



Diagnostic du cancer du sein à l'aide de l'algorithme des k-plus proches voisins

Réalisé par : Samira Mohammadi

Encadré par: Pr. Yacine Yaddaden



Introduction

- Le cancer du sein est un cancer répandu chez les femmes dans le monde entier, et une détection précoce est cruciale pour un traitement réussi[1].
- Notre étude propose une méthode de diagnostic du cancer du sein en utilisant l'algorithme des k plus proches voisins (KNN)[2], une méthode de classification d'apprentissage automatique largement utilisée. Nous visons à utiliser l'algorithme KNN pour diagnostiquer efficacement et avec précision le cancer du sein, et à développer un outil de diagnostic fiable pour aider les professionnels de la santé à détecter le cancer du sein tôt.



Objectifs

- 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 classifier les échantillons en malins ou bénins.
- Évaluer le modèle en utilisant un ensemble de données distinct d'échantillons de sein pour tester sa précision.

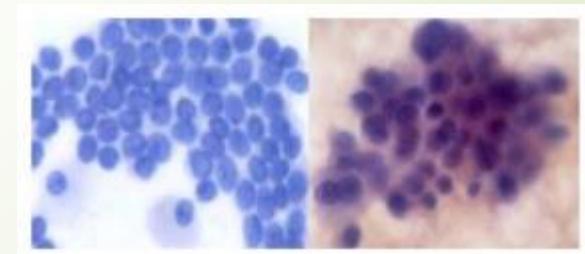


Méthodologie

- L'algorithme des K plus proches voisins 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 métrique de distance pour calculer la similarité entre les échantillons.
- L'algorithme inclut également une matrice de confusion pour évaluer les performances du modèle

Ensemble de données et fonctionnalités

- Le jeu de données contient 7 attributs, dont 6 mesures de la tumeur du sein et une étiquette de classe. Les mesures de la tumeur comprennent le rayon, la surface, le périmètre, les points concaves et d'autres caractéristiques du tissu.
- La colonne de classe indique si la tumeur est bénigne ou maligne, avec des étiquettes "B" pour bénigne et "M" pour maligne.
- Les données ont été collectées à partir de scans d'imagerie médicale de tumeurs mammaires, et sont souvent utilisées dans la recherche sur le cancer pour entraîner des modèles de classification afin d'aider les professionnels de la santé à diagnostiquer le cancer du sein.



[3]



Évaluation

- ▶ La performance de l'algorithme est évaluée en utilisant un ensemble de données de test La précision 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 (VP), de faux positifs (FP), de vrais négatifs (VN) et de faux négatifs (FN)
- ▶ L'impact du changement de la valeur de K sur la performance de l'algorithme est également analysé



Résultats

- La précision du modèle KNN a été évaluée pour des valeurs de K allant de 1 à 15, avec $K = 9$ atteignant la plus haute précision de 90,71% .
- La figure montre les valeurs de VP, VN, FP et FN en fonction des différentes valeurs de K Diapositive 8: Conclusion et travaux futurs.



Conclusion et travaux futurs

- L'algorithme KNN est une méthode prometteuse pour le diagnostic du cancer du sein. L'algorithme a atteint un taux de précision élevé en utilisant l'ensemble de données fourni. Cependant, la performance de l'algorithme est sensible à la valeur de K , et des études supplémentaires sont nécessaires pour optimiser la valeur de K et évaluer la performance de l'algorithme sur de plus grands ensembles de données.
- Les travaux futurs comprennent le test de l'algorithme sur différents ensembles de données et la comparaison de sa performance avec d'autres algorithmes d'apprentissage automatique.

References

- ▶ [1] 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
- ▶ [2] 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
- ▶ [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