

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Abou Bekr Belkaid



جامعة أبي بكر بلقايد

Faculté des sciences  
Département d'informatique

Mémoire de fin d'étude pour l'obtention du diplôme de master en informatique  
Option : modèle intelligent et décision

Thème

## **Classification des cellules tumorales à base de l'algorithme immunitaire modified-AIRS**

Réalisé par  
SEMMOUD Abderrazak  
MOGTIT Qaddour

Encadré par  
Mr CHIKH Mohamed El Amine

Soutenu le 03 juillet 2011, devant le jury composé de :

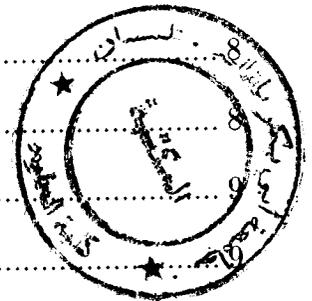
---

Mme. ILES Nawel  
Mr. SMAHI Mohammed Ismail  
Mr. BENZAOUZ Mortada  
Mr. HADJILA Fethallah  
Mr. LAHSAINI Mohammed  
Mr. BENZIANE Yaghmoracen

---

# Table de matières

Remerciements .....	iii
Table de matières .....	iv
Liste de figures .....	viii
Liste des tableaux.....	xi
Introduction générale.....	1
Le cancer du sein.....	4
1.1. Anatomie du sein.....	5
1.2. Le cancer du sein.....	5
1.2.1. Facteurs de risque .....	8
1.2.1.1 L'âge.....	8
1.2.1.2 Les facteurs génétiques.....	8
1.2.1.3 Risques hormonaux.....	8
1.2.1.4 L'alimentation.....	8
1.2.1.5 Le tabac.....	8
1.2.1.6 La pilule.....	8
1.2.1.7 Le Traitement Hormonal de la Ménopause (THM) .....	10
1.2.1.8 La présence du cancer .....	10
1.2.2. Symptôme et signes.....	10
1.2.3. Types de cancer du sein .....	11
1.2.3.1 Le cancer du sein non-infiltrant (in situ).....	11
1.2.3.2 Le cancer du sein infiltrant .....	12
1.2.4. Classification TNM.....	14
1.2.5. Evolution.....	16
1.2.6. Diagnostic .....	17
1.2.6.1 Mammographie.....	18



## Liste des figures

1.2.6.2 Echographie mammaire.....	19
1.2.6.3 Cytoponction .....	20
1.2.6.4 La microbiopsie échoguidée.....	21
1.2.6.5 L'examen histologique .....	22

	36
2.1.6. Propriétés du système immunitaire .....	37
2.1.7. Théories immunitaires .....	38
2.1.7.1 Théorie de la sélection Négative/Positive.....	38
2.1.7.2 Théorie de la sélection clonale.....	40
2.1.7.3 Théorie des réseaux immunitaires .....	41
2.1.7.4 Théorie du danger.....	41
2.1.8. Déroulement d'une réponse immunitaire .....	43
2.2. Systèmes immunitaires artificiels (SIAs).....	43
2.2.1. Historique.....	43
2.2.2. Définitions des SIAs .....	44
2.2.3. Modélisation des systèmes immunitaires artificiels.....	45
2.2.4. Les mesures d'affinités.....	47
2.2.5. Algorithmes immunitaires .....	47
2.2.5.1 Algorithme de la sélection négative.....	48
2.2.5.2 Algorithme de la sélection positive.....	48
2.2.5.3 Algorithme de la sélection clonale.....	55
2.2.5.4 Algorithme du réseau immunitaire.....	59
2.3. Conclusion.....	60
Classification des cellules tumorales à base de m-AIRS .....	60
3.1. La base de Wisconsin .....	62
3.1.1. Prétraitement de la base.....	63
3.1.2. Sélection des attributs .....	64
3.2. Algorithme modified-AIRS .....	64
3.2.1. Principe.....	66
3.2.2. Etapes d'apprentissage .....	66
3.2.2.1 Initialisation.....	66
3.2.2.2 Identification des cellules B et génération des ARBs.....	66

3.2.2.3	Compétition des ressources et développement des cellules	
	mémoires candidates .....	69
3.2.2.4	Introduction des cellules mémoires.....	70
3.2.3.	Classification .....	70
3.3.	Application.....	70
3.3.1.	Choix des paramètres .....	70
3.3.1.1	Pas de mutation.....	71
3.3.1.2	Seuil de stimulation .....	72
3.3.1.3	Scalaire du seuil d'affinité (ATS) .....	72
3.3.1.4	Taux de clonage.....	73
3.3.1.5	Hyper taux de clonage .....	73
3.3.1.6	Nombre de ressources.....	73
3.3.2.	Résultats .....	74
3.3.2.1	L'efficience .....	74
3.3.2.2	L'efficacité .....	75
3.4.	Conclusion.....	78
	Conclusion générale et perspectives .....	79
	Références bibliographiques .....	81
	Annexe : Interface et manuel d'utilisation .....	85
1.	Environnement de travail.....	85
2.	Codage.....	86
3.	Présentation de l'Interface Homme Machine (IHM) .....	87
3.1.	La fenêtre principale .....	87
3.2	Apprentissage .....	89
3.3.	Test .....	89
3.4.	Classification.....	90
3.5.	Aide .....	91

# Liste de figures

Figure 1.1 : Structure anatomique du sein .....	5
Figure 1.2 : Taux d'incidence et de mortalité des différents types du cancer chez les femmes [CIR, 08] .....	7
Figure 1.3 : Cancer canalaire in situ (CCIS) .....	11
Figure 1.4 : Cancer lobulaire in situ (CLIS) .....	12
Figure 1.5 : Cancer canalaire infiltrant .....	13
Figure 1.6 : Cancer lobulaire infiltrant .....	14
Figure 1.7 : Mammographie montrant un cancer du sein à droite .....	18
Figure 1.8 : Échographie montrant 9mm masse bénigne .....	20
Figure 1.9 : Cytoponction d'une tumeur .....	20
Figure 1.10 : Microbiopsie echoguidée .....	21
Figure 1.11 : Tissus prelevé et préparé pour les analyses .....	22
Figure 1.12 : Vue microscopiques des cellules .....	23
Figure 1.13 : Cellule cancéreuse .....	24
Figure 1.14 : Cellule cancéreuse .....	24
Figure 2.1 : Organes immunitaires .....	29
Figure 2.2 : Antigène .....	31
Figure 2.3 : Structure d'un anticorps .....	32
Figure 2.4 : Les forces contribuant à l'affinité de l'interaction anticorps/antigène .....	33
Figure 2.5 : Architecture du système immunitaire .....	34
Figure 2.6 : Processus de la sélection clonale .....	39
Figure 2.7 : représentation du réseau immunitaire idiotypique .....	41
Figure 2.8 : Réponse immunitaire .....	42

Figures 2.9 : Un framework de SIAs pour l'ingénierie et sa structure multi niveau [TIM, 03].....	45
Figure 2.10 : Espace des formes [CAS, 01] .....	46
Figure 2.11 : Affinité en nombre de bits contigus .....	47
Figure 2.12 : Algorithme de sélection Négative en deux phases.....	48
Figure 2.13 : Représentation de l'algorithme de la sélection clonale.....	49
Figure 2.14 : CLONALG pseudo code .....	52
Figure 2.15 : AIRS pseudo code (apprentissage) .....	53
Figure 2.16 : BCA pseudo code.....	54
Figure 2.17 : SIA pseudo code .....	55
Figure 2.18 : RAIN pseudo code .....	56
Figure 2.19 : aiNET pseudo code.....	58
Figure 3.1 : Moyenne des differences entre les valeurs dans chaque classe .....	63
Figure 3.2 : Taux de classification calculé par le KNN après chaque elimination .....	64

Figures 2.9 : Un framework de SIAs pour l'ingénierie et sa structure multi niveau [TIM, 03].....	45
Figure 2.10 : Espace des formes [CAS, 01].....	46
Figure 2.11 : Affinité en nombre de bits contigus .....	47
Figure 2.12 : Algorithme de sélection Négative en deux phases.....	48
Figure 2.13 : Représentation de l'algorithme de la sélection clonale.....	49
Figure 2.14 : CLONALG pseudo code .....	52
Figure 2.15 : AIRS pseudo code (apprentissage).....	53
Figure 2.16 : BCA pseudo code.....	54
Figure 2.17 : SIA pseudo code .....	55
Figure 2.18 : RAIN pseudo code .....	56
Figure 2.19 : aiNET pseudo code.....	58
Figure 3.1 : Moyenne des differences entre les valeurs dans chaque classe .....	63
Figure 3.2 : Taux de classification calculé par le KNN après chaque elimination .....	64
Figure 3.3 : Représentation de l'algorithme m-AIRS.....	65
Figure 3.4 : Algorithme original de mutation.....	67
Figure 3.5 : Nouvel algorithme de mutation .....	68
Figure 3.6 : Temps d'exécution en milisecons et taux de classification calculé avec KNN.....	71
Figure 3.7 : Taux de classification calculé avec KNN.....	72
Figure 3.8 : Taux de classification calculé avec KNN.....	73
Figure 3.9 : Taux de classification calculé avec KNN.....	74
Figure 3.10 : Résultats de classification de KNN et Centroïde.....	77
Figure 3.11 : Comparaison des résultats obtenus avec ceux des différentes approches .....	77
Figure A.1 : Logo du langage JAVA.....	85

Figure A.2 : Packages du projet .....	87
Figure A.3 : Interface de l'application.....	88
Figure A.4 : Onglet d'apprentissage.....	89
Figure A.5 : Onglet de test.....	89
Figure A.6 : Résultats de test.....	90
Figure A.7 : Onglet de classification .....	90
Figure A.8 : Présentation d'une image des cellules .....	91
Figure A.9 : Onglet d'aide .....	91

# Liste des tableaux

Tableau 1.1 : Classification TNM .....	16
Tableau 1.2 : Classification de cancer du sein par stade (Union Internationale Contre le Cancer) .....	17
Tableau 2.1 : Algorithme de sélection clonale .....	50
Tableau 2.2 : CLONALG paramètres .....	51
Tableau 2.3 : BCA paramètres .....	54
Tableau 2.4 : SIA paramètres .....	55
Tableau 3.1 : Attributs de la base de Wisconsin du cancer du sein.....	62
Tableau 3.2 : Résultats après différents prétraitements [ABE, 08] .....	62
Tableau 3.3 : m-AIRS paramètres .....	65
Tableau 3.4 : Temps d'apprentissage en millisecondes .....	69
Tableau 3.5 : Espace mémoire et temps d'exécution pour Apprentissage, Test et Classification .....	74
Tableau 3.5 : Matrice de contingence de la classe $C_i$ .....	76

# Introduction générale

La médecine est la science qui a accompagné l'humanité durant des millénaires, c'est la discipline qui a donné beaucoup de soulagement aux être humains. Mais avec le temps, de nouvelles maladies sont apparues et d'autres sont devenues plus fréquentes et plus dangereuses ce qui a remis les capacités de la médecine en question. Parmi les maladies les plus fréquentes et les plus dangereuses de nos jours ; nous trouvons le cancer, cette maladie qui se développe et s'accroît dans le corps humain silencieusement jusqu'à ce qu'elle le bat. Aujourd'hui, 8 millions de personnes meurent chaque année dans le monde à cause du cancer [CIR, 08], malgré que la possibilité de guérir un cancer atteint 50%. Tous ces risques ont exigé l'introduction des outils pour aider le médecin à mieux diagnostiquer cette maladie afin de la guérir.

Le cancer du sein est le plus fréquent (22.9% [CIR, 08]) chez les femmes, son diagnostic passe par plusieurs étapes en commençant par un examen clinique systématique soit par un gynécologue ou bien par la patiente elle-même, la détection d'une tumeur est confirmée par une mammographie et une échographie mammaire. Après la détection de la tumeur, un traitement sera pris selon le type de la tumeur. L'analyse cytologique, des cellules tirées par une cytoponction ou une microbiopsie, permet de prendre la décision si la tumeur est cancéreuse ou non, et c'est dans cette phase sensible et décisive qu'on peut rencontre des erreurs qui peuvent fausser les analyses et par conséquent risquer de perdre une vie. Cette analyse peut être effectuée aussi au moment de l'intervention chirurgicale pour savoir quelle sera la future action, cette analyse extemporané doit être rapide et précise et c'est ce qui n'est pas toujours fournie par une analyse manuelle. Tous cela nous conduit vers une vraie exigence d'un traitement automatique rapide est précis.

Dans ce mémoire, nous concevons un système semi-automatique pour la classification des cellules tumorales du sein. Ce système peut être précédé par un module de prétraitement pour l'automatisation totale de cette phase de diagnostique.

---

Les concepteurs informatiques ont essayé toujours de fournir des modèles intelligents et performants afin de résoudre des problèmes quotidiens. Il y a une partie importante de ces modèles inspirés des systèmes biologiques tels que le fonctionnement du cerveau humain (Réseaux de neurones), les systèmes génétiques.

Le mémoire est clôturé par une conclusion et des perspectives. Elle donne un aperçu général du travail accompli. Une annexe est présentée en fin de mémoire, décrivant l'environnement de programmation, avec un manuel d'utilisation via une description fondamentale de notre application.

# Chapitre 1

## Le cancer du sein

Le cancer est une tumeur ayant tendance à s'accroître, à détruire les tissus voisins et à donner d'autres tumeurs à distance de son lieu d'origine. Cette maladie véhicule encore beaucoup de peur et reste crainte par la majorité des malades.

A cause des progrès incessants faits tant en matière de traitement que de détection et de prévention, les possibilités de guérir le cancer atteignent aujourd'hui plus de 50%, malgré cela, la mortalité mondiale à cause du cancer en 2008 était près de 8 millions de personnes [CIR, 08] ce qui est supérieure à celle du sida, tuberculose et paludisme réunis.

Le cancer du sein est le type de cancer le plus fréquent chez les femmes. Une femme sur onze est actuellement touchée par le cancer du sein [LAR]. Le cancer du sein affecte habituellement les femmes de plus de 50 ans. Cependant, les femmes de tous les âges peuvent avoir un cancer du sein et, dans de rares cas, le cancer du sein peut aussi affecter les hommes.

En Algérie comme dans la majorité des pays, le cancer du sein est la première tumeur chez les femmes. Il représente 37,7% de l'ensemble des nouveaux cas en 2006. Le nombre de nouveaux cas par an est 7500. En revanche, le nombre de décès annuel par le cancer du sein est d'environ 3500 personnes.

Dans ce chapitre nous aborderons la maladie de cancer du sein, en la définissant et en citant ses symptômes et signes, son évolution ainsi que les différentes étapes de diagnostique et les méthodes utilisé pour le traitement.

### 1.1. Anatomie du sein

Les seins sont centrés par un mamelon, lui-même entouré d'une zone pigmentée, l'aréole. La peau de l'aréole, très fine, est légèrement déformée par les orifices des glandes sébacées, des glandes sudoripares et des follicules pileux. La glande mammaire est constituée d'une vingtaine de lobes glandulaires noyés dans du tissu graisseux. Les canaux excréteurs de ces lobes, appelés canaux galactophores, débouchent sur le mamelon. Les seins reposent en arrière sur le muscle pectoral [LAR].

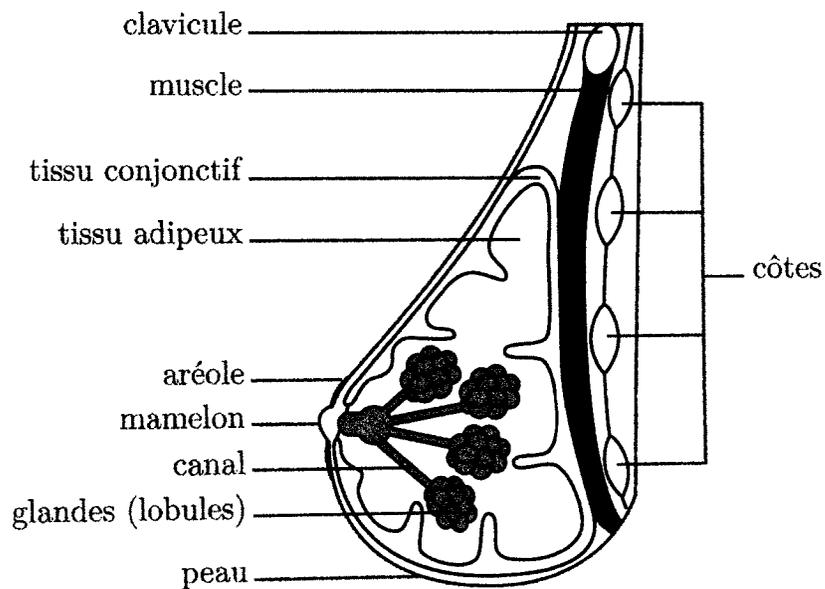


Figure 1.1 : Structure anatomique du sein

### 1.2. Le cancer du sein

Le cancer survient à la suite de mutations ou des changements anormaux, dans les gènes responsables de la régulation de la croissance des cellules et à les garder en bonne santé. Les gènes sont dans le noyau de chaque cellule, qui agit comme la salle de contrôle. Normalement, les cellules de notre corps se remplacent par un processus ordonné de la croissance cellulaire: de nouvelles cellules prennent en charge alors que les anciens disparaissent. Mais au fil du temps, les mutations peuvent allumer certains

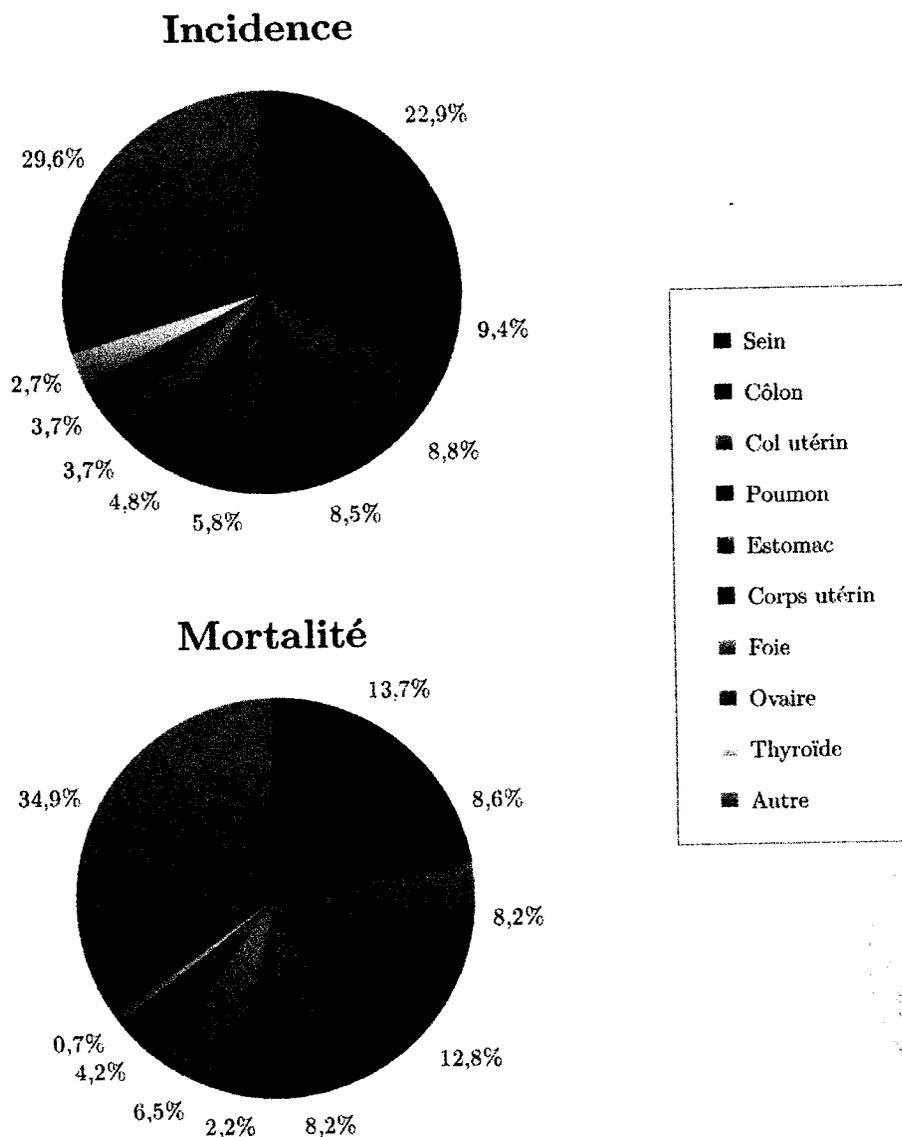


Figure 1.2 : Taux d'incidence et de mortalité des différents types du cancer chez les femmes [CIR, 08]

On pense que le cancer du sein n'arrive qu'aux femmes, mais environ 1% des cancers du sein se produisent chez les hommes [IND, 08]. Les hommes ont une petite quantité de tissu derrière le mamelon et, bien qu'elle ne soit pas aussi importante que chez les femmes, un cancer peut quand même s'y développer. Le cancer du sein masculin apparaît surtout chez les hommes de plus de 60 ans. Peu d'hommes en sont conscients, si bien qu'il est généralement diagnostiqué plus tard que chez les femmes. Par conséquent, le cancer du sein masculin est généralement plus avancé quand il est diagnostiqué et donc plus difficile à traiter.

### **1.2.1. Facteurs de risque**

Malgré les progrès qui ont permis de mieux connaître les mécanismes de développement des cancers, les causes du cancer du sein ne sont actuellement pas connues. Néanmoins, les études ont mis en évidence certains facteurs de risque qui favorisent le cancer du sein. Il existe différents types de facteurs de risque de survenue d'un cancer du sein.

#### **1.2.1.1 L'âge**

L'âge est le facteur de risque le plus important vis-à-vis du cancer du sein [KEL, 96]. La maladie est rare chez les femmes de moins de 30 ans. Le risque augmente entre 50 et 75 ans (près des deux tiers des cancers du sein).

#### **1.2.1.2 Les facteurs génétiques**

On parle de gènes de prédisposition ou de risques de développer un cancer du sein. Plusieurs femmes dans une même famille (mère, grand-mère, tante, sœurs) peuvent présenter un cancer du sein, on soupçonne alors une origine génétique à cette maladie.

Grâce à la recherche génétique, on sait aujourd'hui détecter certaines anomalies chromosomiques qui pourraient être responsables de la survenue d'un cancer du sein (5 à 10 % des cas) [IND, 08] en rapport avec les gènes BRCA1 ou BRCA2. Si une telle anomalie est retrouvée chez une femme, il est possible de faire une recherche chez les autres femmes de la même famille mais il faut savoir que les résultats ne sont jamais affirmatifs.

#### **1.2.1.3 Risques hormonaux**

Il semble que les femmes ayant une puberté précoce, une ménopause tardive, n'ayant pas d'enfants ou ayant des enfants tardivement, n'ayant pas allaité, auraient un risque plus important d'avoir un cancer du sein.

#### **1.2.1.4 L'alimentation**

- L'effet le plus net est celui de la quantité de calories absorbées, plus elle est élevée, plus le risque augmente.

- L'obésité.
- Une consommation importante d'alcool (équivalent de 1/2 L à 1 L de vin par jour),
- La faible consommation de fruits et de légumes sont également suspectées d'augmenter le risque.

Toutes ces données ne sont valables que sur un plan statistique et non pas individuel [LCC, 06]

#### **1.2.1.5 Le tabac**

La fumée du tabac est une importante source de substances cancérigènes. Pourtant, la cigarette n'est pas considérée comme un facteur de risque établi du cancer du sein. Certains investigateurs ont trouvé que les fumeuses présentent un risque réduit, d'autres aucun risque, d'autres ont rapporté une augmentation de risque associé au tabagisme.

Le tabagisme passif semble associé à un risque augmenté d'environ 60 % ; ce risque est multiplié par trois chez les femmes après la ménopause [JOH, 00].

L'effet protecteur de la cigarette dans le cancer du sein serait dû à une diminution des oestrogènes circulants et à l'action anti-oestrogénique du tabac. Il a été rapporté que les fumeuses ont une ménopause précoce et une concentration urinaire réduite d'oestrogènes pendant la phase lutéale du cycle menstruel [MAC, 80].

#### **1.2.1.6 La pilule**

Pour les contraceptifs oraux (CO), la somme des études faites aux États-Unis conclut à une augmentation minime du risque (+ 14 %). Le risque serait un peu plus élevé (+ 20 %) pour les femmes ayant pris des CO pendant plus de 5 ans avant une première grossesse [LCC, 06]

### **1.2.1.7 Le Traitement Hormonal de la Ménopause (THM)**

Il existe un sur-risque de cancer du sein chez les femmes utilisant un THM. Ce sur-risque de cancer augmente avec la durée du traitement. Aucune différence n'est démontrée en fonction de la voie d'administration (orale, extra-digestive), ou du schéma d'administration (séquentiel ou continu). Selon les données d'une étude observationnelle, le sur-risque de cancer du sein lié au traitement estroprogestatif pourrait varier en fonction du type de progestatif mais ces données nécessitent d'être confirmées par d'autres études.

### **1.2.1.8 La présence du cancer**

Les femmes qui ont déjà présenté un cancer du sein ont plus de risques que les autres de présenter un autre (risque de deuxième cancer de 10 %) [IND, 08]. Elles seront en général suivies attentivement après.

Même si la présence d'un ou de plusieurs de ces facteurs favorise le développement d'un cancer du sein, il faut savoir que, mis à part les facteurs familiaux ou génétiques, ces facteurs augmentent généralement assez peu le risque : une femme qui possède une ou plusieurs de ces caractéristiques peut ne jamais développer un cancer du sein. En revanche, il est possible qu'une femme n'ayant aucune de ces caractéristiques soit atteinte d'un cancer du sein. Les facteurs de risque ne permettent donc pas de déterminer les causes exactes à l'origine d'un cancer du sein chez une femme [SOR, 07].

### **1.2.2. Symptôme et signes**

Un cancer du sein peut se manifester par une grosseur (nodule), un écoulement de liquide clair ou sanglant par le mamelon, une déformation du galbe du sein ou du mamelon (rétraction), plus rarement par une douleur. Il ne présente parfois aucun signe et est, à l'heure actuelle, souvent mis en évidence par une mammographie (examen radiologique du sein) réalisée à titre systématique dans le cadre du suivi médical individuel ou lors d'une campagne de dépistage. [LAR]

### 1.2.3. Types de cancer du sein

#### 1.2.3.1 Le cancer du sein non-infiltrant (in situ)

Le cancer du sein non-infiltrant (in situ) se réfère à un cancer qui est encore dans le conduit de lait et/ou les lobules du sein. En d'autres termes, le cancer n'a pas envahi les murs des canaux galactophores ou des lobules.

##### 1.2.3.1.1 Le cancer canalaire in situ (CCIS)

Le cancer canalaire in situ, également appelé cancer intracanaulaire (Figure 1.3), fait référence aux cellules cancéreuses qui se sont formés dans les conduits doux et sont encore confinés dans cet espace. Les conduits se bouchent et élargie à cause des cellules cancéreuses qui s'accumulent à l'intérieur. Le calcium a tendance à s'accumuler dans les canaux bloqués et est visible sur la mammographie.

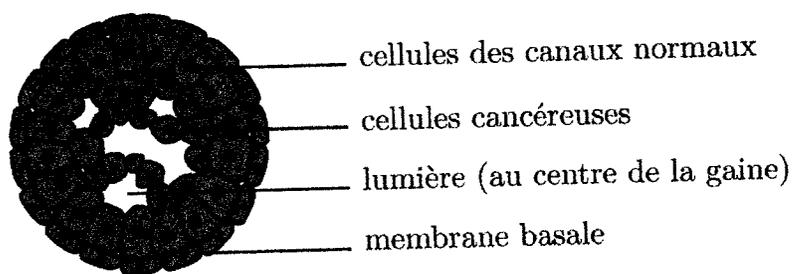
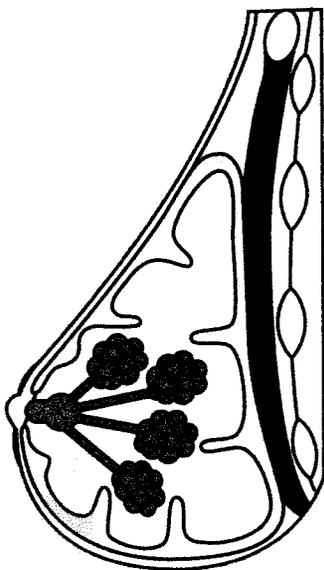


Figure 1.3 : Cancer canalaire in situ (CCIS)

### 1.2.3.1.2 Le cancer lobulaire in situ (CLIS)

Le cancer lobulaire in situ (Figure 1.5) fait référence à des cellules cancéreuses qui se sont formés dans les glandes mammaires et qui sont encore confinées dans cet espace. Ce type de cancer est souvent constaté chez les femmes autour de la ménopause. CLIS est différente de CCIS en ce sens qu'il ya un risque élevé que l'ensemble du tissu des deux seins puissent développer un cancer. Par conséquent, le traitement de CLIS doit viser deux seins.

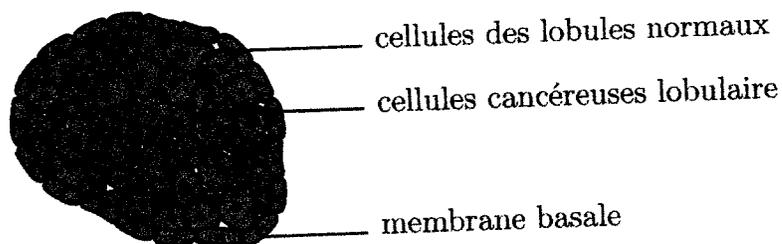
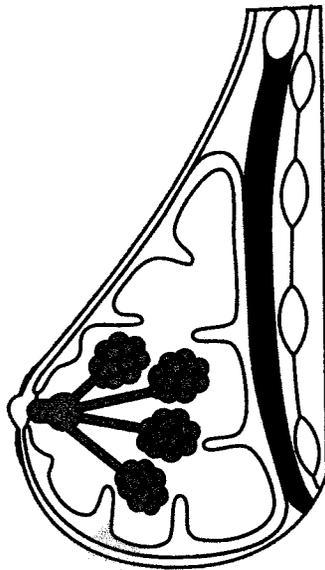


Figure 1.4 : Cancer lobulaire in situ (CLIS)

### 1.2.3.2 Le cancer du sein infiltrant

Le cancer du sein infiltrant se réfère à des cellules qui se sont développées à travers les parois des canaux galactophores et des glandes dans le tissu normal graisseux du sein. Les cellules continuent de croître entraînant une masse ou une grosseur. Ils peuvent alors former des métastases dans la circulation sanguine ou les vaisseaux

lymphatiques vers d'autres parties du corps. Les principaux types de cancer du sein infiltrant sont les suivants:

### 1.2.3.2.1 Le cancer canalaire infiltrant (CCI)

CCI est le plus populaire et est également appelé cancer mammaire ou adénocancer infiltrant (Figure 1.6).

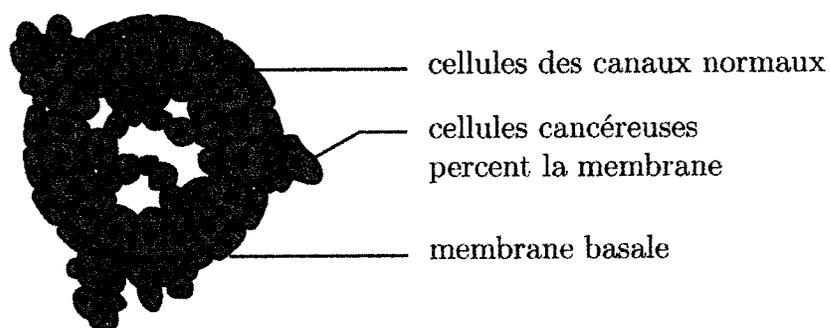
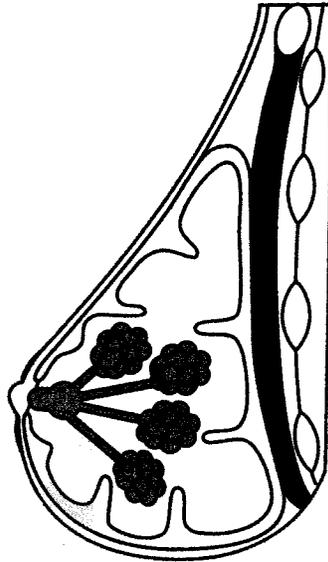


Figure 1.5 : Cancer canalaire infiltrant

### 1.2.3.2.2 Le cancer lobulaire infiltrant (CLI)

Moins fréquents et difficiles à diagnostiquer sur la mammographie, risque légèrement plus élevé d'être dans les deux seins (Figure 1.7).

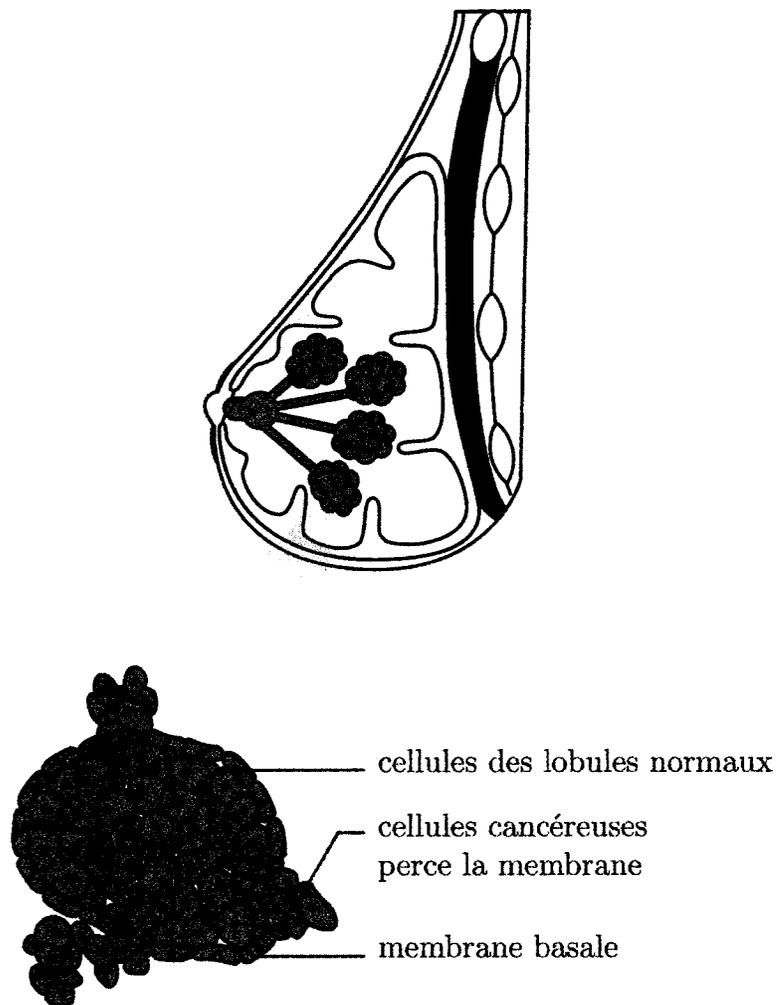


Figure 1.6 : Cancer lobulaire infiltrant

Il existe d'autres noms du cancers du sein que nous pouvons les rencontrer (environ 10% des cas) : Cancer mucineux (colloïde), Cancer médullaire, Cancer papillaire, Cancer tubuleux, Cancer adénoïde kystique, Cancer sécrétant juvénile, Cancer apocrine, Cancer métaplasique de type épidermoïde, Cancer métaplasique de type à cellules fusiformes, Cancer métaplasique de type chondroïde et osseux, Cancer métaplasique de type mixte, Maladie de Paget du mamelon.

#### 1.2.4. Classification TNM

La classification TNM est un système international, proposé par le chirurgien français Pierre Denoix de l'Institut Gustave-Roussy entre 1943 et 1952, de façon à classer les cancers selon leur extension anatomique.

Les trois lettres symbolisent la propagation de la maladie cancéreuse sur le site de la tumeur primitive (Tumor), dans les ganglions lymphatiques voisins (Node) et à distance pour d'éventuelles métastases (Metastasis).

<b>T = Tumeur primitive</b>
<p>Tx : Détermination de la tumeur primitive impossible</p> <p>T0 : Pas de signe de tumeur primitive (non palpable)</p> <p>Tis : Carcinome in situ : carcinome intracanalair, ou carcinome lobulaire in situ, ou maladie de Paget du mamelon sans tumeur décelable</p> <p>T1 : tumeur de moins de 20 mm, avec la sous classification suivante :</p> <p style="padding-left: 40px;">T1mic : micro-invasion &lt; 1 mm</p> <p style="padding-left: 40px;">T1a : moins de 5 mm</p> <p style="padding-left: 40px;">T1b : de 5 à 10 mm</p> <p style="padding-left: 40px;">T1c : de 10 à 20 mm</p> <p>T2 : tumeur de 20 à 50 mm</p> <p>T3 : tumeur de plus de 50 mm</p> <p>T4 : quelle que soit la taille, avec :</p> <p style="padding-left: 40px;">T4a : extension à la paroi thoracique</p> <p style="padding-left: 40px;">T4b : extension à la peau y compris, ulcération cutanée ou nodules de perméation</p> <p style="padding-left: 40px;">T4c : T4a + T4b</p> <p style="padding-left: 40px;">T4c : cancer inflammatoire</p>
<b>N = Adénopathies régionales</b>
<p>Nx : Appréciation impossible de l'atteinte ganglionnaire (du fait, par exemple, d'une exérèse antérieure)</p> <p>N0 : Absence de signe d'envahissement ganglionnaire régional</p> <p>N1 : Ganglions axillaires homolatéraux suspects mobiles</p> <p>N2 : Ganglions axillaires homolatéraux suspects fixés entre eux ou à d'autres structures, ou présence clinique d'adénopathies mammaires internes en l'absence d'adénopathies cliniques axillaires</p> <p style="padding-left: 40px;">N2a : Ganglions axillaires homolatéraux fixés</p> <p style="padding-left: 40px;">N2b : Ganglions mammaires internes homolatéraux</p>

cliniquement apparents sans adénopathies axillaires cliniques
N3 : Ganglions sous-claviculaires homolatéraux (niveau III axillaire) ou mammaires internes avec présence d'adénopathies axillaires ou ganglions sus-claviculaires présents (avec ou sans la présence de ganglions axillaires ou mammaires internes)
N3a : Ganglions suspects sous-claviculaires et axillaires homolatéraux
N3b : Ganglions mammaires internes et ganglions axillaires homolatéraux suspects
N3c : Ganglions sus-claviculaires homolatéraux suspects
<b>M = Métastases</b>
Mx : Renseignements insuffisants pour classer les métastases à distance
M0 : Absence de métastases à distance
M1 : Présence de métastase(s) à distance

Tableau 1.1 : Classification TNM

### 1.2.5. Evolution

Comme la plupart des cancers, le cancer du sein évolue d'abord localement, avec extension aux organes de voisinage et aux ganglions lymphatiques axillaires, puis se propage par voie sanguine (métastases), surtout aux os, au cerveau, au foie et aux poumons. Ces métastases peuvent apparaître tardivement, jusqu'à 10 ans après la découverte du cancer initial. [LAR]

Stade	TNM	Chances de Survie
Stade 0	Tis N0 M0	>95%
Stade I	T1 N0 M0	85%
Stade IIA	T0 N1 M0	66%
	T1 N1 M0	
	T2 N0 M0	
Stade IIB	T2 N1 M0	66%

	T3 N0 M0	
Stade IIIA	T0 N2 M0 T1 N2 M0 T2 N2 M0 T3 N1 M0 T3 N2 M0	41%
Stade IIIB	T4 N0 M0 T4 N1 M0 T4 N2 M0	41%
Stade IIIC	Tous T N3 M0	41%
Stade IV	Tous T Tous N M1	10%

Tableau 1.2 : Classification de cancer du sein par stade (Union Internationale Contre le Cancer)

#### 1.2.6. Diagnostic

La multiplicité des cancers et leur spécificité propre rendent difficile le dénombrement de tous les symptômes de la maladie. Néanmoins, une perte de poids importante et plus ou moins rapide, un manque d'appétit, une fatigue intense, une perte de sang dans les selles ou par la bouche, enfin des douleurs diverses sont des signes fonctionnels qui peuvent être associés à la présence d'un cancer. Le développement souvent silencieux des cancers tend à en retarder le diagnostic et pose des problèmes aux médecins, qui ne voient le patient qu'à un stade déjà avancé de la maladie. Parfois, la maladie est décelée par hasard, au cours d'une visite médicale ou d'un examen de sang. Le diagnostic repose sur l'examen clinique, des examens de laboratoire, des examens radiologiques et endoscopiques, des biopsies. [LAR]

Un cancer du sein est découvert soit à l'occasion d'un examen systématique par le gynécologue, soit par la patiente elle-même au cours d'un auto-examen du sein. Le

diagnostic est confirmé par une mammographie et éventuellement une échographie mammaire, et par une ponction du kyste ou du nodule (biopsie), dont le liquide ou les cellules sont examinés au microscope afin de rechercher des cellules tumorales. [LAR]

### 1.2.6.1 Mammographie

La mammographie est un examen aux rayons X, qui emploie de faibles doses de rayonnements mais qui, néanmoins, doit être évité chez la femme enceinte.

La mammographie permet le dépistage du cancer du sein et des lésions précancéreuses, ainsi que le diagnostic de certains symptômes (douleurs, grosseurs, écoulement du mamelon).

L'examen doit être de préférence réalisé durant la première moitié du cycle. Il est toujours bilatéral et comporte au moins deux clichés (face et profil) de chaque sein, parfois trois (un cliché de biais permettant l'exploration du creux axillaire).

Aucune préparation n'est requise, sauf lorsqu'un écoulement par le mamelon a été constaté, auquel cas une opacification préalable des canaux galactophores par injection d'un produit de contraste est effectuée (galactographie). L'examen ne s'accompagne d'aucun effet secondaire.

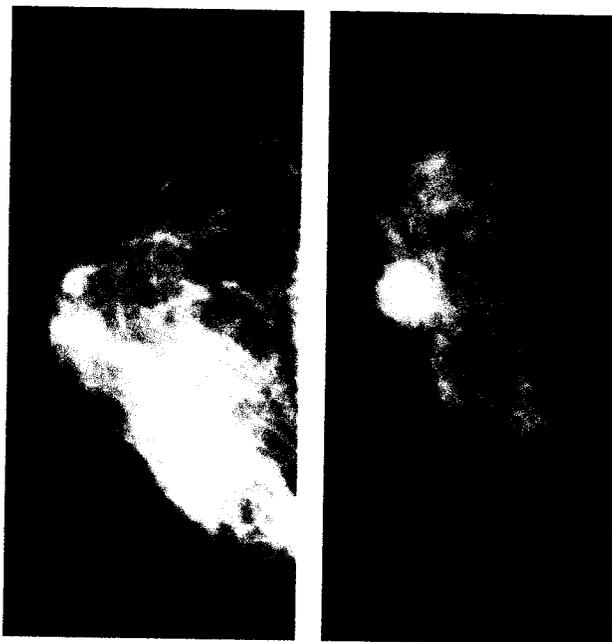


Figure 1.7 : Mammographie montrant un cancer du sein à droite.

Une mammographie systématique est conseillée tous les 2 ans chez les femmes à partir de 50 ans. Elle est aussi effectuée sur avis médical lorsqu'une femme présente un risque accru de cancer du sein (famille à risque, antécédents personnels). Elle ne dispense ni de l'autopalpation des seins ni de la consultation médicale effectuée régulièrement.

#### **1.2.6.2 Echographie mammaire**

Une échographie mammaire est un examen pratiqué par un radiologue qui utilise des ultrasons pour produire des images de l'intérieur du sein. Elle est très utile pour voir la nature, liquide ou solide, des nodules palpés ou découverts sur une mammographie.

Cet examen complète et précise les images obtenues par mammographie. Il ne remplace pas une mammographie qui est l'examen de référence pour le sein. Pour les adolescentes, les jeunes femmes et les femmes enceintes, la mammographie est moins performante : l'échographie mammaire est donc l'examen de référence. Pour les femmes plus âgées, la mammographie reste l'examen de référence.

Une échographie fournit des images "en temps réel", c'est-à-dire visibles immédiatement.

L'échographie permet ainsi de comparer ce que l'on sent avec les doigts (lors de la palpation) et ce que l'on voit sur l'écran.

Une échographie permet aussi de guider un prélèvement.

L'échographie mammaire permet de voir s'il existe certaines anomalies dans le sein, mais elle ne permet pas toujours de déterminer avec certitude s'il s'agit ou non d'un cancer. Pour cela, il faut réaliser d'autres examens complémentaires.

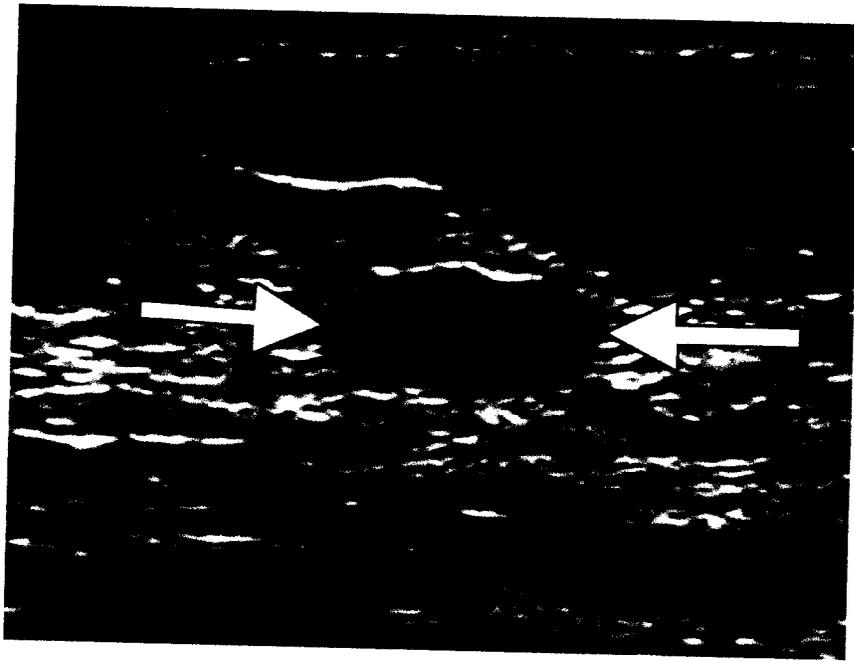


Figure 1.8 : Échographie montrant 9mm masse bénigne.

### 1.2.6.3 Cytoponction

Lorsqu'une lésion est visible à la fois en mammographie et en échographie, la cytoponction est souvent la technique de choix parce qu'elle est plus simple et plus rapide.

C'est un prélèvement qui se fait à l'aide d'une aiguille fine introduite dans la masse.

S'il s'agit d'un kyste, un éventuel liquide est aspiré. S'il s'agit d'un nodule de nature solide, il est prélevé quelques cellules aspirées de la même façon par la seringue.

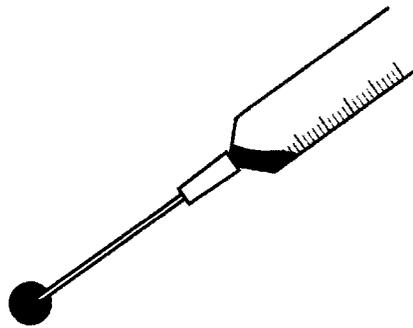


Figure 1.9 : Cytoponction d'une tumeur

#### 1.2.6.4 La microbiopsie échoguidée

Est également un prélèvement qui se fait cette fois ci avec une aiguille de gros diamètre sous anesthésie locale. On prélève un fragment de tissu (une carotte).

Pour certains auteurs, les microbiopsies résultent essentiellement des échecs des cytoponctions notamment pour les lésions suspectes purement échographiques. Pour d'autres, elles pourraient se substituer aux cytoponctions

Le diagnostic de cancer du sein doit toujours être confirmé par un examen histologique d'un prélèvement tissulaire (biopsie). Ces biopsies sont réalisées par voie percutanée, avec des aiguilles spéciales. Elles sont guidées sous échographie ou par un procédé d'imagerie couplant un mammographe et un système de guidage informatisé appelé mammotome. [LAR]

La microbiopsie échoguidée est une méthode confortable, sûre et peu coûteuse. Elle permet avantageusement d'éviter des biopsies diagnostiques, assure un diagnostic histologique précis. Elle doit donc contribuer à améliorer la valeur prédictive positive de la biopsie au cours du dépistage sénologique de masse.

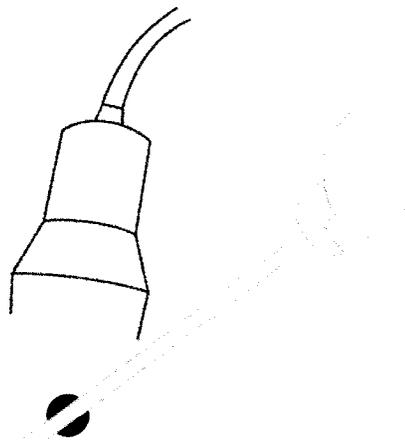


Figure 1.10 : Microbiopsie échoguidée.

### 1.2.6.5 L'examen histologique

L'anatomo-pathologiste est le médecin spécialiste qui examine et analyse les tissus prélevés. Il détermine s'il s'agit, ou non, d'un cancer et analyse les caractéristiques de la tumeur, ce qui contribuera à choisir ensuite le traitement à entreprendre.

1. Le médecin anatomo-pathologiste commence par réaliser un examen macroscopique (c'est à dire à l'œil nu).
2. Le prélèvement est divisé en coupes, dans un premier temps pour l'examen extemporané et, dans un second, pour l'examen anatomo-pathologique définitif, après que les prélèvements aient été fixés pour éviter qu'ils se détériorent, et colorés pour faciliter l'interprétation.

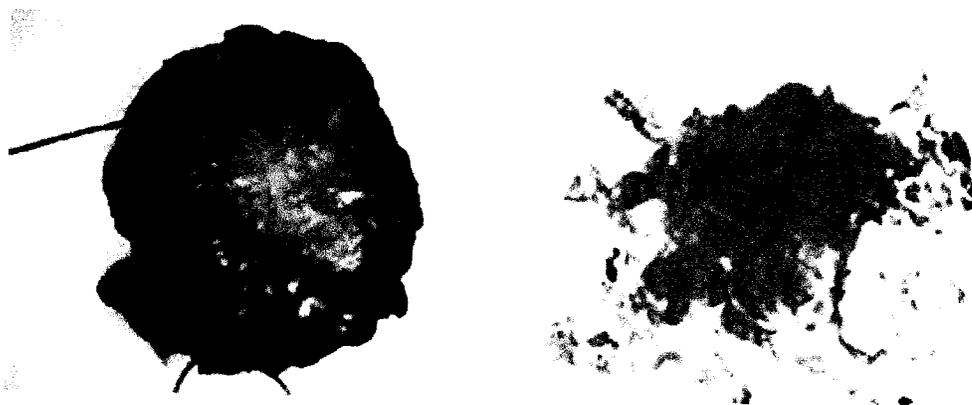


Figure 1.11 : Tissus prélevés et préparés pour les analyses

3. Ensuite, si nécessaire, l'examen extemporané aura lieu dont les résultats sont disponibles en quelques minutes. Une infime partie de la tumeur est analysée immédiatement au microscope. Les résultats sont transmis au chirurgien, qui poursuit ensuite l'intervention en fonction de ce diagnostic, et de ce qu'il avait convenu avec la patiente.
4. Le reste du prélèvement est préparé pour l'examen définitif, qui établira les conclusions définitives concernant les caractéristiques de la tumeur. Plusieurs jours sont nécessaires pour cela. L'eau qui reste dans le tissu est retirée et remplacée par de la paraffine. Au bout de 24 heures, les tissus ont suffisamment durci pour être préparés à l'analyse microscopique.

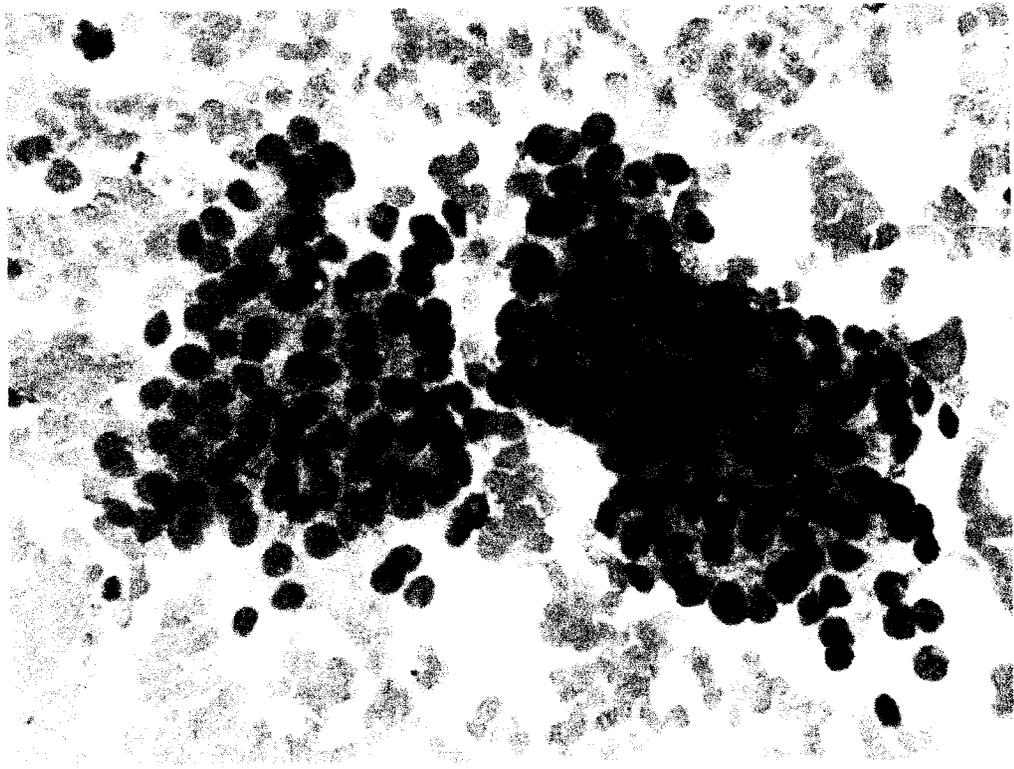


Figure 1.12 : Vue microscopiques des cellules

#### **1.2.6.6 Critères cytologiques de malignité**

Les anomalies génétiques responsables de la transformation maligne se répercutent sur l'apparence de la cellule cancéreuse. Idéalement il devrait être possible de faire le diagnostic de malignité sur une seule cellule. En pratique ceci s'avère difficile voire dangereux. En effet, la lésion maligne possède malgré tout une architecture (plus ou moins différenciée) et présente des relations avec les tissus normaux voisins. Si les caractéristiques architecturales tissulaires sont importantes pour le diagnostic, il n'empêche que dans de nombreuses situations on est amené à se déterminer par rapport à l'aspect des cellules.

##### **1.2.6.6.1 Anomalies des noyaux :**

- Taille : augmentation, inégalité (anisocaryose)
- Contenu : chromatine irrégulièrement répartie (mottes), hyperchromatisme
- Forme : contours irréguliers, incisures
- Nucléole : augmentation de taille, multiplicité, anomalies de forme

- Mitoses : augmentation de nombre, anomalies de forme (mitoses tripolaires, asymétriques)

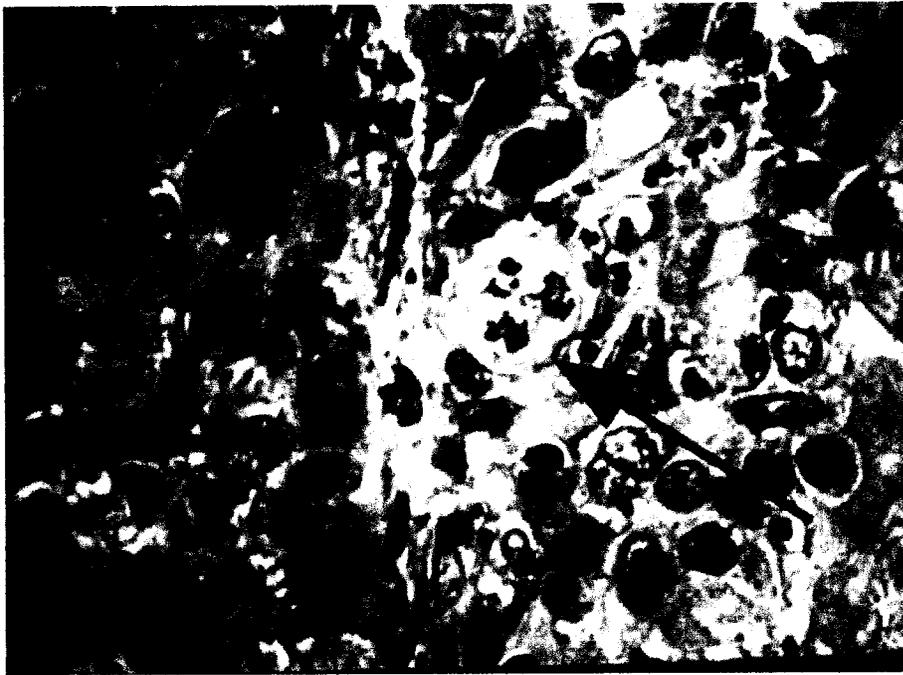


Figure 1.13 : Cellule cancéreuse

#### 1.2.6.6.2 Anomalies des cytoplasmes :

- Diminution de taille : augmentation du rapport nucléocytoplasmique
- Basophilie

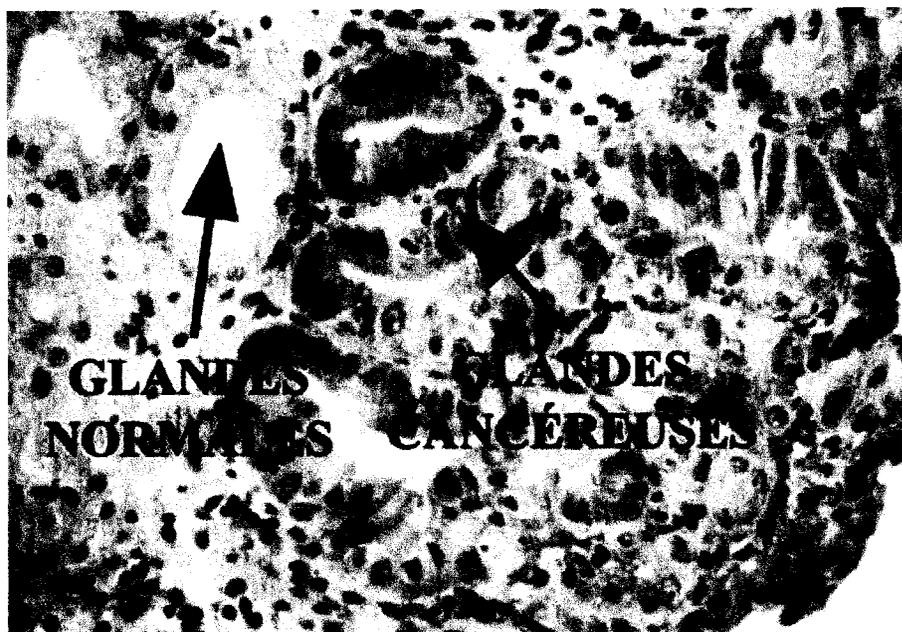


Figure 1.14 : Cellule cancéreuse

### **1.2.7. Traitement**

Quatre traitements peuvent être entrepris, parfois isolément, parfois en association : la chirurgie, la chimiothérapie, la radiothérapie, l'hormonothérapie.

#### **1.2.7.1 La chirurgie**

Est généralement le premier traitement envisagé. Elle consiste le plus souvent, aujourd'hui, en l'ablation de la tumeur (tumorectomie), associée à l'examen histologique immédiat (dit extemporané) de celle-ci, la femme étant toujours sous anesthésie générale : si l'examen histologique confirme la nature maligne de la tumeur, l'ablation des ganglions de l'aisselle (curage ganglionnaire axillaire) est réalisée, l'accord pour cette éventuelle extension de l'acte chirurgical devant avoir été préalablement obtenu. Pour les tumeurs volumineuses ou multiples, la mastectomie (ablation du sein) est encore pratiquée. Pour les tumeurs de petite taille, la technique du ganglion sentinelle peut être proposée de façon à minorer les conséquences fonctionnelles au niveau du bras, tout en sachant qu'un curage axillaire complémentaire s'impose en cas d'atteinte du ganglion sentinelle.

#### **1.2.7.2 La radiothérapie**

Pratiquée après une tumorectomie, consiste à irradier le sein et le pourtour de la zone retirée pour éviter les récurrences locales. Le traitement comprend généralement 4 ou 5 séances par semaine pendant 5 à 6 semaines.

#### **1.2.7.3 La chimiothérapie**

Peut être utilisée en complément de l'ablation chirurgicale du cancer primitif. Elle est utilisée en première intention en cas de cancer du sein avec signes inflammatoires. Les métastases sont aussi traitées par la chimiothérapie. Le traitement comporte l'association de plusieurs médicaments en hôpital de jour, sur un mode ambulatoire. Compte tenu de la causticité des produits administrés, il nécessite la pose préalable d'une chambre implantable pour éviter une extravasation (diffusion du produit en dehors d'un vaisseau), source de nécrose cutanée.

#### **1.2.7.4 L'hormonothérapie**

Très souvent associée aux autres traitements, consiste à prendre, pendant 5 ans, par voie orale des antiœstrogènes lorsque le cancer est hormonodépendant, c'est-à-dire que la tumeur contient des récepteurs hormonaux (éléments situés à l'intérieur de la cellule).

## Chapitre 2

# Les systèmes immunitaires

La biologie était toujours une source d'inspiration pour de nombreux développements en informatique, par exemple, la neurologie est à la base des travaux sur les réseaux de neurones. Ces derniers permettent de classifier des données grâce à des capacités d'apprentissage et de généralisation.

La génétique est à la source des algorithmes génétiques. Ainsi, les logiciels utilisant ces algorithmes possèdent des stratégies de sélection et de mutation pour optimiser leur fonctionnement.

Les sociétés d'insectes telles que les fourmilières apportent des solutions originales et adaptées à certains traitements informatiques où la recherche d'un chemin quasi-optimal est importante.

Plus récemment, le système immunitaire, avec sa capacité de distinguer le « soi » du « non-soi », a permis de mettre au point des systèmes informatiques capables de résoudre des problèmes complexes.

Nous présentons dans ce chapitre le système immunitaire biologique, en citant ses différents composants et mécanismes immunitaires et les propriétés intéressantes de ce système. Dans un second temps, nous abordons le système immunitaire artificiel (SIA) et son processus ainsi que les différents algorithmes et modèles immunitaires.

## **2.1. Le système immunitaire naturel**

Le système immunitaire biologique est un système complexe, robuste et adaptatif qui protège l'organisme contre les maladies ; il détruit les virus, lutte contre les bactéries, attaque les champignons, tue les parasites ainsi que les cellules tumorales. Ce système est constitué de plusieurs cellules, molécules et organes travaillant ensemble pour construire un mécanisme d'identification capable de catégoriser toutes les cellules dans le corps comme étant de cellules appartenant au soi ou de non-soi. Il reste jusqu'à nos jours un grand centre d'intérêt des biologistes et chercheurs.

### **2.1.1. Historique**

Les années soixante sont en général considérées comme le début de l'époque moderne de l'immunologie. Rodney Porter et Gerald Edelman réussirent à élucider la structure des anticorps entre 1959 et 1961, et furent lauréats du prix Nobel de médecine en 1972. Vers 1960, la communauté scientifique découvrait, grâce aux travaux de Jacques Miller, d'autres caractéristiques fondamentales des cellules immunitaires. En 1989, Charles Janeway propose un modèle selon lequel l'immunité innée serait la véritable gardienne des clefs du déclenchement d'une réponse immunitaire.

### **2.1.2. Concepts immunologiques**

Le système immunitaire est fait d'un système d'interactions complexes mettant en œuvre de nombreux organes, cellules et molécules différentes.

### 2.1.2.1 Organes immunitaires

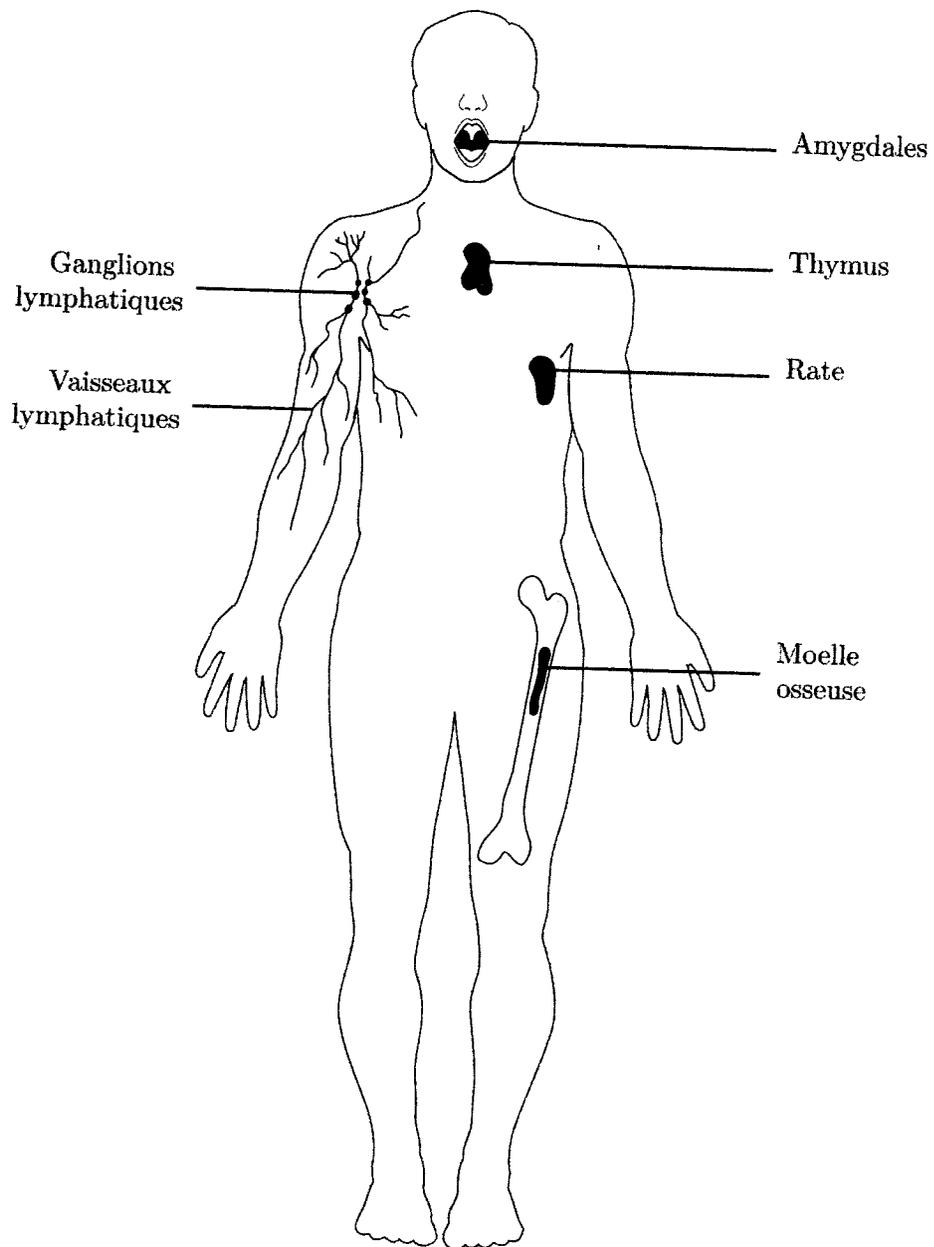


Figure 2.1 : Organes immunitaires

#### 2.1.2.1.1 Organes primaires

Les organes lymphoïdes primaires sont les organes dans lesquels les lymphocytes T et B vont être fabriqués et acquérir leurs différenciations en cellules équipées pour reconnaître un antigène.

- Moelle osseuse : Site de maturation des lymphocytes B.

- Thymus : Situé dans le bas du cou, le thymus est le site de maturation des lymphocytes T. Surtout actif pendant l'enfance, le thymus s'atrophie avec l'âge.

#### **2.1.2.1.2 Organes secondaires**

Ce sont les lieux de passage, d'accumulation, et de rencontre des antigènes et des cellules de l'immunité.

- Amygdales : Situées dans la cavité buccale, les amygdales servent à détruire la plus grande partie des envahisseurs qui entrent par l'air ou les aliments.
- Ganglions lymphatiques : Situés dans tout le corps le long des vaisseaux lymphatiques, les ganglions filtrent la lymphe et en éliminent les envahisseurs.
- Rate : Située sur le côté gauche, la rate est un siège de prolifération des lymphocytes. C'est le plus gros des organes lymphoïdes.
- Vaisseaux lymphatiques : Situés dans tout le corps, les vaisseaux lymphatiques transportent la lymphe.

#### **2.1.2.2 Cellules immunitaires**

Les cellules du système immunitaire impliquées dans la défense sont les globules blancs appelés leucocytes. Ces dernières se trouvent dans le sang et la lymphe.

Il existe différents types de globules blancs (Granulocytes neutrophiles, Granulocytes éosinophiles, Granulocytes basophiles, Granulocytes mastocytes, Lymphocytes B, Lymphocytes T, Monocytes), les plus importantes sont les lymphocytes. Chez les adultes les lymphocytes représentent 20 à 40% [DOC, 11] des lymphocytes sanguin et se compose de :

##### **2.1.2.2.1 Lymphocytes B**

Les lymphocytes B tirent leur nom de la moelle osseuse chez les mammifères (Bone marrow en anglais). Ils possèdent deux fonctions essentielles :

a) Leur activation par un corps étranger induit leur transformation en cellules sécrétrices d'immunoglobulines. b) Les lymphocytes B ont également la capacité de se comporter en cellule présentant le corps étranger. [NAN, 05].

#### 2.1.2.2.2 Lymphocytes T

Il existe différentes variantes des lymphocytes T. Leur rôle est d'attaquer les cellules infectées (réponse cellulaire). Les cellules T aideuses sont essentiellement chargées de l'activation des cellules B ; Les cellules T tueuses quant à elles s'attachent aux anticorps et leurs injectent des produits toxiques pour les tuer ; Une autre variante de cellules T, les suppresseurs, qui servent à éviter les réactions immunitaires non appropriées (maladies auto-immune).

#### 2.1.2.3 Antigène

Substance étrangère, identifiée comme telle par le système immunitaire qui produit des anticorps dirigés spécifiquement contre cette substance. Les antigènes sont généralement des protéines contenues dans des cellules ou des corps étrangers (globules rouges transfusés, organes greffés, bactéries, virus...), ou présentes dans l'environnement (pollens). La réaction antigène-anticorps est la base de l'immunité. Elle assure notre protection contre les infections. Mais cette réaction peut également être nocive lorsqu'elle est disproportionnée ou inappropriée : réaction allergique [ITE, 09].

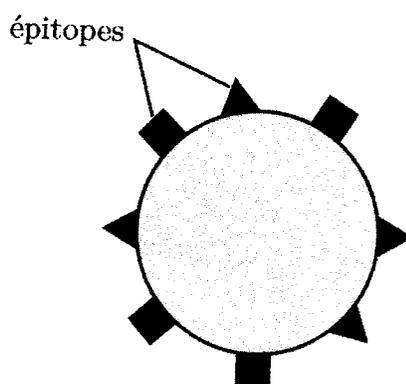


Figure 2.2 : Antigène

### 2.1.2.4 Anticorps

On définit un anticorps comme étant une protéine complexe synthétisée par les cellules du système immunitaire en réponse à la pénétration d'un corps étranger (antigène) [SAL, 09].

Les anticorps sont dérivés des lymphocytes B et les plasmocytes, ils sont capables de détecter et de neutraliser des substances étrangères. Il est important de signaler qu'il y a plus de dix millions d'anticorps différents dans un organisme, ce qui explique leur spécificité. Les immunoglobulines ou anticorps sont formés de deux chaînes polypeptidiques lourdes et de deux autres chaînes légères, assemblées sous forme d'un Y par des ponts disulfures [ALL, 06].

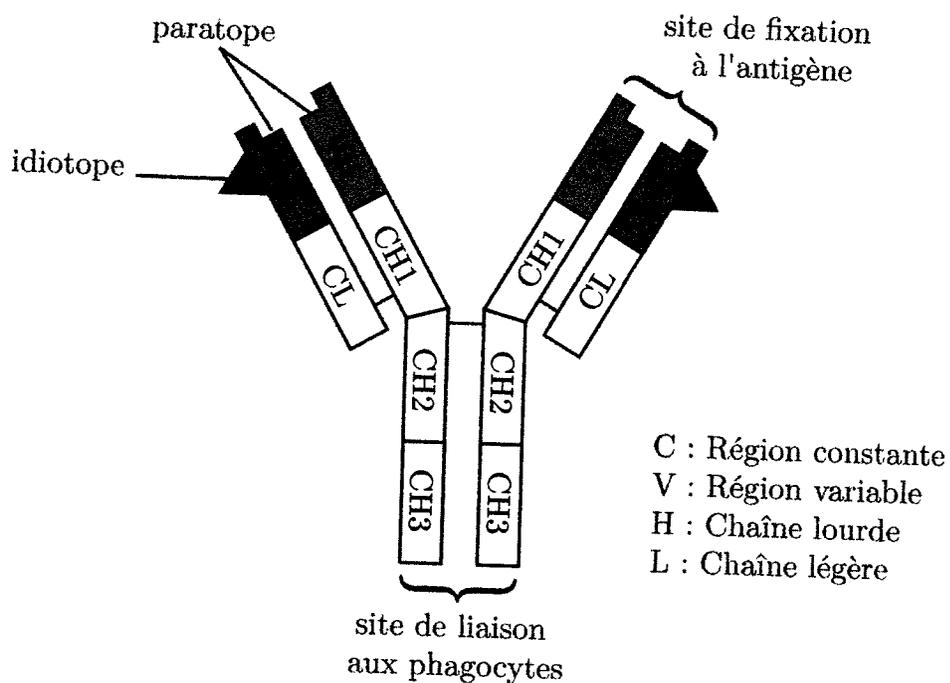


Figure 2.3 : Structure d'un anticorps

### 2.1.3. La réaction anticorps/antigène

La réaction antigène-anticorps dépend du point de vue thermodynamique de l'affinité.

L'affinité d'un anticorps mesure la force de liaison entre le paratope de cet anticorps et l'épitope de l'antigène. Elle est la somme des forces attractives et répulsives opérant

entre le déterminant antigénique et le site anticorps telle qu'elle est illustrée dans la Figure suivante

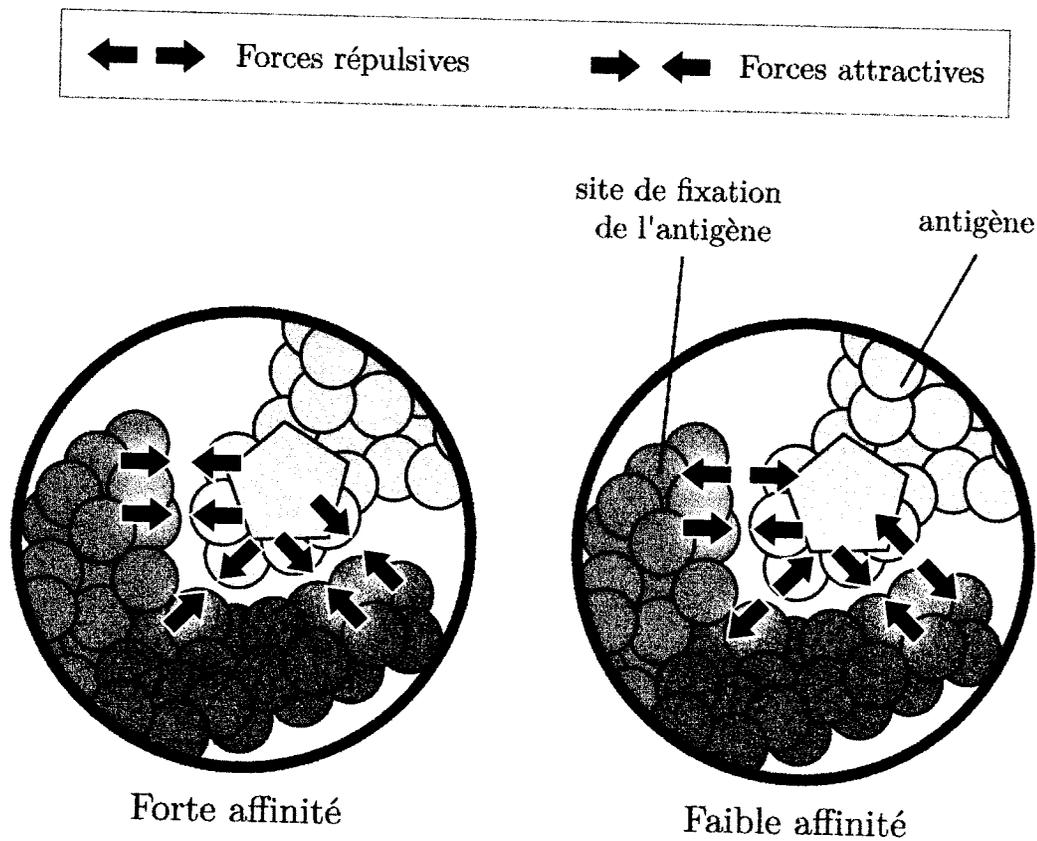


Figure 2.4 : Les forces contribuant à l'affinité de l'interaction anticorps/antigène

#### 2.1.4. Immunité innée et immunité adaptative

La défense de l'organisme contre le milieu extérieur comporte une immunité dite innée ou naturelle, c'est-à-dire existante en absence de tout contact avec un antigène, et une immunité dite adaptative ou acquise, c'est-à-dire apparaissant après contact de l'organisme avec des molécules étrangères qui sont des antigènes [ALL, 06].

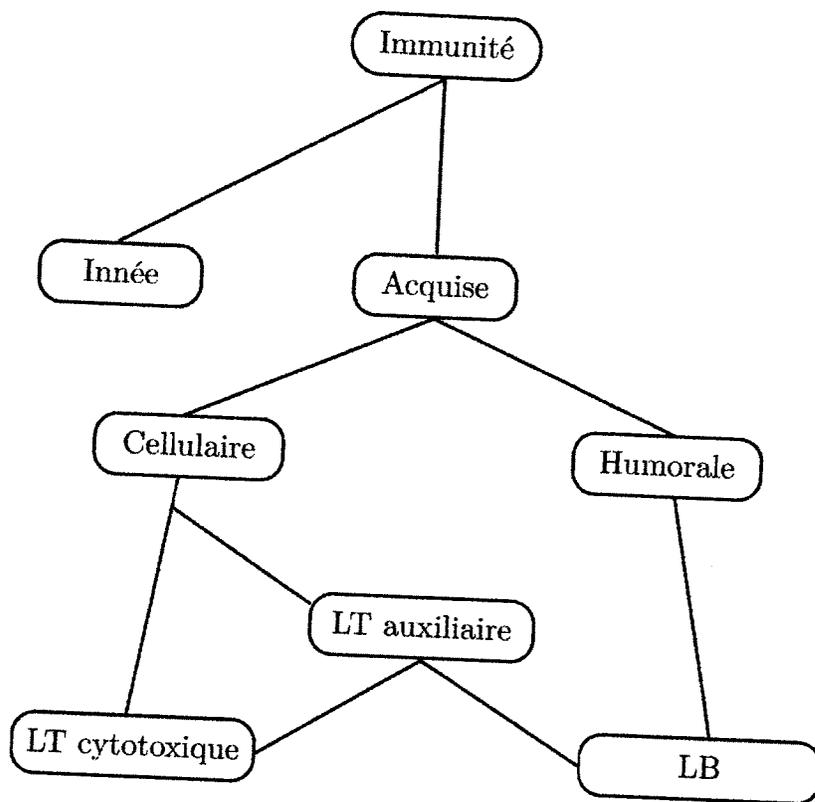


Figure 2.5 : Architecture du système immunitaire

#### 2.1.4.1 Immunité innée

Deux principaux niveaux existent, la barrière anatomique (la peau et le système respiratoire), l'immunité innée (non spécifique).

L'immunité innée est la première ligne de défense vis-à-vis des agents infectieux et pathogènes qui nous entourent. Elle existe depuis la naissance de l'être humain et elle reste invariable tout au long de sa vie. Elle définit les grandes lignes de défenses. Elle est déclenchée en premier lieu dès la pénétration d'un corps étranger, et elle agit sans se soucier du type de ce dernier.

Dans la plupart des cas, l'immunité innée est capable de détruire les différentes pathogènes au premier contact.

#### 2.1.4.2 Immunité adaptative

L'immunité adaptative résulte du contact du système immunitaire avec les antigènes grâce à la caractéristique d'apprentissage et mémorisation du système immunitaire qui sera détaillée plus loin. La première intrusion d'un antigène entraîne

une réponse lente et une réaction difficile du système immunitaire, cependant elle permet de mémoriser l'antigène grâce à ses marqueurs (épitopes) FIG 2.2. Si le même antigène pénètre une seconde fois le corps, la réponse sera plus rapide et bien spécifique avec production d'anticorps particuliers pour cet antigène. Les lymphocytes T, les lymphocytes B et les immunoglobulines constituent les principaux acteurs de l'immunité adaptative. L'immunité adaptative est dite immunité à mémoire [KHE, 06]. La réponse de l'immunité adaptative est lancée après la réponse de l'immunité innée, les deux types d'immunité sont liés et se complètent.

### **2.1.5. Réactions immunitaires**

Les deux grandes lignes de défense du système immunitaire étant l'immunité innée et l'immunité adaptative. Chacune emploie un certain nombre de méthodes et met en œuvre des organes complémentaires. Cependant, la plus efficace et plus précise est sans doute l'immunité adaptative (spécifique) qui réagit selon l'antigène en question. Selon le type d'infection, la réponse spécifique se décompose en :

#### **2.1.5.1 Immunité humorale**

Capable de neutraliser les bactéries et les toxines, elle est assurée par le transfert d'humeurs (vaccin ou sérum) d'un patient à un autre (ou d'un animal expérimental à un autre). Généralement l'immunité humorale est utilisée pour les maladies infectieuses [MOU, 07].

#### **2.1.5.2 Immunité cellulaire**

Nécessite le transfert de cellules. L'immunité cellulaire est transmise par des immunoglobulines synthétisées par des cellules [MOU, 07].

Les phénomènes immunitaires à médiation cellulaire impliquent différents types de cellule, regroupés dans deux concepts: les cellules de l'immunité innée (les lymphocytes NK, les granulocytes, les macrophages, les cellules dendritiques) et celles de l'immunité adaptative (les lymphocytes T).

### 2.1.6. Propriétés du système immunitaire

Le système immunitaire est une source d'inspiration pour les nouvelles branches de l'informatique. Avec des propriétés très importantes, il devient une référence précieuse. Beaucoup de travaux de recherches ont vu le jour en s'inspirant du fonctionnement de ce dernier, voici quelques unes des propriétés les plus importantes du système immunitaire :

- **Multicouche** : Le système immunitaire possède une architecture multicouche qui consiste en deux sous systèmes inter-liés qui sont le système immunitaire inné et le système immunitaire adaptatif. Ces deux systèmes combinent leurs tâches et responsabilités pour assurer la protection et la sécurité globale.
- **Unicité** : Chaque élément dans le système immunitaire assume des responsabilités particulières.
- **Autonomie** : Le système immunitaire humain ne possède aucun contrôle central ou un gestionnaire particulier. Il possède une autonomie globale dans la détection et l'élimination des intrus.
- **Distribution** : Les cellules immunitaires et les molécules sont distribuées dans le corps humain pour assurer sa protection. Il n'existe pas un point de contrôle centralisé.
- **Parallélisme** : Le système immunitaire est capable de produire plusieurs réponses immunitaires en même temps à des endroits dispersés.
- **Tolérance au soi** : Le système immunitaire humain peut différencier entre les cellules de soi et les cellules de non-soi.
- **Apprentissage** : Le système immunitaire augmente la capacité d'identification des anticorps à un antigène sélectif (les réponses primaire et secondaire). Il apprend continuellement les structures de pathogènes.

- **Adaptabilité** : Le système immunitaire humain permet la production des cellules de plus en plus spécialisées pour l'identification des antigènes. Cela est garanti par la théorie de la sélection clonale suivie par le mécanisme de l'hypermutation somatique.
- **Dynamique** : Le système immunitaire change continuellement par la création de nouvelles cellules et molécules, l'élimination des cellules vieilles ou endommagées. Un bon exemple de la dynamique du système immunitaire est la théorie du réseau idiotypique.
- **Diversité** : Le système immunitaire naturel est capable d'identifier n'importe quels intrus qui envahissent le corps. Le répertoire cellulaire est complet. Cette particularité est due à plusieurs mécanismes qui sont par exemple : l'hypermutation somatique, la reproduction de récepteur, l'identification approximative des intrus, etc.
- **Mémorisation** : Après une réponse immunitaire à un antigène donné, un ensemble de cellules constituent l'ensemble de cellules mémoires qui seront dotées par une durée de vie longue afin de fournir des réponses immunitaires plus rapides et plus puissantes aux rencontres suivantes du même antigène.
- **Coopération** : Les cellules immunitaires coopèrent leurs capacités pour assurer une meilleure détection et également une protection puissante par exemple les cellules T d'aide, les molécules MHC, etc.
- **Détection** : Le système immunitaire est capable d'identifier et détecter les intrus dans le corps sans aucune connaissance antérieure de la structure de ces intrus.

#### 2.1.7. Théories immunitaires

Le comportement et les réactions du système immunitaire sont principalement régi par des théories immunitaires.



### **2.1.7.1 Théorie de la sélection Négative/Positive**

Le thymus est une barrière contre les antigènes du non-soi. Les cellules T qui reconnaissent les antigènes du soi, sans les lier fortement sont sélectionnées positivement. Les cellules ne remplissant pas ces critères de sélection, soit parce qu'elles lient trop fortement les antigènes du soi ou parce qu'elles ne les reconnaissent pas, sont détruites dans cet organe. Une fois les deux sélections terminées avec succès les lymphocytes deviendront des cellules immunocompétentes circulant dans le sang et colonisant les organes lymphatiques secondaires.

### **2.1.7.2 Théorie de la sélection clonale**

Une autre théorie tout aussi importante que la sélection négative est la théorie de la sélection clonale. Cette théorie a été exposée en 1959 par Burnet [CAS, 03]. Elle met en avant la réaction du système immunitaire à une stimulation antigénique [CAS, 02].

Dans le sang, nous possédons déjà des millions de lymphocytes qui sont spécifiques pour des millions d'antigènes différents. La plupart de ces lymphocytes ne rencontreront jamais l'antigène qu'ils peuvent reconnaître.

Mais que se passe-t-il lorsque nous rencontrons un antigène commun, par exemple le virus de la grippe? Le processus que nous appelons sélection clonale permettra de sélectionner les lymphocytes qui sont spécifiques pour les antigènes du virus, et ainsi, les bons anticorps seront produits.

Les différentes étapes de la sélection clonale en prenant l'exemple d'un antigène présent dans le sang, et qui sera reconnu par un lymphocyte B :

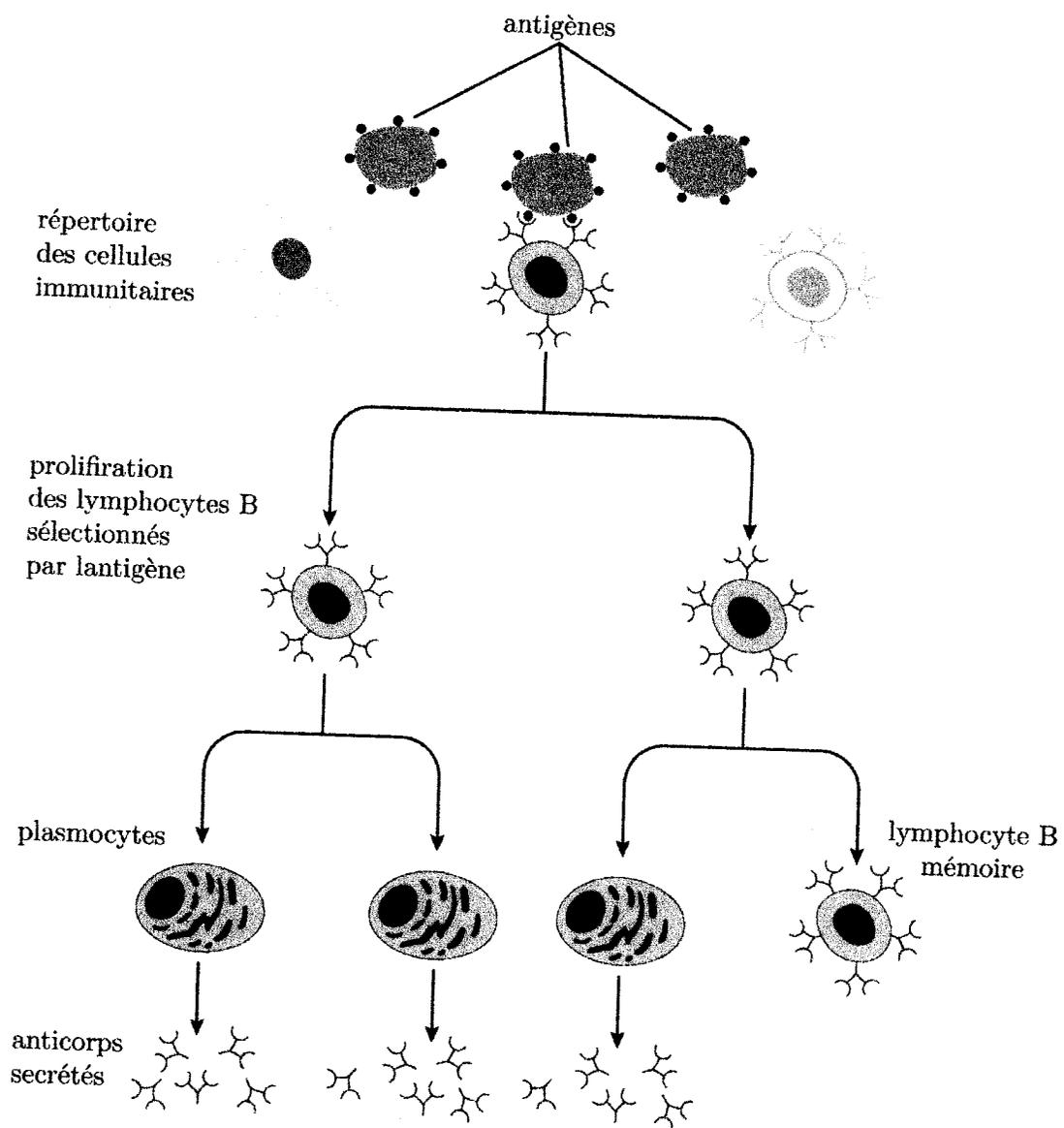


Figure 2.6 : Processus de la sélection clonale

- 1) L'antigène se fixe au récepteur antigénique du lymphocyte B qui est spécifique pour cet antigène.
- 2) Le lymphocyte B sélectionné se multiplie par une série de mitoses (formation de clones).
- 3) Une armée de lymphocytes B spécifiques pour l'antigène est ainsi produite.
- 4) Les lymphocytes B sont activés par les lymphocytes T auxiliaires.

5) Les lymphocytes B activés se différencient en plasmocytes sécréteurs d'anticorps. (Les anticorps sont semblables au récepteur antigénique du lymphocyte B, donc ils sont eux aussi spécifiques pour l'antigène de départ.)

6) Une certaine proportion des lymphocytes B activés se transformera en lymphocytes B mémoire.

Les lymphocytes mémoires restent très longtemps dans le sang, contrairement aux plasmocytes. De plus, les lymphocytes mémoires sont plus rapidement activés si l'antigène rentre à nouveau dans le sang, et les anticorps qu'ils produisent sont plus efficaces. C'est grâce aux lymphocytes mémoire que l'on n'attrape pas deux fois la varicelle. C'est ce qu'on appelle la mémoire immunitaire.

### **2.1.7.3 Théorie des réseaux immunitaires**

La sélection clonale est le moyen utilisé par le système immunitaire pour vaincre des antigènes du non-soi tandis que la sélection négative lui permet de se protéger contre des antigènes du soi. Le réseau immunitaire quant à lui définit la façon dont les cellules réagissent entre elles dans le système immunitaire. Le système immunitaire d'un individu donné maintient un certain état d'équilibre. Cet état d'équilibre se trouve perturbé par une sollicitation immunogène qui provoque la multiplication et différenciation des cellules immunologiques. Cet état d'excitation n'est que provisoire et progressivement un nouvel état d'équilibre est trouvé enrichi de la mémoire de la sollicitation passée.

Cette théorie du réseau immunitaire, a été proposée par Jerne en 1974. Elle est très difficile à prouver expérimentalement, c'est pourquoi elle fut souvent l'objet de réfutations de la part de quelques immunologistes.

Le procédé de cette interaction est assuré par des récepteurs spécialisés présents sur la surface des anticorps appelés « idiotope ». Alors, le système immunitaire est formellement défini par un réseau énorme et complexe de paratopes qui reconnaissent

un ensemble d'idiotopes. Ainsi, chaque élément pourrait reconnaître aussi bien qu'être reconnu. Ce réseau est appelé le réseau immunitaire idiotypique.

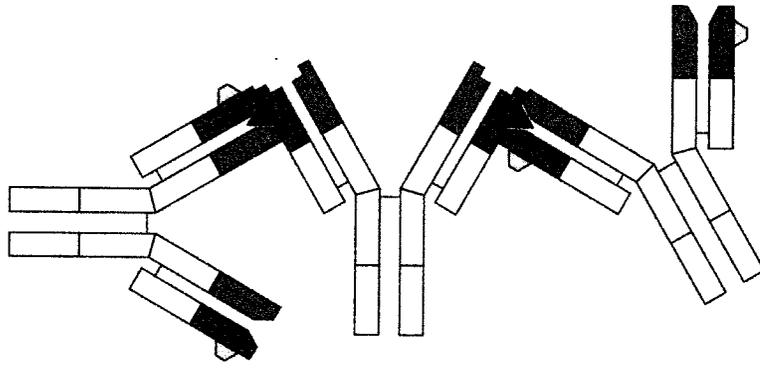


Figure 2.7 : représentation du réseau immunitaire idiotypique

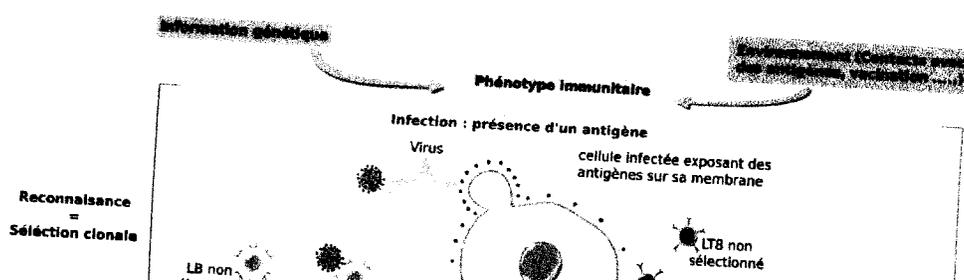
#### 2.1.7.4 Théorie du danger

Cette théorie qui est proposée en 1994 par Matzinger et qui propose des nouvelles conditions pour le déclenchement de la réponse immunitaire. Elle propose que la réponse immunitaire soit déclenchée suite à l'existence de danger et non suite à l'existence d'un élément étranger.

Nous constatons que dans la théorie du danger il ya une corrélation entre les signaux d'alarmes. Il ne suffit pas de détecter les cellules étrangères, il faudrait également savoir si ces dernières sont dangereuses ou pas [AIC, 03] .

#### 2.1.8. Déroulement d'une réponse immunitaire

On appelle réponse immunitaire l'activation des mécanismes du système immunitaire face à la reconnaissance de « non-soi », agressive ou pas, face à une agression ou à une dysfonction de l'organisme. L'ensemble de ces systèmes permet la résilience immunitaire : notion qui recouvre la somme des mécanismes efficaces de défense d'un organisme vis-à-vis d'un agent pathogène.



## 2.2. Systèmes immunitaires artificiels (SIAs)

### 2.2.1. Historique

Les travaux sur les SIAs ont apparus [BOR, 05] au milieu des années 80 avec l'article de Farmer, Packard et Perelson sur les réseaux immunitaires (1986). Cependant, c'est seulement dans le milieu des années 90 que les SIAs deviennent un sujet à part entière. Les travaux de Forrest sur la sélection négative commencèrent en 1994, tandis que Dasgupta menait des études sur les algorithmes de sélection négative. Hunt et Cooke commencèrent leurs travaux sur les modèles de réseaux immunitaires en 1995. Timmis et Neal continuèrent ces travaux en y apportant des améliorations.

Le premier livre sur les Systèmes Immunitaires Artificiels a été écrit par Dasgupta en 1999. Les travaux de De Castro & Von Zuben et Nicosia & Cutello sur la sélection clonale (CLONALG) furent remarqués en 2002. De nouvelles voies, comme la théorie du danger (observation des dégâts plutôt que celle des agents pathogènes) et des algorithmes inspirés par le système immunitaire inné ont également été explorés. Le fait qu'elles apportent quelque chose de nouveau au delà des algorithmes existants est actuellement le sujet de débats qui animent le développement des SIAs.

Au départ, les travaux sur les SIAs visaient à trouver des abstractions efficaces des phénomènes découverts dans le système immunitaire. Plus récemment, les praticiens des SIAs se sont aussi intéressés à la modélisation du système immunitaire et à l'application des résultats issus des SIAs aux problèmes d'immunologie (ce qui entre dans le cadre de l'immuno-informatique) [DAS, 05].

### 2.2.2. Définitions des SIAs

Selon Timmis : « Un système immunitaire artificiel est un système informatique basé sur les métaphores du système immunitaire naturel ».

Dasgupta a défini le système immunitaire artificiel comme suit : « Le système immunitaire artificiel est la composition de méthodologies intelligentes inspirées par le système immunitaire naturel afin de résoudre des problèmes du monde réel ».

Tandis que Timmis et De Castro ont donné la définition suivante : « Les systèmes immunitaires artificiels sont des systèmes adaptatifs inspirés par des théories immunologiques et des observations de fonctions immunitaires, de ses principes et des modèles, qui seront appliqués à la résolution des problèmes ».

Un système immunitaire artificiel (SIA) est une catégorie d'algorithme inspirée par les principes et le fonctionnement du système immunitaire naturel des vertébrés. Ces algorithmes exploitent typiquement les caractéristiques du système immunitaire pour ce qui est de l'apprentissage et de la mémorisation comme moyens de résolution de problèmes.

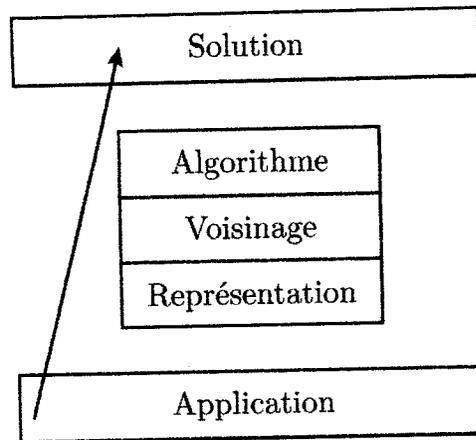
### **2.2.3. Modélisation des systèmes immunitaires artificiels**

Dans le monde de l'ingénieur, il est devenu commode de définir des principes et des processus de base ainsi que le développement de processus génériques aidant à la résolution de problèmes dans un domaine donné.

De Castro et Timmis [TIM, 03], propose un framework pour simplifier le processus de correspondance entre les systèmes immunitaires biologiques et les systèmes immunitaires artificiels. Le corps de ce Framework est composé de trois éléments :

1. Représentation des composants du système : espace de forme, binaire, entier, réel, symbolique.
2. Un ensemble de mécanismes pour l'évaluation de l'interaction de composants avec l'environnement et entre eux (une fonction de l'affinité) : Euclidienne, Manhattan, Hamming.
3. Procédure d'adaptation qui gouverne la dynamique du système : sélection clonale, sélection négative, sélection positive, modèle de réseau immunitaire.

Les différentes étapes sont illustrées dans La Figure 2.9.



Figures 2.9 : Un framework de SIAs pour l'ingénierie et sa structure multi niveau [TIM, 03].

Ce framework utilise une approche multi niveau, dont le domaine d'application influe sur la représentation. Cette dernière influe sur la mesure de l'affinité. Cette approche multi-niveaux a été ensuite généralisée par Stepney et al. Un framework conceptuel générique a été proposé [STE, 04].

#### 2.2.4. Les mesures d'affinités

Les SIAs utilisent la notion "D'espace de forme" [PER, 79] pour décrire les interactions entre les cellules et les antigènes. Chaque antigène et anticorps sont représentés par un point dans "l'espace de forme". La distance entre ces points définit donc l'affinité entre l'antigène et l'anticorps. Il existe deux méthodes différentes pour calculer l'affinité suivant sa représentation.

La première, nommée recognition ball, consiste à calculer la distance entre chaque épitope de l'antigène et les récepteurs qui les ont détectés. Si cette distance est comprise dans le rayon  $\varepsilon$  alors l'affinité est considérée comme suffisante pour générer une réponse immunitaire. Cette distance peut être obtenue en utilisant les distances Euclidienne, de Manhattan ou de Hamming.

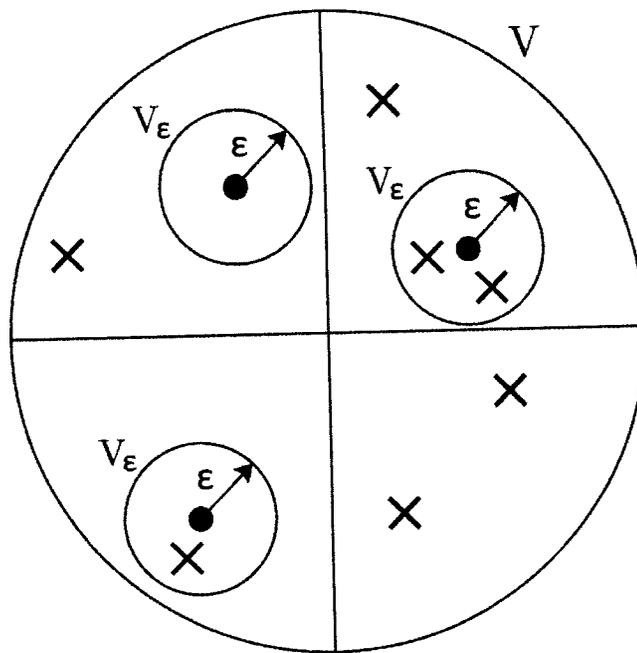


Figure 2.10 : Espace des formes [CAS, 01]

Et les différentes distances pour le calcul d'affinité anticorps-antigène sont les suivantes :

La distance Euclidienne : 
$$D = \sqrt{\sum_{i=1}^L (Ab_i - Ag_i)^2}$$

La distance de Manhattan : 
$$D = \sum_{i=1}^L |Ab_i - Ag_i|$$

La distance de Hamming : 
$$D = \sum_{i=1}^L \delta_i, \text{ avec } \delta_i = \begin{cases} 1 & \text{si } Ab_i \neq Ag_i \\ 0 & \text{sinon} \end{cases}$$

La seconde méthode exposée dans [TIM, 02] est la correspondance r-contigües, c'est-à-dire « deux éléments ont la même longueur de correspondance si au moins r caractères contigües sont identiques ». Dans l'exemple qui suit, il y a deux vecteurs de huit paramètres dont leur valeur est soit 0 ou 1. Le calcul de l'affinité se fait en appliquant un ou exclusif entre les deux vecteurs. L'affinité entre ces deux vecteurs est 4 car il y a 4 valeurs contigües à 1.

0	0	1	1	0	0	1	1
0	1	1	0	1	1	0	1
1	1	0	1	1	1	1	0

Figure 2.11 : Affinité en nombre de bits contigus

### 2.2.5. Algorithmes immunitaires

Les systèmes immunitaire artificiels (SIAs) sont l'étude des modèles et des abstractions des systèmes immunitaires du vertébré (mammifères en général) et l'application de ces modèles et algorithmes dans les domaines de l'intelligence artificielle (classification, optimisation...). Bien que la source d'inspiration pour les modèles informatiques dans le système immunitaire est quasiment illimité, quatre grands sous domaines de recherche sont apparus dans les SIAs : les algorithmes de sélection négative (NSA), les algorithmes de réseau immunitaire (INA), les algorithmes de la théorie de danger (DTA), et des algorithmes de sélection clonale (CSA).

#### 2.2.5.1 Algorithme de la sélection négative

D'un point de vue informatique, l'algorithme de la sélection négative est basé sur l'idée de générer un ensemble de détecteurs et de comparer l'affinité qu'ils ont avec chaque élément du soi [TIM, 02]. L'action suivant la détection d'un élément du non-soi varie selon le problème que l'on souhaite résoudre. Cette discrimination du soi/non-soi impose donc une connaissance des paramètres des éléments.

Le premier algorithme de la sélection négative a été proposé par [FOR, 94] afin de détecter les anomalies causées par des virus dans un système informatique. Cet algorithme se base sur deux phases principales :

- Construction d'un ensemble de détecteurs : cette phase consiste à générer un ensemble de détecteurs « D » aléatoirement. Ensuite, éliminer les détecteurs qui reconnaissent les éléments du soi « S ».
- Mise en place des détecteurs : les détecteurs construits dans la première phase peuvent être utilisés pour classer les nouvelles données.

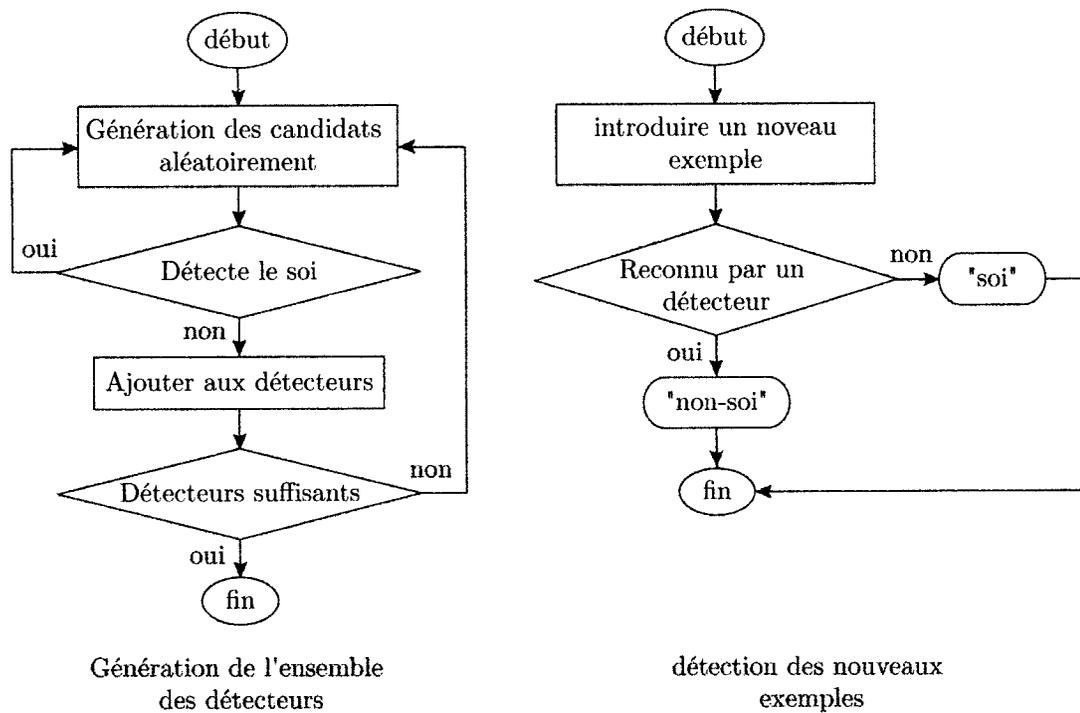


Figure 2.12 : Algorithme de sélection Négative en deux phases

### 2.2.5.2 Algorithme de la sélection positive

L'algorithme de la sélection positive est proposé par Forrest et al [FOR, 94]. Cet algorithme est la solution alternative de l'algorithme de la sélection négative. La différence principale est la génération des détecteurs qui détectent des éléments de soi au lieu la génération des détecteurs qui détectent des éléments de non-soi. Selon cet algorithme, un élément de non-soi suspect doit être comparé avec tout l'ensemble des détecteurs de soi, s'il n'est pas détecté alors il est considéré comme un élément de non-soi.

### 2.2.5.3 Algorithme de la sélection clonale

La théorie de la sélection clonale est le principe fondamental de l'immunologie moderne, il est donc étroitement interconnecté avec d'autres théories immunologiques. Cela fait suite à des algorithmes inspirés par ces théories. Par exemple, les algorithmes de sélection négative se prévaloir du principe de la sélection clonale afin d'améliorer les exemplaires d'une manière itérative. Les algorithmes de regroupement et d'optimisation basée sur la théorie des réseaux immunitaires utilisent le principe de la sélection clonale de raffinement itératif de leurs modèles.

Un algorithme de sélection clonale est principalement concentré sur le principe clonal de choix qui se compose de : choix clonal, expansion clonale, et maturation d'affinité par l'intermédiaire de hypermutation somatique.

Cutello et Nicosia [VIN, 05] proposent que les algorithmes de la sélection clonale prennent en considération deux mécanismes principaux; l'hypermutation et l'expansion clonale. Ils décrivent l'hypermutation comme une procédure de recherche locale mène à une maturation rapide et que la phase d'expansion clonale déclenche la croissance d'une nouvelle population des B-cellules utiles concentrées sur l'antigène de déclenchement. Ils proposent également que la réponse immunitaire primaire peut être considérée comme phase d'apprentissage, tandis que la réponse secondaire peut être prise comme une phase de test.

De Castro et Timmis [TIM, 02] proposent également que les deux mécanismes principaux des algorithmes de la sélection clonale sont les propriétés de la mutation et de clonage.

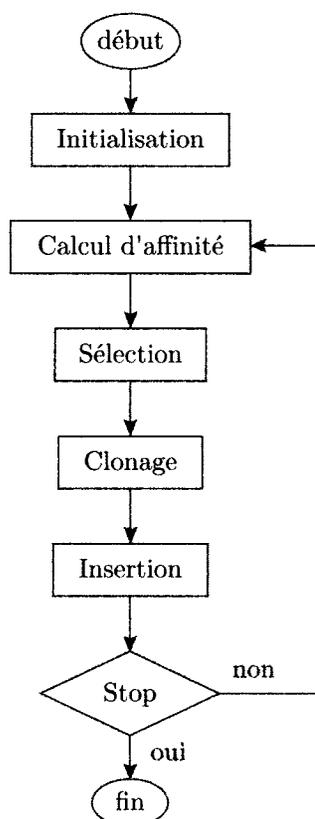


Figure 2.13 : Représentation de l'algorithme de la sélection clonale

Les algorithmes de la sélection clonale sont affiliés à différentes catégories :

<b>Filière</b>	<b>Algorithmes</b>	<b>Domaine d'utilisation</b>
AIRS	AIRS, AIRS2, Parallel AIRS	Classification
BCA	BCA	Optimisation
CLONALG (CSA)	CSA, CLONALG, CLONALG (1, 2), ACS, CLONCLAS, RCSA, MOCSA, IMCSA, AISMM, SACSA, ECA	Optimisation
IA (SIA)	IA, SIA, I-PAES, CLIGA, CLIGA+, NC- IA, READ-Alg, opt-IA, opt-IMMALG, Par-IA, Dyn-IMMALG	Optimisation
MISA	MISA	Multi-objective Optimisation

Tableau 2.1 : Algorithme de sélection clonale

Dans le rapport technique sur les applications des systèmes immunitaires artificiels De Castro et Von Zuben [CAS, 02] ont proposé l'algorithme de la sélection clonale (CSA) comme implémentation informatique du principe de la sélection clonale pour le groupement et l'optimisation. Cet algorithme qui est devenu peut-être le plus populaire dans le domaine de l'SIA, où il a été nommé CLONALG (CLONal selection ALGORITHM).

Parameter	Description
P	Répertoire des anticorps
N	taille de répertoire des anticorps
n	Le nombre d'anticorps à choisir pour le clonage
L	Longueur de chaîne binaire pour chaque anticorps
Nc	Nombre de clones créés par chaque anticorps choisi
d	Nombre d'anticorps aléatoires à insérer à la fin de chaque génération. ils remplacent les anticorps qui ont une faible affinité dans le répertoire.
Stop condition	Le nombre spécifique de générations ou d'évaluations de fonction.
affinity	L'évaluation de solution, typiquement la solution est décodée selon une représentation spécifique de domaine et a assigné un calcul des coûts de qualité
clone	Duplication d'une chaîne binaire.
hypermutate	Modification d'une chaîne binaire où le basculement d'un peu est régi par une distribution de probabilité proportionnée d'affinité

Tableau 2.2 : CLONALG paramètres

```
P <- rand(N, L)
tanque Non StopCondition faire
  pour Chaque p of P faire
    Affinity(p)
  Fin faire
P1 <- select(P, n)
pour Chaque p1 op P1 faire
  C <- clone(p1)
Fin faire
Pour chaque c of C faire
  Hypermutate(c)
Fin faire
Pour chaque c of C faire
  Affinity(c)
Fin faire
P <- insert(C, n)
Pr <- rand(d, L)
P <- replace(P, d, Pr)
Fin faire
```

Figure 2.14 : CLONALG pseudo code

Après CLONALG, AIRS (the Artificial Immune Recognition System) est le deuxième algorithme populaire de la sélection clonale, malgré que cet algorithme a été conçu seulement pour les problèmes de classification. Les premiers travaux sur AIRS étaient dans la thèse de Watkins [WAT, 01] . AIRS est un algorithme pour la classification supervisé, il utilise l'idée d'une boule artificielle d'identification (ARB) présentée dans les premiers travaux sur l'algorithme de réseau immunitaire (AINE).

```

memory = initialiseRandomRepertoire()
Pour chaque (x,y) dans trainingData faire
    best = memory.bestMatchingCell(x,y)
    pool = [best]
    Tanque pool.avgStimulation() < threshold faire
        Pour chaque cell dans pool faire
            pool.add(cell.mutations(cell.stimulation(x)))
        fin
    pool.cellsCompeteForResources()
fin
fit = pool.fittestCell()
Si fit > best Alors memory.add(t)
    Si ||best-fit|| < ε Alors
        memory.remove(best)
    Fin
Fin
Fin

```

Figure 2.15 : AIRS pseudo code (apprentissage)

Kelsey et Timmis [KEL, 03] ont proposé l'algorithme BCA (B-Cell Algorithm) comme un SIA conçu pour l'optimisation des fonctions continues. L'algorithme maintient un ensemble de cellules-B qui sont soumises au clonage et à la mutation.

Paramètre	Description
P	Répertoire d'anticorps
N	Nombre de population d'anticorps
Nc	Nombre de clones pour chaque anticorps
nR	Nombre aléatoire d'anticorps à créer et insérer pour chaque génération
Stop condition	Typiquement si aucun progrès n'est accompli pour un certain nombre de générations

hypermutation	Emploie des processus appelés l'hypermutation contiguë, un endroit aléatoire dans la chaîne binaire est choisi, et une longueur aléatoire de sub-corde est choisie. Chaque peu dans la sous-chaîne est renversé avec la probabilité P
replace	Un parent est remplacé seulement si un membre de son clone a une forte affinité.

Tableau 2.3 : BCA paramètres

```

P <- rand(N, L)
Tanque non StopCondition faire
  Pour chaque p of P faire
    Affinity(p)
  Fin Faire
  Pour chaque p op P faire
    C <- clone(p, Nc)
    C <- rand(nR, L)
    Pour chaque c of C faire
      Hypermutation(c)
    Fin faire
    Pour chaque c of C faire
      Affinity(c)
    Fin faire
    C' <- best(C)
    P <- replace(c, p)
  Fin faire
Fin faire

```

Figure 2.16 : BCA pseudo code

Un algorithme inspiré de la sélection clonale a été proposé par Cutello et Nicosia cet algorithme est appelé IA (Immunological Algorithm), et puis il a été renommé plus tard par SIA (Simple Immunological Algorithm) [GIU, 02]. L'algorithme maintient une population des cellules-B qui sont exposées à un processus d'expansion clonale. Ce processus d'expansion implique le clonage des cellules et l'application d'un opérateur de hypermutation.

Paramètre	Description
P	Population d'anticorps
l	Longueur de la représentation binaire
d	Taille de population (répertoire)
dup	Le nombre de clones créés pour chaque anticorps
clone	Duplication de la chaîne binaire
hypermutation	La modification probabiliste d'une chaîne binaire

Tableau 2.4 : SIA paramètres

```

P <- rand (N, L)
Pour chaque p of P faire
  Affinity (p)
Fin faire
Tanque non StopCondition faire
  Pour chaque p op P faire
    Pc <- clone(p, dup)
  Fin faire
  Pour chaque c of Pc faire
    Hypermutation(c)
  Fin faire
  Pour chaque c of Pc faire
    Affinity(c)
  Fin faire
  P <- select (Pc, P, d)
Fin faire
    
```

Figure 2.17 : SIA pseudo code

#### 2.2.5.4 Algorithme du réseau immunitaire

Il existe deux grands modèles de réseaux immunitaires artificiels : un modèle continu et un autre discret. Comme tout modèle continu, le principal obstacle à son

utilisation est de lui trouver une solution analytique, ce qui n'est pas forcément possible. Les modèles discrets sont généralement adaptatifs, que ça soit dans le nombre d'éléments les constituant ou dans la forme de ces éléments d'accroître leur affinité.

La théorie du réseau immunitaire est utilisée surtout pour produire un système de reconnaissance et de classification des formes. Parmi les Algorithmes qui sont inspirés par les réseaux immunitaires on trouve : RAIN (Resource-limited Artificial Immune Network) [NEA, 01] et aINET (artificial immune NETwork) [CAS, 00].

En se basant sur cette théorie, Timmis & Neal proposent en 2000 l'algorithme d'apprentissage du réseau immunologique RAIN, qui utilise une équation pour déterminer des niveaux de stimulation pour chaque cellule dans une population durant chaque itération de l'algorithme.

- 1. Initialisation:**  
initialiser un réseau de cellules immunitaires
- 2. Pour chaque antigène**
  - Reconnaissance antigénique:**  
faire correspondre les cellules immunitaires et l'antigène.
  - Interactions Réseaux:**  
faire correspondre les cellules immunitaires entre elles.
  - Meta dynamisme:**  
introduire des nouvelles cellules dans le réseau et éliminer les moins utilisées (selon un critère défini).
  - Niveau de Stimulation:**  
évaluer le niveau de stimulation de chaque cellule de réseau en prenant en compte les résultats des étapes précédentes.
  - Dynamisme de Réseaux:**  
MAJ de la structure du réseau et libérer les paramètres selon le niveau de stimulation
- 3. Répéter** la deuxième étape **jusqu'à ce que** le critère de convergence soit atteint

Figure 2.18 : RAIN pseudo code

aiNet comme vu précédemment, est un système immunitaire artificiel s'inspirant de la théorie des réseaux immunitaires et du principe de la sélection clonale. On peut décomposer le fonctionnement de aiNet en deux parties :

- La première est l'application de "CLONALG", un algorithme inspiré de la sélection clonale immunitaire, qui génère des anticorps.
- La seconde étape repose sur l'interaction des anticorps, et leur capacité à se reconnaître les uns les autres dans un réseau immunitaire artificiel. De cette manière, si deux cellules sont trop similaires, on les supprime pour éviter la redondance de données. Ensuite, on introduit de nouvelles cellules dans le réseau afin de forcer la diversification des anticorps.

**Initialization:**  
create an initial random population of network antibodies;

**Antigenic presentation:**  
**for each** antigenic pattern, **do:**

**Clonal selection and expansion:**  
    **for each** network element, determine its affinity with the antigen presented. Select a number of high affinity elements and reproduce (clone) them proportionally to their affinity;

**Affinity maturation:**  
    mutate each clone inversely proportional to affinity. Re-select a number of highest affinity clones and place them into a clonal memory set.

**Clonal interactions:**  
    determine the network interactions (affinity) among all the elements of the clonal memory set.

**Clonal suppression:**  
    eliminate those memory clones whose affinity is less than a pre-specified threshold;

**Metadynamics:**  
    eliminate all memory clones whose affinity with the antigen is less than a pre-defined threshold;

**Network construction:**  
    incorporate the remaining clones of the clonal memory with all network antibodies;

**Network interactions:**  
    determine the similarity between each pair of network antibodies

**Network suppression:**  
    eliminate all network antibodies whose affinity is less than a pre-specified threshold;

**Cycle:**  
repeat steps 2 to 4 until a pre-specified number of iterations is reached

Figure 2.19 : aiNET pseudo code

### 2.3. Conclusion

Les algorithmes s'inspirant du système immunitaire biologique, peuvent s'appliquer à un très grand nombre de domaines de recherche. Les modèles les plus utilisés de cette approche sont la sélection clonale, la sélection négative et les réseaux immunitaires, d'autres modèles comme la théorie de dangers et les bibliothèques des gènes restent en cours de développement et sont moins utilisés. Il n'existe pas d'algorithmes standards bien définis pour chaque modèle. Souvent, les implémentations varient selon le problème traité.

Nous présentons dans le chapitre suivant un modèle immunologique basé sur la théorie de la sélection clonale (modified-AIRS) pour la classification des cellules tumorales du sein.

## **Chapitre 3**

# **Classification des cellules tumorales à base de m-AIRS**

Dans le présent chapitre, nous abordons le problème de la classification automatique en utilisant le principe des systèmes immunitaires l'une des plus récentes approches. Notre but est de se bénéficier des capacités de ce système pour pouvoir concevoir un système intelligent de classification.

Notre travail est basé sur l'algorithme AIRS, inspiré du principe de la sélection clonale, développé par A. Watkins [WAT, 01]. Nous essayerons d'améliorer l'efficacité et l'efficacité de cet algorithme en intervenant principalement au niveau de l'algorithme de mutation et la méthode de classification.

Ce chapitre présente aussi une comparaison des résultats obtenus avec les travaux réalisés dans la littérature.

### **3.1. La base de Wisconsin**

La base de données « Wisconsin Breast Cancer Database » (WBCD) a fait l'objet de nombreuses études dans le domaine du diagnostic du cancer du sein. La thématique de recherche des créateurs de cette base de données, Wolberg, Street, Heisey et Managasarian, était l'automatisation du diagnostic par le traitement d'images et l'apprentissage automatique.

Au-delà du domaine médical, statisticiens et informaticiens ont proposé un large éventail de modèles de classification basé sur WBCD. Ces nouvelles méthodes ont constamment amélioré la référence en termes de performance du diagnostic.

Notre choix de la base de Wisconsin a été motivé par cette large utilisation ce qui nous permet de se situer par rapport aux autres travaux et de comparer nos résultats avec d'autres.

Cette base de données de cancer du sein a été créée par le travail clinique de Dr. William H. Wolberg à l'université de Wisconsin à Madison en 1992.

La base est constituée de 699 instances distribuées sur deux classes : 458 cas bénins (65.5%) et 241 cas malins (34.5%). Parmi les 699 instances, il y a 16 incomplètes.

Chaque instance comporte une valeur pour chacun des 11 attributs (Tableau 3.1).

N°	Code	Attribut	Domaine
01	Id	Sample code number	id number
02	att1	Clump Thickness	1 - 10
03	att2	Uniformity of Cell Size	1 - 10
04	att3	Uniformity of Cell Shape	1 - 10
05	att4	Marginal Adhesion	1 - 10
06	att5	Single Epithelial Cell Size	1 - 10
07	att6	Bare Nuclei	1 - 10
08	att7	Bland Chromatin	1 - 10
09	att8	Normal Nucleoli	1 - 10

10	att9	Mitoses	1 - 10
11	Cl	Class	2 for benign, 4 for malignant

Tableau 3.1 : Attributs de la base de Wisconsin du cancer du sein

### 3.1.1. Prétraitement de la base

Le prétraitement de cette base se concentre sur les valeurs manquantes. En statistique, on parle des données manquantes lorsqu'on n'a pas d'observation pour une variable donnée pour un individu donné. Le problème de la gestion des données manquantes est un vaste sujet. Les données manquantes ne peuvent pas être ignorées lors d'une analyse statistique. Mais selon leur proportion et leur type, des solutions différentes vont être choisies. On pourra soit :

- Retirer les variables ou les individus présentant des données manquantes.
- Imputer des valeurs aux données manquantes : plusieurs méthodes sont appliquées pour le remplacement des valeurs manquantes. Il y a des méthodes mathématiques pour faire ce remplacement, parmi lesquelles nous citons : remplacement par la moyenne, remplacement par tirage conditionnel et remplacement par analyse factorielle.
- Développer des méthodes qui permettent de mener les analyses en présence de données manquantes.

Les résultats des trois méthodes sont présentés dans le tableau suivant :

	Base complète	Manquants remplacés	Manquants supprimés
Taux de classification	94%	95%	96%

Tableau 3.2 : Résultats après différents prétraitements [ABE, 08]

Les résultats obtenus sont très proches et démontre que la suppression des valeurs manquantes donne le plus bon taux de classification.

### 3.1.2. Sélection des attributs

L'ensemble des attributs utilisés dans cette base sont jugés utiles par ses créateurs. Dans cette section, nous essayerons de faire une étude pour évaluer chaque attribut de la base en déterminant son utilité ou non et son importance informationnelle.

D'un point de vue statistique, plus la différence entre l'apparition d'une valeur d'un attribut dans chaque classe est grande, plus cette valeur est significative. Pour évaluer chaque attribut, nous calculons la moyenne des différences pour chaque valeur. L'attribut qui aura la plus grande moyenne est le plus significatif.

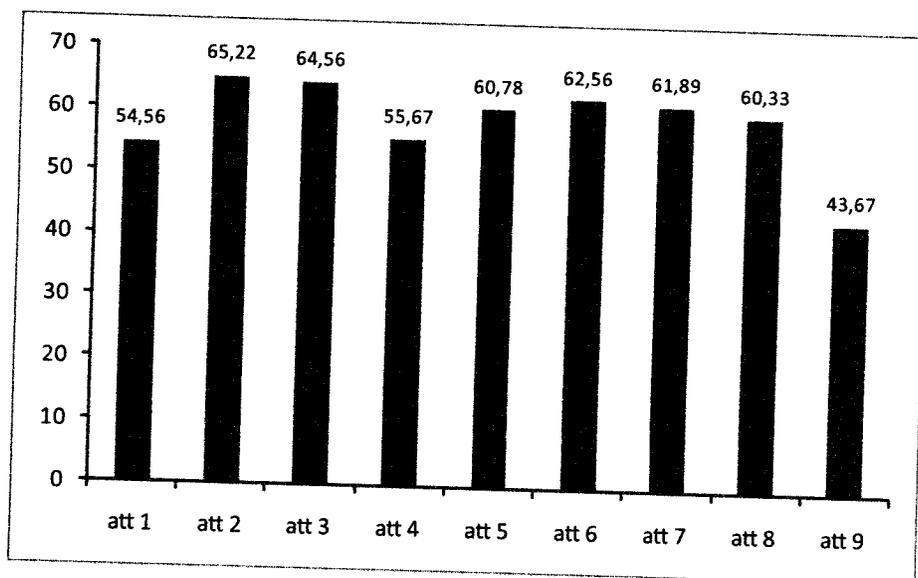


Figure 3.1 : Moyenne des différences entre les valeurs dans chaque classe

Les résultats statistiques montrent les meilleurs attributs par ordre décroissant, att2, att3, att6, att7, att5, att8, att4, att1, att9.

La deuxième méthode que nous avons utilisé pour l'évaluation de chaque attribut consiste à faire une classification en éliminant chaque fois un attribut. Le degré d'importance d'un attribut est lié à son influence sur le taux de classification. Les résultats de classification sont illustrés dans la Figure 3.2 :

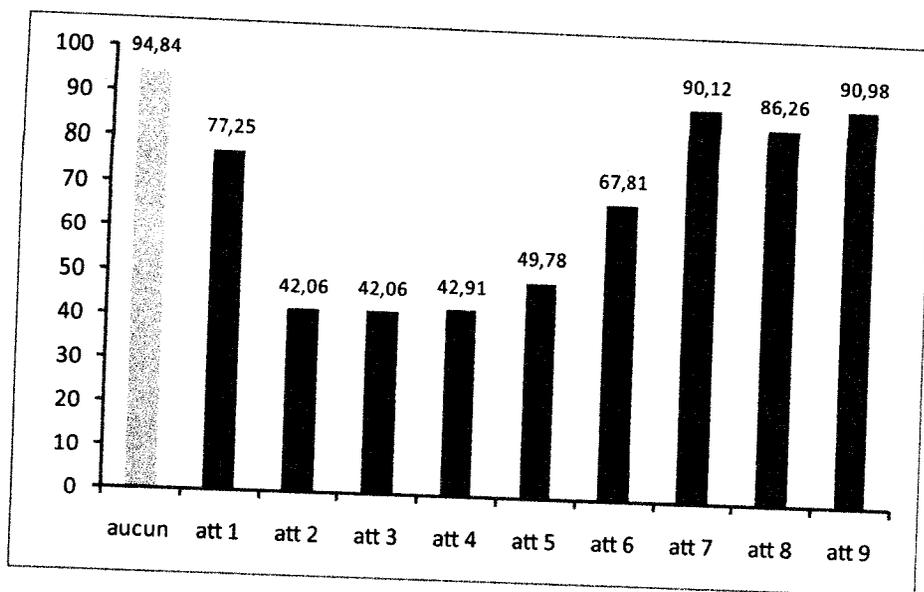


Figure 3.2 : Taux de classification calculé par le KNN après chaque élimination

Les expérimentations montrent que les attributs att2 et att3 sont les plus influents sur le taux de classification ensuite att4, att5, att6, att1, att8, att7, att9.

A partir de l'étude statistique et les résultats expérimentaux, nous pouvons tirer deux conclusions :

- Nous confirmons l'utilité de tous les attributs choisis par les créateurs de la base mais avec différents degrés d'importances.
- Nous pouvons classer les attributs att2 et att3 comme les attributs les plus pertinents et att9 comme l'attribut le moins important.

## 3.2. Algorithme modified-AIRS

### 3.2.1. Principe

Le présent algorithme est inspiré des travaux de A. Watkins [WAT, 01] [WAB, 02a] [WAB, 02b] [WAT, 02] nommé AIRS pour la reconnaissance des formes. Tout d'abord, l'ensemble d'apprentissage représente les antigènes dont il aura besoin pour le déroulement du programme, et le résultat d'apprentissage est l'ensemble d'anticorps (cellules B mémoire) qui seront utilisés pour la phase de test (les nouveaux antigènes introduits dans le système).

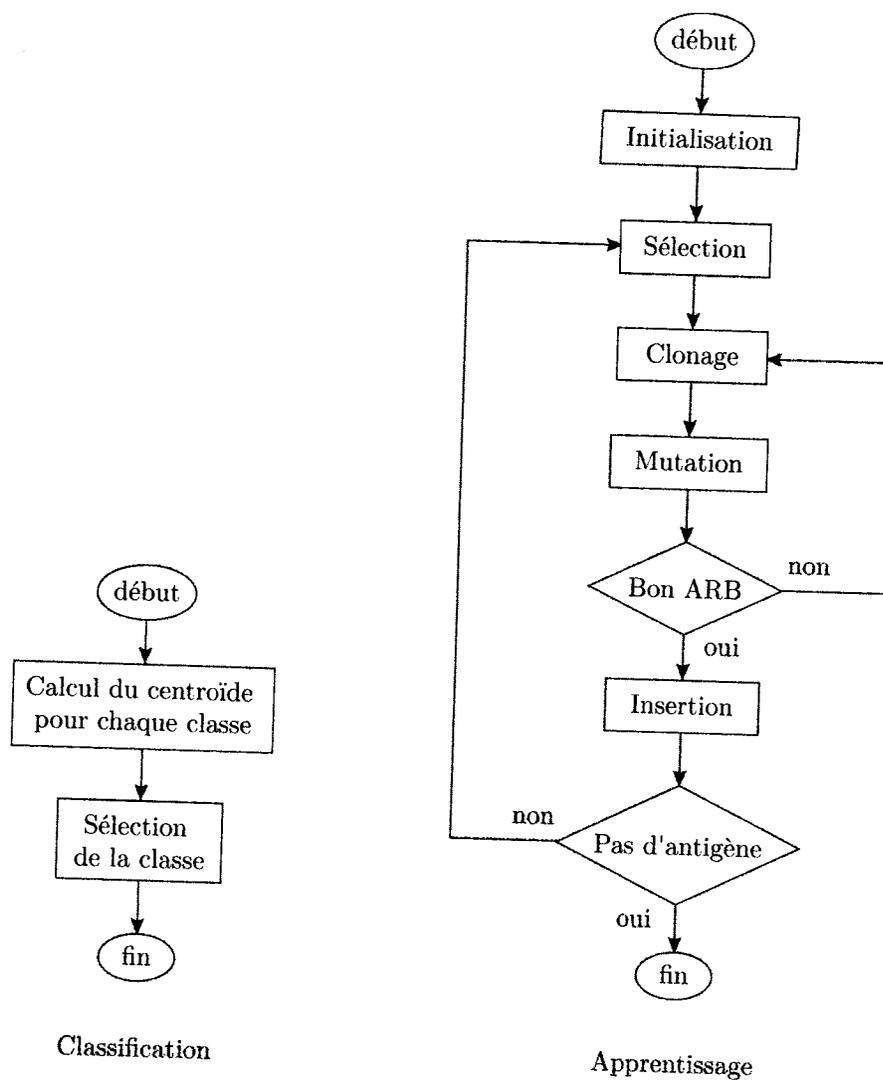


Figure 3.3 : Représentation de l'algorithme m-AIRS

Pour ce qui va suivre nous utilisons la notation suivante :

Paramètre	Description
$MC$	représente l'ensemble des cellules mémoires, $mc$ est un individu de cet ensemble
$ag$	est l'antigène en cours de traitement
$AB$	est l'ensemble des ARBs, $ab$ représente un ARB où $ab \in AB$
$nc$	est le nombre de classes

Tableau 3.3 : m-AIRS paramètres

L'apprentissage du système immunitaire artificiel se déroule en quatre étapes, chaque sortie est une introduction à l'étape suivante. Les étapes sont détaillées dans les points qui suivent:

### 3.2.2. Etapes d'apprentissage

#### 3.2.2.1 Initialisation

La phase d'initialisation peut être divisée en deux parties principales, le prétraitement de la base d'apprentissage et le calcul du seuil d'affinité. Durant cette phase les valeurs de la base sont toutes normalisées dans l'intervalle [0,1]. Après la normalisation, le seuil d'affinité est calculé par la moyenne de toutes les valeurs d'affinité entre les paires de vecteurs appartenant à la base d'apprentissage. Le seuil d'affinité est calculé par l'équation suivante :

$$\text{seuil d'affinité} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{affinite}(ag_i, ag_j)}{\frac{n(n-1)}{2}}$$

Où  $n$  est le nombre d'antigènes dans la base d'apprentissage,  $ag_i$  et  $ag_j$  sont respectivement le  $i^{\text{ème}}$  et le  $j^{\text{ème}}$  antigène de la base d'apprentissage et  $\text{affinite}(x,y)$  retourne la distance Euclidienne entre les vecteurs des deux antigènes.

Il faut noter ici qu'il est possible d'utiliser d'autres mesures de stimulation telle que la distance de Manhattan ou d'autres.

La dernière étape de l'initialisation est la génération de l'ensemble initiale des cellules mémoires. L'initialisation de cet ensemble se fait par l'introduction de 0 ou plus d'antigènes choisis aléatoirement de la base d'apprentissage.

#### 3.2.2.2 Identification des cellules B et génération des ARBs

Une fois que l'initialisation est achevée, cette étape aura lieu pour chaque nouvel antigène introduit. Une cellule mémoire est sélectionnée de l'ensemble entier de cellules B est nommée  $mc_{match}$ , cette dernière est tirée de telle sorte qu'elle ressemble le plus à l'antigène en cours de traitement selon la formule :

$$mc_{match} = \arg \max_{mc \in MC} stimulation(ag, mc)$$

Où  $mc$  est une cellule mémoire appartenant à l'ensemble  $MC$ ,  $ag$  est l'antigène en cours de traitement et  $stimulation(x, y)$  retourne  $1-affinite(x, y)$ .

Une fois que la cellule  $mc_{match}$  est sélectionnée, elle sera utilisée pour générer l'ensemble ARB. L'ensemble ARB contient les clones de  $mc_{match}$  correspondante à l'antigène en cours de traitement.

Le nombre de clones de la cellule  $mc_{match}$  généré par mutation est limité selon la formule:

$$nombre\_de\_clones = hyper\_taux\_clonage * taux\_clonage * stimulation(mc_{match}, ag)$$

```

Mutation(X,b){
  Pour chaque xi de X faire
    change = random()
    changer_vers = random()
    Si(change < taux_mutation)
      xi = changer_vers
      b = vrai;
    FinSi
  FinPour
  Si(b == vrai)
    change = random()
    changer_vers = lrandom() mod nc
    Si(change < taux_mutation)
      x.c = changer_vers
    FinSi
  FinSi
  Retourner X
}

```

Figure 3.4 : Algorithme original de mutation

L'algorithme original de la mutation [WAB, 02a] consiste à générer à chaque fois des nouvelles valeurs aléatoires pour la cellule mutée. Cette méthode ne permet pas de savoir ni d'estimer le temps nécessaire pour avoir une cellule qui a une bonne

stimulation avec l'antigène car elle utilise des modifications totalement aléatoire pour avoir la cellule mutée. Dans notre travail nous allons s'inspirer de la mutation du système immunitaire naturel pour essayer de résoudre ce problème.

L'algorithme de mutation utilisé dans ce travail est le suivant :

```
Mutation(X){
  Pour chaque xi de X faire
    opération = random()
    Si(opération dans [0, 0.33])
      xi = xi+pas_mutation
    FinSi
    Si(opération dans ]0.33, 0.66])
      xi = xi
    FinSi
    Si(opération dans ]0.66, 1])
      xi = xi-pas_mutation
    FinSi
  FinPour
  Retourner X
}
```

Figure 3.5 : Nouvel algorithme de mutation

Cet algorithme consiste à générer des cellules voisines à la cellule initiale en modifiant partiellement les valeurs de cette dernière ce qui nous permet de se rapprocher de la meilleure stimulation avec l'antigène.

Seuil de stimulation	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Algorithme original	3300	3300	3300	6570	224070	3.5x10 <sup>6</sup>	10.8x10 <sup>6</sup>
Nouvel algorithme	313	344	328	328	344	469	1031

Tableau 3.4 : Temps d'apprentissage en millisecondes

Le résultat obtenu démontre que la durée d'apprentissage appliqué sur la base WBCD a diminué à un temps moyen égal à 2s pour un seuil de stimulation égale à 0.9 et c'est une valeur qui peut être négligée par rapport au temps nécessaire pour l'algorithme original qui dépasse 180mn pour la même valeur du seuil.

### 3.2.2.3 Compétition des ressources et développement des cellules mémoires candidates

Cette partie se base sur les résultats de la précédente, elle complète les informations des ARBs générés en calculant les ressources pour chaque anticorps avec l'antigène en cours de traitement. Ces ressources sont mises à jour durant l'apprentissage, et chaque ARB n'ayant pas de ressources sera supprimé de l'ensemble des ARBs. Le nombre de ressources pour chaque ARB est calculé par la formule :

$$ressources = stimulation(ag, ab) * taux\_clonage$$

Les étapes 2 et 3 seront exécutées pour un antigène donné jusqu'à ce que la condition  $s \geq Seuil\_stimulation$  soit vérifiée, avec

$$s = \frac{\sum_{i=1}^{|AB|} ab_i.stim}{|AB|}, ab_i \in AB$$

Une fois que la condition précédente est vérifiée, on peut dire que nous avons pu avoir un ensemble d'ARBs optimal.

Une cellule candidate sera ajoutée à l'ensemble des cellules mémoires seulement si elle retourne une stimulation plus élevée que la cellule de  $mc_{match}$  avec l'antigène en cours de traitement. Si l'affinité entre  $mc_{match}$  et  $mc_{candidate}$  est inférieure au produit entre le seuil d'affinité et le scalaire du seuil d'affinité (ATS) définis par l'utilisateur, dans ce cas  $mc_{match}$  sera supprimée de l'ensemble des cellules mémoires.

### 3.2.3. Classification

A la fin de la phase d'apprentissage, l'ensemble des cellules mémoires est classé en fonction de leur affinité avec l'antigène en cours de traitement.

### 3.3.1.1 Pas de mutation

Le pas de modification de chaque attribut est défini par *pas\_mutation*. Puisque les valeurs de chaque cellule appartiennent à  $[0,1]$  la valeur de *pas\_mutation* doit appartenir au même intervalle. La valeur prise définit le degré de voisinage entre la cellule originale est la cellule mutée, plus la valeur est grande, plus les deux cellules sont plus éloignées. Elle a aussi un impacte sur le temps d'exécution. En se basant sur les propriétés précédentes, nous avons choisis les valeurs suivantes pour être testées  $\{0.05, 0.1, 0.2, 0.3, 0.4\}$ .

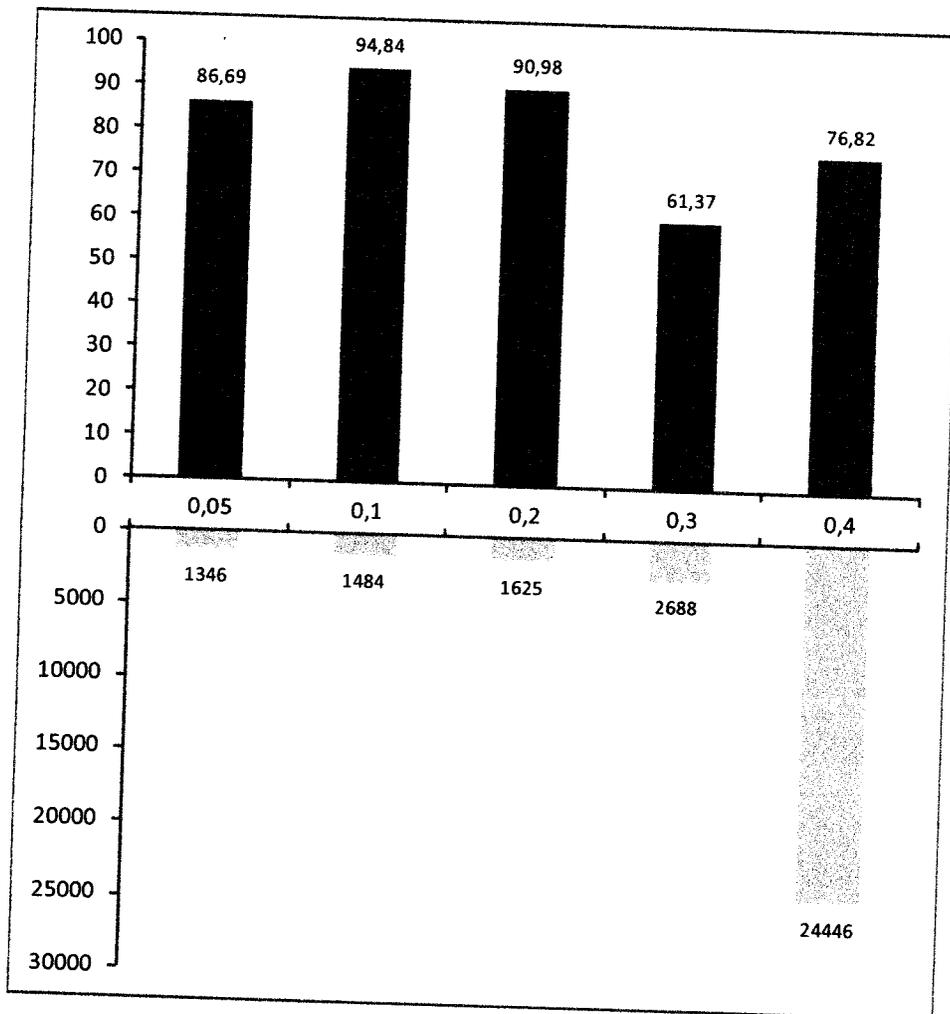


Figure 3.6 : Temps d'exécution en millisecondes et taux de classification calculé avec KNN

Les résultats obtenus montrent que la valeur 0,1 pour le paramètre *pas\_mutation* est la meilleure soit pour le taux de classification soit pour le temps d'exécution.

### 3.3.1.2 Seuil de stimulation

Le seuil de stimulation est le paramètre qui raffine l'ensemble des cellules mémoires et il définit aussi la qualité de l'ensemble ARB. Une grande valeur du seuil de stimulation veut dire que les cellules ont une bonne stimulation avec l'antigène. Les valeurs choisies pour être testées sont {0.6, 0.7, 0.8, 0.9, 0.99}

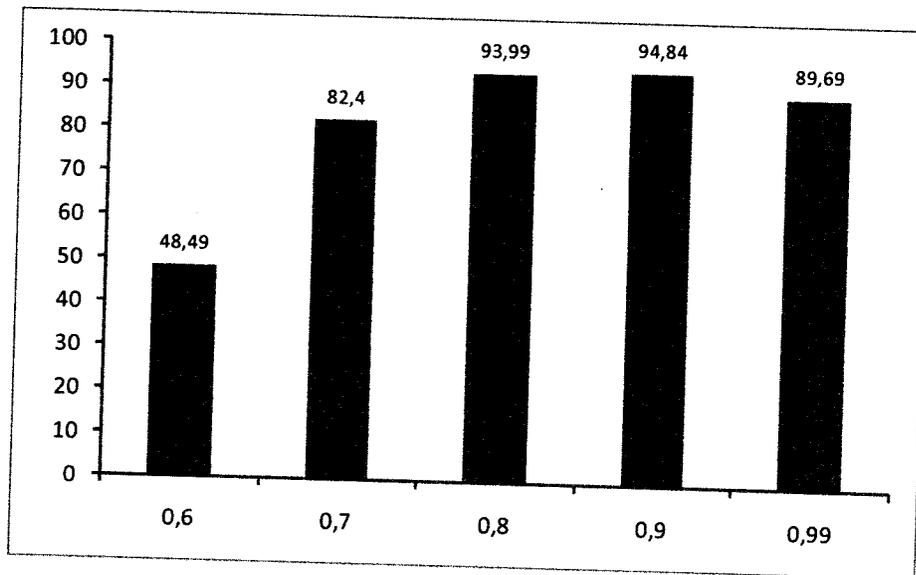


Figure 3.7 : Taux de classification calculé avec KNN

La meilleure valeur pour le seuil de stimulation est 0,9.

### 3.3.1.3 Scalaire du seuil d'affinité (ATS)

ATS est le scalaire du seuil de l'affinité, il est utilisé pour comparer la distance entre la nouvelle cellule mémoire et  $mc_{match}$ , si cette distance est petite (i.e. les deux cellules reconnaissent le même antigène) la  $mc_{match}$  est supprimé. Plus le scalaire est grand plus les cellules mémoires obtenus auront la possibilité de reconnaître divers antigènes et c'est ce qu'on veut dans notre cas.

Les valeurs choisies pour être testées sont {0.5, 0.6, 0.7, 0.8, 0.9}.

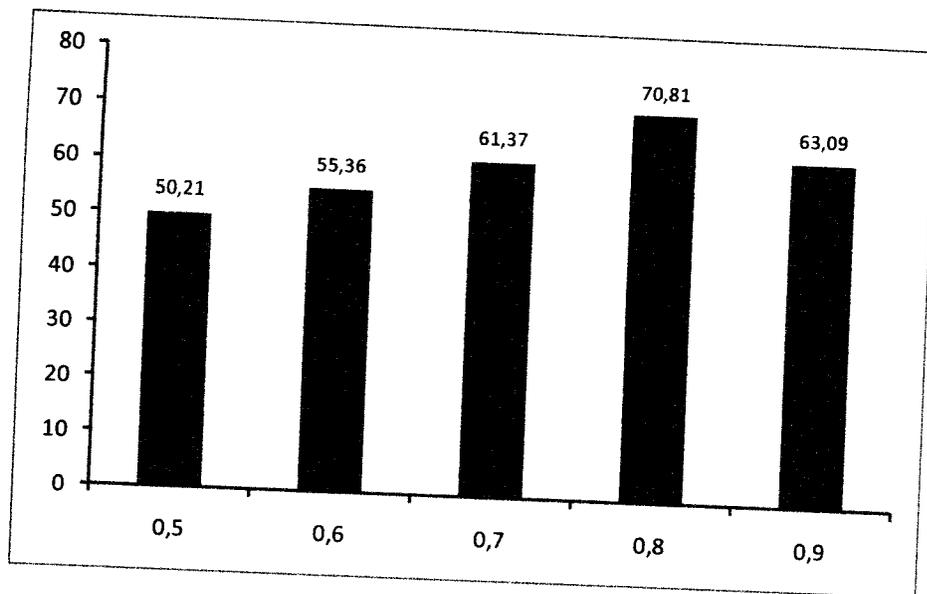


Figure 3.8 : Taux de classification calculé avec KNN

Les résultats obtenus montrent que la valeur 0,8 est la meilleure pour ce paramètre.

#### 3.3.1.4 Taux de clonage

Le taux de clonage va déterminer à la fois le nombre de clones pour chaque cellule et le nombre de ressources prises par chaque cellule dans ARB. Nous avons fixé le taux de clonage à 10 pour que le nombre de ressources de chaque cellule soit un entier entre 1 et 10.

#### 3.3.1.5 Hyper taux de clonage

L'hyper taux de clonage est un scalaire du taux de clonage. Il est utilisé pour déterminer le nombre de clones pour chaque cellule. La valeur choisie est égale à 2.

#### 3.3.1.6 Nombre de ressources

Le nombre de ressource définit le volume de l'ensemble ARB par rapport à son qualité. Plus le nombre de ressources est grand plus le nombre de cellules de ARB est grand.

Si on veut avoir par exemple, un ensemble de cellules avec une stimulation entre 0,8 et 0,9, le nombre de cellules ARB, pour un taux de clonage égale à 10, sera compris

$$\text{entre } \frac{\text{Nbr\_ressources}}{10 \times 0,9} \text{ et } \frac{\text{Nbr\_ressources}}{10 \times 0,8}.$$

Dans notre cas, nous avons choisis les valeurs suivantes pour être testées sont {200, 250, 300, 350, 400}.

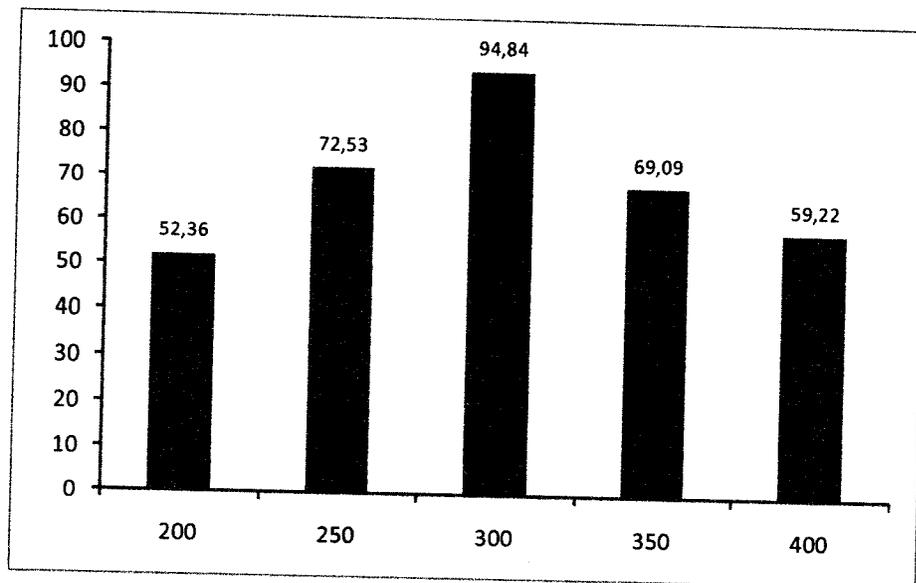


Figure 3.9 : Taux de classification calculé avec KNN

### 3.3.2. Résultats

L'évaluation d'un système de classification peut être abordée selon deux aspects :

#### 3.3.2.1 L'efficience

L'efficience est la qualité d'un rendement permettant de réaliser un objectif avec l'optimisation des moyens engagés. Dans notre cas, nous parlons d'espace mémoire et surtout du temps d'exécution.

	Espace mémoire occupé	Temps d'exécution
Apprentissage	3,28 Mo	1,984 s
Test	301,35 Ko	0,016
Classification	94,22 Ko	$\epsilon$ s

Tableau 3.5 : Espace mémoire et temps d'exécution pour Apprentissage, Test et Classification

En tenant compte que les résultats d'apprentissage+test présentés dans le tableau ci-dessus ont été calculés pour une base de 683 instances (450 pour l'apprentissage et 233 pour le test) et en tenant compte aussi les capacités à l'heure actuelle des machines

en terme de mémoire et de vitesse d'exécution, nous pouvons conclure que notre système a une très bonne efficacité.

### 3.3.2.2 L'efficacité

L'efficacité mesure la pertinence de classification. Les performances liées à l'efficacité du classifieur sont évaluées en utilisant la mesure d'exactitude.

Notre choix de cette mesure a été fait pour pouvoir comparer notre résultat avec les autres approches appliquées sur la base WBCD.

#### 3.3.2.2.1 Matrice de contingence

Pour évaluer un système de classification de ce type, nous utilisons un base étiquetée pour laquelle on connaît la vraie catégorie de chaque instance, et le résultat obtenu par le classifieur. Pour cette base, nous pouvons construire la matrice de contingence pour chaque classe (Tableau 3.5), qui fournit 4 informations essentielles :

- Vrai Positif (VP) : Le nombre d'instances attribuées à une classe convenablement. (Instances attribuées à leurs vraies classes)
- Faux Positif (FP) : Le nombre d'instances attribuées à une classe inconvenablement. (Instances attribuées à des mauvaises classes)
- Faux Négatif (FN) : Le nombre d'instances inconvenablement non attribuées. (Qui auraient dû être attribuées à une classe mais qui ne l'ont pas été).
- Vrai Négatif(VN) : Le nombre d'instances non attribuées à une classe convenablement. (Qui n'ont pas à être attribuées à une classe, et ne l'ont pas été).

Classe Ci		Jugement Expert	
		Oui	Non
Jugement Classifieur	Oui	<b>VPi</b>	FPi
	Non	FNi	<b>VNi</b>

Tableau 3.5 : Matrice de contingence de la classe Ci

### 3.3.2.2.2 Exactitude

Dans les domaines de la science, l'ingénierie, l'industrie et des statistiques, l'exactitude d'un système est le degré de proximité de mesures d'une quantité à la quantité réelle.

L'exactitude est la proportion de vrais résultats (deux vrais positifs et vrais négatifs) dans la population.

$$\text{Taux\_de\_classification} = \frac{VP + VN}{VP + FP + VN + FN}$$

La mesure d'exactitude est la mesure utilisée dans les travaux traitant la base WBCD. Dans ce travail nous avons choisi cette mesure pour pouvoir comparer notre résultat avec les autres.

Deux méthodes ont été proposées pour faire la classification, la première est les K plus proches voisins et la deuxième est le centroïde. Pour pouvoir choisir entre les deux, nous avons fait une comparaison entre ces deux méthodes. Les résultats obtenus sont présentés dans la Figure 3.10.

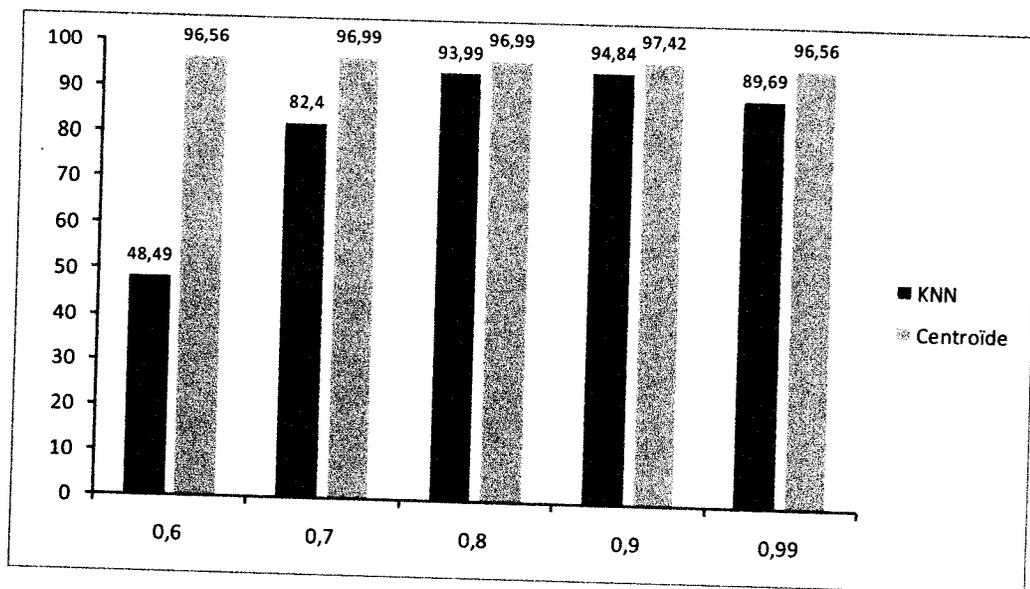


Figure 3.10 : Résultats de classification de KNN et Centroïde

Il est très apparent que le calcul du centroïde des deux classes augmente les performances du classifieur jusqu'à 50%, ce qui nous a conduit vers l'utilisation du centroïde pour la phase de classification.

Les résultats obtenus dans ce travail sont comparés avec les algorithmes C4.5 (Arbres de décision, Quinlan, 1996), SVM (Support Vector Machine, Bennett and Blue, 1997), Fuzzy-GA1 (Fuzzy-Genetic Algorithms, Pena-Reyes and Sipper, 1999), Neuro-Rule 2a (Neuro-Rule, Setiono, 2000), Big-LVQ (Big-Learning Vector Quantization, Goodman et al., 2002), SFC (Supervised Fuzzy Clustering, Abonyi and Szeifert, 2003), AIRS (Artificial Immune Recognition System, Goodman et al., 2002).

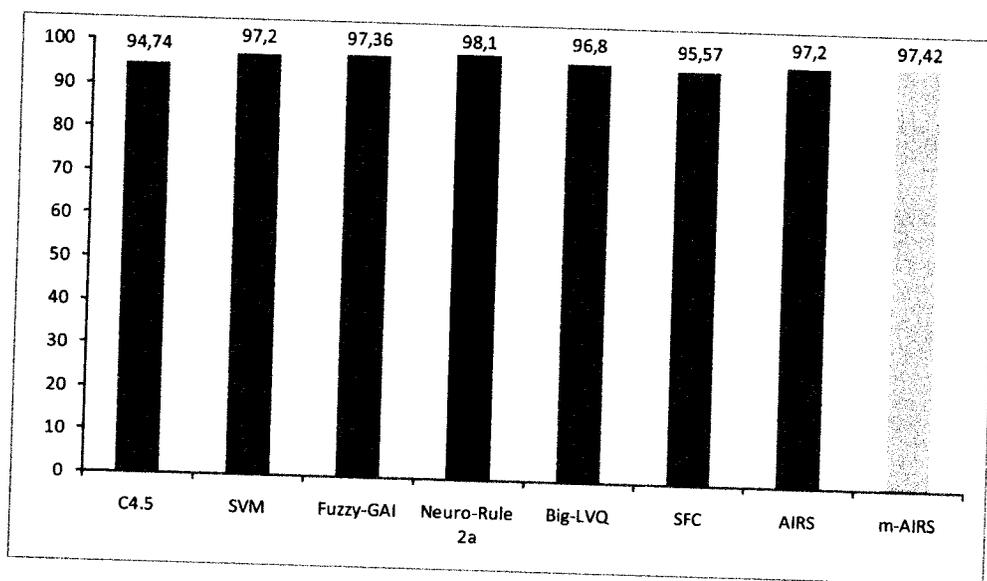


Figure 3.11 : Comparaison des résultats obtenus avec ceux des différentes approches

### 3.4. Conclusion

Dans ce chapitre, nous avons présenté une des approches des systèmes immunitaires. Le point de départ était les travaux de A.Watkins nommés AIRS qui est un algorithme de la sélection clonale. Dans notre travail, nous avons essayé de faire quelques améliorations sur cet algorithme, soit en termes de rapidité d'exécution, ou bien, en termes de taux de classification. Le présent travail a été appliqué sur la base du cancer du sein de l'université du Wisconsin afin de pouvoir comparer nos résultats avec les travaux déjà appliqués sur cette base.

Les points essentiels touchés dans ce chapitre sont les suivants :

1. Nous avons modifié l'algorithme original de mutation, ce qui nous a permis d'améliorer le temps d'exécution de l'algorithme et le taux de classification.
2. Nous avons aussi changé la méthode classification ce qui nous a donné jusqu'à 50% d'améliorations de taux de classification.
3. Une étude statistique et une autre expérimentale ont permis de classer les attributs de la base selon leur importance informationnelle.
4. En se basant sur l'impact de chaque paramètre de l'algorithme, le déroulement de ce dernier et sur le résultat obtenu, nous avons essayé de guider le paramétrage et choisir les meilleures valeurs.

Notre algorithme m-AIRS a atteint son objectif puisque nous avons eu une amélioration, soit en efficience ou bien en efficacité, par rapport à l'algorithme original AIRS, nous avons eu aussi le meilleur taux de classification après l'approche hybride neuro-rule.

# Conclusion générale et perspectives

Les développements les plus appréciables dans le domaine de l'intelligence artificielle se basent principalement sur l'extraction des métaphores utiles à partir des systèmes biologiques afin de créer des solutions aux problèmes complexes.

Dans notre mémoire de fin d'études, nous concevons un modèle de classification inspiré d'une approche immunitaire et nous suivons un chemin de développement d'un algorithme particulier, inspiré principalement de la théorie de la sélection clonale nommé m-AIRS.

Les algorithmes immunitaires sont des outils très efficaces pour l'optimisation, mais ils ont l'inconvénient de former des systèmes indéterministes (i.e. les même entrées ne donnent pas forcément les même sorties), et cela dû à la fonction de mutation qui génère les cellules aléatoirement.

Nous avons obtenu des résultats très satisfaisants et prometteurs par l'application de l'algorithme m-AIRS. Notre classifieur présente des résultats très similaires en terme de précision (97,42%) à d'autres classifieurs cités dans la littérature, nous citons en particulier le modèle hybride neuro-rule 2a réalisé par [Setiono, 2000].

Même avec une erreur de classification de quelques cellules (2,58%) nous jugeons les performances de notre modèle comme efficaces, sachant que pour identifier la malignité d'une tumeur, l'anatomo-pathologiste analyse la cytologie de plusieurs cellules avec un taux de classification acceptable et le type de la tumeur sera extrait parfaitement.

Les résultats obtenus de cette application sont encourageants et ils peuvent être améliorés en explorant d'autres extensions des SIAs. Ces améliorations peuvent être appliquées au niveau des points suivants :

- Le premier point consiste à automatiser totalement cette phase de diagnostique (Analyse cytologique) en ajoutant un module de traitement d'image pour l'extraction des caractéristiques cytologiques correspondantes aux 09 attributs de la base WBCD.
- Le deuxième point concerne les attributs de la base, où le degrés d'importance varie d'un attribut à un autre. Nous proposons de pondérer donner les attributs selon leurs degrés d'importance.
- Le dernier point porte sur l'hybridation des SIAs avec l'approche floue afin d'augmenter l'interprétabilité des classifieurs.

# Références bibliographiques

- [[ABE, 08] Abel Medhanie Gebreyesus, Sunil Nair. Classification of Breast Cancer dataset using Decision Tree Induction, Dalhousie University, November 25, 2008
- [AIC, 03] U Aickelin, P Bentley, S Cayzer, J Kim, J McLeod, Danger Theory: The Link between AIS and IDS?. Proceedings ICARIS-2003, 2nd International Conference on Artificial Immune Systems 2003.
- [ALL, 06] P. ALLAIN, <http://www.pharmacorama.com/ezine/index.php> , 2006
- [BOR, 05] Rafael H. BORDINI, Jiirgen Dix, Mehdi Dastani, Amal El Fallah Seghrouchni, Multi-Agent Programming, Languages, Platforms and Applications, ISBN-10:0-387-24568-5, 2005.
- [CAS, 00] Castro, L. N. and Von Zuben, F. N. (2000). An evolutionary immune network for data clustering. In 6th Brazilian Symp. Neural Networks, SBRN '00, 84-89. IEEE
- [CAS, 02] Castro L.N., Von Zuben F.J.: "Learning and optimization using the clonal selection principle". IEEE Trans. on Evolutionary Computation, vol 6, no 3, pp.239-251, (2002).
- [CAS, 01] Leandro Nunes DE CASTRO, An Introduction to the Artificial Immune Systems, ICANN -Prague, 22-25th April, 2001.
- [CAS, 02] L. N. DE CASTRO, J. TIMMIS, In Artificial Neural Networks in Pattern Recognition Artificial Immune Systems: A Novel Paradigm to Pattern Recognition, University of Paisley, UK, 2002.
- [CAS, 03] Leandro Nunes DE CASTRO, Fernando J. VON ZUBEN, Getúlio A. de Deus Jr, The Construction of a Boolean Competitive Neural Network Using Ideas from Immunology, Neurocomputing, 50C, 2003

- [CIR, 08] Le Centre international de Recherche sur le Cancer (CIRC), 2008
- [DAS, 05] Mehdi DASTANI and Jorge J. Gomez-Sanz Programming Multi Agents Systems, 2005.
- [DOC, 11] Analyses médicales , Hémogramme : formule leucocytaire
- [FOR, 94] Stephanie Forrest, Alan S. Perelson, Lawrence Allen, and Rajesh Cherukuri. Self-nonsel self discrimination in a computer. In SP '94 : Proceedings of the 1994 IEEE Symposium on Security and Privacy, page 202, Washington, DC, USA, 1994. IEEE Computer Society
- [GIU, 02] Giuseppe Nicosia AND Vincenzo Cutello , "An Immunological Approach to Combinatorial Optimization Problems," Proceedings of the 8th Ibero-American Conference on AI: Advances in Artificial Intelligence, Seville, Spain, 2002.
- [IND, 08] Institut National Du Cancer, Cancer du sein, mise à jour avril 2008
- [ITE, 09] [http://etechpharma.com/details-lexique-public.php?id\\_lexique=100](http://etechpharma.com/details-lexique-public.php?id_lexique=100)
- [JOH, 00] Johnson KC, Hu J, Mao Y. Passive and active smoking and breast cancer risk in Canada, 1994-97. The Canadian Cancer Registries Epidemiology Research Group. Cancer Causes Control 2000 ; 11 : 211-21.
- [KEL, 03] Johnny KELSEY and Jon TIMMIS, "Immune Inspired Somatic Contiguous Hypermutation for Function Optimisation," Proceedings, Part I Genetic and Evolutionary Computation Conference (GECCO 2003), Chicago, IL, USA, pp. 207-218, 2003.
- [KEL, 96] Kelsey JL, Bernstein L. Epidemiology and prevention of breast cancer. Annu Rev Publ Health 1996 ; 17 : 47-67.

- Parole SIMPA – Université des sciences et de la technologie d'Oran, 2006.
- [LAR] Larousse Médical
- [LCC, 06] La ligue contre le cancer : le cancer du sein : information et prévention - septembre 2002 mise à jour février 2006
- [MAC, 80] MacMahon B, Andersen AP, Brown J, et al. Urine estrogen

- [TIM, 02] Jon Timmis and Leandro N. de Castro. *Artificial Immune Systems: A new computational intelligence approach*, Great Britain: Springer-Verlag, 2002
- [TIM, 03] J. Timmis and L.N. Castro, "Artificial Immune Systems as a Novel Soft Computing Paradigm". In the *Soft Computing Journal*, vol.7, Issue 7, July 2003.
- [VIN, 05] Vincenzo Cutello and Giuseppe Nicosia. Chapter VI. The Clonal Selection Principle for In Silico and In Vivo Computing. In: *Recent Developments in Biologically Inspired Computing*, eds. Leandro Nunes de Castro and Fernando J. Von Zuben. Hershey, London, Melbourne, Singapore: Idea Group Publishing, 2005. pp. 104-146.
- [WAB, 02a] Watkins A., Boggess L. (2002a). A New Classifier Based on Resources Limited Artificial Immune Systems. *Procéding congress de Evolutionary Computation, IEEE World Congress on Computational Intelligence held en Honolulu*, pp 1546-1551. IEEE. USA.
- [WAB, 02b] Watkins A., Boggess L. (2002b). A Resource Limited Artificial Immune System. *Journal IEEE World Congress on Computational Intelligence help in Honolulu, USA* pp 926-931.
- [WAT, 01] Andrew B. Watkins AIRS : Watkins A. (2001). AIRS : A resource limited artificial immune classifier. Thèse de l'université de Mississipi du département d'informatique. Mississipi
- [WAT, 02] Watkins A., Timmis J. (2002). Artificial Immune Recognition System (AIRS): Revisions and Refinements. 1ère Conférence international du Système immunitaire artificiel ICARIS'2002, pp 173-181, Université de Kent à Canterbury.

# Annexe : Interface et manuel d'utilisation

## 1. Environnement de travail

Le langage Java est un langage de programmation informatique orienté objet créé par James Gosling et Patrick Naughton, employés de Sun Microsystems, avec le soutien de Bill Joy (cofondateur de Sun Microsystems en 1982).



Figure A.1 : Logo du langage JAVA

La particularité principale de Java est que les logiciels écrits dans ce langage sont très facilement portables sur plusieurs systèmes d'exploitation tels que UNIX, Windows, Mac OS ou GNU/Linux, avec peu ou pas de modifications.

Le langage reprend en grande partie la syntaxe du langage C++, très utilisé par les informaticiens. Néanmoins, Java a été épuré des concepts les plus subtils du C++ et à la fois les plus déroutants, tels que les pointeurs et références, et l'héritage multiple remplacé par l'implémentation des interfaces.

Il existe deux types de programmes en Java : les applets et les applications. Une application autonome (stand alone program) est une application qui s'exécute sous le contrôle direct du système d'exploitation. Une applet est une application qui est chargée par un navigateur et qui est exécutée sous le contrôle d'un plug in de ce dernier.

Aujourd'hui, Java (sous la forme de JavaFx) trouve une nouvelle niche dans la création d'applications RIA (Rich Internet Applications), des applications qui proposent des fonctionnalités, notamment des interfaces, plus évoluées à la fois sur Internet et sur les téléphones portables. Le langage JavaFx est un langage agile dérivé de Java, sous le contrôle de Sun Microsystems, qui met à profit la portabilité de Java ainsi que les vastes bibliothèques déjà disponibles dans le langage Java et ce qui a guider notre choix du langage et d'environnement de développement.

## 2. Codage

Notre projet est constitué de d'environ de 2000 lignes de code, 12 classes divisé es en trois packages :

- App : il contient le contient le *main* du projet et les interfaces graphiques de l'application.
- Noyau : contient le noyau du système, la classe AIRS responsable de l'apprentissage, le test et la classification.
- Tools : contient des classes personnalisées pour notre application.

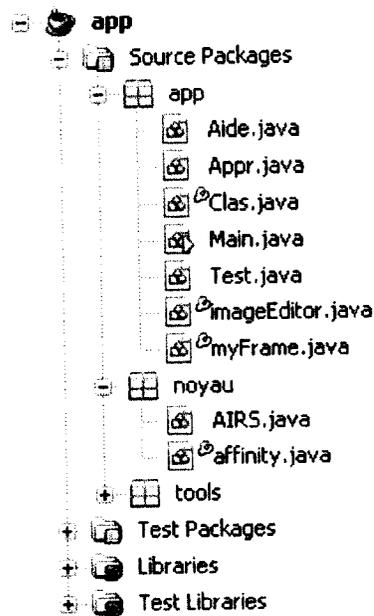


Figure A.2 : Packages du projet

### 3. Présentation de l'Interface Homme Machine (IHM)

#### 3.1. La fenêtre principale

La fenêtre principale est constituée de quatre onglets :

- Apprentissage : il permet de faire l'apprentissage avec la possibilité de changer les paramètres et de sélectionner les différents attributs.
- Test : il affiche les résultats de test pour chaque classe et pour le classifieur.
- Classification : permet de classer les nouvelles cellules en sélectionnant les valeurs des attributs.
- Aide : manuel d'utilisation.

On trouve aussi deux boutons :

- Le bouton  permet d'ouvrir un répertoire de cellules mémoires enregistré pour faire le test ou la classification.
- Le bouton  permet l'enregistrement d'un répertoire de cellules mémoires pour une future utilisation (test, classification).

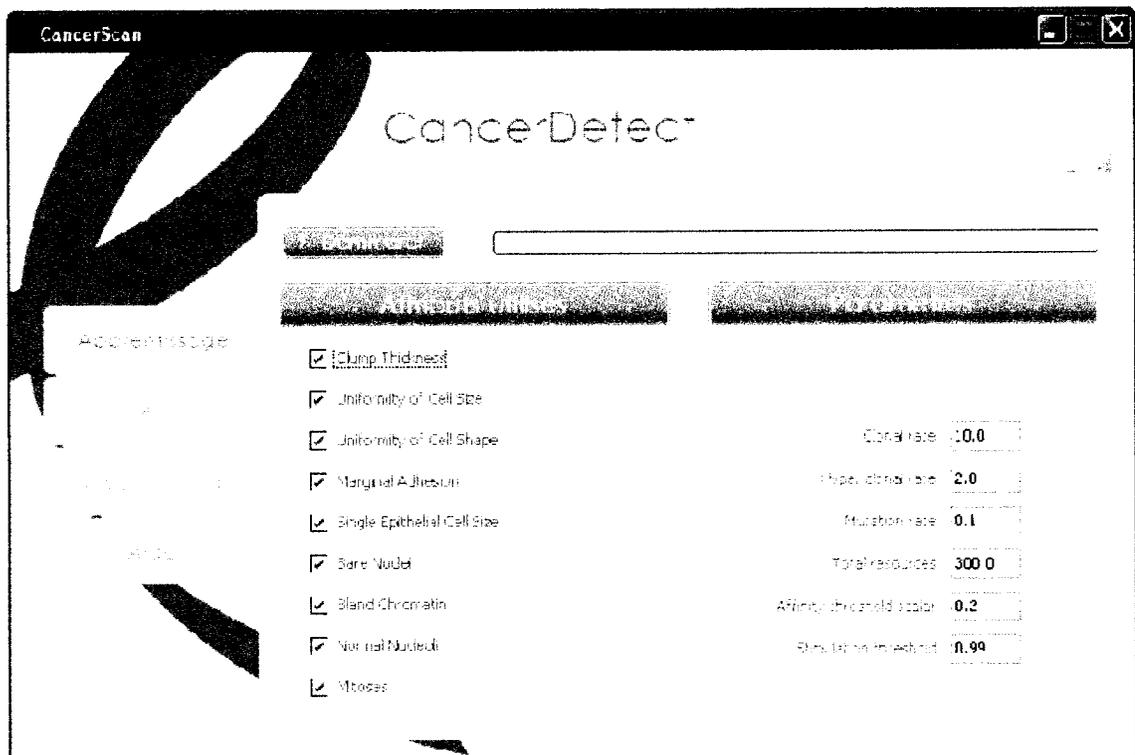


Figure A.3 : Interface de l'application.



### 3.2. Apprentissage

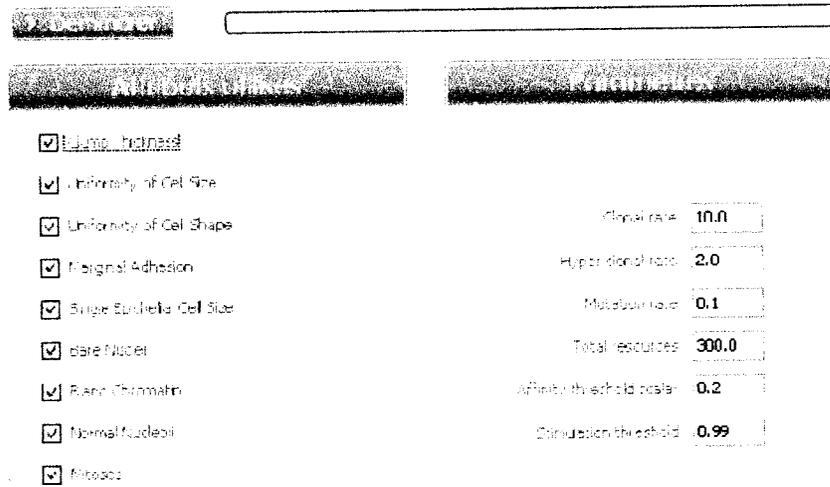


Figure A.4 : Onglet d'apprentissage

L'onglet d'apprentissage nous permet de sélectionner les attributs utilisés pendant l'apprentissage. La sélection d'un attribut se fait par un simple clique sur .

L'onglet apprentissage nous permet aussi de changer les différents paramètres.

Pour lancer l'apprentissage il suffit de cliquer sur le bouton , le progrès de l'apprentissage peut être suivie à partir de *progress bar*.

### 3.3. Test

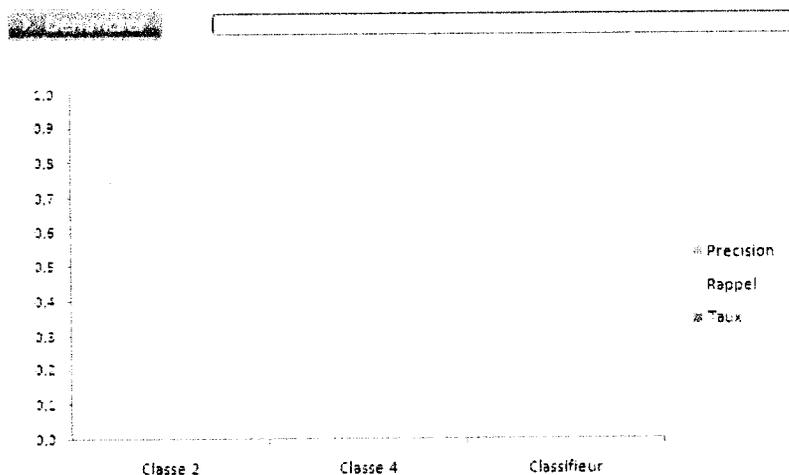


Figure A.5 : Onglet de test

L'onglet test nous permet de lancer le test en cliquant sur le bouton . Il nous donne comme résultats la précision, le rappel et le taux de

classification de chaque classe et du classifieur sous forme d'un histogramme (Figure A.6).

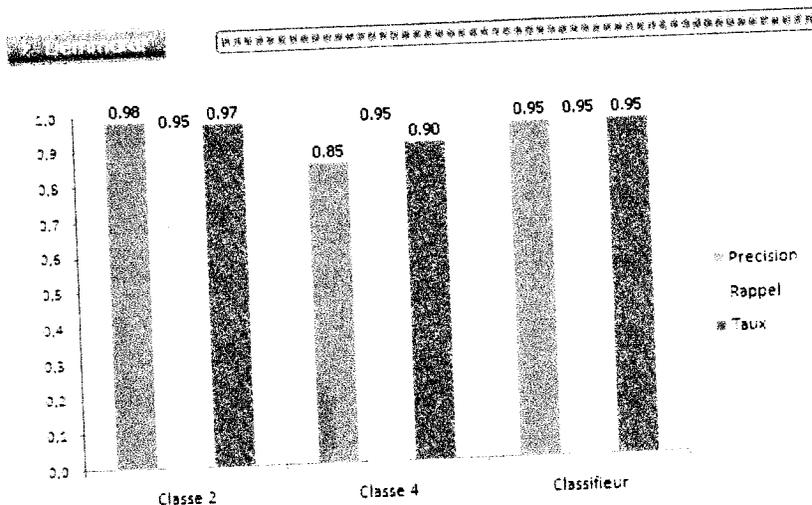


Figure A.6 : Résultats de test

### 3.4. Classification

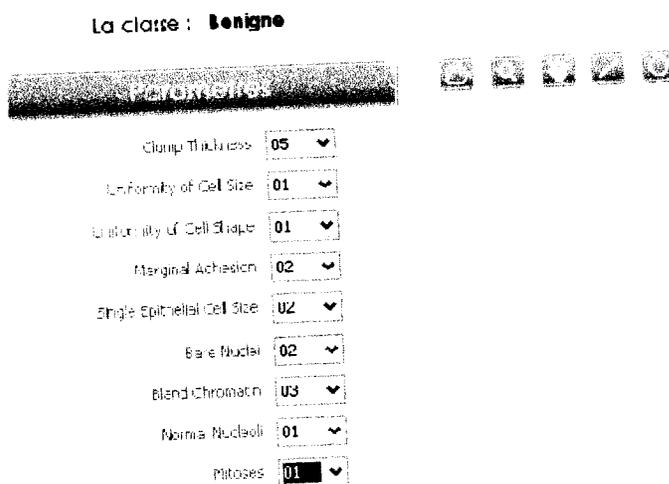


Figure A.7 : Onglet de classification

L'onglet classification est utilisé pour la classification des nouvelles cellules. Le classifieur donne la classe de la nouvelle cellule après le choix de toutes les valeurs des paramètres.

La partie à droite (Figure A.8) est utilisée pour aider l'anatomo-pathologiste à extraire les caractéristiques de la cellule, comme elle peut être utilisé dans un futur travail pour la segmentation est l'extraction des caractéristiques d'une cellule

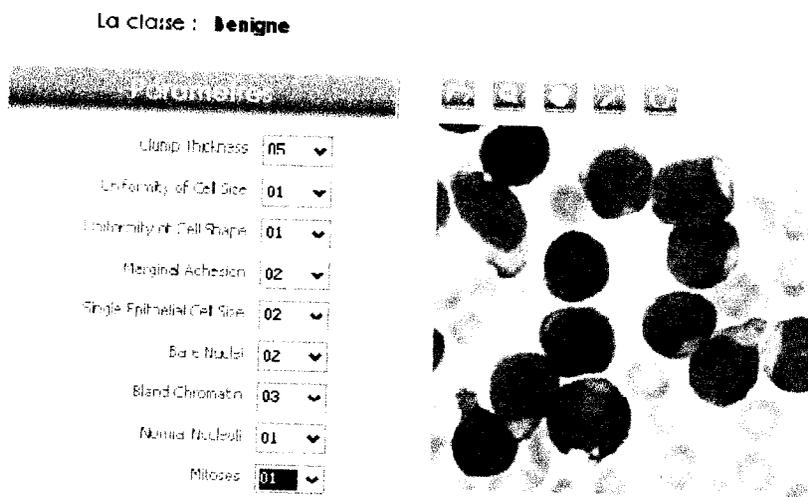


Figure A.8 : Présentation d'une image des cellules

### 3.5. Aide

Fenêtre principale  
Apprentissage  
Test  
Classification



- La fenêtre principale est constituée de quatre onglets :
- Apprentissage : il permet de faire l'apprentissage avec la possibilité de changer les paramètres et de sélectionner les différents attributs.
  - Test : il affiche les résultats de test pour chaque classe et pour le classifieur.
  - Classification : permet de classer les nouvelles cellules en sélectionnant les valeurs des attributs.
  - Aide : manuel d'utilisation.

On trouve aussi deux boutons :

- Le bouton  permet d'ouvrir un répertoire de cellules mémoires enregistré pour faire le test ou la classification.

Figure A.9 : Onglet d'aide

L'onglet Aide nous permet d'avoir une vue générale sur les fonctionnalités de l'application.

Il donne la méthode d'utiliser les différents onglets d'apprentissage, de test et de classification.