



**FACULTE DES SCIENCES**  
**Département d'Informatique**

**MEMOIRE**

*SOUTENU EN VUE DE L'OBTENTION DU DIPLOME DE*  
**MASTER EN INFORMATIQUE**  
**OPTION : Modèles Intelligents et Décision**

**THEME**

*« Système immunitaire artificiel  
pour la reconnaissance du diabète »*

Présenté par :

BENDIABDELLAH Mohamed Hakim  
Sous la Direction du Dr Mohamed Amine CHIKH

**Soutenu devant le Jury composé de Messieurs**

Mr BENAMAR A.	Président
Mr CHIKH M.A	Encadreur
Mr LAHASAINI M.	Examineur
Mr BENZIANE Y.	Examineur
Mr BENZAOUZ M.	Examineur
Mr HADJILA F.	Examineur
Mr SMAHI M.I	Examineur

*Juillet 2011*

# **REMERCIEMENTS**



Mes profonds remerciements et mon entière reconnaissance à  
mon encadreur

**Pr Amine CHIKH**

Pour toute l'attention qu'il a bien voulu m'accorder  
et pour les précieux conseils qu'il m'a prodigué tout  
au long de ce mémoire. Profond respect.



**Aux honorables membres de mon jury**

qui ont bien voulu sacrifier de leur temps  
à la lecture de ce modeste travail et à son évaluation.

Qu'ils en soient remerciés.



**Enfin, à tous mes enseignants**

qui m'ont fait profiter de leur savoir,  
sans lequel, un tel travail n'aurait pu voir le jour.

## **DEDICACES**



*Ma profonde reconnaissance va d'abord*

**A MON PERE ET A MA MERE**

*pour toute l'affection dont ils m'ont entouré  
et pour l'encouragement moral qu'ils m'ont témoigné*



*A mon frère et mes sœurs*

**AMEL, YACINE ET WAFAA**

*Et a **HICHEM** et **INES***

*pour leur affectueuse complicité*



**A mes amis**

**et à toute la promotion MID**

## **TABLE DES MATIERES**

TABLE DES MATIERES .....	1
LISTE DES TABLEAUX .....	3
LISTE DES FIGURES .....	4
INTRODUCTION GENERALE .....	5

### **CHAPITRE I : LE DIABETE**

1. Introduction.....	7
2. Historique.....	7
3. Définition.....	8
3.1. Diabète de type 1.....	8
3.2. Diabète de type 2 .....	8
4. Principaux facteurs de risque de diabète de type 2 .....	9
5. Symptômes et complication.....	10
6. Critères diagnostiques.....	12
7. Conclusion .....	12

### **CHAPITRE II : LES SYSTEMES IMMUNITAIRES**

Introduction.....	14
A. SYSTEMES IMMUNITAIRES NATURELS.....	15
1. Historique.....	15
2. Concepts immunologiques.....	16
2.1 Organes immunitaires .....	16
2.2 Cellules immunitaires .....	16
2.3 Antigène.....	17
3. Architecture du système immunitaire .....	18
3.1 Immunité innée .....	18
3.2 Immunité adaptative.....	19
4. Propriétés du système immunitaire .....	19
5. Théories immunitaires .....	21
5.1 La sélection Négative/Positive .....	21
5.2 La sélection clonale .....	21
5.3 Théorie du danger.....	22
6. Déroulement d'une réponse immunitaire .....	22
B. SYSTEME IMMUNITAIRE ARTIFICIEL.....	23

1. <b>Historique</b> .....	23
2. <b>Définition</b> .....	24
3. <b>Modélisation des systèmes immunitaires artificiels</b> .....	24
3.1 La représentation.....	25
3.2 Les mesures d'affinités.....	25
3.3 Les algorithmes immunitaires.....	26
3.3.1 L'algorithme de la sélection négative/positive.....	26
3.3.2 L'algorithme de la sélection clonale.....	27
3.3.3 L'algorithme du réseau immunitaire.....	28
3.4 Etude comparative entre les différents systèmes inspirés de la biologie.....	29
4. <b>Domaines d'application de l'AIS</b> .....	29
5. <b>L'algorithme AIRS</b> .....	30
5.1 Introduction.....	30
5.2 Vue d'ensemble de l'algorithme AIRS.....	32
5.2.1 Définition.....	32
5.2.2 Déroulement de l'algorithme.....	35
5.3 Classification.....	44
5.3.1 K plus proche voisin.....	44
5.3.2 Algorithme de naïf Bayes.....	47
6. <b>Conclusion</b> .....	50

### **CHAPITRE III : APPLICATION ET RESULTATS**

1. <b>Introduction</b> .....	51
2. <b>Base de donnée PIMA</b> .....	51
3. <b>Analyse statistique des données</b> .....	52
4. <b>Les paramètres d'apprentissage</b> .....	53
5. <b>Critère d'évaluation</b> .....	56
6. <b>Langages et environnement de développement</b> .....	57
7. <b>Résultat et discussion</b> .....	59
7.1. Première expérimentation.....	59
7.2. Deuxième expérimentation.....	62
7.3. Troisième expérimentation.....	64
8. <b>Application de l'algorithme AIRS2 sur la base du cancer du sein et la base Iris</b> .....	67
9. <b>Conclusion</b> .....	68
<b>Conclusion générale</b> .....	69
<b>Références Bibliographiques</b> .....	71

## **LISTE DES TABLEAUX**

Tableau 1 Comparaison entre les différents systemes inspirés de la biologie.....	29
Tableau 2 Correspondance entre AIRS et le systeme immunitaire .....	32
Tableau 3 Corrélation mutuelle entre quelques descripteurs.....	52
Tableau 4 : La matrice de confusion .....	57
Tableau 5 Les classes principales du programme.....	59
Tableau 6 performance d'apprentissage avec AIRS2 .....	60
Tableau 7 Comparaison de classification avec et sans AIRS.....	65
Tableau 8 Classification des bases : breast cancer wisconsin et Iris .....	67

## LISTE DES FIGURES

Figure 1: Structure d'un antigène/anticorps.....	18
Figure 2 : Architecture du système immunitaire .....	18
Figure 3 théorie de la sélection clonale .....	22
Figure 4 Structure de conception d'un AIS .....	25
Figure 5 La structure générale de l'algorithme de la sélection négative.....	27
Figure 6 Hyper-Mutation pour la génération d'ARB .....	37
Figure 7 Fonction de mutation.....	38
Figure 8 Stimulation,Allocation de ressources,Suppression ARB .....	41
Figure 9 Mutation des ARB survivants .....	42
Figure 10 Introduction d'une cellule mémoire .....	43
Figure 11 Classification de deux classe avec Knn, a gauche k=1 , a droite k=3.....	44
Figure 12Algorithme Kppv .....	45
Figure 13 Fonction discriminantes et décision .....	48
Figure 14.Frontière de décision dans l'espace des observations.....	49
Figure 15 l'evolution d'apprentissage avec 1,3 plus proches voisins et naif bayes .....	61
Figure 16 l'évaluation de la classification par rapport a le scalaire du suil de stimulation .....	63
Figure 17 histogramme des performance des différents classificateur avec et sans AIRS2.....	64
Figure 18 Classification avec Knn .....	66
Figure 19 Classification Knn avec et sans AIRS sans AIRS .....	66
Figure 20 Aperçu de notre application pour la classification de la base breast cancer wisconsin avec L'algorithme AIRS2.....	68



# INTRODUCTION GENERALE

# Introduction générale

---

## Introduction générale

Le diabète est une maladie chronique qui apparaît lorsque le pancréas ne produit pas suffisamment d'insuline ou que l'organisme n'utilise pas correctement l'insuline qu'il produit. L'insuline est une hormone qui régule la concentration de sucre dans le sang. L'hyperglycémie, ou concentration sanguine élevée de sucre, est un effet fréquent du diabète non contrôlé qui conduit avec le temps à des atteintes graves de nombreux systèmes organiques et plus particulièrement des nerfs et des vaisseaux sanguins.

Les Indiens Pimas ont la plus forte prévalence connue dans le monde de diabète non insulino-dépendant: elle atteint 70 % entre 50 et 64 ans. Ce diabète est associé à une fréquence particulière de l'obésité mais celle-ci ne suffit pas à expliquer la prévalence record du diabète. Les modifications du mode de vie au cours du siècle dernier sont à l'origine de l'apparition du diabète chez les Indiens Pimas mais elles ne peuvent être tenues pour responsables de son caractère épidémique. Cette population présente une susceptibilité génétique particulière au diabète non insulino-dépendant. A degré d'obésité et niveau glycémique similaires, les Indiens Pimas, comparés aux Caucasiens, présentent une résistance à l'action de l'insuline. Cette caractéristique est très certainement en partie génétiquement déterminée et la recherche de (ou des) anomalie(s) en cause est très activement menée. De façon surprenante pour une population caractérisée par un tel degré d'insulino résistance, les Indiens Pimas ont une mortalité cardiovasculaire particulièrement faible. Comprendre les différences et similitudes entre les diabétiques Caucasiens et Pimas est une des voies pour faire progresser nos connaissances sur le diabète non insulino-dépendant et l'insulinorésistance.

Selon l'estimation de l'Organisation Mondiale de la Santé (OMS), il y aura 370 millions de diabétiques en 2030, soit 5,4% de la population mondiale. Il devient donc de plus en plus important de diagnostiquer si une personne a, ou est susceptible d'acquérir, le diabète.

Un diagnostic médical est un processus de classification. L'utilisation des méthodes dites intelligentes pour effectuer ce classement est de plus en plus fréquente.

## Introduction générale

---

Même si la décision de l'expert est le facteur le plus important dans le diagnostic, les systèmes de classification fournissent une aide substantielle, car elles réduisent les erreurs dues à la fatigue et le temps nécessaire pour le diagnostic. Il y a une grande variété de méthodes relatives à la classification et le diagnostic de la maladie du diabète.

Notre étude est basée sur l'approche dites System immunitaire artificiel (SIA).

AIRS (Artificial Immune Recognition System) a été proposé par Andrew Watkins (2001). Ce chercheur se sert de la métaphore du système immunitaire des vertébrés à reconnaître les antigènes, à mémoriser les cellules, et à sélectionner les clones.

Ce mémoire est organisé comme suit:

- Le chapitre 1 présente des généralités sur la maladie du diabète.
- Le chapitre 2 présente les méthodes utilisées: le système immunitaire artificiel et des algorithmes Knn et Naive Bayes.
- Analyse et discussion des résultats obtenus dans le chapitre 3.

Chapitre I

# LE DIABETE

## **Chapitre1** **Le Diabète**

### **1. Introduction :**

Le diabète est une maladie dans laquelle la production d'insuline est insuffisante pour répondre aux besoins de l'organisme et ou la réponse des cellules à l'insuline est anormale. Le carburant principal de l'organisme est un type de sucre appelé glucose, qui provient des aliments (une fois ceux-ci dégradés). Le glucose passe dans le sang, puis les cellules l'utilisent pour obtenir de l'énergie. Pour pouvoir se servir du glucose, l'organisme a besoin d'une hormone appelée insuline, qui est synthétisée par le pancréas. L'insuline est importante, car elle permet au glucose de quitter le sang pour pénétrer dans les cellules. Le diabète se manifeste lorsque l'organisme ne peut pas synthétiser d'insuline ou ne parvient pas à utiliser correctement l'insuline produite. [1]

Chez les personnes qui ont un diabète de type 2, l'organisme est devenu résistant à l'insuline. Dans ces cas, la production d'insuline continue, mais l'organisme ne répond plus aussi bien à ses effets. Après un certain temps, la capacité du pancréas à produire de l'insuline diminue. C'est pour cette raison que la maîtrise du diabète de type 2 nécessite une augmentation graduelle des types de médicaments que la personne prend, ce qui inclut parfois l'insuline. Que ce soit en raison du manque d'insuline ou de l'incapacité à utiliser correctement l'insuline, le résultat est une élévation du taux de sucre sanguin (de la glycémie), c'est-à-dire une hyperglycémie. Dans 90 % des cas de diabète, il s'agit d'un diabète de type 2. Ce type de diabète est également appelé diabète non insulino-dépendant (DNID) ou diabète de l'adulte. Bien que le diabète de type 2 puisse également survenir chez l'enfant, il commence généralement après l'âge de 40 ans [2] et devient plus courant avec le vieillissement. Environ 15 % des personnes de plus de 70 ans présentent un diabète de type 2.

### **2. Historique:**

Les indiens Pimas constituent une population originaire du Mexique. Il y a environ 30000 ans, ils ont migré aux Etats-Unis où ils se sont installés au sud de

# Chapitre I: Le Diabète

---

l'Arizona (dans le désert de Sonoran). Ils ont pu adapter leur mode de vie au désert, en créant des systèmes d'irrigation qui leur ont permis de cultiver du maïs, des haricots et du coton, et de se nourrir principalement de poissons et de gibier. A la fin du XIXe siècle, des colons se sont emparés des eaux de rivière propres aux indiens pimas. Ces derniers ont perdu leur activité physique et ont orienté leur nourriture essentiellement vers le sucre, le lard et la farine. Cette population possède, dès l'origine, des gènes de prédisposition au diabète. Au début, lorsque l'environnement dans lequel ils évoluaient était favorable, l'obésité et le diabète étaient rares. Après modification de leur environnement, plus de la moitié de la population indienne a développé un diabète de type 2, et plus de 75% des indiens ont contracté la maladie.

## **3. Définition**

Il existe différents types de diabète sucré : le diabète insulino-dépendant ou le diabète de type 1 et le diabète non insulino-dépendant ou diabète de type 2 (plus fréquent que le premier)

### **3.1 Le Diabète de type 1**

Le diabète insulino-dépendant représente 10 à 15% des cas de diabète. Il survient le plus souvent chez un sujet non obèse, avant l'âge de 30 ans.

Ce type de diabète se caractérise par une émission d'urine excessive (polyurie), une soif intense (polydipsie) et un appétit anormalement augmenté (polyphagie). Il a aussi pour conséquence un amaigrissement malgré une prise de nourriture abondante, une hyperglycémie (c'est-à-dire un excès de glucose dans le sang) supérieure à 1,26 g/l de sucre dans le sang à jeun, ou supérieure à 2 g/l (11 mmol/l) à n'importe quel moment de la journée, avec parfois une présence d'acétone dans les urines, accompagnée d'une haleine «de pomme reinette» caractéristique. [3]

### **3.2 Diabète de type 2**

Le diabète de type 2 qui touche les deux sexes avec une discrète majorité masculine, concerne surtout des individus âgés de plus de 40 ans. Il est toutefois en progression chez les sujets âgés entre 30 et 50 ans et il apparaît de plus en plus comme une complication fréquente de l'obésité de l'enfant, surtout aux U.S.A. On estime que

# Chapitre I: Le Diabète

---

80 % de l'ensemble des diabétiques ont des diabètes de type 2, 80 % d'entre eux étant en surpoids ou obèses.

Le terme diabète non insulino-dépendant a été abandonné pour celui de diabète de type 2, car si au moment du diagnostic et pendant au moins une décennie, l'insulinothérapie n'est pas nécessaire au traitement, celle-ci au contraire devient indispensable ou très utile (diabète insulino-requérant ou insulino-nécessitant) après plusieurs années d'évolution, du fait d'une insulino-pénie progressive et inéluctable.

Cette forme de diabète, par sa fréquence, est responsable de la très grande majorité des complications liées à l'ensemble des diabètes. Ainsi, ce diabète longtemps qualifié de « petit diabète », est en réalité un redoutable et « silencieux » pourvoyeur de complications.

## **4 Principaux facteurs de risque de diabète de type 2 :**

Dans les cas de diabète de type 2, soit le pancréas ne synthétise pas suffisamment d'insuline ou bien l'organisme n'utilise pas l'insuline correctement. Bien que la cause exacte du diabète de type 2 demeure mal comprise, il a été établi que ce trouble est plus fréquent chez les personnes présentant les caractéristiques suivantes :

- ✓ *Age supérieur à 40 ans* : Même si de plus en plus de jeunes sont atteints de diabète de type 2, l'âge reste un des principaux facteurs de risque. Ceci est dû au fait qu'avec l'âge, l'effet de l'insuline diminue, surtout chez les personnes en surpoids. Pour y remédier, le pancréas augmente sa production d'insuline. Lorsque l'organe a atteint ses limites de production, le diabète se manifeste.
- ✓ *Obésité.*
- ✓ *Antécédents familiaux de diabète* : les facteurs génétiques jouent également un rôle important dans le diabète de type 2. Par exemple, si un des parents est diabétique, les enfants ont un risque de 30 à 40 % de développer un diabète de type 2.
- ✓ *Appartenance à une population à risque élevé* (exemple : personnes d'origine autochtone, hispanique, sud-asiatique, asiatique ou africaine),
- ✓ *Maladie cardiaque.*
- ✓ *Taux de cholestérol anormal.*
- ✓ *Hypertension artérielle* : Souvent l'hypertension artérielle va de pair avec la surcharge pondérale et reste longtemps méconnue parce qu'elle ne présente pas de

## Chapitre I: Le Diabète

---

symptômes. Les personnes diabétiques souffrent trois fois plus souvent d'hypertension que les non-diabétiques. Si elle n'est pas traitée, l'hypertension augmente le risque d'infarctus du myocarde et d'accident vasculaire cérébral.

- ✓ *Diabète de grossesse.*
- ✓ *Accouchement d'un bébé de plus de 4 kg.*

Un bon nombre des facteurs énumérés (par exemple : taux de cholestérol anormal, hypertension artérielle, excès de poids) ont tendance à se manifester à la même période, ce qui crée un groupe de facteurs de risque auquel on donne le nom de « syndrome métabolique ». L'apparition de ces facteurs de risque a probablement une cause commune, soit une accumulation excessive de graisse abdominale. On croit qu'une diminution de cette réserve « entrale » de gras peut réduire la résistance à l'insuline et donc avoir un effet bénéfique sur un bon nombre des facteurs de risque énumérés.

### **5 Symptômes et Complications :**

Les personnes atteintes du diabète de type 2 peuvent ne présenter aucun symptôme pendant des années ou des dizaines d'années, mais, au fur et à mesure que la maladie évolue, les symptômes apparaissent. Les signes et les symptômes possibles sont les suivants :

- ✓ Soif et faim accrues,
- ✓ besoin fréquent d'uriner,
- ✓ fatigue,
- ✓ peau sèche accompagnée de démangeaisons,
- ✓ cicatrisation lente des coupures et des lésions,
- ✓ vision trouble,
- ✓ infections urinaires et vaginales fréquentes,
- ✓ diminution des sensations ou engourdissement dans les mains et les pieds,
- ✓ impuissance chez l'homme (dysfonction érectile).

## Chapitre I: Le Diabète

---

Malheureusement, dans beaucoup de cas, le diabète de type 2 passe inaperçu pendant plusieurs années et il est diagnostiqué seulement lorsque la personne consulte le médecin en raison de complications liées au diabète.

Un taux élevé de sucre dans le sang (hyperglycémie) peut mener à un trouble appelé intoxication au glucose, qui aggrave l'atteinte du pancréas et réduit la capacité de l'organisme à produire de l'insuline. Sans insuline, le taux de sucre continue d'augmenter et atteint un niveau qui provoque des lésions dans divers organes tels que les yeux, les nerfs et les reins. Une atteinte des yeux, plus particulièrement de la rétine, s'appelle rétinopathie diabétique et représente la première cause de cécité. Une atteinte des reins, un trouble appelé néphropathie diabétique, peut mener à l'insuffisance rénale et à la dialyse. Une atteinte des nerfs qui innervent les bras, les jambes et les voies digestives est qualifiée de neuropathie. Enfin, certains diabétiques qui ont une neuropathie périphérique (atteinte des nerfs dans les jambes) et une mauvaise circulation sanguine dans les jambes auront besoin d'une amputation un jour ou l'autre.

Les personnes diabétiques sont exposées à un risque accru de durcissement des grosses artères (athérosclérose), ce qui peut provoquer une crise cardiaque, un accident vasculaire cérébral ou une mauvaise circulation sanguine dans les jambes. Les lésions des petits vaisseaux sanguins peuvent nuire aux yeux, aux reins et aux nerfs.

Si la glycémie devient très élevée, plus particulièrement en présence d'autres facteurs de stress telle qu'une infection, les personnes atteintes du diabète de type 2 peuvent présenter des épisodes de confusion et d'étourdissements, voire des convulsions. On parle alors d'un coma hyperosmolaire, sans acidocétose et avec hyperglycémie, qui exige des soins médicaux immédiats.

Heureusement, il est possible de prévenir les complications du diabète, de retarder leur apparition ou de freiner leur évolution en maîtrisant la glycémie et en la maintenant le plus près possible des valeurs normales (glycémie à jeun normale : de 4 µl/L à 6 µl/L). [4]

# Chapitre I: Le Diabète

---

## 6 Critères diagnostiques

Seule une prise de sang faite en laboratoire déterminera avec certitude l'état de santé. Elle mesurera le taux de glucose (sucre) dans le sang.

Les critères diagnostiques du diabète proposés par *Les Lignes Directrices de Pratique Clinique 2008 pour la prévention et le traitement du diabète au Canada* sont :

- ▶ Glycémie, à jeun : 1,26 g/L (7,0 mmol/L) et plus OU
- ▶ Glycémie, 2 heures après l'ingestion de 75 g de glucose : 2 g/L (11,1 mmol/L) et plus OU
- ▶ Glycémie, à tout moment de la journée : 2 g/L (11,1 mmol/L) et plus, avec les symptômes classiques

Chez une personne non-diabétique, on considère normal:

- ▶ Glycémie, à jeun: moins de 5,6 mmol/L (entre 0,7 et 1,10 g/l)
- ▶ Glycémie, 2 heures après l'ingestion de 75 g de glucose : moins de 7,8 mmol/L

Parfois, une deuxième analyse est nécessaire afin de confirmer les résultats obtenus.

À partir de 40 ans, il est recommandé que toute personne qui ne présente pas de facteurs de risques fasse mesurer sa glycémie à jeun à tous les trois ans. [4]

## 7 Conclusion :

Le diabète touche des millions de personnes dans le monde. Si cette maladie n'est pas maîtrisée, elle peut entraîner d'autres troubles graves et invalidants. Un régime alimentaire sain peut faire une différence importante dans la prise en charge quotidienne du diabète.

A ce jour, on n'a pas encore trouvé de cure permettant de guérir le diabète, mais une médication adéquate, un bon régime alimentaire et quelques modifications du mode de la vie, peuvent permettre aux personnes diabétiques de mener une vie pratiquement normale tout en évitant à long terme les problèmes et les complications souvent associés à cette maladie.

## Chapitre I: Le Diabète

---

Le traitement du diabète par l'insuline consiste à normaliser la glycémie, d'une part, pour garantir un bien-être physique optimal et, d'autre part, pour éviter l'apparition de complications tardives. Un contrôle continu et régulier (c'est-à-dire le maintien de la glycémie proche de la zone normale), peut empêcher ou en tout cas freiner l'installation de ces complications potentiellement invalidantes.

Notre travail, dans le cadre de ce mémoire, consiste à implémenter un algorithme basé sur le Système immunitaire naturel pour la reconnaissance automatique du diabète.

Chapitre II

**LES SYSTEMES  
IMMUNITAIRES**

### Chapitre2 Les systèmes immunitaires

#### Introduction

L'étude du corps humain a toujours suscité une grande attention du fait de sa grande complexité. Notre domaine de recherche se base principalement sur l'extraction des métaphores utiles, à partir des systèmes biologiques, afin de créer des solutions informatiques efficaces aux problèmes complexes. Les développements les plus appréciables tirés du modèle humain, ont été les « réseaux de neurones » inspirés par *le fonctionnement du cerveau*, et les « algorithmes évolutionnaires » inspirés par *la théorie de l'évolution darwinienne*.

Cependant, plus récemment, on assiste à un intérêt croissant accordé à l'utilisation d'un autre système biologique qui est « le système immunitaire » comme source d'inspiration *pour résoudre des problèmes complexes*.

Le système immunitaire biologique est doté de capacités de traitement de l'information y compris l'identification du modèle, l'apprentissage, la mémorisation et le traitement parallèle distribué. Pour tout cela et pour d'autres raisons, le système immunitaire a suscité un intérêt significatif en vue de son emploi comme métaphore d'inspiration dans le calcul. Ce domaine de recherche est connu sous l'appellation des *systèmes immunitaires artificiels*.

Ce chapitre est composé de deux sections principales :

- Une première section est consacrée à la présentation du système immunitaire biologique, en montrant les différents composants immunitaires et les différents mécanismes utilisés par ce système. Nous aurons alors à récapituler les propriétés intéressantes du système immunitaire, qui constituent d'un point de vue informatique une source d'inspiration très riche.

- Dans la deuxième section, présente le système immunitaire artificiel (SIA) et le processus de conception d'un SIA. Nous présentons également les différents algorithmes immunitaires.

## Chapitre II: Les systèmes immunitaires

---

### SECTION 1- SYSTEMES IMMUNITAIRES NATURELS

Le NIS (naturel immun system) est un système complexe qui peut être vu sous différents angles : molécules, cellules et organes. Le système doit protéger le corps des entités dangereuses appelées antigènes. Les éléments de base du NIS sont les globules blancs ou lymphocytes. Pour pouvoir identifier les autres molécules, des lymphocytes particuliers (cellules B) produisent des récepteurs, appelés anticorps (paratope), responsables de la reconnaissance des antigènes. Le paratope se lie à une partie spécifique de l'antigène appelée épitope. Le degré de cette liaison ou affinité est très fort si seulement les deux formes sont complémentaires.

Ainsi, la reconnaissance d'un antigène par une cellule B est en fonction de l'affinité entre les anticorps de la cellule B et cet antigène. Les cellules B qui reconnaissent mieux l'antigène vont être proliférées en se clonant, selon le principe de la sélection clonale. Les clones subissent alors des mutations somatiques qui vont promouvoir leur variation génétique.

Lorsque la population atteint la maturité, les clones se différencient en cellules mémoires et cellules plasma. Cette expansion clonale confère au système immunitaire sa mémoire. D'un autre côté, les cellules B avec une faible affinité seront mutées, ou détruites par sélection négative.

#### **1. Historique**

L'immunologie est une science relativement nouvelle. Son origine est attribuée à Edward Jenner qui a découvert en 1796 (il y a un peu plus de 2 siècles), que la vaccine ou cow-pox, induit une protection contre la variole humaine, une maladie souvent mortelle (Janeway Jr. & Travers, 1997). Ce chercheur la baptise « vaccination hisprocess », expression qui décrit l'inoculation des individus en bonne santé par des échantillons affaiblis ou atténués d'agents qui provoquent des maladies, en vue d'obtenir une protection contre ces maladies.

Lorsqu'Edward Jenner introduisit la vaccination, on ne savait encore rien sur l'agent étiologique de l'immunologie. Au XIX<sup>e</sup> siècle, Robert Koch confirma que les maladies infectieuses sont causées par des microorganismes pathogènes, dont chacun est responsable d'une pathologie donnée.

## Chapitre II: Les systèmes immunitaires

---

Au début des années 1900, Jules Bordet et Karl Landsteiner portent la discussion sur la notion de « spécificité immunologique ». Il montre que le système immunitaire est capable de produire des anticorps spécifiques contre les produits chimiques synthétisés artificiellement, et qui donc n'ont jamais existé à l'état naturel.

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, ce serait l'immunité innée qui serait la véritable gardienne des clefs du déclenchement d'une réponse immunitaire.

### 2. Concepts immunologiques

Le système immunitaire étant très complexe, il doit être vu sous différents angles. Nous pouvons le considérer en tant qu'un ensemble d'organes, molécules ou cellules.

#### 2.1 Les organes immunitaires :

Le système immunitaire est constitué principalement des organes suivants:

- *une Moelle osseuse* : c'est le lieu de maturation des lymphocytes B.
- *un Thymus* dans le bas du cou ; il constitue le site de maturation des lymphocytes T.
- *des Vaisseaux lymphatiques* : ils transportent la lymphe ; les vaisseaux lymphatiques sont situés dans tout le corps.

#### 2.2 Les cellules immunitaires :

Les globules blancs appelées leucocytes sont impliquées dans la défense. Elles se trouvent dans le sang et la lymphe. Il existe différents types de globules blancs, les plus importantes étant les lymphocytes. Chez l'homme les lymphocytes représentent 5 à 15% des lymphocytes sanguin [5] et se composent de :

**a/Les lymphocytes B**, ou cellules B, appelés également bursocytes. Ces lymphocytes ont pour rôle de fabriquer des immunoglobulines appelées anticorps : ils sont responsables de l'immunité humorale. Ils possèdent deux fonctions essentielles :

- Leur activation par un corps étranger induit leur transformation en cellules sécrétrices d'immunoglobulines.

## Chapitre II: Les systèmes immunitaires

---

- Les lymphocytes B ont également la capacité de se comporter en cellule présentant le corps étranger.

**b/Lymphocytes T:** c'est une catégorie de lymphocytes qui joue un grand rôle dans la réponse immunitaire primaire mais également secondaire. « T » est l'abréviation de thymus. Il existe différentes variantes des lymphocytes T. Leur rôle est d'attaquer les cellules infectées. 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*, servent à éviter les réactions immunitaires non appropriées (maladies auto-immune).

**c/Anticorps:** Un anticorps est une protéine complexe utilisée par le système immunitaire pour détecter et neutraliser les agents pathogènes de manière spécifique. Les anticorps constituent l'immunoglobuline principale du sang. Ils sont sécrétés par des cellules dérivées des lymphocytes B : 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 [5] .

### 2.3 Les antigènes :

Un antigène est une macromolécule naturelle ou synthétique, reconnue par des anticorps ou des cellules du système immunitaire et capable d'engendrer une réponse immunitaire. Les antigènes sont généralement des protéines, des polysaccharides et leurs dérivés lipidiques. On distingue deux classes d'antigènes :

- ✓ Les antigènes de classe 1, présents dans toutes les cellules;
- ✓ Les antigènes de classe 2, présents uniquement dans les cellules immunitaires.

Les antigènes, en tant que marqueurs des agents étrangers à l'organisme, sont à la base de la réponse immunitaire adaptative. C'est la reconnaissance de l'antigène par les cellules immunocompétentes, directement ou via les cellules présentatrices d'antigène (CPA), qui active l'immunité spécifique. [6]

## Chapitre II: Les systèmes immunitaires

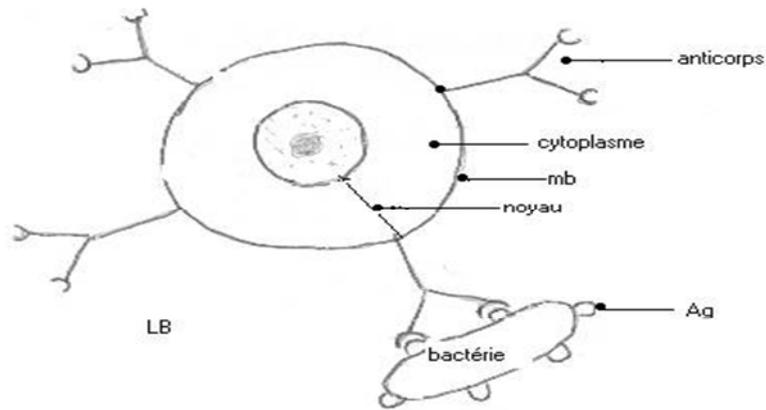


Figure 1: Structure d'un antigène/anticorps

### 3. Architecture du système immunitaire :

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.

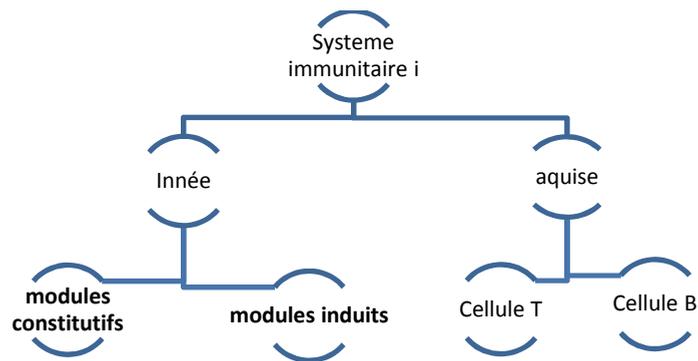


Figure 2 : Architecture du système immunitaire

#### 3.1 Immunité innée :

L'immunité innée est la première ligne de défense vis-à-vis des agents infectieux et pathogènes qui nous entourent. Elle est mise en jeu immédiatement et est fonctionnelle 4 jours (96 heures). Elle met en jeu différents modules de défense :

## Chapitre II: Les systèmes immunitaires

---

- ✓ Des modules constitutifs comme la barrière peau-muqueuse.
- ✓ Des modules induits comme la phagocytose et la réponse inflammatoire, qui nécessite les cellules phagocytaires et les cytokines.

La réponse immunitaire innée est induite par un signal danger émis suite à l'interaction spécifique entre des récepteurs du soi appelés PRR (pour « Pattern Recognition Receptors ») et des molécules du non-soi appelées PAMP (pour « Pathogen Associated Molecular Patterns ») présentes au niveau des microorganismes qu'ils soient pathogène ou non.

La réponse du système immunitaire inné est non spécifique à un type particulier d'intrus mais elle est identique contre tous les pathogènes qui envahissent le corps. Il joue un rôle vital pour l'initialisation et la régularisation de la réponse immunitaire adaptative. [7]

### 3.2 Immunité adaptative

L'immunité adaptative est constituée de types différents de cellules dont chacune joue un rôle important. Elle résulte du contact du système immunitaire avec les antigènes grâce à la caractéristique d'apprentissage et mémorisation du système immunitaire.

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. 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. La réponse de l'immunité adaptative est lancée après la réponse de l'immunité innée, les deux types d'immunités sont liées et se complètent. [8]

#### 4. Propriétés du système immunitaire :

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

## Chapitre II: Les systèmes immunitaires

---

**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'hyper mutation 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.

**Mémorisation** : Après une réponse immunitaire à un antigène donné, un ensemble de cellules constituent l'ensemble des cellules mémoires qui seront dotées pour une durée de vie longue afin de fournir des réponses immunitaires plus rapides et plus puissantes aux rencontres suivantes d'un 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.

## Chapitre II: Les systèmes immunitaires

---

**Détection**: Le système immunitaire est capable d'identifier et de détecter les intrus dans le corps sans aucune connaissance antérieure de la structure de ces intrus.

### 5. **Théories immunitaires** :

Le comportement et les réactions du système immunitaire sont principalement régis par des théories immunitaires :

#### 5.1 La sélection Négative/Positive :

Cette théorie gère le processus de création au niveau de la discrimination entre soi et non soi. Les lymphocytes ont sur leurs surfaces des récepteurs (paratopes). Les lymphocytes, issus de la moelle osseuse, migrent vers le thymus ; à ce stade ils sont appelés cellules T naïves ou immatures. Leurs paratopes subissent un processus de réarrangement génétique pseudo aléatoire, puis un test très important est mis en place [9]. Le test en question consiste à vérifier si les nouveaux récepteurs s'attaquent aux cellules du soi ; dans ce cas ces lymphocytes sont détruits et purgés de la population des nouveaux lymphocytes : on parle de sélection négative. Le reste de la population est autorisé à quitter le thymus pour circuler dans le sang et effectuer leurs tâches de surveillance [8].

#### 5.2 La sélection clonale :

La théorie de la sélection clonale décrit les conséquences de la réponse immunitaire suite à un stimulus antigénique en assurant que seules les cellules qui reconnaissent l'antigène subissent des proliférations et différenciations.

Quand un antigène envahit le corps, des cellules immunitaires reconnaissent cet antigène avec des degrés d'affinité différents. La réponse des cellules B est la production des anticorps dont chaque cellule sécrète un seul type d'anticorps qui est relativement spécifique à l'antigène. L'appariement fort entre les récepteurs des anticorps et l'antigène, produit la stimulation des cellules B, c'est-à-dire la prolifération (clone) et la maturation des cellules de plasma.

Le taux de prolifération d'une cellule est directement proportionnel à son affinité. Avec l'antigène les cellules qui ont les plus grandes affinités seront les plus proliférées et réciproquement. En plus, les lymphocytes qui ont une forte affinité peuvent se différencier en des cellules mémoires. [9]

## Chapitre II: Les systèmes immunitaires

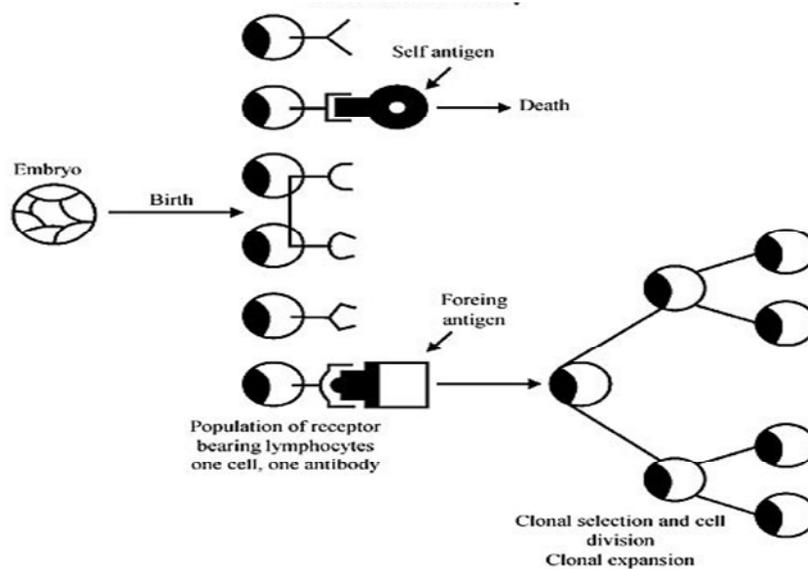


Figure 3 théorie de la sélection clonale

### 5.3 Théorie du danger:

La théorie du danger gère le comportement du système et sa réaction selon les nouvelles conditions suivantes :

- Le système immunitaire ne doit pas réagir contre le soi, sauf si ce dernier est dangereux.
- Le système immunitaire réagit contre le non soi, sauf si ce dernier n'est pas dangereux.

### **6. Déroulement d'une réponse immunitaire :**

Une importante collection de cellules et de molécules protège notre corps. Ces dernières collaborent contre n'importe quelle molécule étrangère comme les bactéries ou d'autres envahisseurs :

- ✓ Quand un intrus envahit le corps, les cellules de présentation antigénique comme les macrophages procèdent à l'ingestion et la digestion de l'antigène rencontré pour le présenter comme des fragments de peptides antigéniques, ces peptides seront liées avec les molécules MHC pour permettre leurs liaisons avec les cellules T qui ont la capacité de reconnaître la combinaison de peptide / MHC.
- ✓ Les cellules T activées par cette identification produisent et sécrètent des lymphokines ou des signaux chimiques pour mobiliser d'autres composants du système immunitaire.

## Chapitre II: Les systèmes immunitaires

---

- ✓ Les cellules B qui ont aussi des molécules de récepteur complémentaires répondent à ces signaux. À la différence des récepteurs de cellules T, ceux de cellules B peuvent reconnaître les parties d'antigènes libres sans les molécules MHC. Après cette activation, les cellules B prolifèrent et se différencient et sécrètent des protéines d'anticorps.
- ✓ La liaison entre les anticorps et les antigènes disponibles mènent à la destruction et la suppression des antigènes.
- ✓ Certaines cellules B et T deviennent des cellules mémoires qui ont une durée de vie illimitée, en permettant l'élimination rapide de l'antigène s'il se présente une autre fois dans l'avenir.

### **SECTION 2- SYSTEMES IMMUNITAIRES ARTIFICIELS**

#### **1. Historique :**

C'est seulement dans le milieu des années 90 que les *SIA* devinrent 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.

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 explorées. 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 *AIS*. Au départ, les travaux sur les *AIS* visaient à trouver des abstractions efficaces des phénomènes découverts dans le système immunitaire.

Plus récemment, les praticiens des *AIS* se sont aussi intéressés à la modélisation du système immunitaire et à l'application des résultats issus des *SIA* aux problèmes d'immunologie (ce qui entre dans le cadre de l'immuno-informatique). [9]

## Chapitre II: Les systèmes immunitaires

---

### 2. Définition :

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* » [10].

De leur côté, Timmis et De Castro [11] ont donné la définition suivante : « *Les systèmes immunitaires artificiels sont des systèmes adaptatifs inspirés par des théories immunologiques et des observations de fonctions immunitaires, des principes et des modèles, qui seront appliqués à la résolution des problèmes* ».

Les systèmes immunitaires artificiels sont une nouvelle branche de l'intelligence artificielle. Destinés à résoudre des problèmes divers, ils s'inspirent des propriétés et concepts remarquables du système immunitaire biologique.

Les systèmes immunitaires artificiels sont une implémentation mathématique ou informatique du fonctionnement du système immunitaire naturel. Cette implémentation reprend les plus grandes lignes de son fonctionnement. Cependant il reste presque impossible de modéliser le comportement complet des systèmes biologiques. [12]

### 3. Modélisation des systèmes immunitaires artificiels :

Le modèle commun connu sous le nom du Framework des systèmes immunitaires artificiels, définit les règles que doit respecter un AIS ainsi que les processus à suivre pour l'élaboration de nouvelles approches. Les conditions nécessaires sont :

- ✓ La représentation des composants systèmes (modèles abstraits des cellules immunitaires).
- ✓ L'utilisation des mesures d'affinité pour évaluer l'affinité entre les composants systèmes.
- ✓ Un ensemble d'algorithmes pour contrôler l'évolution et la dynamique d'AIS.

Les trois conditions citées ci-dessus sont indispensables pour l'élaboration d'un Framework pour définir un système immunitaire artificiel.

*Ce schéma représente le processus de modélisation d'un AIS :*



Figure 4 Structure de conception d'un AIS

### 3.1 La représentation :

Les cellules B et T sont les cellules les plus importantes dans le système immunitaire. Elles présentent des récepteurs superficiels utiles pour la reconnaissance des intrus, récepteurs dont les formes sont complémentaires à la forme d'antigène. Les cellules et les molécules immunitaires sont alors les éléments qui doivent être modélisés et employés dans les modèles proposés par le système immunitaire artificiel.

Il est assumé que chaque antigènes agit spécifiquement avec tous les anticorps dont les compléments existent dans une petite région d'encerclement. Cette région est caractérisée par un paramètre «S » appelé le seuil d'affinité.

Le résultat de la définition du seuil d'affinité est le volume  $V_s$  qui est appelé la région d'identification.

### 3.2 Les mesures d'affinités :

L'affinité entre un anticorps et un antigène est relative à leur distance, Un antigène est représenté par un vecteur  $Ag = \langle Ag_1, Ag_2, \dots, Ag_L \rangle$ , un anticorps est à son tour représenté par un vecteur  $Ab = \langle Ab_1, Ab_2, \dots, Ab_L \rangle$ .

Pour mesurer le degré de complétude entre l'antigène et l'anticorps, plusieurs techniques peuvent être utilisées. Le plus souvent on recourt à l'utilisation des distances. Différentes distances existent dont voici les plus utilisées :

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

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

La distance de Hamming :  $D = \sum_{i=1}^n \delta^i$  où  $\delta = \begin{cases} 1, & Ab_i \neq Ag_i \\ 0, & \text{sinon} \end{cases}$

## Chapitre II: Les systèmes immunitaires

---

Plus la distance antigène-anticorps est petite, plus l'affinité entre ces derniers, est grande.

Il ne reste plus qu'à implémenter les théories immunitaires que nous allons voir dans la suite. Les principales théories immunitaires sont la théorie de la sélection clonale, la théorie de la sélection négative/positive et la théorie du danger.

### 3.3 Les algorithmes immunitaires:

Il existe différents algorithmes des théories immunitaires selon le contexte et le problème à résoudre.

#### 3.3.1 L'algorithme de la sélection négative/positive

L'idée sur laquelle se base cette théorie est que seules les cellules T qui ne s'attaquent pas aux cellules du soi sont autorisées à quitter le thymus et auront pour tâche de reconnaître les cellules du non soi.

Cette notion est très intéressante, surtout pour les applications de surveillance des systèmes et la détection d'utilisations anormales ou inhabituelles. Forrest et al ont proposé l'algorithme de la sélection négative qui reflète ce principe. Ils ont considéré l'algorithme de la sélection négative comme un processus de détection d'anomalies composé de trois phases principales :

- ✓ La définition du soi.
- ✓ La génération des détecteurs.
- ✓ Le contrôle d'occurrence des anomalies.

L'algorithme se déroule comme suit : On considère un ensemble donnée de modèles de soi (P). Générer un ensemble de détecteurs (M) qui n'identifient aucun élément appartenant à l'ensemble (P).

La génération des détecteurs (M) se passe par les étapes itératives décrites comme suit :

-générer un ensemble de cellules candidats aléatoirement de l'ensemble P.

- Calculer l'affinité entre chaque cellule C et tout l'ensemble de soi P.

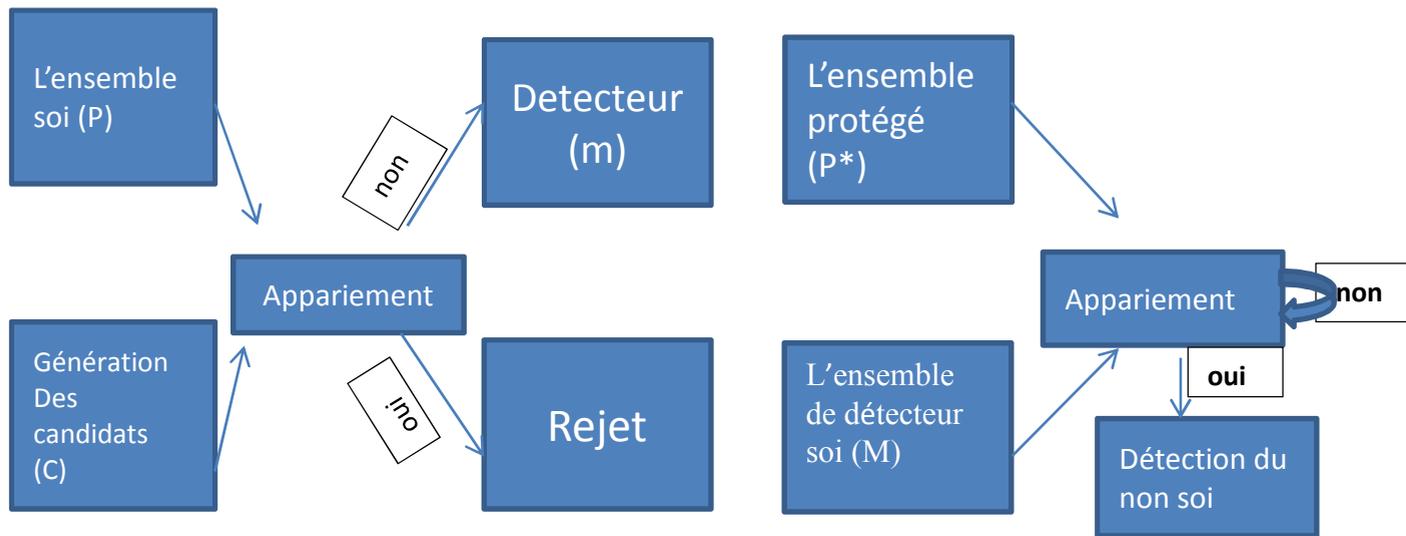
-**Si** l'affinité entre un élément C et au moins un élément P est supérieur ou égal à un seuil d'affinité prédéfini, alors cet élément C sera supprimé (il est considéré comme un élément de soi) .

-**Si non** il sera considéré comme un détecteur de non soi et sera ajouté à l'ensemble de détecteur M.

## Chapitre II: Les systèmes immunitaires

Après avoir obtenu l'ensemble de détecteur, la prochaine étape sera de détecter la présence du modèle de non soi.

Ce schéma résume l'algorithme de sélection négative:



**Figure 5 La structure générale de l'algorithme de la sélection négative**

L'algorithme de la sélection positive est une alternative de l'algorithme de la sélection négative. La différence principale est la génération des détecteurs qui détectent des éléments de soi au lieu de ceux 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.

### 3.3.2 L'algorithme de la sélection clonale :

Cet algorithme se base sur le principe que seules les cellules ayant reconnu l'antigène, seront clonées et mutées.

CLONALG est un algorithme proposé par De Castro & Von Zuben. Les étapes de base de cet algorithme peuvent être résumées en 3 étapes, comme suit :

*Initialisation :*

- Créer un ensemble d'anticorps de taille N.
- Créer un sous ensemble de cellules mémoires MC qui sera initialisé aléatoirement à partir de l'ensemble d'anticorps.

## Chapitre II: Les systèmes immunitaires

---

*Clonage et mutation :*

- Sélectionner un antigène ag de l'ensemble des Ags.
- Pour chaque anticorps ac de l'ensemble des anticorps AC, calculer son affinité avec l'antigène ag de l'ensemble AG.
- Sélectionner les n anticorps qui ont la plus grande affinité et les cloner.
- Muter ces clones pour augmenter leur degré de correspondance avec l'antigène a.
- Exposer ces clones mutés de nouveau à l'antigène ag et recalculer leurs affinités.
- Placer les meilleurs clones dans l'ensemble des cellules mémoires mc.
- Remplacer les cellules de faible affinité de N par des nouvelles cellules aléatoires.
- Répéter cette étape tant que certains critères sont vérifiés selon le problème à résoudre.

*Fin :*

- L'ensemble MC représente la solution de l'algorithme
- La solution globale peut être un individu ou une collection d'individus de l'ensemble MC.

### ***3.3.3 L'algorithme du réseau immunitaire :***

La théorie du réseau immunitaire a suggéré un système immunitaire avec un comportement dynamique même en absence d'un antigène de non soi. Le déroulement de cet algorithme peut être résumé comme suit :

- Initialiser le réseau avec une population d'anticorps générés aléatoirement.
- Déterminer l'affinité entre chaque modèle d'antigènes et chaque élément l.
- Sélectionner les anticorps avec la plus grande affinité et les cloner selon leurs degrés d'affinité.
- Les clones produits subissent l'hyper-mutation somatique qui est inversement proportionnelle à leur affinité antigénique.
- Les clones avec la plus haute affinité sont maintenus dans le réseau pour constituer la mémoire clonale.
- Eliminer les anticorps restant qui ont une affinité inférieure à un seuil donné.
- Eliminer tous les clones mémoires dont l'affinité avec l'antigène est inférieure à un seuil prédéterminé.
- Incorporer les clones restants de l'ensemble mémoire avec les anticorps du réseau.
- déterminer la similitude entre chaque paire d'anticorps du réseau.
- éliminer tous les anticorps du réseau dont l'affinité est moins d'un seuil prédéterminé.
- Répéter ces étapes tant que certains critères sont vérifiés.

## Chapitre II: Les systèmes immunitaires

### 3.4 Etude comparative entre les différents systèmes inspirés de la biologie :

Ce tableau récapitule la comparaison entre les différents systèmes inspirés de la biologie qui sont : les systèmes immunitaires artificiels inspirés du système immunitaire humain, les réseaux de neurones inspirés du fonctionnement du cerveau et les algorithmes évolutionnaires inspirés par la théorie de l'évolution darwinienne.

<b>Système Caractéristique</b>	<b>AIS</b>	<b>RNA</b>	<b>AG</b>
<b>Composants</b>	Chaîne d'attribut	Neurones artificiels	Chaînes de chromosomes
<b>Structure</b>	Ensemble d'éléments discrets/gérés en réseau	Neurones gérés en réseau	Élément discret
<b>Stockage de la Connaissance</b>	Chaînes d'attributs / connexion réseau	Poids de connexion	Chaînes Chromosomiques
<b>Dynamique</b>	Apprentissage/Evolution	Apprentissage	Evolution
<b>Interactions avec d'autres composants</b>	Par l'identification des chaînes d'attribut ou des connexions réseau	Par des connexions du réseau	Par des opérateurs de recombinaison et/ou la fonction d'évaluation
<b>Seuil</b>	Influence l'affinité des éléments	Influence l'activation de neurone	Influence les Variations génétiques
<b>Etat</b>	Concentration et affinité	Niveau d'activation des neurones de sortie	L'information génétique dans les chromosomes

Tableau 1 Comparaison entre les différents systèmes inspirés de la biologie

### 4. Domaines d'application des SIA :

Le système immunitaire artificiel peut être utilisé dans plusieurs domaines d'application. Selon le type de problème, on peut citer parmi ces domaines :

- ✓ La classification des données.
- ✓ Reconnaissance de formes.
- ✓ Robotique.
- ✓ Optimisation.
- ✓ La détection et l'élimination des virus informatiques.
- ✓ La sécurité informatique.

## Chapitre II: Les systèmes immunitaires

---

### 5. L'algorithme A.I.R.S (Artificial Immune Recognition System):

#### 5.1 Introduction

AIRS (système de reconnaissance immunitaire artificiel) est inspiré d'algorithme d'apprentissage supervisé.

Notre motivation pour cette étude vient du fait qu'il y a un manque important de recherches qui ont exploré l'utilisation de la métaphore du système immunitaire pour l'apprentissage supervisé. En effet, le seul travail identifié est celui de Carter [13]. Toutefois, il a été noté que dans la communauté AIS, il y a eu un certain nombre d'enquêtes sur l'exploitation des mécanismes immunitaires à l'apprentissage non supervisé.

Dans son étude, Castro [14] a examiné le rôle du processus de sélection clonale au sein du système immunitaire. Il a développé un apprentissage non supervisé connu sous le nom CLONALG (cité précédemment).

Nous décrivons brièvement les aspects les plus pertinents de l'immunologie qui nous ont servi de source d'inspiration pour ce travail.

Tout au long de la vie d'une personne, le corps est exposé à une grande variété de pathogènes (potentiellement dangereux). Le système immunitaire contient des lymphocytes connus sous le nom de cellules B et T ; chacune de ces cellules a un type unique de récepteur moléculaire (localisation dans un espace de forme). Les récepteurs dans cet espace de forme, permettent la liaison des Antigènes avec une plus grande affinité entre le récepteur et l'antigène qui indique une forte liaison.

Dans AIRS, l'idée de liaison antigène / anticorps est connue et employée comme la présentation antigénique. Lors du traitement des algorithmes d'apprentissage, elle est utilisée pour mettre en œuvre l'idée de l'adéquation entre les données d'apprentissage (antigènes) et les solutions possibles (cellules B). Dans leur travail, Timmis, J. and M. Neal [15] ont employé l'idée de « artificial recognition ball » (ARB), qui a été inspirée par le travail de Farmer [16] décrivant l'interaction antigénique au sein d'un réseau immunitaire.

Autrement dit, un ARB peut être considéré comme représentant un certain nombre de cellules B identiques ; c'est un mécanisme utilisé pour réduire les redondances. Une fois l'affinité entre une cellule B et un antigène déterminé, la cellule B concernée se transforme en une cellule plasma et va être clonée. Pendant le processus

## Chapitre II: Les systèmes immunitaires

---

d'expansion clonale, les cellules B subissent une prolifération rapide (clonage) en proportion de la façon dont elles correspondent à l'antigène. Cette réponse est l'antigène spécifique.

Ces clones passent par la maturation d'affinité, où certains engagent une hyper mutation somatique (la mutation est ici inversement proportionnelle à l'affinité antigénique) et finiront par passer par un processus de sélection par lequel une cellule donnée peut devenir une cellule mémoire. Ces cellules mémoire sont conservées pour permettre une réponse plus rapide au même antigène, ou à un antigène similaire. Dans l'algorithme AIRS, l'idée de l'expansion clonale et la maturation d'affinité, sont employées pour encourager la production de cellules de mémoire potentiel.

Ces cellules mémoires sont ensuite utilisées pour la classification. S'inspirant des travaux de Timmis [15], AIRS utilise l'idée de niveau de stimulation pour un ARB, qui, encore une fois, a été calculé à partir des équations d'un réseau immunitaire. Bien que AIRS a été inspiré par ce travail sur les réseaux immunitaires, le développement du classificateur conduit à l'abandon des principes du réseau.

Toutefois, AIRS ne prend pas en compte le principe de la mutation d'affinité proportionnelle. Lorsque les ARB ont été créés, ils ont été soumis à un processus de mutation aléatoire avec une certaine probabilité et ont ensuite été intégrés à l'ensemble des cellules de mémoire lorsque leur affinité satisfait à certains critères. Dans le système AIRS, les ARB survivants reposent sur l'idée d'un système de ressources limitées.

Un nombre prédéfini de ressources existent, pour lesquels les ARB participent en fonction de leur niveau de stimulation: plus la valeur de stimulation d'un ARB est haute plus le nombre de ressources allouées est grand. Les ARB qui ne peuvent pas résister à la concurrence pour les ressources, seront retirés du système. Le terme métadynamique du système immunitaire se réfère à l'évolution constante des cellules B à travers la prolifération cellulaire et la mort. Cela a été présenté dans AIRS à la production continue et à l'élimination des ARB de la population. Ce tableau résume la correspondance entre le système immunitaire et AIRS :

## Chapitre II: Les systèmes immunitaires

Système immunitaire	AIRS
Anticorps	Vecteur de caractéristique
Expansion clonale	Reproduction des ARB qui correspondent le mieux avec des antigènes
Antigènes	Donnée d'apprentissage
Maturation d'affinité	Mutation aléatoire de l'ARB et le retrait des ARB les moins stimulés
Mémoire immunitaire	L'ensemble mémoire des ARB mutés
Metadynamics	L'enlèvement continu et la création des ARB et des cellules de mémoire

Tableau 2 Correspondance entre AIRS et le système immunitaire

### **5.2 Vue d'ensemble de l'algorithme AIRS :**

Cette section présente un aperçu de l'algorithme AIRS. On définit d'abord (5.2.1) quelques-uns des principaux termes et concepts importants pour la compréhension de l'algorithme. On détermine ensuite (5.2.2) le déroulement de l'algorithme d'apprentissage. Enfin, on présente une vue d'ensemble plus conceptuelle de l'algorithme (5.2.3).

#### ***5.2.1 Définition :***

Ce paragraphe présente les définitions des principaux termes et concepts qui s'appliquent à l'algorithme AIRS. Ces concepts sont les suivants :

- ✓ *affinité*: c'est une mesure de «proximité» ou de similarité entre deux anticorps ou antigènes. Cette valeur est assurée d'être entre 0 et 1 et se calcule simplement avec la distance euclidienne de vecteurs de caractéristiques de deux objets. Ainsi, les valeurs d'*affinité petite* indiquent une forte affinité.
- ✓ *seuil d'affinité (AT)*: c'est la valeur d'affinité moyenne parmi tous les antigènes dans l'ensemble d'apprentissage ou au sein d'un sous-ensemble sélectionné de ces antigènes d'apprentissage.
- ✓ *scalaire du seuil d'affinité (ATS)*: c'est une valeur comprise entre 0 et 1 ; quand elle est multipliée par le seuil d'affinité, elle fournit une valeur seuil pour le remplacement de cellules mémoire dans la boucle d'apprentissage AIRS.
- ✓ *Anticorps* : c'est un vecteur de caractéristique associé avec la classe. La combinaison vecteur de caractéristique/classe est appelée un anticorps quand elle fait partie d'un ARB ou d'une cellule mémoire.

## Chapitre II: Les systèmes immunitaires

---

- ✓ *Antigène*: dans sa représentation, c'est similaire à un anticorps, mais la combinaison de vecteurs caractéristiques-classe est appelée un antigène quand elle est présentée à l'ARB pour la stimulation.
- ✓ *Artificiel Immune Recognition System(AIRS)* : c'est l'algorithme de classification inspirée du système immunitaire naturel.
- ✓ *Artificielle Reconnaissance Ball (ARB)*: également connu sous le nom cellule B, l'ARB se compose des anticorps, un comptage du nombre de ressources détenues par la cellule, et la valeur de stimulation de la cellule courante.
- ✓ *Cellule mémoire candidate*: l'anticorps d'un ARB, de la même classe que l'antigène d'apprentissage, qui était le plus stimulée après l'exposition à l'antigène donné.
- ✓ *Classe*: il s'agit de la catégorie d'un vecteur caractéristique donné. Ceci est aussi appelé la sortie d'une cellule.
- ✓ *Taux de clonage* : une valeur entière utilisée pour déterminer le nombre de clones mutés pour un ARB donné. Dans l'implémentation actuelle, un ARB sélectionné est autorisée à produire jusqu'à (taux de clonage x valeur stimulation) clones mutés après avoir répondu à un antigène donné. Ce produit est également utilisé dans l'attribution des ressources à un ARB. Par conséquent, le taux de clonage a un double rôle :c'est un facteur d'allocation des ressources et un facteur de mutation clonale de la population des cellules.
- ✓ *Cellule mémoire établi(mc match)*: c'est l'anticorps d'un ARB qui a survécu à la concurrence pour les ressources et a été le plus stimulé à un antigène d'apprentissage et qui a été ajouté à l'ensemble des cellules de mémoire.
- ✓ *Vecteur de caractéristiques*: c'est une instance de données représentées par une séquence de valeurs, et où chaque élément a ses propres valeurs.
- ✓ *Taux d'hyper-mutation*: une valeur entière utilisée pour déterminer le nombre de clones mutés pour une cellule mémoire donné, autorisée à injecter dans la population cellulaire. Dans l'implémentation actuelle, la cellule mémoire sélectionnée injecte au moins (taux d'hyper-mutation \* taux de clonage \*valeur de stimulation ) clones mutés dans la population cellulaire au moment de l'introduction d'antigène.
- ✓ *Cellules mémoire (MC)*: l'anticorps d'un ARB qui était le plus stimulé par un antigène d'apprentissage donné à la fin de l'exposition à cet antigène. Il est utilisé pour l'hyper-mutation en réponse à des antigènes d'apprentissage entrant. Un mc

## Chapitre II: Les systèmes immunitaires

---

peut être remplacé, toutefois. Cela se produit uniquement lorsque le mc candidat est plus stimulé à un antigène d'apprentissage donné que la cellule mémoire établie la plus stimulée et l'affinité entre le mc établi et le mc candidat est inférieure au produit du seuil d'affinité et de l'affinité Seuil scalaire.

- ✓ *Taux de mutation* : c'est un paramètre entre 0 et 1 qui indique la probabilité que toute caractéristique donnée (ou la sortie) d'un ARB sera muté.
- ✓ *Sortie*: il s'agit de la catégorie de classification associée à une cellule. La même que la classe du vecteur de caractéristiques correspondant à cette cellule
- ✓ *Ressource*: un paramètre qui limite le nombre d'ARB permis dans le système. A chaque ARB est attribué un certain nombre de ressources en fonction de sa valeur de stimulation et le taux de clonage. Le nombre total des ressources du système est fixé à une certaine limite. Si plus de ressources sont consommées que ne sont autorisées à exister dans le système, alors les ressources sont retirées du moins jusqu'à ce que le nombre de ressources d'un ARB stimulé dans le système retourne au nombre autorisé. Si toutes les ressources de données ARB sont retirées, alors l'ARB est retiré de la population de cellules.
- ✓ *cellule initial* (seed cell) : c'est un anticorps, tiré de l'ensemble d'apprentissage, utilisé pour initialiser les cellules mémoire et ARB au début d'apprentissage.
- ✓ *fonction de stimulation*: Dans la formulation actuelle du classificateur AIRS, cette fonction doit avoir une valeur entre 0 et 1 inclusivement. Pour la mise en œuvre de l'AIRS présenté dans cette étude, la fonction de stimulation est inversement proportionnelle à la distance euclidienne entre les vecteurs de caractéristiques de l'ARB et l'antigène.
- ✓ *valeur de stimulation*: la valeur retournée par la fonction de stimulation.
- ✓ *seuil de stimulation*: un paramètre entre 0 et 1 utilisé comme critère d'arrêt pour l'apprentissage sur un antigène spécifique.
- ✓ *ensemble de test*: ensemble d'antigènes utilisés pour évaluer les performances de classification du classificateur AIRS.
- ✓ *Ensemble d'apprentissage* : la collecte des antigènes utilisés pour l'entraînement du classificateur AIRS.

## Chapitre II: Les systèmes immunitaires

---

### *b/ Déroulement de l'algorithme :*

Cette section présente un aperçu des routines primaires, les méthodes et les équations utilisées dans l'apprentissage et le renforcement d'un système immunitaire basé sur un classificateur.

Les étapes principales impliquées dans l'algorithme de AIRS sont :

- La normalisation des données et l'initialisation.
- L'identification des cellules mémoires et la génération ARB.
- La compétition pour les ressources dans le développement d'une cellule mémoire candidate.
- L'introduction potentielle de la cellule mémoire candidate dans l'ensemble des cellules de mémoire mis en place.

On va établir les notations suivantes :

- ✓ MC représente l'ensemble des cellules de mémoire et mc représente un membre individuel de cet ensemble.
- ✓ Ag.c représente la classe d'un antigène donné ag, et  $ag.c \in C = \{1, 2, \dots, nc\}$  et nc est le nombre de classe de cet ensemble d'antigènes.
- ✓ Mc.c représente la classe d'une cellule mémoire donnée mc,  $mc.c \in C = \{1, 2, \dots, nc\}$ .
- ✓ On définit  $MC_c = \{MC_1, MC_2, \dots, MC_n\}$  et  $mc \in MC_c$  si  $mc.c \equiv c$ .
- ✓ Ag.f et mc.f représente le vecteur de caractéristique d'un antigène et cellule mémoire donnée ag et mc, respectivement, et soit  $ag.fi$  représente la valeur du ieme caractéristique dans ag.f et  $mc.fila$  valeur du du ieme caractéristique dans ag.f.
- ✓ AB représente l'ensemble des ARB, et MU représente l'ensemble des clones mutés des ARB. Par ailleurs, ab représente un élément d'un ARB quand  $ab \in AB$ .  
Ab.c représente la classe de l'ARB donné ab, quand  $ab.c \in C = \{1, 2, \dots, nc\}$
- ✓ On définit  $AB_c = \{AB_1, AB_2, \dots, AB_N\}$  et  $ab \in AB_c$  si  $ab.c \equiv c$ .
- ✓ Soit ab.stim représente le niveau de stimulation d'un ARB ab.
- ✓ Soit ab.ressources représente le nombre de ressources tenues par un ARB ab.
- ✓ TotalNumRessources représente le nombre totale des ressources allouées par le système.

## Chapitre II: Les systèmes immunitaires

---

### - Initialisation :

La première étape de l'algorithme, l'initialisation, peut surtout être considérée comme une base de données de pré-traitement combiné avec une étape de découverte de paramètres. Lors de l'initialisation, d'abord, tous les éléments dans l'ensemble de données sont normalisés telle que la distance euclidienne entre les vecteurs de caractéristiques des deux éléments est dans l'intervalle [0,1].

Ceci peut être réalisé à travers une variété de méthodes et pourrait également être réalisé comme une véritable étape de pré-traitement avant que l'algorithme commence. Il est important à noter que, la distance euclidienne est la principale mesure d'affinité et de stimulation, d'autres fonctions pourraient être utilisées ainsi. Ce qui est important à propos de cette normalisation, c'est que l'éventail des réactions possibles de l'interaction de cellule à cellule reste dans l'intervalle de [0,1]. Après la normalisation, le seuil d'affinité est calculé.

Le seuil d'affinité est la valeur d'affinité moyenne sur tous les vecteurs de caractéristiques des éléments de données d'apprentissage. Le seuil d'affinité est calculé comme décrit dans l'équationci-dessous:

$$\text{Seuil d'affinité} = \frac{\sum_{i=1}^n \sum_{j=i+1}^n \text{affinity}(agi, agj)}{\frac{n(n-1)}{2}}$$

Où n est le nombre d'éléments de données d'apprentissage(antigènes) en question,  $agi$  et  $agj$  sont les  $i^{\text{ème}}$  et  $j^{\text{ème}}$  antigène d'apprentissage dans le vecteur d'apprentissage d'antigène, et l'affinité(x, y) retourne la distance euclidienne entre vecteurs de caractéristiques des deux antigènes.

La dernière étape d'initialisation est l'introduction des cellules de mémoire. Ceci est fait en choisissant aléatoirement 0 ou plusieurs antigènes de l'ensemble des vecteurs d'apprentissage à ajouter à l'ensemble des cellules mémoire.

### - L'identification de cellule mémoire et la génération d'ARB :

Une fois l'initialisation terminée, le processus d'apprentissage est comme un algorithme one-shot incrémental. Autrement dit, chaque élément de données d'apprentissage est présenté à l'algorithme d'apprentissage AIRS exactement une fois.

## Chapitre II: Les systèmes immunitaires

La première étape de cette partie de l'algorithme est l'identification de cellules mémoire et la génération d'ARB. Étant donné un antigène spécifique d'apprentissage,  $ag$ , trouvez la cellule de mémoire,  $mc_{match}$ , qui a la propriété suivante:

$$mcmatch = \operatorname{argmax}_{mc} MCag.c \text{ stimulation}(ag, mc)$$

d'où  $\text{stimulation}(x, y) = 1 - \text{affinité}(x, y)$

Si  $MCag.c \equiv \Phi$ , alors  $mc_{match} \leftarrow ag$  et  $MCag.c \leftarrow MCag.c \cup ag$ .

En d'autres termes, si la classe d'ensemble des cellules mémoire de la même classe que l'antigène est vide, alors ajoutez cet antigène à cet ensemble et le désigner comme  $mc_{match}$ .

Une fois  $mc_{match}$  identifiée, cette cellule mémoire est utilisée pour générer de nouveaux ARB pour être placés dans la population des (éventuellement) ARB préexistants. Cela se fait par la méthode de la figure 6, où la fonction  $\text{makeARB}(x)$  retourne un ARB avec  $x$  comme anticorps de cet ARB et  $\text{muter}(x, b)$  est défini dans la figure 7.

```
MU ← ∅
MU ← MU ∪ makeARB(mcmatch)
Stim ← stimulation(ag, mcmatch)
NumClones ← hyper_clone_rate * clonal_rate * stim
Tantque (|MU| < NumClones) Faire
  Mut ← faux
  MCclone ← mcmatch
  MCclone ← muter(mcclone, mut)
  Si (mut ≡ vrai)
    MU ← MU ∪ makeARB(mcclone)
Fintanque
AB ← AB ∪ MU
```

Figure 6 Hyper-Mutation pour la génération d'ARB

## Chapitre II: Les systèmes immunitaires

---

Dans la figure 7, la fonction `drandom` retourne une valeur aléatoire. Le rang de mutation pour un antigène donné = 1 - valeur de stimulation normalisé de la cellule.

```
Muter(x,b)
{
  Range ← 1-x.stim
  Pour chaque(x.fi dans x.f)
  Faire
    Change ← drandom()
    Change_to ← drandom()
    Bottom ←  $\frac{x.f_i}{\text{valeur de normalisation}} - \frac{\text{range}}{2}$ 
    Si (bottom < 0)
      Bottom ← 0
    Change_to ← (change_to * range) + bottom
    Si (change_to > 1)
      Change_to ← 1
    Si (change < taux_mutation )
      x.fi ← change_to * valeur_de_normalisation
      b ← vraie
  fin
  retourner x
}
```

Figure 7 Fonction de mutation

## Chapitre II: Les systèmes immunitaires

---

### - **Compétition pour les ressources et le développement d'une cellule mémoire candidate:**

À ce stade, un ensemble de ARB (AB) existe qui comprend  $m_{cmatch}$ ,  $m_{cmatch}$  mutée, et (éventuellement) ARB restants à partir des réponses à antigènes rencontrés auparavant. Rappelons que l'algorithme de AIRS est un algorithme de one-shot (monocoup), donc même si la discussion a été divisée en étapes distinctes, un seul antigène passe par ce processus à un moment (l'exception évidente étant la phase d'initialisation qui se déroule sur l'ensemble des données avant le début d'apprentissage).

Le but de la partie suivante de l'algorithme est de développer une cellule mémoire candidate qui est la plus réussie à classer correctement avec un antigène donné,  $ag$ . Cela se fait principalement par le biais de trois mécanismes. **Le premier mécanisme** est à travers la compétition pour les ressources système. Après les premières méthodes décrites par [17] et plus complète réalisée par [15], les ressources sont allouées à une donnée ARB basé sur sa valeur de stimulation normalisé, qui est utilisée comme une indication de son affinité comme une reconnaissance de l'Antigène  $ag$ . **Le second mécanisme** est l'utilisation de la mutation à la diversification et l'exploration de l'espace forme (shape-space). **Le troisième mécanisme** est l'utilisation d'un seuil de stimulation moyen comme critère d'arrêt d'apprentissage sur l'antigène  $ag$ .

Similaires à des principes dans les algorithmes génétiques, l'algorithme AIRS emploie un concept d'affinité pour la survie des individus au sein de la population ARB. La survie d'un ARB donnée est déterminée en deux étapes, de manière étroitement liés. D'abord, chaque ARB à l'AB population est présenté avec l'antigène  $AG$  pour déterminer le niveau de stimulation de l'ARB. Cette stimulation sera ensuite normalisée dans la population ARB basée sur le niveau de stimulation à la fois brut et la classe de l'ARB donné ( $ab: c$ ).

Sur la base de cette valeur de stimulation normalisée, chaque  $ab$  AB se voit attribuer un nombre fini de ressources. Si cette allocation des ressources aboutirait à plus de ressources allouées dans la population que ne l'autorise, alors les ressources sont retirées des ARB les plus faibles (les moins stimulés) jusqu'à ce que le nombre total des ressources dans le système retourne au nombre de ressources autorisées. Ceux qui ont ARB à zéro ressources sont retirés de la population ARB. Ce processus est formalisé dans la Figure 8.

## Chapitre II: Les systèmes immunitaires

---

Deux aspects essentiels de cette allocation des ressources sont notés ici :

Premièrement, la valeur de stimulation d'un ARB n'est pas seulement déterminée par la fonction de stimulation dans l'équation 3, mais est également basée sur la classe de de l'ARB. La méthode de calcul de stimulation indiquée dans la figure 8 permet de renforcer pour les ARB de la même classe que ag qui sont fortement stimulées par AG.

Deuxièmement, la répartition des ressources est également basée sur la classe de l'ARB. Cela se fait en fournissant des renforcements supplémentaire pour les ARB de de la même classe que ag sans perdre les qualités potentiellement positifs de ces ARB restants pour la réaction aux antigènes avenir. A cette étape de l'algorithme, la population ARB dans AB se compose des seuls ARB qui étaient les plus stimulés par l'antigène donné, ag, ou plus précisément, AB se compose désormais des ARB qui ont réussi à obtenir des ressources.

L'algorithme se poursuit d'abord en déterminant si l'ARB dans AB a été assez stimulé par ag pour arrêter d'apprentissage sur cet élément. Ceci est fait en définissant un vecteur  $\vec{s}$  qui est de longueur  $n_c$  qui contient la valeur de stimulation moyenne pour chaque sous-ensemble de la classe AB. Cette valeur est calculée par :

$$S_i \leftarrow \frac{\sum_{j=1}^{|AB_i|} ab_j.stim}{|AB_i|}, ab_j \in AB_i$$

Le critère d'arrêt est atteint si  $s_i \geq$  seuil stimulation pour tout les éléments dans  $\vec{s} = \{s_1, s_2, \dots, s_{n_c}\}$

```

minStim ← MAX
maxStim ← Min
pour chaque (ab ∈ AB)
faire
    stim ← stimulation(ag, ab)
    si ( stim < minStim )
        minStim ← stim
    si ( stim > maxStim )
        maxStim ← stim
    ab.stim ← stim
fin pour
    
```

## Chapitre II: Les systèmes immunitaires

---

```
pour chaque (ab ∈ AB)
faire
    ab.stim ←  $\frac{ab.stim - minStim}{maxStim - minStim}$ 
    ab.ressources ← ab.stim * taux_clonage
finpour
resAlloc ←  $\sum_{j=1}^{|AB_{ag.c}|} ab_j.ressources, ab_j \in AB_{ag.c}$ 
NumResAllouer ← TotalNumRessources
Tant que (resAlloc > NumResAllouer)
Faire
    NumResSupprimer ← resAlloc - NumResAllouer
    ab_remove ← argmin_{ab ∈ AB_{ag.c}} (ab.stim)
    Si (ab_remove.ressources ≤ NumResSupprimer)
         $AB_{ag.c} \leftarrow AB_{ag.c} - ab_{remove}$ 
        restAlloc ← resAlloc - ab_remove.ressources
    sinon
        ab_remove.ressources ← ab_remove.ressources - NumResSupprimer
        resAlloc ← resAlloc - NumResSupprimer
Fin pour
```

Figure 8 Stimulation, Allocation de ressources, Suppression ARB

## Chapitre II: Les systèmes immunitaires

---

Indépendamment du fait que le critère d'arrêt est atteint ou pas, l'algorithme procède en permettant à chaque ARB dans AB la possibilité de produire des descendants mutés. Bien que cette adjonction de descendants mutés, est similaire à la méthode décrite dans la figure 6, il existe quelques différences. Cette modification de génération de mutation est présentée dans la figure 9.

```
MU ← ∅
Pour chaque (ab ∈ AB)
Faire
    rd ← drandom()
    si (ab.stim > rd)
        NumClones ← ab.stim * taux_clonage
        i ← 1
        tant que (i ≤ NumClones)
            faire
                mut ← faux
                ab_clone ← ab
                ab_clone ← mutate(ab_clone, mut)
                si (mut ≡ vrai)
                    MU ← MU ∪ ab_clone
                i ← i + 1
            faire tanque
        fin pour
AB ← AB ∪ MU
```

**Figure 9 Mutation des ARB survivants**

## Chapitre II: Les systèmes immunitaires

Après l'autorisation à chaque survivant ARB la possibilité de produire une descendance muté, le critère d'arrêt est examiné. Si le critère d'arrêt est atteint, alors l'apprentissage sur cet antigène s'arrête. Si le critère d'arrêt n'a pas été atteint, alors tout ce processus, en commençant par la méthode décrite dans la figure 8, est répété jusqu'à ce que le critère d'arrêt soit atteint.

La seule exception à cette répétition, c'est que sur chaque passage dans cette partie de l'algorithme, à l'exception du premier passage déjà discuté, si le critère d'arrêt est atteint après la stimulation et la phase d'allocation des ressources, alors la production d'une descendance mutée n'est pas effectuée. Une fois le critère d'arrêt atteint, alors la cellule mémoire candidate est choisie.

La cellule mémoire candidate,  $mc_{candidate}$ , est le vecteur de caractéristiques et de la classe de l'ARB qui existaient dans le système avant la plus récente série de mutations qui a été la plus stimulée avec l'ARB de la même classe que l'ag d'antigène d'apprentissage.

### - Introduction d'une cellule mémoire :

L'étape finale d'apprentissage est l'introduction de la cellule mémoire candidate  $mc_{candidate}$ , dans l'ensemble de cellules mémoire MC existants. C'est au cours de cette étape que le seuil d'affinité calculée pendant l'initialisation devient critique car il détermine si la  $mc_{candidate}$  remplace  $mc_{match}$  qui a été précédemment identifié.

La cellule mémoire candidate est ajoutée à l'ensemble des cellules mémoire que si elle est plus stimulée par l'antigène d'apprentissage, ag, que  $mc_{match}$ . Si ce test est réussi, alors, si l'affinité entre  $mc_{candidate}$  et  $mc_{match}$  est inférieur au produit du seuil d'affinité et le scalaire de seuil d'affinité, alors  $mc_{candidate}$  remplace  $mc_{match}$  dans l'ensemble des cellules mémoires. Cette méthode d'introduction de cellules de mémoire est présentée dans la figure 10.

```
CandStim ← stimulation( ag, mccandidate )
MatchStim ← stimulation(ag, mcmatch )
CellAff ← affinité(mccandidate, mcmatch )
Si ( CandStim > MatchStim )
    Si ( CellAff < AT * ATS )
        MC ← MC - mcmatch
        MC ← MC + mccandidate
```

Figure 10 Introduction d'une cellule mémoire

## Chapitre II: Les systèmes immunitaires

Une fois la cellule mémoire candidate évaluée pour l'ajout dans l'ensemble des cellules de mémoire établies, l'apprentissage sur cet antigène est complet. L'antigène suivant dans l'ensemble d'apprentissage poursuit le processus d'apprentissage avec l'identification des cellules mémoire et la génération d'ARB. Ce processus s'achève quand tous les antigènes seront présentés au système.

### 5.3 Classification

Après la fin apprentissage les cellules mémoire générées sont prêtes pour une utilisation pour la classification. La classification est réalisée avec deux approches au choix : k-plus proches voisin et Naïve Bayes.

#### 5.3.1 K plus proche voisin :

- *Définition :*

L'algorithme le plus proche voisin k-(k-NN) est une méthode de classification des éléments à partir des exemples d'apprentissage. k-NN est un type d'apprentissage paresseux où la fonction est seulement approchée localement et tout le calcul est reporté à la classification. L'algorithme du k plus proche voisin est parmi les plus simples de tous les algorithmes d'apprentissage machine: un objet est classé par un vote à la majorité de ses voisins, dans le but d'être affecté à la classe la plus commune parmi ses k plus proches voisins (k est un entier positif, généralement de petite taille). Si  $k = 1$ , alors l'objet est simplement attribué à la classe de son plus proche voisin. Les voisins sont pris à partir d'un ensemble d'éléments pour lesquels la classification correcte est connue. Cela est considéré comme l'ensemble d'apprentissage de l'algorithme. [18]

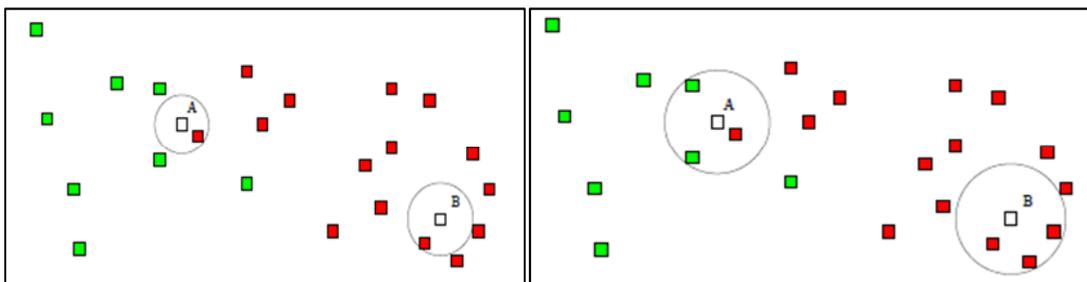


Figure 11 Classification de deux classe avec Knn, a gauche k=1 , a droite k=3

## Chapitre II: Les systèmes immunitaires

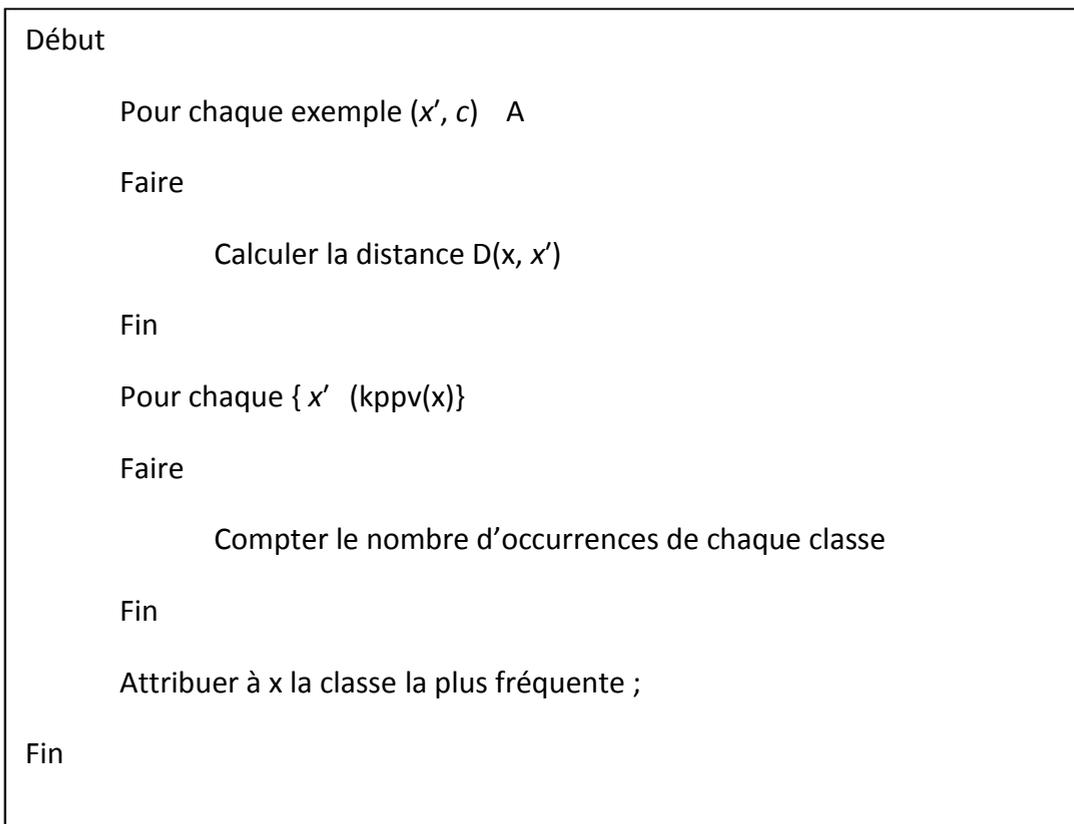
---

### - Algorithme :

Les exemples d'apprentissage sont des vecteurs en espace de caractéristiques multidimensionnelles, chacun avec une classe. La phase d'apprentissage de l'algorithme se limite à stocker les vecteurs de caractéristiques et les classes des échantillons d'apprentissage.

Dans la phase de classification,  $k$  est une constante définie par l'utilisateur, et un vecteur non marqué (une requête ou d'un point de test) est classé en attribuant la classe qui est plus fréquente parmi les  $k$  échantillons d'apprentissage plus proche de ce point de requête.

- ✓ Soit  $A = \{(x', c) | x' \in \mathbb{R}^{Rd}, c \in C\}$  l'ensemble d'apprentissage
- ✓ Soit  $x$  l'exemple dont on souhaite déterminer la classe
- ✓  $D(x, x')$  la distance entre le vecteur de test et d'apprentissage, généralement on utilise la distance Euclidienne



**Figure 12** Algorithme Kppv

## Chapitre II: Les systèmes immunitaires

---

### - Choix des paramètres :

Le meilleur choix de  $k$  dépend des données, généralement, les grandes valeurs de  $k$  réduisent l'effet du bruit sur les classifications, mais les frontières entre les classes seront moins distinctes. Une bonne valeur de  $k$  peut être choisie par diverses techniques heuristiques, par exemple, la validation croisée.

Le cas particulier où la classe est prévue pour être la classe de l'échantillon d'apprentissage le plus proche (c'est à dire lorsque  $k=1$ ) est appelé l'algorithme du plus proche voisin.

La précision de l'algorithme  $k$ -NN peut être gravement altérée par la présence d'éléments bruyants ou non pertinents, ou si les échelles de caractéristiques ne sont pas compatibles avec leur importance. Des efforts de recherche considérables ont été mis dans la sélection ou l'extension des caractéristiques pour améliorer la classification. Dans les problèmes de classification binaire (deux classes comme pour le cas du PIMA Diabète), il est utile à choisir  $k$  à un nombre impair car cela évite l'égalité de vote. En cas d'égalité de classification on a le choix d'appliquer l'une des décisions suivantes :

- Augmenter la valeur de  $k$  de 1 pour trancher. L'ambiguïté peut persister
- Tirer au hasard la classe parmi les classes ambiguës.
- Pondération des exemples par leur distance au point  $x$

L'inconvénient pour le « vote à la majorité » de classification est que les classes avec les exemples les plus fréquents ont tendance à dominer les prédictions du nouveau vecteur, car ils ont tendance à venir dans les  $k$  plus proches voisins quand les voisins sont calculés en raison de leur grand nombre. Une façon de surmonter ce problème consiste à pondérer les classements en tenant compte de la distance entre les points d'essai à chacun de ses  $k$  plus proches voisins. [18]

### - Amélioration de l'algorithme :

On peut réduire l'ensemble d'apprentissage (méthodes de) sans changer (à la limite) le résultat des futures décisions par  $K$ -ppv .

## Chapitre II: Les systèmes immunitaires

---

### b/ Algorithme de naïf bayes :

#### - définition :

La classification naïve bayésienne est un type de classification Bayésienne probabiliste simple basée sur le théorème de Bayes avec une forte indépendance (dite naïve) des hypothèses. Elle met en œuvre un classifieur bayésien naïf, ou classifieur naïf de Bayes, appartenant à la famille des classifieurs linéaires.

Un terme plus approprié pour le modèle probabiliste sous-jacent pourrait être « modèle à caractéristiques statistiquement indépendantes ».

En termes simples, un classifieur bayésien naïf suppose que l'existence d'une caractéristique pour une classe, est indépendante de l'existence d'autres caractéristiques. Un fruit peut être considéré comme une pomme s'il est rouge, arrondi, et fait une dizaine de centimètres. Même si ces caractéristiques sont liées dans la réalité, un classifieur bayésien naïf déterminera que le fruit est une pomme en considérant indépendamment ces caractéristiques de couleur, de forme et de taille.

Selon la nature de chaque modèle probabiliste, les classifieurs bayésiens naïfs peuvent être entraînés efficacement dans un contexte d'apprentissage supervisé. Dans beaucoup d'applications pratiques, l'estimation des paramètres pour les modèles bayésiens naïfs repose sur le maximum de vraisemblance. Autrement dit, il est possible de travailler avec le modèle bayésien naïf sans se préoccuper de probabilité bayésienne ou utiliser les méthodes bayésiennes.

Malgré leur modèle de conception « naïf » et ses hypothèses de base extrêmement simplistes, les classifieurs bayésiens naïfs ont fait preuve d'une efficacité plus que suffisante dans beaucoup de situations réelles complexes. En 2004, un article a montré qu'il existe des raisons théoriques derrière cette efficacité inattendue. Toutefois, une autre étude de 2006 montre que des approches plus récentes (arbres renforcés, forêts aléatoires) permettent d'obtenir de meilleurs résultats.

L'avantage du classifieur bayésien naïf est qu'il requiert relativement peu de données d'entraînement pour estimer les paramètres nécessaires à la classification, à savoir moyennes et variances des différentes variables. En effet, l'hypothèse

## Chapitre II: Les systèmes immunitaires

d'indépendance des variables permet de se contenter de la variance de chacune d'entre elle pour chaque classe, sans avoir à calculer de matrice de covariance. [19]

### b/ Théorie de la décision bayésienne :

Soit les définitions suivantes :

- $W=\{w_1,w_2,\dots,w_s\}$  l'ensemble des  $s$  classes d'appartenance possibles.
- $x=\{x_1,x_2,\dots,x_d\}$  vecteur de paramètres décrivant une forme.
- $P(x/w_j)$  fonction de densité de probabilité de la classe  $w_j$ .
- $P(w_j)$  probabilité à priori de la classe  $w_j$ .

Pour disposer d'une règle de décision chaque classe recherchée est caractérisée par une fonction discriminante :  $g_i(x)$  pour  $i=1 \dots s$ , et

Une forme  $x$  est classée dans la classe  $w_i$  si  $g_i(x) > g_j(x)$  pour tout  $j \neq i$ .

Le classifieur obtenu calcule donc pour un vecteur  $x$  donné les fonctions discriminantes puis recherche le maximum qui permet de choisir la classe. Ce classifieur peut être schématisé de la façon suivante :

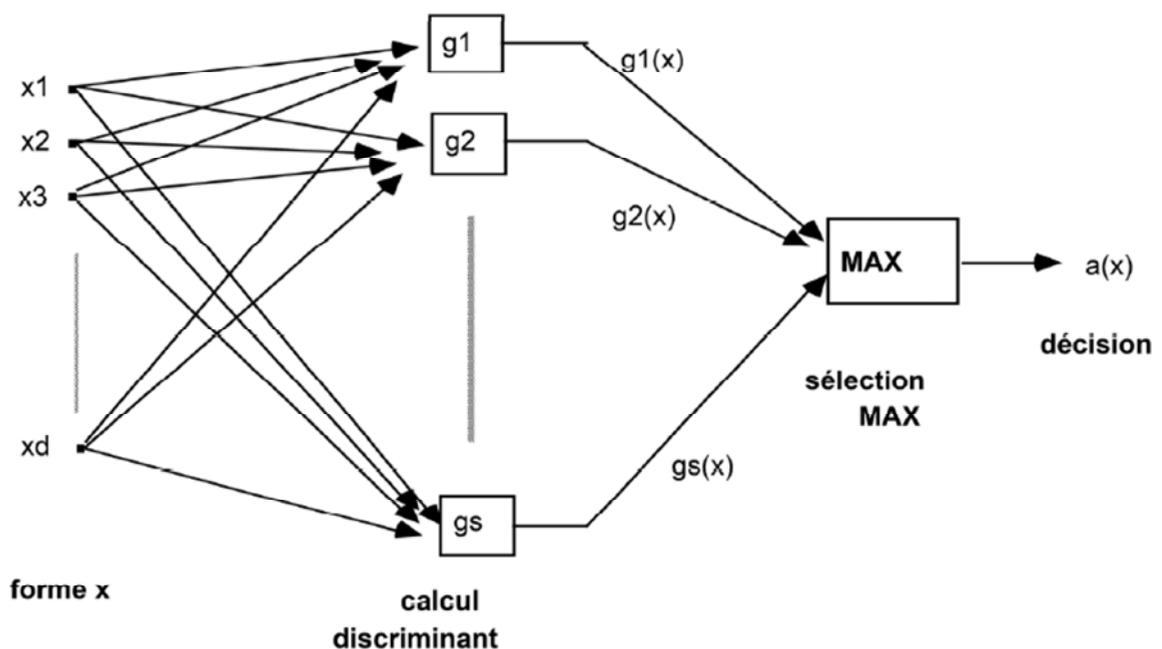


Figure 13 Fonction discriminantes et décision

## Chapitre II: Les systèmes immunitaires

La fonction discriminante est définie comme suit :

$$G_i(x) = P(x|w_i)P(w_i)$$

L'utilisation des  $s$  fonctions discriminantes pour réaliser la classification de forme amène à découper l'espace des observations (espace des paramètres  $x$ ) en  $s$  régions  $R_1 R_2 \dots R_s$ .

Chaque région est définie comme suit :

Si  $g_i(x) > g_j(x)$  pour tout  $j \neq i$  alors  $x$  appartient à  $R_i$ , donc assigner  $x$  à  $w_i$ .

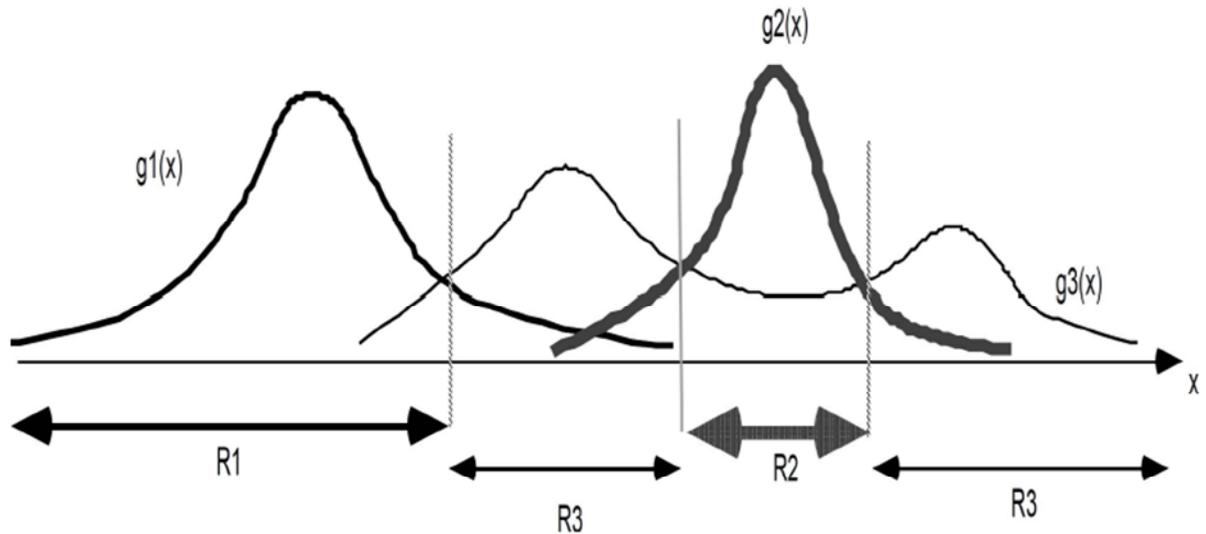


Figure 14. Frontière de décision dans l'espace des observations

### - Analyse :

Fait étonnant, malgré les hypothèses d'indépendance relativement simplistes, le classifieur bayésien naïf a plusieurs propriétés qui le rendent très pratique dans les cas réels. En particulier, la dissociation des lois de probabilités conditionnelles de classe entre les différentes caractéristiques aboutit au fait que chaque loi de probabilité peut être estimée indépendamment en tant que loi de probabilité à une dimension.

Cela permet d'éviter nombre de problèmes venant du fléau de la dimension, par exemple le besoin de disposer d'ensembles de données d'entraînement dont la quantité augmente exponentiellement avec le nombre de caractéristiques. Comme tous les classifieurs probabilistes utilisant la règle de décision du maximum a posteriori, il classifie correctement du moment que la classe adéquate est plus probable que toutes les autres.

## Chapitre II: Les systèmes immunitaires

---

Par conséquent les probabilités de classe n'ont pas à être estimées de façon très précise. Le classifieur dans l'ensemble est suffisamment robuste pour ne pas tenir compte de sérieux défauts dans son modèle de base de probabilités naïves.

### **Conclusion :**

Ce chapitre a mis l'accent sur un système d'apprentissage supervisé basé sur des principes immunologiques. Le système de reconnaissance immunitaire artificiel (AIRS) présenté comme un succès initial d'algorithme de classification. Cependant, comme avec n'importe quel système initial, il y a des révisions et des améliorations qui peuvent être apportées au AIRS qui réduisent la complexité du système

La seconde partie de ce projet sera consacrée à la conception de notre application basée sur l'algorithme AIRS inspiré des systèmes immunitaires artificiels. Nous finirons par une série de tests qui nous permettront d'évaluer notre algorithme et surtout de mettre en relief l'apport des systèmes immunitaires pour la classification médicale.

Chapitre III

**APPLICATION ET  
RESULTATS**

### **Chapitre3** **Résultats et discussions**

#### **1. Introduction :**

L'objectif de notre travail consiste dans l'analyse du comportement et de la capacité de l'algorithme immunitaire AIRS à reconnaître une personne diabétique à partir des caractéristiques, et la présentation d'une analyse sur l'effet des différents paramètres d'apprentissage sur la performance de cet algorithme.

La base de données utilisée dans ce chapitre nous a permis d'étudier les performances de divers systèmes de classification différents.

Ce chapitre est organisé comme suit :

- La section 1 décrit la base de données utilisée pour la classification.
- La section 2 présente une analyse des cinq paramètres d'apprentissage visés par l'analyse et les résultats obtenus lors de la variation des paramètres les plus importants.

Les paramètres examinés sont les suivants :

- Le nombre de ressources autorisées dans le système ;
- Le seuil de stimulation utilisé comme un critère d'arrêt pour l'apprentissage ;
- Le taux de mutation ;
- Le scalaire du seuil d'affinité employé pour le remplacement des cellules mémoires;
- La valeur de k utilisée du classifieur Knn pour la classification.

#### **2. Base de données Pima :**

La classification a été appliquée sur la base de données médicales PIMA (indian data bases) détenue par l'Institut National du Diabète et des Maladies Digestives et Rénales. La base contient les caractéristiques physiologiques de 392 patients (âgés d'au moins 21 ans) de la tribu indienne Pima vivant près de Phoenix (USA).

Tous les échantillons ont huit descripteurs et deux classes.

## Chapitre III : Résultats et discussions

Ces descripteurs sont :

1. Le Prénant: Nombre de grossesses.
2. Le Glucose : Glycémie après un test oral de tolérance au glucose TOTG (cg/l).
3. La Tension artérielle diastolique (mm Hg).
4. L'épaisseur de pli cutané au niveau du triceps (mm).
5. La concentration d'insuline (mu U/ml).
6. L'Indice de masse corporelle (poids en Kg/taille en m<sup>2</sup>).
7. L'Indice d'antécédents familiaux pour le diabète.
8. L'âge (année).

Les Classes sont : absence (0) ou présence (1) du diabète.

### 3. Analyse statistique des données :

Après le calcul de la corrélation mutuelle entre les différents descripteurs avec la

formule suivante :  $C = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$ , Nous avons obtenu les résultats suivants:

Descripteur	corrélation
L'âge et le nombre de grossesse	0.68
L'indice de masse corporelle et l'épaisseur du triceps	0.66
La glycémie et l'insuline	0.58
La glycémie et l'âge	0.34
La pression diastolique et l'indice de masse corporelle	0.3

Tableau 3 Corrélation mutuelle entre quelques descripteurs

#### 3.1 La corrélation entre l'âge et le nombre de grossesse :

Le test montre que le pourcentage des femmes non diabétiques ayant un âge inférieur à 30 ans est de 75%, et les femmes ayant un nombre de grossesses inférieur ou égal à 3 est de 79% femmes.

Le pourcentage des femmes non diabétiques ayant un âge supérieur à 30ans est de 85% et le pourcentage des femmes ayant un nombre de grossesses supérieur ou égal à 1 est de 87%.

## Chapitre III : Résultats et discussions

---

Ces résultats démontrent que ses deux descripteurs sont corrélés entre eux et on peut conclure de ces pourcentages que la majorité des femmes jeunes ont peu de grossesses.

### **3.2 La corrélation entre l'indice de masse corporelle et l'épaisseur du triceps :**

Le pourcentage des femmes non diabétiques ayant un IMC inférieur à  $36\text{kg/m}^2$  est de 74% et le pourcentage des femmes ayant une épaisseur de la peau inférieure ou égale à 34mm est de 74%.

Le pourcentage des femmes non diabétiques ayant un IMC supérieur à  $30\text{kg/m}^2$  est de 84% et le pourcentage des femmes ayant une épaisseur de la peau supérieure à 25 mm est de 77%. Ce qui démontre que ces deux descripteurs sont proportionnels entre eux.

### **3.3 La corrélation entre la concentration d'insuline et la glycémie après un test oral de tolérance au glucose :**

Le pourcentage de femmes non diabétiques ayant un taux d'insuline inférieur à  $180\mu\text{U/ml}$  est de 78% et le pourcentage des non diabétiques ayant une concentration du glucose inférieure à  $130\text{mg/l}$  est de 80%.

D'autre part, le pourcentage des femmes diabétiques ayant une concentration du glucose supérieure à  $130\text{mg/l}$  est de 64% alors que le pourcentage de celles ayant un taux d'insuline supérieur à  $130\mu\text{U/ml}$  est de 73%.

On peut conclure de cela que les 2 descripteurs sont corrélés entre eux.

## **4. Les paramètres d'apprentissage :**

Les paramètres par défaut de ces expériences se présentent comme suit:

- le nombre de ressources autorisées dans le système est 200,
- le seuil de stimulation a été fixé à 0,9,
- le taux de mutation a été fixé à 0,1,

## Chapitre III : Résultats et discussions

---

- le scalaire seuil d'affinité (ATS) a été fixé à 0,2
- les valeurs de k ont été utilisées pour la classification 1 et 3.

### 4.1 Les ressources :

Le nombre de ressources à un effet direct sur le nombre des ARB présents dans le système. Des ARB à zéro ressource sont retirés du système après chaque itération d'apprentissage. L'accroissement du nombre de ressources devrait augmenter la diversité des cellules mémoire qui sont générés à partir des ARB pool, avec comme résultat : plus d'ARB seront autorisés à rester en vie.

Il a été émis l'hypothèse que l'augmentation du nombre de ressources conduirait à une augmentation de la précision de la classification. Une explication possible est que l'augmentation du nombre de ressources résulte potentiellement d'une augmentation du nombre d'ARB. Tous ces ARB doivent être très stimulés. Ainsi, cette augmentation de population est en réalité une augmentation de population très stimulée à l'ensemble d'apprentissage. Par conséquent, il y aurait une amélioration de classification.

### 4.2 seuil de stimulation :

Nous étudions, ensuite, l'effet du seuil de stimulation sur la précision de classification du système. Le seuil de stimulation est utilisé comme un critère d'arrêt pour l'apprentissage de la population ARB sur un antigène spécifique.

On définit un vecteur  $s$  qui est le nombre de classes, de longueur  $n$ , pour contenir la valeur moyenne de stimulation pour chaque sous-ensemble de classe de la

population ARB. A savoir:  $S_i \leftarrow \frac{\sum_{j=1}^{|ABi|} abj.stim}{|ABi|}, abj \quad ABi$

Le critère d'arrêt est atteint si et seulement si  $S_i$  est supérieur au seuil de stimulation. Ainsi, le seuil de stimulation détermine comment la population ARB stimulée doit être, et cela, avant la fin d'apprentissage d'un antigène particulier. Dans ce cas, il est prévu que le seuil de stimulation augmente la précision de classification.

## Chapitre III : Résultats et discussions

---

### 4.3 le taux de mutation :

Le taux de mutation permet d'affecter la diversité de la population ARB évolué. Toutefois, étant donné que les mutations sont aléatoires, un taux de mutation élevé peut avoir un résultat dans la production de nombreux enfants qui ne seront pas fortement stimulés par l'antigène d'apprentissage.

Il est possible alors que l'augmentation du taux de mutation peut avoir un effet négatif sur les performances de l'algorithme à la fois en termes de précision de classification et en temps d'exécution.

### 4.4 Affinity threshold scalar (ATS) :

Le scalaire du seuil d'affinité (ATS) a un effet direct sur le remplacement des cellules mémoire. Rappelons que le seuil d'affinité est défini comme l'affinité moyenne (distance euclidienne) de l'affinité entre les diverses éléments de l'ensemble d'apprentissage.

L'ATS est un scalaire défini par l'utilisateur qui est multiplié par le seuil d'affinité pour déterminer le remplacement de cellules de mémoire. Ceci dit, une cellule mémoire nouvellement générée remplace une cellule mémoire établie si et seulement si elle est plus stimulée par l'antigène d'apprentissage, de même l'affinité entre les deux cellules mémoire (candidate et match) doit être inférieure au produit du seuil d'affinité et l'ATS. Ceci implique que plus la valeur ATS augmente, plus le nombre de cellules mémoire dans le système diminuera. L'augmentation de l'ATS (et donc la diminution du nombre de cellules mémoire) devrait avoir un impact considérable sur la précision de classification. Cela devrait être particulièrement vrai lorsque la classification du système dans le cas du classifieur k plus proche voisin est basé sur une valeur k supérieure à 1.

### 4.5 La valeur K

Le paramètre d'utilisateur final dans cette série d'expériences est la valeur k utilisée pour la classification dans le cas du classificateur k plus proche voisin. La classification d'un élément donné est basée sur un vote majoritaire des k cellules mémoire les plus proches de l'élément donné. En général, il n'existe aucun moyen évident de prévoir la "meilleure" valeur de k pour un ensemble de données. Elle est extrêmement dépendante de la nature des données elles-mêmes fixés.

### 5. Critère d'évaluation :

#### a. Taux de classification :

$$\text{taux}(T) = \frac{\sum_{i=1}^{|T|} \text{assess}(t_i)}{|T|}$$
$$\text{assess}(t) = \begin{cases} 1 & \text{if } \text{classif}(t)=t.c \\ 0 & \text{otherwise} \end{cases}$$

$T$  est l'ensemble de tests ;  $t \in T$  ;  $t.c$  est la classe de l'élément  $t$  ; et  $\text{classif}(t)$  retourne la classification de  $t$  par l'algorithme.

#### b. La sensibilité :

Elle représente le taux de classification des cas diabétiques (détectés correctement par le classifieur) par rapport au nombre total des cas diabétiques réels.

$$\text{Sensibilité (\%)} = \frac{vp}{vp+fn} \times 100$$

#### c. La spécificité :

Elle représente le taux de classification des cas non diabétiques (détectés correctement par le classifieur) par rapport au nombre total de cas non diabétiques réels.

$$\text{Spécificité (\%)} = \frac{vn}{vn+fn} \times 100$$

VP, VN, FP et FN désignent respectivement:

- Vrai positif : un exemple diabétique prédit diabétique.
- Vrai négatif : un exemple non diabétique prédit non diabétique.
- Faux positif : un exemple non diabétique prédit diabétique.
- Faux négatif : un exemple diabétique prédit non diabétique.

#### d. Matrices de confusion :

En utilisant cette matrice de confusion nous avons évalué les classificateurs : la matrice de confusion contient les informations sur les classifications réelles et nous avons évalué par les classificateurs :

## Chapitre III : Résultats et discussions

	Prédit	Négative	Positive
Actuelle			
Négative		VN	FP
Positive		FN	VP

Tableau 4 : La matrice de confusion

### 6. Langages et environnements de développement :

Pour le développement de notre application, nous avons utilisé le langage JAVA avec l'IDE NetBeans. Ceci nous a permis de développer un programme pouvant fonctionner sur différents systèmes d'exploitation.

#### 6.1 Découpage des classes :

Dans cette partie nous allons citer les classes principales constituant l'application :

- ✓ AirsAlg : c'est la classe qui implémente l'algorithme AIRS.
- ✓ Knn : Cette classe est utilisée pour la classification avec l'approche du K plus proche voisin.
- ✓ Bayes : Cette classe définit l'algorithme de Naïf Bayes.
- ✓ Matrice : Les trois classes précédentes font appel à cette classe pour faire les différentes opérations (avoir la classe ou les caractéristiques d'un cas donné, calcul de la valeur de similarité, et les différentes fonctions nécessaires pour les classifieurs)
- ✓ Chrono : cette classe évalue de temps d'exécution des différents classifieurs , ainsi que le temps d'apprentissage.
- ✓ Classifieur : c'est dans cette classe que s'effectue le calcul du taux de classification, la spécificité, la sensibilité ... à partir des 2 classes de classification.
- ✓ Evaluation : cette classe permet d'évaluer les deux classifieurs et de choisir le meilleur classifieur en fonction de son taux de classification et, en cas de similitude, du temps nécessaire pour la classification.
- ✓ AIRS : c'est la classe principale où s'effectue l'apprentissage et la classification.

Nous allons détailler quelques méthodes et classes principale dans le tableau ci-dessous :

## Chapitre III : Résultats et discussions

Classe	Méthode	Paramètres	Valeur retournée	Rôle :
AirsAlg	AirsAlg	(Matrice :ag,double Ats,double TotalRes ,double ClonalRate ,double mutationRate ,double stimulationThreshold,doub le hyperClonalRate)	void	Applique l'algorithme AIRS.
	getMC	()	Matrice	Retourne les cellules mémoire générées à la fin d'apprentissage.
	mutate	(Matrice mc_clone,boolean mut)		Fonction de mutation
Matrice	Classe	(int i)	double	Obtenir la classe d'un antigène spécifique i
	Add	(Matrice m ,int i)	Matrice	Ajouter une cellule mémoire ou un arb i
	Affinite	(Matrice a ,int i ,Matrice j,int j)	double	Calculer l'affinité entre deux antigènes i,j
	DistanceB	(Matrice cov,Matrice inv, Matrice m l)	double	Calcule de la distance bayésienne
	DistanceK	(Matrice appr)	double	Calcule de la distance euclidienne pour l'algorithme Knn
	Feature()	()	Matrice	Avoir les caractéristiques d'un antigène sans la classe
	getRes	(Int i)	double	Obtenir le nombre de ressources d'un ARB i
	getStim	(Int i)	double	Obtenir la valeur de stimulation d'un ARB i
	getnormalise()	()	Matrice	Normaliser une matrice (l'étape d'initialisation de l'algorithme)
	setRes /setStim	(Int i,double r)	Matrice	Affecter le nombre de ressources/la valeur de stimulation à un ARB i
	seuilAffinite	()	double	Calculer le seuil d'affinité
	supprimer	(int i)	Matrice	Suppression d'un ARB i ou une cellule mémoire i
AIRS	Classification N	()	void	Effectuer une classification manuelle(définir la taille de la base de teste)
	CrossValidation	(Matrice m,int kf)	void	Effectuer une classification par

## Chapitre III : Résultats et discussions

				CrossValidation avec k-fold=kf (nombre de partition)
	Evaluation	()	void	Evaluation du meilleur classifieur et comparer les résultat .
	MalClasser	()	void	Afficher les patient mal classer avec leur caractéristique et leurs classe négative .
	ResumerApp	()	void	Avoir le résumer d'apprentissage de l'algorithme AIRS (taille de cellule mémoire générer ,les differentes valeur d'évaluation , temps d'exécution ...)
	aperçu	()	void	Afficher la base de donner (et la base de teste pour dans le cas de la classification manuelle)

Tableau 5 Les classes principales du programme

### 7. Résultats et discussions :

#### 7.1 Première expérimentation :

Dans cette première expérimentation nous effectuons la classification sur notre base de diabète qui contient 392 cas. C'est-à-dire que nous effectuons l'apprentissage de cette base avec l'algorithme AIRS2. Les cellules mémoire générées seront ensuite utilisées comme base d'apprentissage pour la classification ; et la base de test sera notre base d'apprentissage initial (392 cas).

Le but de cette expérimentation est de savoir si les cellules mémoires générées restent toujours pertinentes, malgré la réduction de leurs nombres (diminution de la taille de la base). Les performances de classification obtenues ont été évaluées en utilisant les classificateurs Knn et naif Bayes (la classification s'effectue en même temps). La relation entre l'ATS, le seuil de stimulation,

la précision de classification et le temps d'apprentissage ,est indiqué dans le tableau ci-dessous :

## Chapitre III : Résultats et discussions

**Tableau 6 performance d'apprentissage avec AIRS2**

ATS	Seuil de stimulation	Taille de MC	Classifieur	Taux de classification (%)	Sens (%)	Spec (%)	Temps d'apprentissage (millisecond)
<b>0.1</b>	0.90	294	Knn(k=1)	85.96	75.20	90.33	13744
			Knn(k=3)	82.65	73.29	87.76	
			Bayes	78.82	71.89	82.23	
<b>0.1</b>	0.94	293	Knn(k=1)	86.98	77.23	91.34	18349
			Knn(k=3)	83.41	67.58	91.43	
			Bayes	79.33	68.46	84.35	
<b>0.1</b>	0.96	302	Knn(k=1)	88.01	83.46	90.34	27018
			Knn(k=3)	83.41	67.35	91.65	
			Bayes	78.31	69.42	82.34	
<b>0.1</b>	0.98	317	Knn(k=1)	93.62	92.30	94.27	45052
			Knn(k=3)	83.92	79.23	86.25	
			Bayes	78.82	70.76	82.82	
<b>0.1</b>	1	380	Knn(k=1)	98.97	97.93	99.50	137340
			Knn(k=3)	83.67	62.25	94.34	
			Bayes	77.55	69.34	81.35	
<b>0.2</b>	0.90	195	Knn(k=1)	83.92	72.30	89.69	11754
			Knn(k=3)	82.90	65.38	91.60	
			Bayes	78.31	63.84	85.49	
<b>0.2</b>	0.94	221	Knn(k=1)	83.92	77.69	87.02	19189
			Knn(k=3)	82.14	67.69	89.31	
			Bayes	78.06	69.23	82.44	
<b>0.2</b>	0.96	227	Knn(k=1)	89.03	81.53	92.74	25163
			Knn(k=3)	82.90	69.23	89.69	
			Bayes	77.80	64.61	84.35	
<b>0.2</b>	0.98	300	Knn(k=1)	93.87	90.00	95.80	43921
			Knn(k=3)	82.90	69.23	89.69	
			Bayes	77.55	71.53	80.53	
<b>0.2</b>	1	359	Knn(k=1)	99.23	100	98.85	142068
			Knn(k=3)	83.92	65.38	93.12	
			Bayes	77.80	68.46	82.44	

## Chapitre III : Résultats et discussions

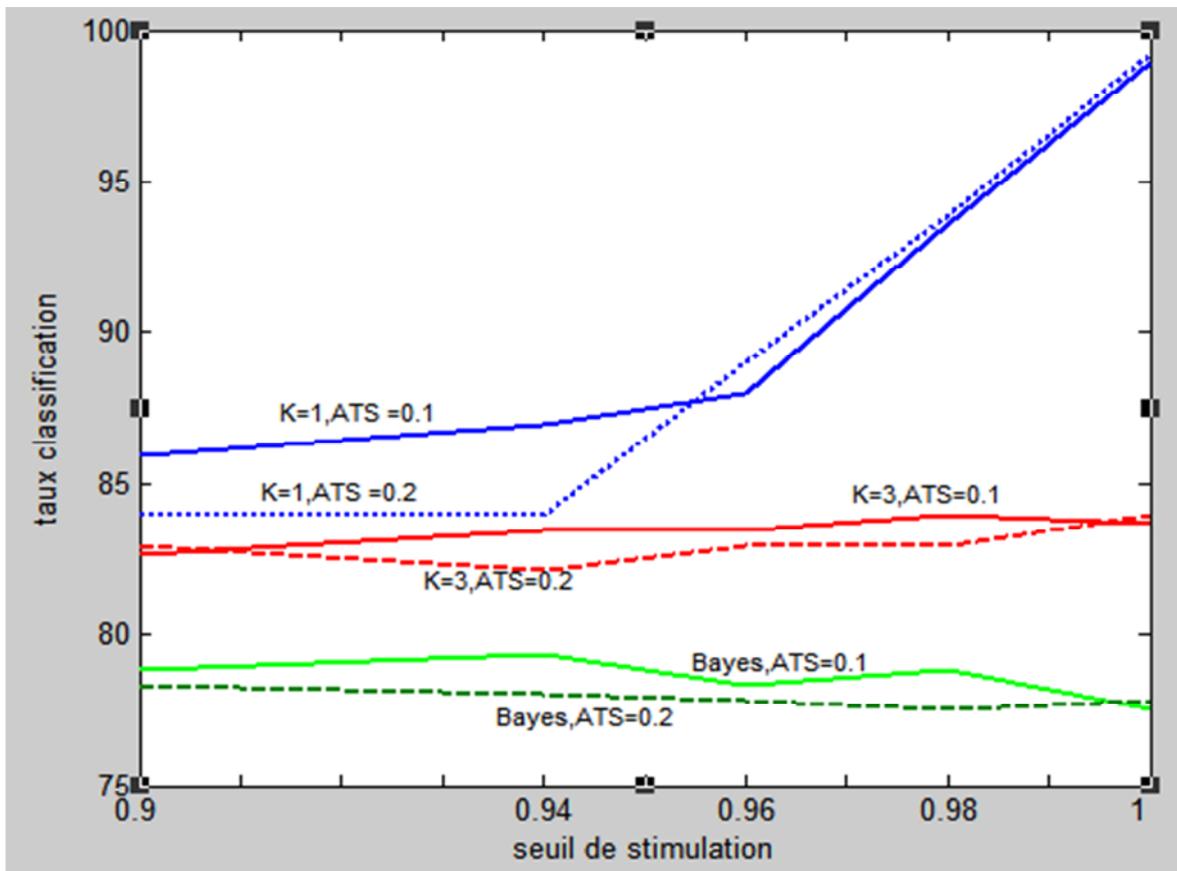


Figure 15 l'evolution d'apprentissage avec 1,3 plus proches voisins et naïf bayes

Suite aux résultats présentés au tableau 3 et à la figure 15, nous constatons que l'augmentation de la valeur du seuil de stimulation améliore le taux de classification plus précisément avec la valeur de  $k=1$  (avec naïf Bayes nous remarquons que le taux de classification ne change pas); mais d'autre part, le temps d'apprentissage devient plus important et le nombre de cellules mémoires augmente. Le meilleur choix pour la base de diabète, c'est de prendre un seuil de stimulation entre 0.90 et 0.94.

Les mêmes résultats montrent que les taux de classification avec l'ATS=0.1 et l'ATS=0.2, sont similaires et ceci malgré la réduction de la taille de cellules mémoires enregistrée entre les deux valeurs. Si on augmente la valeur d'ATS, le nombre de cellules mémoires va diminuer et nous constatons une amélioration du taux de classification (jusqu'à certain seuil).

Nous allons voir dans l'expérimentation suivante les autres effets du changement de la valeur d'ATS.

## Chapitre III : Résultats et discussions

---

### 7.2 Deuxième expérimentation :

Dans cette expérimentation, les performances de classification obtenues ont été évaluées en utilisant la validation croisée. Les relations entre l'ATS, la précision de classification, la sensibilité et la spécificité des valeurs, le temps de classification, sont indiquées dans le tableau 3.

**Tableau 1 évaluation de la classification avec Knn et Naive Bayes**

Nbre de ressources	ATS	K-fold	Classifieur	Taux de classification (%)	Sens (%)	Spec (%)	Temps De classification (millisecond)
200	0.1	10	Knn(k=3)	81,82	67,21	88,09	231
			Bayes	80,45	75,38	82,81	2912
200	0.15	10	Knn (k=3)	81,36	62.36	89.08	31
			Bayes	80.45	69.14	84.81	1623
50	0.15	10	Knn (k=3)	75.00	49.85	85.10	35
			Bayes	75.91	68.71	78.80	2865
200	0.2	10	Knn (k=3)	79.55	59.19	87.78	9
			Bayes	80.91	72.89	84.19	1620
200	0.25	10	Knn (k=3)	76.41	61.38	83.65	2
			Bayes	75.38	62.28	82.37	1545
200	0.1	6	Knn(k=3)	81.08	59.65	89.81	43
			Bayes	82.43	76.63	84.92	1686
200	0.15	6	Knn (k=3)	80.63	63.01	87.88	35
			Bayes	80.27	76.63	81.70	1616
200	0.2	6	Knn(k=3)	80.18	68.71	84.76	16
			Bayes	80.63	68.14	85.37	1496
200	0.25	6	Knn(k=3)	77.27	58.06	84.5	2
			Bayes	75.18	64.18	83.49	1372

## Chapitre III : Résultats et discussions

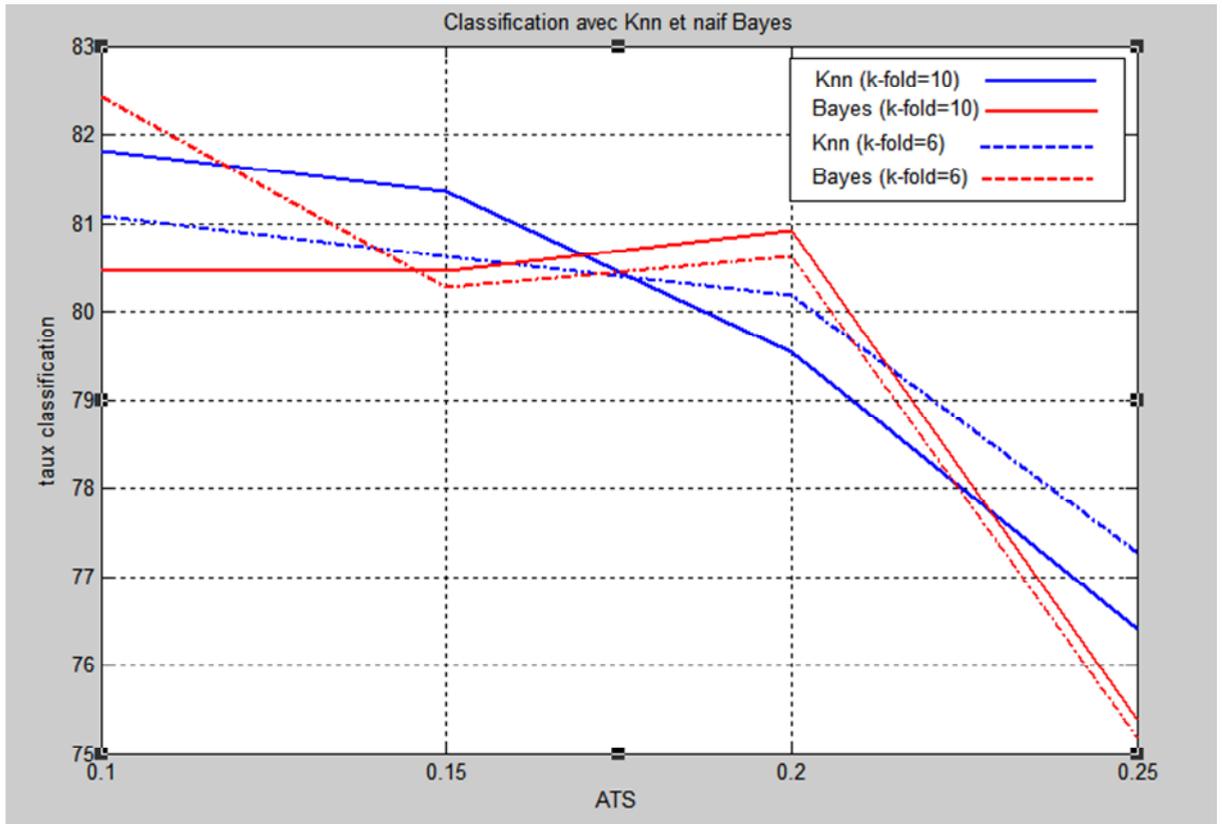


Figure 16 l'évaluation de la classification par rapport a le scalaire du suil de stimulation

Nous avons utilisé respectivement 10 et 6 validations croisées et le taux de classification était presque similaire. Il n'y a pas une grande différence de classification avec une valeur d'ATS égale à 0.15 et 0.2. On voit même qu'avec cette dernière valeur, nous obtenons un taux de classification supérieur à celui avec une valeur 0.1, et ceci, malgré le fait que le nombre de cellules mémoire a été réduit à 39%. Mais avec une augmentation plus importante, par exemple avec une valeur d'ATS égale à 0.25, le taux de classification diminue, parce que le nombre de cellules mémoire d'apprentissage générées devient insuffisant pour la classification.

## Chapitre III : Résultats et discussions

### 7.3 Troisième expérimentation :

Dans cette expérimentation nous allons comparer les performances de classification (avec) et (sans) l'algorithme AIRS2. En d'autres termes, nous allons d'abord évaluer les deux classificateur Knn et Naïf bayes, sans recourir à l'algorithme AIRS2 avec une base d'apprentissage qui contient 352 cas dont 236 cas non diabétiques et 116 cas diabétiques ; ensuite nous allons faire l'apprentissage avec AIRS2 en utilisant les cellules mémoires générées comme base d'apprentissage. Dans les deux cas la base de test contient 40 cas dont 26 sont non diabétiques et 14 sont diabétiques.

La figure 17 nous montre le taux de classification des deux classificateurs Knn et naïf Bayes, avec et sans apprentissage AIRS2 :

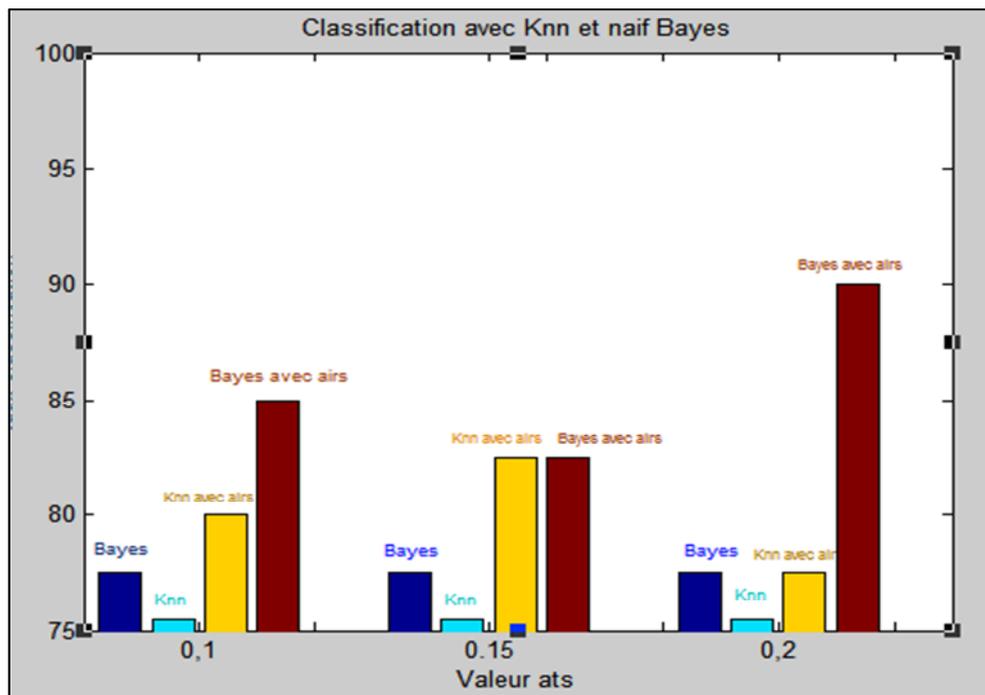


Figure 17 histogramme des performance des différents classificateur avec et sans AIRS2

Les relations entre la valeur d'ATS (scalaire seuil d'affinité), la taille des cellules mémoire, la précision de classification, et le temps de classification, sont indiquées dans le tableau 4 :

## Chapitre III : Résultats et discussions

**Tableau 7 Comparaison de classification avec et sans AIRS**

ATS	Base d'apprentissage	Taille de la base d'apprentissage taille des cellules mémoires générées (avec AIRS)	Classifieur	Taux de classification (%)	durée de classification(en milliseconde)
<b>0.1</b>	Sans AIRS	352	Knn	75.50	26.0
	Avec AIRS	268	(k=5)	80.00	13.0
<b>0.1</b>	Sans AIRS	352	Bayes	77.50	436.0
	Avec AIRS	268		85.00	428.0
<b>0.15</b>	Sans AIRS	352	Knn	72.50	26.0
	Avec AIRS	222	(k=3)	82.50	9.0
<b>0.15</b>	Sans AIRS	352	Bayes	77.50	436.0
	Avec AIRS	222		82.50	428.0
<b>0.2</b>	Sans AIRS	352	Knn	72.50	26.0
	Avec AIRS	197	(k=8)	77.50	8.0
<b>0.2</b>	Sans AIRS	352	Bayes	77.50	436.0
	Avec AIRS	197		90.00	426.0

D'après les résultats du tableau 4, nous remarquons qu'il y aura non seulement une augmentation du taux de classification avec l'algorithme AIRS2, mais on constate aussi une optimisation du temps de classification (ceci devient intéressant avec des grandes bases de données).

Avec une valeur d'ATS de 0.2, on a eu un taux de classification assez satisfaisant avec les deux classifieurs, malgré que la taille de la base d'apprentissage a été réduite de 37,5% (220) par rapport à la base initiale. Ce résultat revient à dire que les cellules mémoire générées sont les plus stimulées.

Les figures 17 et 18 présentent un aperçu de l'interface de notre application et les résultats de classification :

## Chapitre III : Résultats et discussions

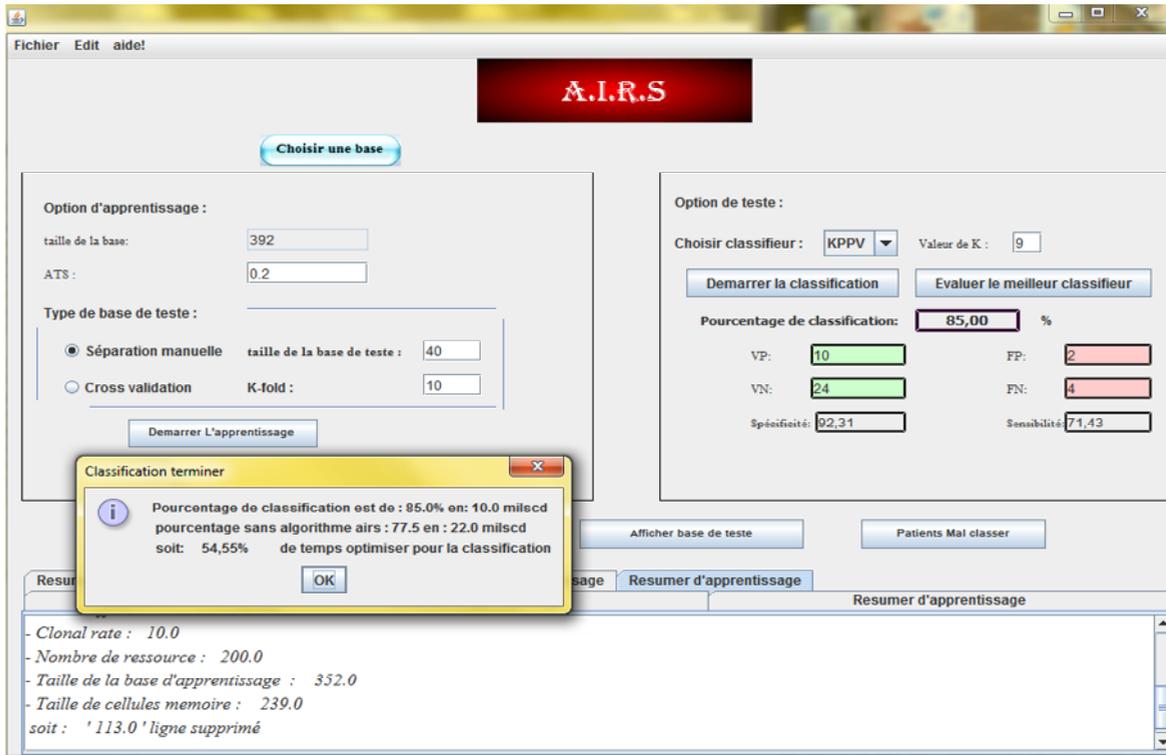


Figure 18 Classification avec Knn

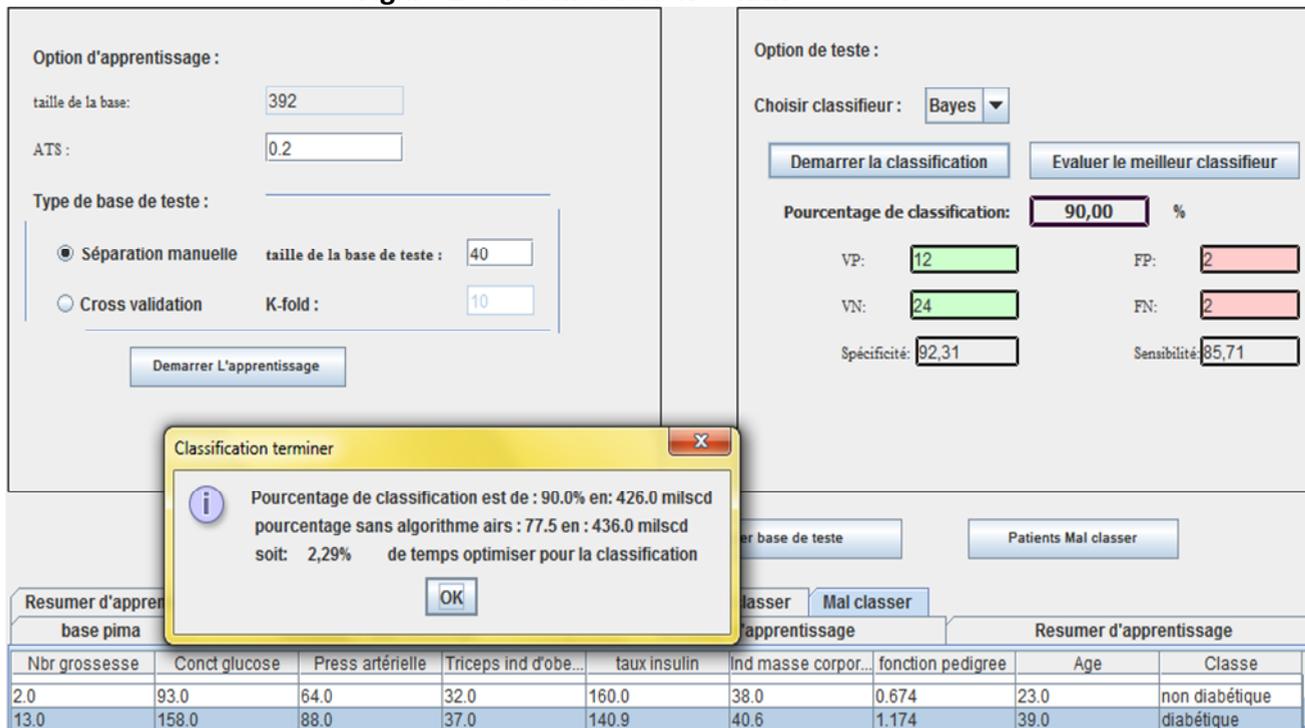


Figure 19 Classification Knn avec et sans AIRS sans AIRS

Dans la figure 19 (voir les 2 dernières lignes), nous remarquons 2 cas mal classés par le système. Pour le 1er cas par exemple, notre système l'a classé non diabétique alors qu'il est diabétique. Il s'agit d'une jeune femme (23ans), qui a une concentration

## Chapitre III : Résultats et discussions

---

de glucose normale (93.0mg /l), et qui, du fait de sa jeunesse et de l'absence de glucose anormal, ne devrait pas avoir les symptômes d'un cas diabétique, ce que reflète notre propre résultat.

Par contre pour le 2eme cas, alors que cette femme est considérée comme NON DIABETIQUE parla base de données Pima, notre propre système l'a classifié comme DIABETIQUE. En effet, notre tableau montre que cette femme a une concentration de glucose anormale (153mg /l contre 120 pour une personne normale) ;son nombre de grossesse est aussi un facteur de risque (13 grossesses pour cette personne) ; son IMC (40.6kg/m<sup>2</sup>) et l'épaisseur de peau (37mm) démontrent également que cette personne obèse réunit tous critères d'une personne diabétique. Il est d'ailleurs difficile de croire que cette personne est non diabétique à partir de ses descripteurs, si nous prenons en compte les seuils et les facteurs de risque.

### **8. Application de l'algorithme AIRS2 sur la base du cancer du sein et la base Iris :**

Afin d'évaluer les performances de notre algorithme AIRS nous l'avons testé sur d'autres bases de données ; le tableau ci-dessus montre les résultats de classification :

**Tableau 8 Classification des bases : breast cancer wisconsin et Iris**

Base de donnée	AT S	K-fold	Classifieur	Taux de classification (%)	Sens (%)	Spec (%)	Temps de classification( millisecond)
Breast cancer wisconsin(683)	0.1	10	Knn(k=3)	96.32	94.19	97.73	415
			Bayes	92.50	83.89	98.29	70901
	0.2	10	Knn (k=3)	97.50	96.61	97.40	300
			Bayes	90.88	80.70	97.99	63750
IRIS (150)	0.1	8	Knn (k=3)	95.14			6
			Bayes	95.83			47
	0.2	8	Knn (k=3)	96.52			5
			Bayes	92.36			16

## Chapitre III : Résultats et discussions

The screenshot shows the AIRS2 application interface. On the left, there is a settings panel titled "Choisir une base" with options for learning type, base size (683), ATS (0.2), and test type (Cross validation, K-fold: 10). A "Demarrer L'apprentissage" button is visible. On the right, a summary window titled "Le meilleur classifieur est : K plus proche voisin" displays classification results for two models: K-Nearest Neighbors (97.50% accuracy) and Naïf Bayes (90.88% accuracy). Below the settings, a table titled "base breast cancer wisconsin" displays 10 features and a class label for 20 samples.

Clump Thickne...	Uniformity of Ce...	Uniformity of Ce...	Marginal Adhes...	Single Epithelia...	Bare Nuclei	Bland Chromatin	Normal Nucleoli	Mitoses	Classe
5.0	1.0	1.0	1.0	2.0	1.0	3.0	1.0	1.0	2.0
5.0	4.0	4.0	5.0	7.0	10.0	3.0	2.0	1.0	2.0
3.0	1.0	1.0	1.0	2.0	2.0	3.0	1.0	1.0	2.0
6.0	8.0	8.0	1.0	3.0	4.0	3.0	7.0	1.0	2.0
4.0	1.0	1.0	3.0	2.0	1.0	3.0	1.0	1.0	2.0
8.0	10.0	10.0	8.0	7.0	10.0	9.0	7.0	1.0	4.0
1.0	1.0	1.0	1.0	2.0	10.0	3.0	1.0	1.0	2.0
2.0	1.0	2.0	1.0	2.0	1.0	3.0	1.0	1.0	2.0

Figure 20 Aperçu de notre application pour la classification de la base breast cancer wisconsin avec L'algorithme AIRS2

### 9. Conclusion :

Le nombre d'informations et de cas à traiter deviennent de plus en plus importants. La façon de décider rapidement et correctement devient un problème complexe, surtout pour les médecins qui doivent diagnostiquer de nombreux patients.

Un nombre de plus en plus grand de maladies doivent être diagnostiquées et traitées rapidement et le diabète ne fait pas exception à cette règle.

D'où l'intérêt de recourir à des algorithmes du genre AIRS pour déterminer que certains types de personnes sont susceptibles de contracter le diabète, car les médecins y trouvent des moyens indispensables pour diagnostiquer leurs patients.

D'après les résultats que nous avons obtenus, AIRS est une des méthodes les plus précises, comparativement aux méthodes traditionnelles. Nous espérons pouvoir affiner ces résultats dans les recherches futures.



# CONCLUSION GENERALE

## Conclusion générale

---

### Conclusion Générale

Ce travail a présenté un nouveau paradigme d'apprentissage supervisé à partir d'observations de fonctionnement du système immunitaire. Plusieurs éléments clés du système immunitaire ont été employés pour le développement de ce modèle. L'un de ces éléments est l'abstraction des cellules B en tant que moyen de représentation des données.

La réaction entre un antigène et un anticorps des cellules B a été simulée grâce à l'utilisation de la distance euclidienne. Le concept d'ARB a également été adopté pour représenter clones exacts dans l'espace de recherche basé sur la concurrence pour les ressources du système.

Le travail fondamental présenté ici est le développement des cellules mémoire ; ces cellules mémoire fournissent le principal mécanisme de reconnaissance des formes dans des classificateurs immunitaires artificiels à ressources limitées. Le développement de ces cellules mémoire est étroitement lié aux idées de la sélection clonale et la maturation.

La notion d'affinité entre les cellules est très importante. L'affinité inter-cellulaire joue un rôle clé dans la réduction des données et des critères de conservation de la mémoire du système ; par conséquent, elle favorise les possibilités de généralisation qui sont importantes pour un classificateur.

Bien que un bon nombre d'algorithmes a utilisé le développement des classificateurs immunitaires artificiels à ressources limitées, aucune n'a été exploitée pour l'apprentissage supervisé. L'apprentissage supervisé a été introduit de plusieurs façons :

*Premièrement*, la stimulation d'une cellule du système en réaction à un antigène d'apprentissage est basée non seulement sur son affinité pour l'antigène mais aussi sur la classification de la cellule par rapport à la classification de l'antigène.

*Deuxièmement*, les ressources sont allouées à des cellules en fonction de leurs classes et le niveau de stimulation avec plus de ressources allouées à des cellules de la même classe que l'antigène donné.

*Troisièmement*, un critère d'arrêt d'apprentissage est basé au niveau de stimulation et de la classe, pour s'assurer que les cellules de la même classification de haute qualité sont conservées dans le système.

## Conclusion générale

---

*Quatrièmement*, l'identification de cellules de mémoire, la conservation, et le remplacement, sont basés principalement sur leurs classes.

AIRS démontre sa capacité à élaborer une représentation des classes dans les données d'apprentissage et ensuite à utiliser cette représentation pour réussir à classer des nouveaux éléments de données à un degré élevé de précision. Enfin, une comparaison de l'algorithme AIRS a montré que les performances de cet algorithme sont comparables, et dans certains cas supérieur, à l'exécution d'autres techniques d'apprentissage supervisé.

### *Perspectives de recherche*

L'exploration des classificateurs immunitaires artificiels à ressources limitées présenté ici est loin d'être achevée. Ce travail a mis en évidence plusieurs domaines qui devraient se révéler fructueux pour des recherches futures. Le paragraphe qui suit va nous permettre de décrire quelques-uns de ces éléments plus en détail.

Il y a plusieurs modifications à l'AIRS qui pourraient être prises en compte. Une modification potentielle serait de déterminer le seuil d'affinité basée sur l'affinité moyenne des éléments d'apprentissage au sein de la même classe, plutôt que la méthode générale de l'utilisation de tous les éléments d'apprentissage. Cela pourrait éventuellement conduire à une représentation plus précise de chaque classe dans l'ensemble de données.

Une autre modification pourrait consister dans une seconde application des données d'apprentissage sur le système, sachant qu'au cours de ce deuxième passage, seules les cellules mémoire seraient impliquées.

Toutefois, un obstacle potentiel pour cette modification pourrait être le développement d'un ensemble de cellules de mémoire trop spécifiques aux données d'apprentissage fixées, ce qui ferait perdre cette capacité de généralisation aux nouveaux éléments de données (de test).

Il y a aussi la question des mutations et des taux de mutation. AIRS emploie un taux de mutation uniforme qui est aveugle à la qualité de l'ARB ou de la cellule mémoire à être muté. En outre, le mécanisme de la mutation dans AIRS est basé sur des changements aléatoires. Ceci pourrait être modifié dans des études futures.

Par exemple, pour les cellules fortement stimulées le taux de mutation sera réduit et pour les cellules mémoire les moins stimulées, le taux de mutation sera augmenté. De cette façon, seuls des changements mineurs sont effectués sur des cellules qui sont connus pour avoir une bonne reconnaissance, ce qui permet également d'optimiser le temps d'apprentissage.

## Références Bibliographiques

1. **Fédération international diabète** *Diabètes Atlas résumé*. s.l. :, 2003.
2. **André, Grimaldi** *EMC référence diabète de type 2*.
3. **Delecroix, Jean-Marie**. *Le diabète : le gérer - le guérir*.
4. **Guillausseau Pierre-Jean et al.**,. *Le diabète de type 2*. s.l. : Ellipses, 2003.
5. **P. Allain**. *Les médicaments 3ème édition*. 2006, Magazine pharmacorama.
6. **CHU de Rouen** *Les Anticorps* . s.l. :, 2009.
7. **Université de médecine de Nantes** *Immunologie, Lymphocytes B*. s.l. :, 2005.
8. **Leandro Nunes De Castro, Fernando J. Von Zuben.**,. *The Construction of a Boolean Competitive Neural Network Using Ideas from Immunology*. s.l. : Neurocomputing, 2003. pp. 51-85.
9. **Mehdi Dastani, Amal El Fallah Sghrouchni**. *Programming Multi Agents Systems*. s.l. : First International Workshop, ProMAS, 2003.
10. **D. Dasgupta**, *Artificial Immune Systems and Their Application* . 1999 : Springer-Verglas.
11. **Castro, J. Timmis & Leandro Nunes De Castro**. *Artificial immune Systems: A New computational Intelligence Approach*. london : Springer-Verglas, 2002.
12. **Leandro Nunes De Castro**. *An Introduction to the Artificial Immune Systems* . s.l. : ICANNGA, April, 2001.
13. **Carter**. *The immune systems as a model for pattern recognition*. s.l. : Journal of the American Medical Informatics Association, 2000.
14. **L. N. De Castro, and F. von Zuben**.*The clonal selection algothm*. pp. 36–37., Las Vegas, USA., : s.n., 2000.
15. **Timmis, J. and M. Neal**. *A Resource Limited Artificial Immune System*. 2001.
16. **Farmer**. *The immune system and adaptation and machine learning*. 1986.
17. **Timmis J., M. Neal, and J. Hunt**. *'An Artificial Immune System for*. 2000.
18. **Bermejo, S. Cabestany, J**. *Adaptive soft k-nearest-neighbour classifiers*. 2000., Vol. Vol. 33.
19. **Wikipedia**.*Classification naïve bayesienne*. [En ligne] [Citation : 2011 05 28.] [http://fr.wikipedia.org/wiki/Classification\\_na%C3%AFve\\_bayesienne](http://fr.wikipedia.org/wiki/Classification_na%C3%AFve_bayesienne).

20. **Kemal Polat, Salih Günes.** *Une approche système expert basé sur l'analyse en composantes principales et d'adaptation du système d'inférence neuro-flous au diagnostic de la maladie du diabète. Traitement numérique du signal.* Digital Signal Processing , p. 2007.
21. **S. W. Purnami, A. Embong, J. M. Zain, S. P. Rahayu.** *A New Smooth Support Vector Machine and Its Applications in Diabetes Disease Diagnosis.* 2009, Journal of Computer Science, pp. 1003–1008.
22. **Seral Şahan, Kemal Polat, Halife Kodaz , and Salih Güneş.** *The Medical Applications of Attribute Weighted Artificial Immune System (AWAIS): Diagnosis of Heart and Diabetes Diseases.*
23. **Gunes, Kemal Polat and Salih.** *An improved approach to medical data sets classification: artificial immune recognition system with fuzzy resource allocation mechanism.* . September 2007, pp. Vol. 24, No. 4.
24. **Simon, Matthieu.** *Immunologie générale.* Université de Liège : s.n., 2009.
25. **Hiba Khelil, Abdelkader Benyettou.** *Application du système immunitaire artificiel ordinaire et amélioré pour la reconnaissance des caractères artificiels.* Université des sciences et de la technologie d'Oran : Laboratoire Signal Image Parole SIMPA, 2006.
26. **De. Castro.L.N, & J.Timmis** *Artificial Immune System as a novel Soft Computing.* s.l. : To appear in the Soft Computing Journal, 2003, Vol. vol7. Issue 7.

## Résumé :

*Le diabète est une affection chronique qui demeure incurable encore aujourd'hui. Elle est la conséquence d'une carence ou d'une mauvaise gestion de l'insuline dans l'organisme, ce qui cause une élévation du taux de sucre dans le sang. En plus le diabète augmente les risques de développer des maladies cardiaques, des maladies rénales, des lésions nerveuses, et peut causer des dommages aux vaisseaux sanguins.*

*L'utilisation de systèmes experts et des techniques d'intelligence artificielle dans le diagnostic des maladies a augmenté progressivement.*

*AIRS (système de reconnaissance immunitaire artificielle) est l'une des méthodes utilisées dans les problèmes de classification médicale.*

*Ce projet de fin d'études présente une approche basée sur le système immunitaire artificiel (SIA) et deux algorithmes de classification : K-plus proches voisins et Naive Bayes (utilisés pour classer les données du diabète de type 2. Les performances obtenue sont très prometteuse .*

*Mots clés : diabète ,intelligence artificielle ,SIA , système immunitaire artificiel.*

## Abstract :

*Diabetes is a chronic condition that still remains incurable. It is caused by the lack or resistance to insulin in the organism causing the blood sugar levels to raise. In addition, diabetes is associated with a high risk of developing cardiac, renal, cerebral as well s vascular diseases.*

*The use of expert systems and artificial intelligence techniques in order to diagnose various conditions has been continuously increasing.*

*Artificial immune recognition system (AIRS) is one of a few methods used in the diagnosis of medical conditions.*

*This project outlines an approach based on AIRS as well as two classification algorithms : K- nearest neighbor and Naive Bayes used to classify diabetic (type 2)subjects. The results show that the implemented algorithm is successful and reliable for the recognition of diabetes in these subjects.*

Keywords: diabetes, artificial intelligence, AIRS, Artificial immune system.

## ملخص:

المرض السكري هو مرض مزمن الذي يزال عضال حتى يومنا هذا. هو نتيجة لنقص أو سوء إدارة الانسولين في الجسم ، مما تسبب في ارتفاع نسبة السكر في الدم. بالإضافة إلى هذا المرض السكري يزيد من خطر الإصابة بأمراض القلب، وأمراض الكلى وتلف الأعصاب، ويمكن أن تتسبب في تلف الأوعية الدموية. ازداد استخدام النظم الخبيرة وتقنيات الذكاء الاصطناعي و النماذج الذكية في تشخيص المرض تدريجيا. AIRS(نظام التعرف المناعي الاصطناعي) هو احدى الأساليب المستخدمة في تصنيف المشاكل الطبية. هذا المشروع يمثل الوسيلة القائمة على النظام المناعي الاصطناعي وخوارزميات التصنيف : ك - أقرب جار ونايف بايز تصنيف البيانات مرضى السكري من صنف 2. وأظهرت النتائج أن تنفيذ الخوارزمية فعال وموثوق به الاعتراف بالمرض

الكلمات الرئيسية : المرض السكري ، الذكاء الاصطناعي ، نظام المناعة الاصطناعية