilité croissante de données multimodales et au développement de nouvelles techniques d'apprentissage automatique.

Les données multimodales sont diverses et variées. On peut citer des données de bio marqueurs et autres récoltées (mammographie, IRM, échographie, bio marqueurs, hérédité, âge, tabagisme, rythme sommeil, ...).

Les approches hybrides et multimodales combinées à l'IA [1] permettent de mieux exploiter la richesse des informations disponibles et offrent des perspectives prometteuses dans le diagnostic du cancer du sein. Par exemple, Shen et al. [2] ont utilisé un modèle CNN pour fusionner des images IRM et mammographiques afin d'améliorer la détection des lésions. Une autre étude [3] a proposé une architecture qui combine un **CNN** pour l'extraction des caractéristiques visuelles à partir des images, avec des algorithmes de Random Forest pour la fusion de données cliniques, telles que l'hérédité ou les habitudes de vie (tabagisme, sommeil). Cette combinaison permet une meilleure prise en compte des corrélations complexes entre les données d'images et les informations non visuelles. Des modèles plus avancés, tels que les réseaux à mécanisme d'attention (attention mechanism), ont été récemment proposés pour pondérer l'importance relative des différentes modalités de données [4].

Travail à réaliser :

- faire une recherche bibliographique sur les méthodes de ML et DL de prédiction
- Programmation basée sur le machine learning et deep learning
- Recherche de base de données multimodale
- Prédiction de la survenue du cancer du sein dans x années (ou pas)
- Concevoir une application web qui permet de tester votre modèle de prédiction.

Bibliographie :

[1] Antropova, N., Huynh, B. Q., & Giger, M. L.. A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. *Medical Physics*, 44(10), 5162-5171 (2018).

[2] Shen, L., Margolies, L., Rothstein, J. H., Fluder, E., McBride, R., & Sieh, W. Deep learning to improve breast cancer detection on screening mammography. *Scientific Reports*, 9(1), 1-12 (2019).

[3] Zhao, W., Zhang, L., & Zhu, X. A multimodal deep learning model for early diagnosis of breast cancer using mammography and clinical data. *IEEE Access*, 8, 200012-200021 (2020).

[4] Huang, S., Yang, J., Fong, S., & Zhao, Q. Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges. *Cancer Letters*, 500, 31-41 (2021).

OUTILS MATÉRIELS / LOGICIELS SUPPORT :

- Langages & APIs : Python, TensorFlow/Keras
- Bibliothèques : OpenCV, NumPy, Matplotlib, Scikit-learn
- Outils de développement : Jupyter Notebook, Google Colab, GitHub

ENCADREMENT :

Imen KACHOURI : imen.kachouri@esiee.fr