

# DIAGNOSTIC DU CANCER DU SEIN AVEC L'ALGORITHME DES K PLUS PROCHES VOISINS (KNN).

Samira Mohammadi

Département de mathématiques, informatique et génie, Université du Québec à Rimouski

## Problématique

Le cancer du sein est un cancer répandu chez les femmes dans le monde entier, et la détection précoce du cancer du sein est cruciale pour un traitement réussi. Les méthodes de diagnostic assistées par ordinateur peuvent aider à détecter et diagnostiquer le cancer du sein [2]. Dans cette étude, nous proposons une méthode de diagnostic du cancer du sein en utilisant l'algorithme des K-Nearest Neighbors (KNN) [1], qui est une méthode de classification largement utilisée en apprentissage automatique. L'algorithme utilise un ensemble de données de la méthode de séparation de motifs à surfaces multiples pour le diagnostic médical appliqué à la cytologie du sein[4]. Plus précisément, l'algorithme vise à prédire si un échantillon de cytologie du sein donné est bénin ou malin en fonction des caractéristiques extraites.

## Objectifs

Les objectifs globaux de ce programme sont d'utiliser l'algorithme KNN pour diagnostiquer efficacement et précisément le cancer du sein. Cela sera réalisé en prenant un ensemble de données d'échantillons de sein en entrée et en formant un modèle pour classer les échantillons comme malins ou bénins. Le modèle sera ensuite évalué en utilisant un ensemble de données séparé d'échantillons de sein. Le programme vise à développer un outil de diagnostic fiable et efficace qui peut aider les professionnels de la santé à identifier le cancer du sein à un stade précoce. Ainsi, le programme espère améliorer le diagnostic et le traitement, en particulier dans les cas où les méthodes de diagnostic traditionnelles ne sont pas faisables.

- Utiliser l'algorithme des K plus proches voisins (KNN) pour diagnostiquer efficacement et avec précision le cancer du sein.
- Prendre un ensemble de données d'échantillons de sein en entrée et entraîner un modèle pour classer les échantillons comme malins (M) ou bénins (B).
- Évaluer le modèle en utilisant un ensemble de données séparé d'échantillons de sein pour tester sa précision.

## Méthodologie suivie

L'algorithme KNN est implémenté en utilisant le langage de programmation C Sharp . L'algorithme prend un ensemble de données d'entraînement en entrée et l'utilise pour prédire les étiquettes d'un ensemble de données de test. La distance euclidienne est utilisée comme mesure de distance pour calculer la similarité entre les échantillons. L'algorithme inclut également une matrice de confusion pour évaluer la performance du modèle.

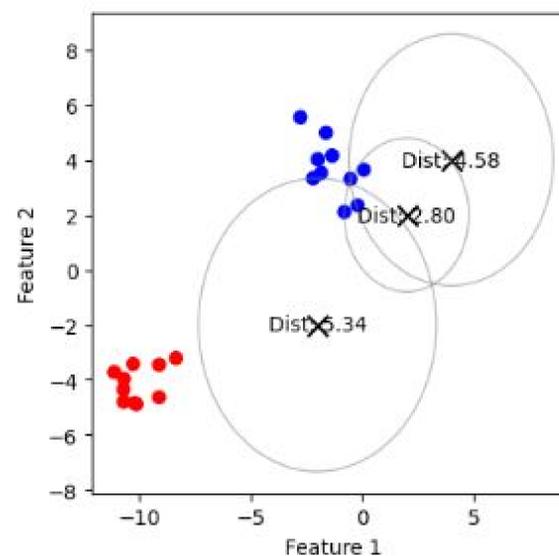


FIGURE 1 – Titre de la figure.

## Évaluation

La performance de l'algorithme est évaluée en utilisant un ensemble de données de test. L'exactitude de l'algorithme est calculée comme le pourcentage d'échantillons correctement classés. La matrice de confusion est utilisée pour calculer les taux de vrais positifs (TP), de faux positifs (FP), de vrais négatifs (TN) et de faux négatifs (FN). L'impact du changement de la valeur de K sur la performance de l'algorithme est également analysé.

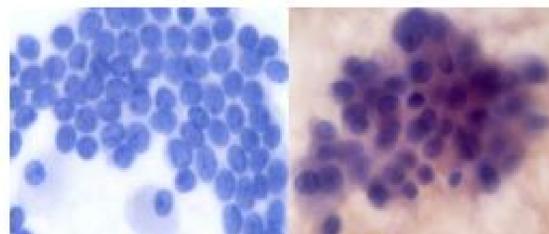


FIGURE 2 – Image cytologique d'échantillons de FNAC du sein (a) Cas bénin (b) Cas malin[3]

## Résultats

Le taux de précision du modèle KNN a été évalué pour des valeurs de K allant de 1 à 15, avec K=9 atteignant la précision la plus élevée de 90,71

K	TP	TN	FP	FN	Performance
3	50	112	3	18	88.52
5	52	112	3	16	89.62
7	53	112	3	15	89.62
9	54	112	3	14	90.71
11	53	111	4	15	89.62
13	53	110	5	15	89.07
15	53	110	5	15	89.07

FIGURE 3 – Les valeurs de TP, TN, FP, FN sont basées sur différentes valeurs de k.

## Conclusion

KNN est une méthode prometteuse pour le diagnostic du cancer du sein avec une grande précision sur les données fournies, mais la performance dépend fortement de la valeur de K. Des études supplémentaires sont nécessaires pour optimiser cette valeur et évaluer la performance de l'algorithme sur des ensembles de données plus vastes.

## Références

- [1] Abdulsalam Alarabeyyat, Mohannad Alhanahnah, et al. Breast cancer detection using k-nearest neighbor machine learning algorithm. In *2016 9th International Conference on Developments in eSystems Engineering (DeSE)*, pages 35–39. IEEE, 2016.
- [2] Afsaneh Jalalian, Syamsiah BT Mashohor, Hajjah Rozi Mahmud, M Iqbal B Saripan, Abdul Rahman B Ramli, and Babak Karasfi. Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound : a review. *Clinical imaging*, 37(3) :420–426, 2013.
- [3] S Issac Niwas, P Palanisamy, and K Sujathan. Complex wavelet based texture features of cancer cytology images. In *2010 5th International Conference on Industrial and Information Systems*, pages 348–353. IEEE, 2010.
- [4] William H Wolberg and Olvi L Mangasarian. Multisurface method of pattern separation for medical diagnosis applied to breast cytology. *Proceedings of the national academy of sciences*, 87(23) :9193–9196, 1990.