



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE  
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE  
UNIVERSITE DES SCIENCES ET DE LA TECHNOLOGIE HOUARI BOUMEDIENE

Faculté d'Informatique

Filière  
**Informatique**

Spécialité  
**Bio-Informatique**

Thème

**Prédiction du cancer de sein avec  
l'apprentissage automatique**

Module: SYSTÈME ET PROGRAMMATION DE SCRIPTS

Présenté par :

-SMAIL Mehrez

Année Universitaire : 2022-2023

# Table des matières

Introduction générale.....	4
Chapitre 1 Le cancer du sein.....	5
1.1. Introduction.....	5
1.2. Le cancer du sein.....	5
1.2.1. Types de cancer de sien.....	5
1.2.2. Anatomie de Sein.....	5
1.2.3. Les symptômes et facteurs de risques de cancer du sein.....	6
1.2.4. Les traitements médicaux du cancer du sein.....	7
1.3. Statistiques de cancer du sein en Algérie.....	7
1.4. Conclusion.....	7
Chapitre 2 L'apprentissage automatique.....	8
2.1. Introduction.....	8
2.2. Définition.....	8
2.3. Types de l'apprentissage automatique.....	9
2.3.1. Apprentissage supervisé.....	9
2.3.2. Apprentissage non supervisé.....	9
2.4. Les algorithmes de l'apprentissage automatique.....	10
2.4.1. La méthode des K plus proches voisins.....	10
2.4.2. la méthode de l'Arbre de décision (Decision Trees).....	10
2.4.3. Forêts aléatoires (Random Forest).....	11
2.4.4. Machines à vecteur de support (SVM) .....	11
2.5. Conclusion.....	11
Chapitre 3 L'approche Proposée.....	12
3.1. Introduction.....	12
3.2. Définition de l'approche proposée Random Forest.....	12
3.3. Principe de fonctionnement Random Forest.....	12
3.4. Algorithme de Random Forest.....	13
3.5. Applications de la Random Forest.....	14
3.6. Conclusion.....	14
Chapitre 4 Tests et résultats expérimentaux.....	15
4.1. Présentation de l'environnement de test.....	15
4.1.1. Collection de test.....	15
4.1.2. Outils et mesures d'évaluation.....	16
4.2. Réglages des paramètres de RandomForest.....	17
4.3. Résultats expérimentaux.....	17
4.4. L' interfaces d'implémentation.....	18
4.4. Conclusion.....	18
Conclusion générale.....	19

# Index des figures

Figure 01: La structure de sein.....	5
Figure 02: L'apprentissage automatique.....	7
Figure 03: types d'apprentissage automatique.....	8
Figure 04: Le fonctionnement de K-NN.....	9
Figure 05: fonctionnement Random Forest.....	12
Figure 06: Collection de test.....	14
Figure 07: les résultats de Score en fonction de nombre d'estimation.....	16
Figure 08: La courbe de caractéristique de fonctionnement du récepteur.....	16
Figure 09: L' interfaces d'implémentation.....	17

# Introduction générale

L'utilisation de l'intelligence artificielle, de la robotique et des technologies connexes dans le domaine de la santé, annonce une révolution majeure dans le domaine de la médecine. Ces nouvelles technologies interviennent dans l'aide au diagnostic « intelligence artificielle », aux actes techniques « robotique chirurgicale », à la consultation « télémédecine », aussi la révolution des « Micro-Robots » circulants injectés dans le sang et capables d'atteindre les endroits les plus reculés du corps humain. Cette évolution indique un brillant avenir pour la médecine face à diverses maladies incurables.

La santé est un état de complet bien-être physique, mental et social, et ne consiste pas seulement en une absence de maladie ou d'infirmité. Une meilleure santé est essentielle au bonheur et au bien-être. Les êtres humains essaient toujours d'être en bonne santé, car il est important de permettre à tous de vivre en bonne santé et de promouvoir le bien-être de tous à tout âge pour construire des sociétés prospères. Pour faire ça il est essentiel de prévenir les maladies, d'autre part le cancer du sein est l'une des maladies qui a suscité un grand intérêt chez les chercheurs, en particulier dans le domaine de l'intelligence artificielle. Parmi les nouvelles méthodes de prévenir la maladie c'est l'utilisation de l'apprentissage automatique ou l'intelligence artificielle car ce n'est pas pratique d'analyser des données énormes manuellement. Le mémoire est organisé comme suit. Dans le chapitre 1, nous présentons le cancer de sein et quelques Types de cancer de sein, les symptômes et facteurs de risques de cancer du sein, Les traitements médicaux du cancer du sein. Dans le chapitre 2, nous passons brièvement en revue l'apprentissage automatique et les types de l'apprentissage automatique. Dans le chapitre 3, nous présentons notre approche nommée Random Forest qui consiste à prédire la maladie de cancer de sein. Les résultats expérimentaux sont présentés dans le chapitre 4. Pour finir, nous tirons des conclusions et présentons des perspectives.

# Chapitre 1 Le cancer du sein

## 1.1. Introduction

La santé est un aspect important de la vie de chacun. Parfois une personne a Les maladies courantes peuvent être guéries, mais parfois elle souffre de maladies malignes, y compris Le taux de guérison est aussi bas que le cancer.

Dans ce chapitre, nous aborderons le cancer du sein, ses types, ses symptômes et ses méthodes de traitement. Ensuite, nous allons présenter également les statistiques relatives à cette maladie en Algérie.

## 1.2. Le cancer du sein

Le cancer du sein est l'un des plus célèbres cancers dans le monde et le plus diagnostiqué chez les femmes. Les statistiques indiquent que le nombre de personnes touchées a légèrement augmenté au cours des trois dernières décennies. D'autre part, le taux de mortalité a diminué régulièrement au cours de la même période, grâce aux progrès du dépistage, du diagnostic et du traitement, mais cette diminution ne s'applique pas aux pays développés en raison de la longévité (Courbiere & Carcopino, 2016), ce cancer prend naissance dans les cellules mammaires, où une tumeur cancéreuse qui est un groupe de cellules cancéreuses envahit et détruit les tissus voisins. Il peut également se déplacer et se propager à d'autres parties du corps. Les cellules mammaires changent parfois, provoquant leur croissance ou leur comportement anormal. Ces changements peuvent entraîner des affections mammaires bénignes non cancéreuses telles que l'hyperplasie atypique et les kystes (Kelley, 2020).

Mentionnons que les hommes peuvent aussi en être touchés, ils représentent 1 % de l'ensemble des cas. (Lahmer, 2018).

### 1.2.1. Types de cancer de sien

Il est important d'identifier le type de cancer du sein afin de planifier le traitement et d'augmenter les chances de succès face à la maladie.

Il existe cinq types du cancer de sein :

- Carcinome canalaire.
- Carcinome lobulaire.
- Le cancer inflammatoire du sein.
- La maladie de Paget du sein .

### 1.2.2. Anatomie de Sein

La fonction biologique du sein est de produire du lait afin de nourrir un nouveau-né, le sein se compose de graisse, de glandes et de canaux . Les glandes agencées en lobules produisent le lait et les canaux servent à transporter le lait jusqu'au mamelon, les tissus mammaires sont influencés par des hormones produites par les femmes en quantité variable tout au long de leur vie. Ces hormones sont l'œstrogène et la progestérone (Institut National Du Cancer).

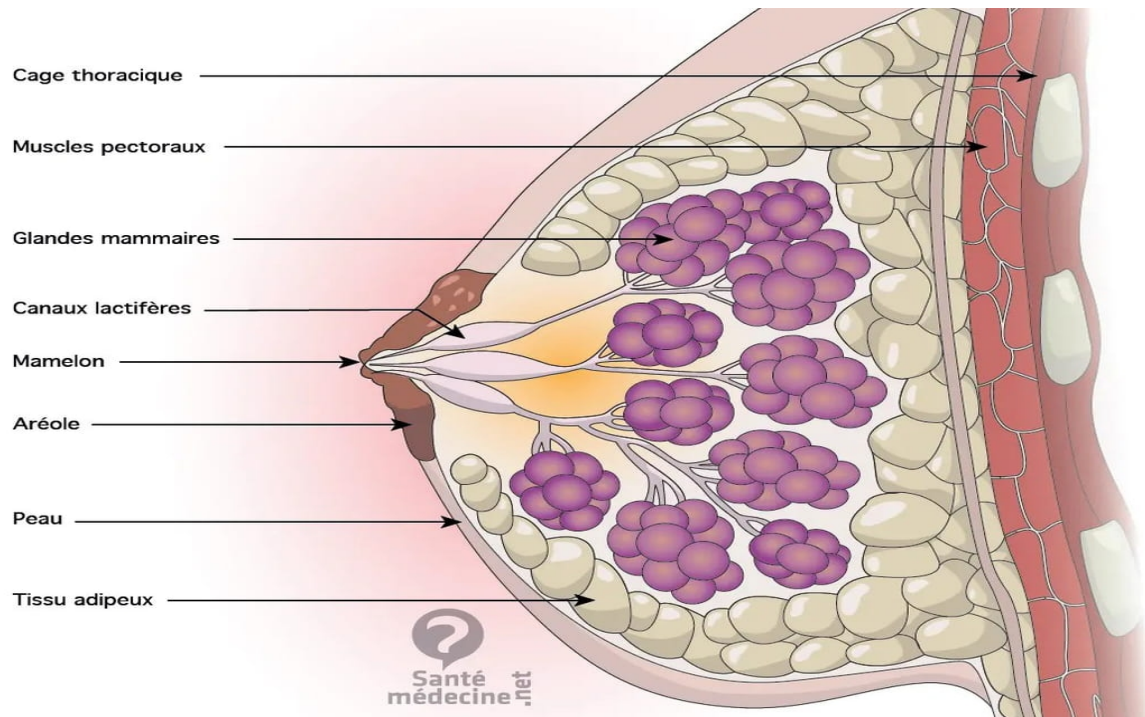


Figure 01: La structure de sein

### 1.2.3. Les symptômes et facteurs de risques de cancer du sein

Le cancer du sein peut ne causer aucun symptôme aux premiers stades de la maladie, mais on peut dénombrer plusieurs signes, qui conduisent directement à des cas sensibles de la maladie.

D'après (Institut National Du Cancer, 2018), les symptômes du cancer du sein sont les suivants :

- Une bosse au sein.
- Des écoulements spontanés.
- Une rétraction du mamelon.
- Un changement d'apparence de la peau d'un sein.
- Un changement de la forme d'un sein.

#### 1.2.4. Les traitements médicaux du cancer du sein

Les tests effectués influenceront le choix du traitement. Différents types de traitements sont utilisés pour traiter le cancer du sein : chirurgie, radiothérapie, hormonothérapie, chimiothérapie et thérapie ciblée.

Le traitement dépend du type du cancer et de son stade d'évolution, la chirurgie fait partie intégrante du traitement de la très grande majorité des cancers du sein.

### 1.3. Statistiques de cancer du sein en Algérie

Selon les statistiques de l'INSP pour 2019, le cancer du sein arrive en tête de liste des cancers prévalents en Algérie, avec plus de 12 000 nouveaux cas sur un total de 50 000 cas de divers types. Le Plan national de lutte contre le cancer 2015-2019 clarifie les types de cancer les plus courants en Algérie, avec le cancer du sein en tête.

Les cas de cancer du sein détectés atteignaient 50 % et 30 % au troisième stade de la maladie, tandis que les cas métastatiques passaient de 50 % à 20 %.

### 1.4. Conclusion

L'évolution de la médecine a contribué pour rendre les traitements de plus en plus efficaces, ainsi l'intégration de l'informatique dans le domaine médical a contribué de créer des méthodes de prédiction de cette maladie. La détection précoce peut permettre de diminuer la durée du traitement et de la convalescence, plus le cancer est détecté rapidement, meilleures sont les chances de survie.

Le prochain chapitre sera consacré à l'apprentissage automatique et leurs différentes approches. Ensuite, nous allons exposer quelques travaux de la littérature qui ont utilisé les algorithmes de l'apprentissage en profondeur pour la prédiction le cancer du sein.

# Chapitre 2

## L'apprentissage automatique

### 2.1. Introduction

L'intelligence artificielle (IA) est un vaste domaine, et c'est un processus d'imitation de l'intelligence humaine qui repose sur la création et l'application d'algorithmes exécutés dans un environnement informatique dynamique, l'apprentissage automatique (Machine Learning) est une sous-catégorie de l'intelligence artificielle. Son but est de permettre aux ordinateurs de penser et d'agir comme des êtres humains. L'intelligence artificielle a besoin d'une quantité de données et d'une capacité de traitement élevée. L'intelligence artificielle présente de nombreux avantages, concernant essentiellement le monde du travail, ses hautes performances et ses perspectives économiques intéressantes. Dans ce chapitre, nous allons présenter d'abord quelques notions concernant l'apprentissage automatique (Machine Learning).

### 2.2. Définition

L'apprentissage automatique est un sous-domaine de l'intelligence artificielle (IA) qui se concentre sur la conception de systèmes qui apprennent ou améliorent le rendement en fonction des données qu'ils consomment.

L'intelligence artificielle est un terme général qui se rapporte aux systèmes ou aux machines qui imitent l'intelligence humaine. L'apprentissage automatique et l'intelligence artificielle sont souvent évoqués ensemble les termes sont parfois utilisés de façon interchangeable, mais ne signifient pas la même chose. Une importante distinction est que même si tout apprentissage automatique repose sur l'intelligence artificielle, cette dernière concerne bien plus que l'apprentissage automatique.

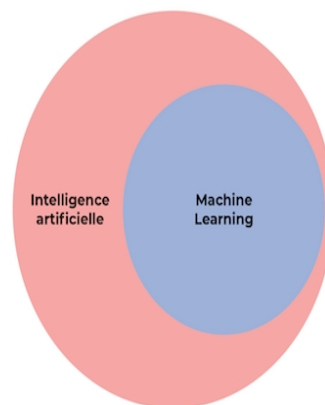


Figure 02: L'apprentissage automatique

## 2.3. Types de l'apprentissage automatique

Nous nous intéressons au premier stade du Machine Learning, qui est la classification de données, il consiste à regrouper les données en classes et il est constitué de deux principaux types :

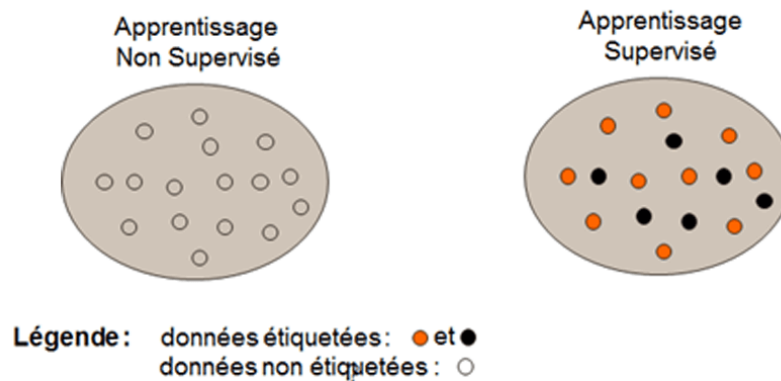


Figure 03: types d'apprentissage automatique

### 2.3.1. Apprentissage supervisé

L'apprentissage supervisé commence généralement par un ensemble de données bien défini et une certaine compréhension de la façon dont ces données sont classifiées, l'apprentissage supervisé a pour but de déceler des modèles au sein des données et de les appliquer à un processus analytique. Ces données comportent des caractéristiques associées à des libellés qui définissent leur signification.

L'apprentissage supervisé peut être regroupé en deux catégories d'algorithmes :

- Classification
- Régression

### 2.3.2. Apprentissage non supervisé

L'apprentissage non supervisé est utilisé pour tirer des conclusions à partir d'ensembles de données constitués de données d'entrée sans réponses étiquetées. Dans cette approche, il y a essentiellement deux tâches d'apprentissage, à savoir l'association et le regroupement, apriori est considéré comme le meilleur algorithme utilisé dans la règle d'association.



## 2.4. Les algorithmes de l'apprentissage automatique

### 2.4.1. La méthode des K plus proches voisins

Les 'k plus proches voisins' ou k-nearest neighbors en anglais (d'où l'appellation knn) est une méthode non paramétrique dans laquelle le modèle mémorise les observations de l'ensemble d'apprentissage pour la classification des données de l'ensemble de test, cet algorithme est qualifié comme paresseux (Lazy Learning) car il n'apprend rien pendant la phase d'entraînement. Pour prédire la classe d'une nouvelle donnée d'entrée, il va chercher ses K voisins les plus proches (en utilisant la distance euclidienne, ou autres) et choisira la classe des voisins majoritaires.

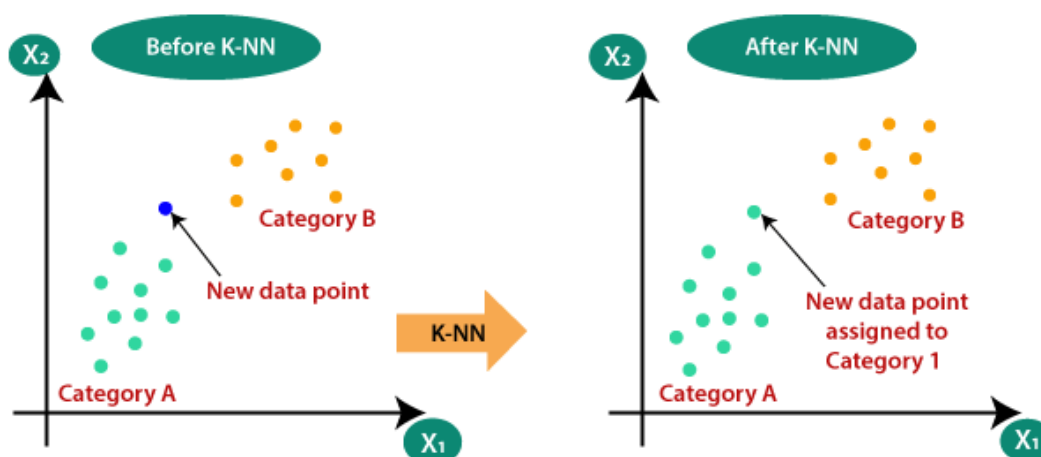


Figure 04: Le fonctionnement de K-NN

Le fonctionnement de K-NN peut être expliqué sur la base de l'algorithme ci-dessous :

- Sélectionnez le nombre K des voisins
- Calculer la distance euclidienne du **nombre K de voisins**
- Prenez les K voisins les plus proches selon la distance euclidienne
- comptez le nombre de points de données dans chaque catégorie.
- Attribuez les nouveaux points de données à la catégorie pour laquelle le nombre de voisins est maximal.
- Notre modèle est prêt.

### 2.4.2. la méthode de l'Arbre de décision (Decision Trees)

Utilisez les modèles d'arbres de décision pour développer des systèmes de classification prévoyant ou classant les observations futures à partir d'un ensemble de règles de décision. Si vos données sont divisées en classes qui vous intéressent (par exemple, prêts à haut risque, prêts à faible risque, abonnés/non-abonnés, votants/abstentionnistes, ou types de bactérie), vous pouvez les utiliser pour construire des règles qui permettront de classer les observations anciennes ou nouvelles avec une exactitude maximale. Par exemple, vous

pouvez construire un arbre qui classe le risque de crédit ou l'intention d'achat en fonction de l'âge et d'autres facteurs.

### 2.4.3. Forêts aléatoires (Random Forest)

L'algorithme des forêts aléatoires est un algorithme de classification qui réduit la variance des prévisions d'un arbre de décision seul, améliorant ainsi leurs performances. Pour cela, il combine de nombreux arbres de décisions dans une approche de type bagging, cet algorithme effectue un apprentissage en parallèle sur de multiples arbres de décision construits aléatoirement et entraînés sur des sous-ensembles de données différents. Le nombre idéal d'arbres est un paramètre important il est très variable et dépend du problème. Concrètement, chaque arbre de la forêt aléatoire est entraîné sur un sous ensemble aléatoire de données selon le principe du bagging, avec un sous ensemble aléatoire de caractéristiques variables des données selon le principe des projections aléatoires. Il a été utilisé dans de nombreuses applications, y compris grand public, comme pour la classification d'images de la caméra de console de jeu Kinect dans le but d'identifier des positions du corps.

### 2.4.4. Machines à vecteur de support (SVM)

Les SVM représentent une famille d'algorithmes de Machine Learning capables de résoudre un problème de discrimination, de régression mathématique ou de détection d'anomalie. Les premières machines à vecteurs de support ont été développées une nouvelle théorie statistique de l'apprentissage automatique connue sous le nom de la théorie de VapnikChervonenkis. Très rapidement, leurs travaux ont suscité beaucoup d'engouement grâce à leur capacité à gérer des données de grandes dimensions. De plus, les SVM ne requièrent que peu de paramètres. Dans l'immense majorité des cas, l'utilisation du SVM ne nécessite que deux hyperparamètres : le choix de la technique de régularisation et celui du noyau (Sobolev, noyaux polynomiaux...).

## 2.5. Conclusion

Dans ce chapitre nous avons défini les principaux concepts d'apprentissage automatique et ses différents types de techniques avec ses algorithmes.

Dans le chapitre suivant, nous allons présenter notre approche proposée, qui a pour rôle la prédiction du cancer de sein, basée sur l'apprentissage automatique.

# Chapitre 3 L'approche Proposée

## 3.1. Introduction

Le cancer du sein est le type de cancer le plus connu chez les femmes, la première phase du traitement est basée sur la détection précoce de la maladie. Cette phase demeure un problème important dans le monde entier, Au cours de ces dernières années divers systèmes basés sur l'apprentissage automatique ont été mis au point pour améliorer la classification et la prédiction du cancer du sein, et ils dépendent sur des données étiquetées Dans ce chapitre, nous allons présenter notre approche nommée Random Forest qui consiste à prédire la maladie en utilisant l'apprentissage automatique .

## 3.2. Définition de l'approche proposée Random Forest

Un algorithme de machine learning qui appartient à la famille des méthodes d'apprentissage ensembliste est le Random Forest , c'est un type d'algorithme d'arbre de décision qui combine les prédictions de plusieurs arbres de décision pour obtenir une prédiction plus précise et stable. il est facile à mettre en œuvre et peut être utilisée pour les tâches de classification et de régression. Contrairement à un seul arbre de décision, qui peut être sujet à le débordement et ne pas généraliser bien aux données non vues, un modèle de forêts aléatoires utilise plusieurs arbres de décision et moyenne leurs prédictions pour faire une prédiction finale. Cela rend le modèle plus robuste et moins sensible au bruit dans les données d'entraînement.

## 3.3. Principe de fonctionnement Random Forest

L'architecture Random Forest est une architecture basée sur l'apprentissage automatique supervisé ,par conséquent nous avons utilisé une base de données étiquetée pour entraînant plusieurs arbres de décision sur différents sous-ensembles des données d'entraînement, puis en moyennant les prédictions de ces arbres de décision pour obtenir une prédiction finale.

Le processus de travail peut être expliqué dans les étapes :

**Étape 1:** Sélectionnez des points de données K aléatoires dans l'ensemble d'apprentissage

**Étape 2:** Construisez les arbres de décision associés aux points de données sélectionnés

**Étape 3:** Choisissez le nombre N pour les arbres de décision que vous souhaitez créer.

**Étape 4:** Répétez les étapes 1 et 2.

**Étape 5:** Pour les nouveaux points de données, recherchez les prédictions de chaque arbre de décision et attribuez les nouveaux points de données à la catégorie qui remporte la majorité des votes.

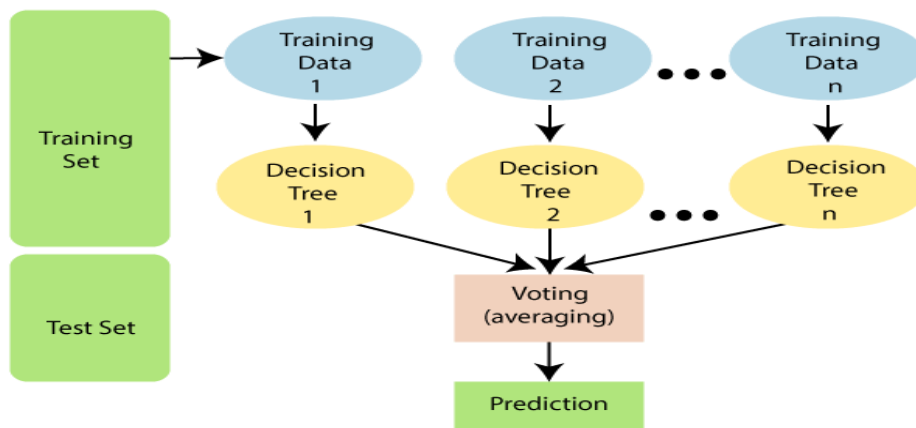


Figure 05: fonctionnement Random Forest

### 3.4. Algorithme de Random Forest

Les principales étapes de l'algorithme Random Forest pour la prédiction de cancer de sein sont présentées dans l'Algorithme 3.1

---

**Algorithme 3.1:** Pseudo code de l'algorithme Random Forest pour la prédiction de cancer de sein

---

```
import numpy as nm
```

```
import matplotlib.pyplot as plt
```

```
import pandas as pd
```

*importer des ensembles de données*

```
Data= pd.read_csv('data.csv')
```

*Extraction de variables indépendantes et dépendantes*

```
y_train= Data["diagnosis"]
```

```
x_train=Data.drop(columns=["diagnosis"])
```

*Diviser l'ensemble de données en ensemble d'entraînement et de test*

```
x_train,x_test,y_train,y_test=train_test_split(x_train,y_train,test_size=0.25)
```

*sélection le model*

```
model = RandomForestClassifier()
```

```
model.fit(X_train,y_train)
```

```
model.score(X_train,y_train)
```

```
model.score(X_test,y_test)
```

---

## 3.5. Applications de la Random Forest

Le Random Forest est principalement utilisée dans quatre secteurs :

- ⊙ Secteur bancaire : Le secteur bancaire utilise principalement cet algorithme pour l'identification du risque de prêt.
- ⊙ Médecine : Avec l'aide de cet algorithme, les tendances et les risques liés aux maladies peuvent être identifiés. Utilisation des terres : Cet algorithme permet d'identifier les zones d'utilisation similaire des sols.
- ⊙ Commercialisation : Les tendances en matière de marketing peuvent être identifiées à l'aide de cet algorithme.

## 3.6. Conclusion

L'algorithme Random Forest est un excellent algorithme, tant pour les problèmes de classification que de régression, pour produire un modèle prédictif. Ses hyperparamètres par défaut donnent déjà d'excellents résultats et le système est très efficace pour éviter le débordement. De plus, il est un assez bon indicateur de l'importance qu'il attribue à vos caractéristiques. Le prochain chapitre sera consacré à l'implémentation de l'approche proposée, afin de montrer son efficacité, et de comparer les résultats obtenus avec les valeurs obtenues par les travaux cités dans la littérature.

# Chapitre 4

## Tests et résultats expérimentaux

Nous présentons dans ce chapitre les expérimentations que nous avons effectuées. Ces expérimentations ont pour objectif de valider notre random forest classifier algorithm *d'apprentissage automatique supervisé* pour la prédiction de cancer de sein. D'abord nous introduisons l'environnement de test (section 4.1). Ensuite, nous présentons les différents tests effectués pour régler les paramètres algorithm *d'apprentissage automatique supervisé* (section 4.2). Enfin, nous décrivons les résultats expérimentaux (section 4.3) et L' interfaces d'implémentation (section 4.4)

### 4.1. Présentation de l'environnement de test

#### 4.1.1. Collection de test

	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave_points_mean
0	1	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.30010	0.14710
1	1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.08690	0.07017
2	1	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.19740	0.12790
3	1	11.42	20.38	77.58	386.1	0.14250	0.28390	0.24140	0.10520
4	1	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.19800	0.10430
5	1	12.45	15.70	82.57	477.1	0.12780	0.17000	0.15780	0.08089
6	1	18.25	19.98	119.60	1040.0	0.09463	0.10900	0.11270	0.07400
7	1	13.71	20.83	90.20	577.9	0.11890	0.16450	0.09366	0.05985
8	1	13.00	21.82	87.50	519.8	0.12730	0.19320	0.18590	0.09353
9	1	12.46	24.04	83.97	475.9	0.11860	0.23960	0.22730	0.08543

Figure o6: Collection de test

Une collection de test est une dataset constituée d'un ensemble de patients et chaque patient est défini par des caractéristiques. Les caractéristiques sont calculées à partir d'une image numérisée d'une aspiration à l'aiguille fine (FNA) d'une masse mammaire. Ils décrivent les caractéristiques des noyaux cellulaires présents dans l'image.

Informations sur les attributs :

- rayon (moyenne des distances du centre aux points du périmètre)
- texture (écart-type des valeurs d'échelle de gris)
- périmètre
- aire
- lissé (variation locale des longueurs de rayon)
- compacité ( $\text{périmètre}^2 / \text{surface} - 1.0$ )
- concavité (sévérité des parties concaves du contour)
- points concaves (nombre de parties concaves du contour)
- symétrie
- dimension fractale ("approximation de la ligne de côte" - 1)

La moyenne, l'erreur standard et la "pire" ou la plus grande (moyenne des trois plus grandes valeurs) de ces caractéristiques ont été calculées pour chaque image, ce qui a donné 20 caractéristiques. Par

exemple, 1 Numéro d'identification, 2 Diagnostic (1 = malin, 0 = bénin) ,le champ 3 est le rayon moyen, le champ 13 est le rayon SE (voir la Figure 06).

#### 4.1.2. Outils et mesures d'évaluation

**Software :** afin de réaliser notre SMA pour le problème de DAU, nous utilisons le langage Python à travers les logiciels suivants :



*PyCharm* est un environnement de développement intégré (IDE) utilisé pour programmer en Python. Il a pour rôle de faciliter l'écriture des codes ainsi que l'analyse des codes.



*NumPy* est une bibliothèque utilisée en Python, elle permet de manipuler des vecteurs à plusieurs dimensions et elle propose des fonctions mathématiques complètes.



*PyQt* est composée d'une bibliothèque Python Qt et d'une application Qt designer permettant de créer des interfaces graphique.



*Pandas* est un outil d'analyse et de manipulation de données open source rapide, puissant, flexible et facile à utiliser, construit sur le langage de programmation Python.



Scikit-learn est une bibliothèque Python destinée à l'apprentissage automatique



permettant aux Data Scientists de transformer des données brutes en graphiques et en diagrammes



*Jupyter* est un outil de développement interactif basé sur le Web pour le code et les données, son interface flexible permet aux utilisateurs de configurer et d'organiser des flux de travail en science des données.

## 4.2. Réglages des paramètres de RandomForest

nous fixons d'abord les paramètres de RandomForest,  $X_{train}$ ,  $y_{train}$ , le nombre maximum d'estimation, notée  $param\_range$   $K$  et la constante utilisée pour Détermine la stratégie de fractionnement de la validation croisée, notée  $CV$ . Dans ces tests initiaux, nous exécutons notre algorithme 50 fois Afin de garantir des résultats de test fiables et précis, puis une moyenne des résultats est calculée. La Figure 4.2 présente les résultats de Score en fonction de nombre d'estimation.

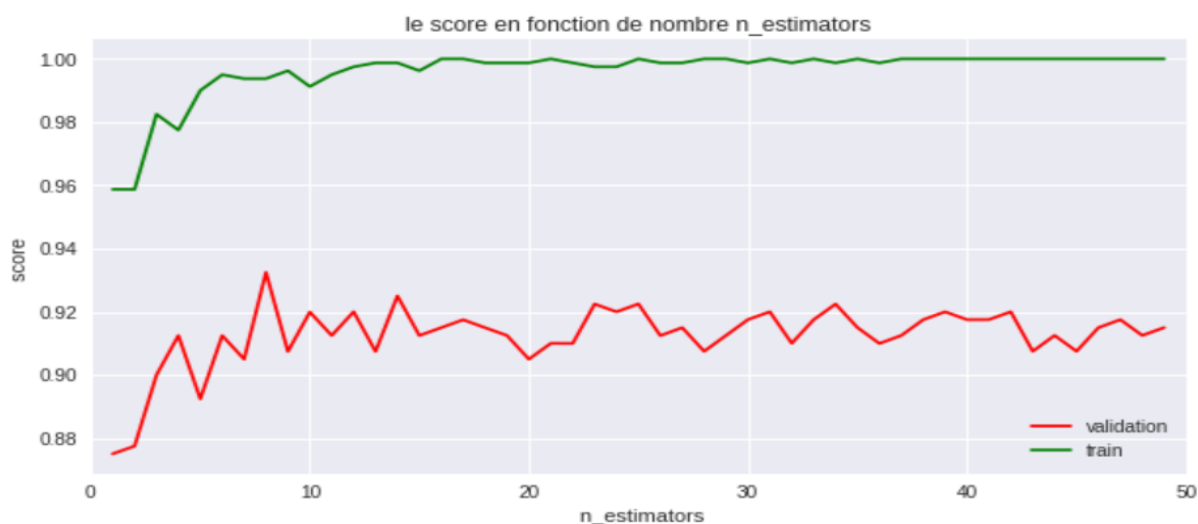


Figure 07: les résultats de Score en fonction de nombre d'estimation

## 4.3. Résultats expérimentaux

Model	Train Accuracy	AUC SCORE
RandomForestClassifier	100%	97%

À partir ce tableau, nous remarquons que les résultats obtenus par notre approche RandomForest sont meilleurs,Ce qui prouve l'efficacité de notre approche.

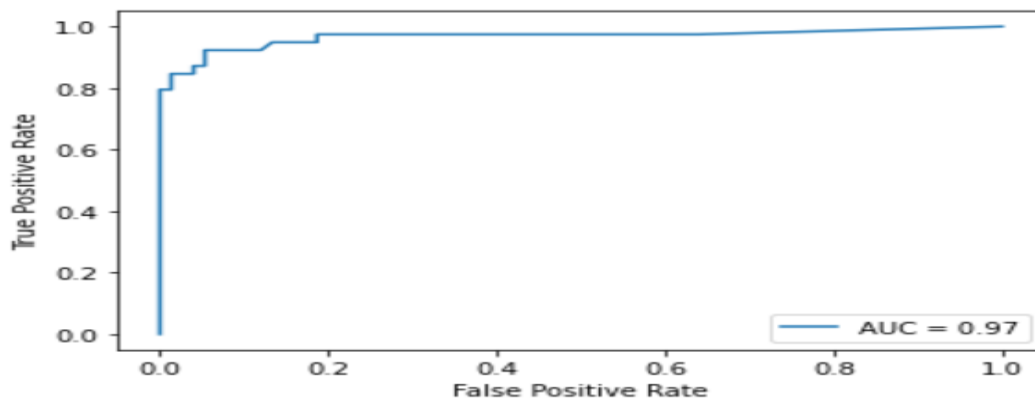


Figure o8: La courbe de caractéristique de fonctionnement du récepteur

Cela est montré par les résultats obtenus à travers la courbe (receiver operating characteristic)



## 4.4. L' interfaces d'implémentation

Dans cette section, nous présentons l' interfaces principales du code d'implémentation de notre approche proposée

The image shows a web application interface for breast cancer prediction. The title bar says "breast cancer". The main header is pink with the text "BREAST CANCER" and two ribbon icons. Below the header, there are three main sections:

- Patient information's**: Contains input fields for Name, N° id, First name, and Age.
- Values for the test procedure**: Contains two columns of input fields for various breast cancer features. The left column includes: radius\_mean, texture\_mean, perimeter\_mean, area\_mean, smoothness\_mean, compactness\_mean, concavity\_mean, concave\_points\_mean, and symmetry\_mean. The right column includes: fractal\_dimension\_mean, radius\_se, texture\_se, perimeter\_se, area\_se, smoothness\_se, compactness\_se, concavity\_se, and concave\_points\_se.
- RESULT**: Contains a large "score" placeholder, a "TEST" button, and a message: "this result remains an expectaion,so please see your doctor". There is also a "Clear" button at the bottom right.

Figure 09: L' interfaces d'implémentation

## 4.4. Conclusion

Dans ce chapitre, nous constatons bien que l'approche proposée a permis d'améliorer considérablement la prédiction. De plus, le RandomForest proposé pour la prédiction du cancer de sein a obtenu les meilleures performances en termes de score.

# Conclusion générale

Dans ce travail, nous avons abordé, dans le premier chapitre, une description du cancer du sein et la nécessité de la détection précoce de cette maladie pour réduire le nombre de la mortalité. Ensuite, dans le deuxième chapitre, nous avons montré le rôle de l'informatique, en particulier l'intelligence artificielle, pour la recherche des solutions et des moyens pour la détection du cancer du sein, dans le but de réduire le risque de cette maladie. Pour cela, nous avons passé brièvement sur les concepts du domaine de l'intelligence artificielle et ses différentes branches « l'apprentissage automatique ». A la fin de ce chapitre, nous avons analysé une série des travaux antérieurs.

D'autre part, le modèle proposé possède la capacité de s'adapter pour la prédiction d'autres maladies, tel que le diabète, les maladies cardiaques et les personnes infectés par le Covid 19.