**Projet Vital Stats – Cancer du Sein**

**Le cancer du sein en chiffres :**

* L’incidence (taux standardisé monde) du cancer du sein chez la femme a augmenté entre 1990 et 2023 (+0,9 % par an en moyenne, passant de 72,8 nouveaux cas de cancer du sein pour 100 000 femmes à 99,2).
* En revanche, la mortalité (taux standardisé monde) a diminué (-1,3 % par an en moyenne, passant de 20,2 décès par cancer du sein pour 100 000 femmes en 1990 à 14,0 en 2018)

Le cancer du sein se situe au **1er rang des cancers incidents chez la femme**, nettement devant le cancer du côlon-rectum et le cancer du poumon. C’est aussi celui qui cause le plus grand nombre de décès chez la femme, avec **14 % des décès féminins par cancer en 2018**. Il représente par ailleurs près de 8 % de l’ensemble des décès par cancer, tous sexes confondus (le cancer du sein chez l’homme ne représente qu’environ 1 % des cas mais il est souvent de mauvais pronostic).

**Source**: Institut National du Cancer (INCa)

**Facteurs de risque :**

* **Non modifiables :** l’âge, antécédents personnels et familiaux, puberté précoce, ménopause tardive
* **Modifiables :** nulliparité ou première grossesse tardive, traitement hormonal de la ménopause pendant plus de 5 ans, consommation d’alcool, surpoids / obésité après la ménopause, tabagisme, manque d’acitivité physique

**Sources :** Vidal, INCa, Institut Curie

**Symptômes :**

|  |  |
| --- | --- |
| Chez la femme | Chez l’homme |
| Une masse ou grosseur dans le sein, généralement indolore | Masse indolore au niveau de la poitrine |
| Changements de la peau du sein ou du mamelon (rougeur, rétraction, aspect peau d’orange) | Douleur ou enflure de la poitrine |
| Modification de la taille ou forme du sein | Augmentation du volume des ganglions lymphatiques |
| Ecoulement au niveau du mamelon, pouvant être teinté de sang | Ecoulement ou saignement du mamelon |
| Rétractation ou inversion du mamelon | Rétractation récente du mamelon |
| Douleur ou sensibilité du sein ou du mamelon | Croûte, desquamation ou ulcération de la peau / mamelon |
|  | Symptômes aréolaires |

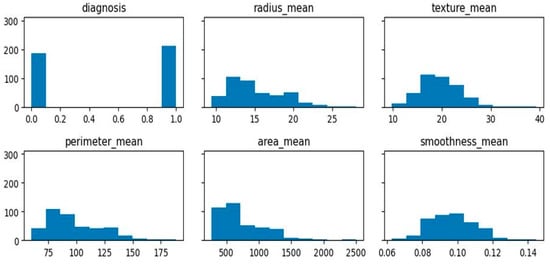
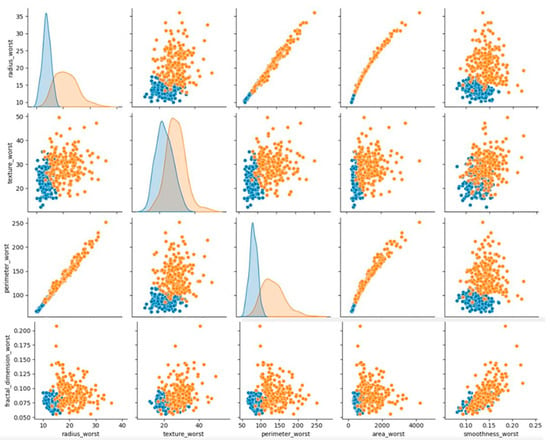
**Sources :** INCa, Fondation Cancer du Sein du Québec

**Marqueurs morphologiques du Cancer du Sein : pré-processing et visualisation**

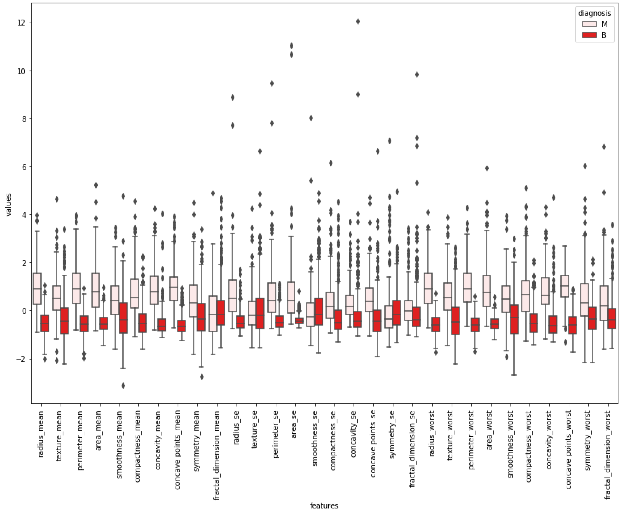
|  |  |  |
| --- | --- | --- |
| Marqueur | Définition | Valeurs |
| Radius Mean | Moyenne des distances du centre aux points sur le périmètre | Pas de seuil spécifique  Cela donne une indication sur la taille et la forme générale de la tumeur |
| Area Mean | Moyenne de la surface de la tumeur | Pas de valeur définie  Une aire plus grande peut indiquer une tumeur de taille plus importante |
| Compactness Mean | Moyenne de (périmètre^2 / surface – 1.0) | > 0.1  Une valeur plus élevée indique une forme moins compacte et plus irrégulière |
| Concavity Mean | Moyenne de la gravité des parties concaves du contour | > 0.2/0.3  Des concavités plus prononcées peuvent être un signe de malignité |
| Concave Points Mean | Moyenne du nombre de parties concaves du contour | > 0.1  Plus il y a de points concaves, plus la forme est irrégulière, ce qui peut indiquer un potentiel malin |
| Area Worst | Aire la plus mauvaise (moyenne des trois plus grandes valeurs) de la tumeur | > 2000mm² |
| Compactness Worst | Compacité la plus mauvaise (moyenne des trois plus grandes valeurs) de la tumeur | > 0.5  Une compacité élevée indique une forme moins lisse et plus irrégulière, ce qui est caractéristique des tumeurs malignes |
| Concavity Worst | Concavité la plus mauvaise (moyenne des trois plus grandes valeurs) de la tumeur | > 0.5  Une concavité prononcée est un signe de malignité, les tumeurs bénignes ayant généralement un contour plus lisse |
| Area SE | Erreur standard de la surface de la tumeur | > 100mm² |
| Fractal Dimension SE | Erreur standard de « l’approximation du littoral » - 1 | > 0.15  Une dimension fractale élevée dans ces zones indique des contours très irréguliers, signe de malignité |
| Symmetry Worst | Symétrie la pire (moyenne des trois plus grandes valeurs) de la tumeur | Entre 0.2 et 0.5, > 0.5 |
| Fractal Dimension Worst | Dimension fractale la pire (moyenne des trois plus grandes valeurs) de la tumeur | > 0.01  Une erreur standard élevée suggère une grande variabilité dans les mesures de dimension fractale entre les différentes régions de la tumeur, ce qui est caractéristique des tumeurs malignes aux contours irréguliers |

Les marqueurs qui nous sont présentés dans le document d’information sont des **marqueurs morphologiques**. Ils indiquent la présence d’une tumeur et nous renseignent sur **sa taille, sa forme, sa complexité et son irrégularité**, autant d’éléments qui permettront de **caractériser son agressivité, son potentiel malin voire son potentiel métastatique**

Dans les étude que nous avons consulté, les datasets utilisés (Motamed cancer institute (ACECR), Tehran, Iran – Breast Cancer Wisconsin Data) ont fait l’objet **d’un important travail de préprocessing et de visualisation** dont voici quelques exemples :



Une image contenant texte, capture d’écran, Police, Tracé

Description générée automatiquement

**Source :** Breast Cancer Detection and Prevention Using Machine Learning by Arslan Khalid, Arif Mehmood, Amerah Alabrah,, Bader Fahad Alkhamees, Farhan Amin, Hussain AlSalman, Gyu Sang Choi

**Machine Learning pour la détection et le diagnostic**

Les analyses basées sur le Machine Learning reposent sur des **modèles de classification** divers. Parmi les plus utilisés on trouve :

|  |  |
| --- | --- |
| * Régression Logistique | * Decision Tree |
| * KNN | * Naïve Bayes |
| * SVM | * Random Forest |

Chaque modèle a été testé puis évalué sur plusieurs jeux de données (toutes issues du même dataset) Les images ci-dessous donnent un aperçu des modèles utilisés et de leurs évaluation c’est-à-dire la qualité de leurs prédictions.

Une image contenant texte, Police, capture d’écran, conception

Description générée automatiquement

Une image contenant texte, capture d’écran, ligne, Tracé

Description générée automatiquement Une image contenant texte, capture d’écran, nombre, Police

Description générée automatiquement

Ces modèles ont été utilisés pour les objectifs suivants :

* Établir un modèle adéquat en **révélant les facteurs prédictifs** des patientes atteintes d'un cancer du sein à un stade précoce dans une perspective plus large et **comparer la robustesse du modèle par des mesures de précision**
* Une comparaison et une analyse plus complètes à l'aide d'applications de visualisation de données et d'apprentissage automatique pour la **détection et la visibilité du cancer du sein afin de valider le modèle**
* Observer **quelles caractéristiques sont les plus efficaces** pour prédire le cancer du sein et **comprendre les tendances générales**
* Une **meilleure prédiction du cancer du sein grâce à des méthodes d’exploration de données**

**Sources :** Prediction of Breast Cancer using Machine Learning Approaches by Reza Rabiei,[Seyed Mohammad Ayyoubzadeh](https://pubmed.ncbi.nlm.nih.gov/?term=Ayyoubzadeh%20SM%5BAuthor%5D), [Solmaz Sohrabei](https://pubmed.ncbi.nlm.nih.gov/?term=Sohrabei%20S%5BAuthor%5D), [Marzieh Esmaeili](https://pubmed.ncbi.nlm.nih.gov/?term=Esmaeili%20M%5BAuthor%5D),and [Alireza Atashi](https://pubmed.ncbi.nlm.nih.gov/?term=Atashi%20A%5BAuthor%5D)

A Comparative Analysis of Breast Cancer Detection and Diagnosis Using Data Visualization and Machine Learning Applications by [Muhammet Fatih Ak](https://pubmed.ncbi.nlm.nih.gov/?term=Ak%20MF%5BAuthor%5D)

Application of Machine Learning Algorithms in Breast Cancer Diagnosis and Classification by [Clement G. Yedjou](https://pubmed.ncbi.nlm.nih.gov/?term=Yedjou%20CG%5BAuthor%5D), [Solange S. Tchounwou](https://pubmed.ncbi.nlm.nih.gov/?term=Tchounwou%20SS%5BAuthor%5D),[Richard A. Aló](https://pubmed.ncbi.nlm.nih.gov/?term=Al%C3%B3%20RA%5BAuthor%5D),[Rashid Elhag](https://pubmed.ncbi.nlm.nih.gov/?term=Elhag%20R%5BAuthor%5D),[BereKet Mochona](https://pubmed.ncbi.nlm.nih.gov/?term=Mochona%20B%5BAuthor%5D),and [Lekan Latinwo](https://pubmed.ncbi.nlm.nih.gov/?term=Latinwo%20L%5BAuthor%5D)

**Conclusion**

Dans la problématique qui nous est proposée pour les prochaines semaines, l’enjeu sera à travers des méthodes de pré-processing et de visualisation d’identifier les relations entre les différents marqueurs et lorsque ce sera possible d’identifier les marqueurs les plus impactant.

Dans la littérature médicale spécifique au cancer du sein, les marqueurs dont nous disposons et leurs relations permettent d’indiquer le stade, voire le grade du cancer et à partir de ces informations définir la stratégie thérapeutique adéquate

L’objectif de notre projet peut donc être de regrouper les patientes en fonction des stades ou grades de cancer.