

# Approche Multi agents pour la reconnaissance du cancer du sein

Nawel Hamdan

► **To cite this version:**

Nawel Hamdan. Approche Multi agents pour la reconnaissance du cancer du sein. [Rapport de recherche] 2011, pp.19. <inria-00633933>

**HAL Id: inria-00633933**

**<https://hal.inria.fr/inria-00633933>**

Submitted on 19 Oct 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Approche Multi agents pour la reconnaissance du cancer du sein

Rapport de Recherche de Master  
Hamdan Nawel  
2010-2011

## Résumé :

Le cancer du sein est le cancer qui touche de nos jours de plus en plus de femmes dans le monde. Ainsi, la lutte contre le cancer est loin d'être achevée. La médecine avance sur tous les fronts afin d'améliorer les soins des patients et vaincre cette maladie du siècle. De ce fait, il est indispensable que plusieurs disciplines continuent à apporter leur contribution et particulièrement l'Intelligence Artificielle. Pour permettre d'établir une aide au diagnostic médical, robuste et fiable, les systèmes multi agents (SMA) peuvent être un outil puissant de diagnostic distribué.

Dans ce mémoire, nous avons testé les performances des réseaux de neurones, des algorithmes génétiques sur la base de données du cancer du sein (WBCD) ensuite nous avons intégré ces derniers avec un agent Contrôleur (SMA) pour augmenter la précision de la classification. Les résultats obtenus de la classification sont très prometteurs.

## Sommaire :

Introduction générale.....	2
1. Contexte .....	2
2. Problématique.....	2
3. Contribution.....	3
1. Approche Multi Agents.....	3
1.1 Etat de l'art.....	4
1.2 Description de la base de la maladie du cancer du sein.....	5
1.3 Les outils de programmation.....	7
1.4 Principe de la classification.....	7
1.5 Phase de test (d'évaluation).....	7
2. Travail effectué.....	9
2.1 Classification mono-agents.....	9
2.1.1 Classification neuronale.....	9
2.1.2 : Classifieur neuro-génétique(NGC.....)	10
2.1.3 : Résultats .....	12
2.2 Approche multi agents.....	13
2.3 Conclusion et interprétation .....	15
2.4 Menu d'utilisation de notre application :.....	15
Conclusion générale.....	17
Références bibliographiques.....	17

## Introduction Générale

### 1. Contexte

Depuis les années quatre-vingt-dix, le taux de mortalité des patients atteints du cancer du sein a diminué et ceci est le fruit de plusieurs recherches effectuées au cours des dernières décennies.

Cependant le cancer du sein reste la maladie la plus commune et malveillante chez les femmes.

Le cancer du sein représente un problème de santé publique majeur, avec plus de 370,000 nouveaux cas et 130,000 morts par an en Europe chez les femmes de 35 à 64 ans et une survie complète à 5 et 10 ans autour de 68 % et 50 %, respectivement

Par ailleurs, il est difficile pour un spécialiste de diagnostiquer chez un patient si ce dernier est atteint d'un cancer du sein ou pas, de confirmer sa présence ou déterminer ses caractéristiques (son extension, son agressivité, ....)

La situation actuelle a motivé la recherche dans ce domaine et la nécessité d'automatiser le diagnostic médical est devenu indispensable

L'utilisation d'un outil intelligent spécialisé dans le diagnostic médical a pour ambition de lever cette difficulté

Ce système a pour ambition de donner plus d'autonomie et d'initiative aux différents modules logiciels spécialisés dans le diagnostic médical et qui peuvent dialoguer pour partager leurs connaissances comme des experts humains.

### 2. Problématique

Le **cancer du sein** reste le premier cancer féminin à l'échelle mondiale, il représente un véritable problème social

La fréquence du cancer du sein est très variable selon les pays et les facteurs qui interviennent sont multiples : facteurs génétiques, rôle de l'alimentation.

Dans les recherches relatives au cancer de façon générale, les études épidémiologiques qui concernent le **cancer du sein** occupent une place de plus en plus importante. Sans nul doute parce que le cancer du sein demeure une pathologie fréquente et grave : 460 000 décès chaque année dans le monde.

La problématique de la détection du cancer du sein conduit les chercheurs, les spécialistes du domaine à se pencher sur d'autres tendances, de nouvelles technologies autres qu'humaines afin de remédier à ce véritable problème social

### **3. Contribution**

L'objectif de notre travail consiste à créer une nouvelle approche qui permettra de savoir si un patient a un cancer bénin ou malin suivant plusieurs descripteurs. Pour atteindre cet objectif, nous proposons une solution basée sur le concept des systèmes multi-agents.

Nous réalisons un couplage entre deux grands domaines la Médecine et l'Informatique pour le diagnostic du cancer du sein et le paradigme agent respectivement.

L'optique de ce couplage est d'améliorer les performances et l'efficacité des systèmes de diagnostic médical.

La particularité de notre approche est le développement et la définition d'un modèle fondée sur une architecture composée de parties distinctes, chaque partie va diagnostiquer si le cancer est malin ou bénin à partir d'une base de données d'une manière spécifique, mais pouvant communiquer pour partager leurs connaissances.

Le paradigme multi agent dans notre approche serait appliqué dans plusieurs contextes :

- Chaque agent dans ce système fera son propre diagnostic pour la même base d'apprentissage (autonomie)
- La décision finale sera prise après un vote majoritaire : le patient sera attribué à la classe qui a été nommé par le grand nombre d'agents.

#### **1. Approche Multi agents**

Le diagnostic médical est un processus de classification. L'utilisation de techniques dite intelligentes pour la réalisation de cette classification devient de plus en plus importante.

Même si la décision de l'expert est le facteur le plus important lors du diagnostic, les systèmes de classification fournissent une aide substantielle, car ils réduisent les erreurs dues à la fatigue et le temps nécessaire pour le diagnostic.

Le but d'une classification médicale de pathologies est de rassembler les cas qui ont des similitudes biologiques et qui sont susceptibles de partager certains facteurs étiopathologiques.

L'identification des classes est importante vu qu'elle permet, d'une part de comprendre le processus de la maladie et d'autre part d'instaurer l'approche thérapeutique adéquate. En outre, elle permet de dégager le pronostic global de la maladie.

Plusieurs méthodes de classification comprenant les statistiques, l'intelligence artificielle et les réseaux de neurones ont été utilisées pour l'aide au diagnostic médical [1] [2] [3]

Dans le cadre de ce mémoire ; nous nous sommes intéressés à étudier la performance des réseaux de neurones, leurs hybridations avec les algorithmes génétiques pour la classification du cancer du sein

Après la phase de la classification mono agent, nous avons instauré le principe multi agent comme suit :

Nous avons utilisé quatre agents :

- ✓ les trois premiers représenteront les classifieurs neuronaux et neuro-génétiques paramétrique et structurel
- ✓ le quatrième sera un agent contrôleur Il aura la tâche de rassembler les résultats respectifs des trois agents pour donner le diagnostic final

### 1.1 Etat de l'art

Plusieurs travaux ont été réalisés dans la littérature pour le diagnostic médical en utilisant les méthodes connexionnistes, évolutionnaires et les approches multi agents.

En ce qui concerne

#### ➤ *l'approche connexionniste*

1/**Cheng et al.** [5] ont utilisé deux modèles des réseaux de neurones artificiels.

- ✓ Le premier est un réseau à rétro propagation d'erreur avec une topologie à trois couches.
- ✓ Le second est un réseau utilisant des fonctions à base radiale.

Le taux de reconnaissance obtenu avec le premier est de **72%**, et avec le second est de **65%**, bien que la démarche est intéressante, le temps d'exécution est long entraînant un sur apprentissage et ceci a donné des résultats faibles.

2/ **Guo et Nandi** [12], (2006), ont proposé un perceptron multicouche (PMC) en tant que classifieur avec l'algorithme de rétro propagation d'erreur pour le diagnostic du cancer du sein avec la base de données *WDBC (Wisconsin Diagnosis Breast Cancer)*

ils ont obtenu un taux de classification de **96.21 %**.

3/**Verma** [4] a proposé un nouvel algorithme d'apprentissage pour les réseaux de neurones appelé **SCNN(Soft cluster neural network)** basé sur la technique de soft clustering pour l'optimisation des poids des réseaux de neurones en utilisant la base de données de mammographie DDSM (Digital Database of Screening Mammography),. Ce système a donné un taux de classification **94%**

#### ➤ **L'approche évolutionnaire.**

1/ **Zhang et al.** [6] ont employé un algorithme génétique avec réseau de neurones pour la classification du cancer du sein avec un taux de reconnaissance de **90.5%**

2/Sekkal et al. [7] ont appliqué les algorithmes génétiques combinés avec les réseaux de neurones pour l'amélioration de l'architecture sur la base de données d'arythmies cardiaques de la base MIT BIH ils ont obtenu un taux de classification de **98.86%**, une sensibilité de 99.09% et une spécificité de 98.66%.

Ils ont utilisé les algorithmes génétiques fondés sur un classifieur neuronal pour les poids de connexions sur la même base de données. Cette approche a donné de très bons résultats avec un taux de classification correcte de 98,72% et une sensibilité de 97,33% par rapport à classificateur classique qui a un taux de classification de 95,71% et 87,98% de sensibilité.[13]

### ➤ L'approche multi agents

**Le système de diagnostic CMDS** (Contract Net Based Medical Diagnosis System) se compose d'agents artificiels et de médecins spécialisés dans différents domaines médicaux. Au début, un problème noté « Pr » est communiqué à un agent médical nommé « Mda » membre du système, qui aura pour rôle de manipuler de façon autonome la résolution des problèmes. en cas de difficultés il fait appel à un agent auxiliaire appelé « Asg » ou un autre agent médecin « Mda1 » avec une autre spécialité. [8]

**A.-J. Fougères** a illustré Une architecture cognitive d'agents communicants dans des systèmes d'information complexes

le système se compose des agents suivants :

Un agent nommé « Agent *infectieux* » est introduit localement (par un agent appelé « Agent de simulation » qui active les maladies (grippe, méningite). Il peut, en raison de sa proximité, infecter un agent nommé « agent *individu* sain ». Ce dernier devient porteur de la maladie et peut la véhiculer dans son environnement évolutif (déplacements), il informe ses symptômes (sous forme de messages) à un agent nommé « AgentMedecin » dans le cas d'une épidémie, l'AgentMedecin informe un agent nommé « Agent INVS » pour assurer le control [11]

**A Multi-Agent System (MAS) Based Scheme for Health Care and Medical Diagnosis System** se compose des agents suivants :

Un agent utilisateur nommé « Agent **UA** » dont le rôle est de prendre des symptômes des nouveaux patients comme des données brutes, ceci à l'aide d'une interface utilisateurs, un agent principal nommé « Agent **MA** » applique une certaine méthode pour convertir les données en connaissances afin de choisir un agent docteur spécialiste (nomme « agent **SDA** ») pour manipuler le cas du patient [10]

**NB** : Les agents « médecins » et les agents « INVS » sont doté d'une base de connaissance agent principal « MA » et l'agent spécialiste « SDA » sont doté d'une base de connaissance.

## 1.2 Description de la base de la maladie du cancer du sein

La base de données du cancer du sein dénommée « Wisconsin Breast Cancer Database » a été obtenue par l'Université du Wisconsin [9]

elle contient les informations médicales de 699 cas cliniques relatifs au cancer du sein classés comme bénin ou malin : 458 patientes (soit 65.5%) sont des cas bénins et 241 patientes (soit 34.5%) sont des cas malins.

La base de données contient 16 données manquantes; les patientes sont caractérisées par 11 attributs : le premier fait référence à l'identificateur de la patiente et le dernier représente la classe: le diagnostic est de 2 si le cas est bénin ,4 si le cas est malin quant aux 9 autres, ils représentent des cas cliniques suivants:

1-Clump Thickness: l'épaisseur de la membrane plasmique d'une cellule cancéreuse est plus importante que celle d'une cellule normale.

2. Uniformity of Cell Size : les cellules cancéreuses sont caractérisées par une anisocytose, à savoir une inégalité au niveau de la taille par comparaison avec les cellules saines.

3. Uniformity of Cell Shape : les cellules cancéreuses sont marquées par des contours irréguliers ainsi que des incisures

4. Shape Marginal Adhesion: une surexpression de la protéine integrin beta3 au niveau de la surface de la cellule cancéreuse.

5 .Single Epithelial Cell Size: étant donné que les cellules épithéliales sont absentes à l'état naturel au niveau de la moelle osseuse et qu'elles ne sont pas détectées chez les individus sains, la moelle osseuse peut, de ce fait, être considérée comme un indicateur de maladie métastatique chez les patients atteints du cancer du sein au stade primaire.

6. Bare Nuclei: à l'état normal, les nucléoles se trouvent à l'intérieur du noyau. Dans le cas où ses derniers se trouvent confondus avec le cytoplasme cela indique que la cellule présente une anomalie et qu'elle est susceptible de devenir cancéreuse.

7 Bland Chromatin : H2az est une protéine qui induit l'expression du gène du récepteur d'œstrogènes.

La surproduction de cette protéine est un marqueur de présence de cellules cancéreuses au niveau du sein étant donné qu'elles son hormono-dépendantes.

8 Normal Nucleoli : L'ADN est naturellement protégé par une membrane nucléaire. Une défaillance observée au niveau de cette membrane peut refléter une croissance tumorale.

9. Mitoses : La mitose est un processus de division cellulaire régulé permettant de reproduire des cellules filles génétiquement identique à la cellule parentale.

Les cellules malignes sont caractérisées par une division cellulaire anarchique et intense par comparaison avec une population cellulaire normale.

Remarque : Etant donné qu'il y a 16 données manquantes, nous nous sommes restreints à travailler sur 683 /699 patientes

### **1.3 Les outils de programmation**

Afin de réaliser ce travail nous avons utilisé des outils suivants :

- 1) pour effectuer la phase d'apprentissage et la phase de test sur la base de données nous avons utilisé le langage Matlab 7
- 2) pour illustrer le principe des systèmes multi agents nous avons sollicité la plateforme jade en ayant recours au langage Java

*\*A noter que le passage de Matlab à Java s'est réalisé à l'aide de la bibliothèque jMatlink*

### **1.4 principe de la classification**

Dans certains cas, il est possible de décrire complètement, de manière linguistique, la démarche de classification; dans ce cas, un algorithme reproduisant cette démarche peut être construit et le problème est résolu. Dans d'autres cas, il est impossible de décrire précisément la classification; une solution consiste alors à demander à un professeur (expert) de classer un échantillon d'objets.

Des méthodes de résolution, qui apprennent par l'exemple, sont capables de reproduire la classification de l'expert et, ensuite, de classer automatiquement de nouveaux exemples inconnus.

Nous avons utilisé ce deuxième type qui se base sur l'apprentissage par l'exemple.

Pour effectuer la phase d'apprentissage il est nécessaire d'avoir une optimisation de la fonction de coût d'où la définition de cette dernière est primordiale.

Car celle-ci sert à mesurer l'écart entre la sortie du modèle et les mesures faites sur les exemples d'apprentissage. Ainsi elle consiste à avoir une maximisation du taux de reconnaissance et en même temps une minimisation de l'erreur.

L'idée est de disposer d'un ensemble permettant de tester la qualité de la procédure de classification induite. On partitionne l'échantillon en un ensemble d'apprentissage et un ensemble test. La répartition entre les deux ensembles doit être faite expérimentalement. L'estimation de l'erreur réelle est alors l'erreur apparente mesurée sur l'ensemble test.

La qualité de l'apprentissage augmente avec la taille de l'ensemble d'apprentissage. Mais, dans la pratique, la taille de l'échantillon est limitée.

Cette méthode donne de bons résultats lorsque l'échantillon est "assez" grand. Il existe peu de résultats théoriques sur les tailles d'échantillon nécessaires pour utiliser cette méthode, nous ne disposons que de résultats empiriques qui dépendent du problème (souvent, plusieurs centaines d'exemples).

La répartition de l'échantillon entre les deux ensembles se fait en général dans des proportions 1/2, 1/2 pour chacun des deux ensembles ou 2/3 pour l'ensemble d'apprentissage et 1/3 pour l'ensemble test.

### **1.5 Phase de test (d'évaluation)**

Cette phase doit permettre l'affectation d'un nouvel objet à l'une des classes, au moyen d'une règle de décision intégrant les résultats de la phase d'apprentissage. L'objectif est d'obtenir une



estimation la plus fidèle possible du comportement du classifieur dans des conditions réelles d'utilisation. Pour cela, des critères classiques comme les taux de classification et les taux d'erreur sont presque systématiquement utilisés. Mais d'autres critères, comme la spécificité et la sensibilité, apportent aussi des informations utiles.

**a-taux de classification** : Les taux de classification et d'erreurs permettent d'évaluer la qualité du classifieur par rapport au problème pour lequel il a été conçu. Ces taux sont évalués grâce à une base de test qui contient des formes étiquetées par leur classe réelle d'appartenance comme celles utilisées pour l'apprentissage afin de pouvoir vérifier les réponses du classifieur.

Pour que l'estimation du taux de reconnaissance soit la plus fiable possible, il est important que le classifieur n'ait jamais utilisé les échantillons de cette base pour faire son apprentissage, de plus cette base de test doit être suffisamment représentative du problème de classification.

En général, quand les échantillons étiquetés à disposition sont suffisamment nombreux, ils sont séparés en deux parties disjointes et en respectant les proportions par classes de la base initiale. Une partie sert pour former la base d'apprentissage et l'autre pour former la base de test.

Les performances en termes de taux de classification sont alors déterminées en présentant au classifieur chacun des exemples de la base de test et en comparant la classe donnée en résultat à la vraie classe.

Le taux de classification correcte est défini par :

$$CC = \frac{VP(i)+VN(i)}{VP(i)+VN(i)+FP(i)+FN(i)}$$

Avec : *VP* : Vrai Positif : nombre de positifs classés positifs.

*VN* : Vrai Négatif : nombre de négatifs classés négatifs.

*FP* : Faux Positif : nombre de négatifs classés positifs.

*FN* : Faux Négatif : nombre de positifs classés négatifs.

## **b. sensibilité et spécificité**

**Sensibilité** : on appelle sensibilité (*Se*) du test sa capacité de donner un résultat positif quand la maladie est présente. Dans le langage des probabilités, la sensibilité mesure la probabilité conventionnelle que le test soit positif lorsque la maladie est présente. La sensibilité est estimée par la proportion de résultats positifs par suite de l'application du test à un groupe d'individus reconnus comme ayant la maladie.

$$S e = \frac{VP}{VP + FN}$$

**Spécificité** : on appelle spécificité (*Sp*) du test cette capacité de donner un résultat négatif quand la maladie est absente. Dans le langage des probabilités, la spécificité mesure la probabilité conventionnelle que le test soit négatif lorsque la maladie est absente. La spécificité est estimée par la proportion de résultats négatifs conséquemment à l'application du test à un groupe d'individus reconnus comme n'ayant pas la maladie.

$$S p = \frac{VN}{VN + FP}$$

## 2 TRAVAIL EFFECTUE

Ce travail se divise en deux parties,

- la première partie étant la classification par les trois méthodes (méthode neuronale, neuro-génétique paramétrique, neuro-génétique structurel)
- la seconde décrit comment nous avons procédé pour illustrer le principe multi agents et ceci à l'aide des outils cités précédemment.

### 2.1 Classification mono-agents

Dans ce qui suit nous détaillons les démarches suivies afin de réaliser la classification

#### 2.1.1 Classification neuronale (CNC)

- **Implémentation d'un RNMC (réseau de neurone multicouches)**

✓ Apprentissage structurel

- ✓ Nombre de neurones d'entrée : nous avons utilisé 9 neurones (les 9 vecteurs d'entrées), nous avons jugé d'enlever le premier (le code de la patiente) vu qu'il n'a pas un impact sur les résultats
  - ✓ Nombre de neurones cachés : après plusieurs expérimentations, nous avons fixé le nombre de neurones à 15
  - ✓ Nombre de neurones de sortie : nous avons utilisé un seul neurone représentant la classe
- Remarque : La fonction d'activation pour les neurones cachés est 'logsig'

La fonction d'activation pour la couche de sortie est 'purelin'

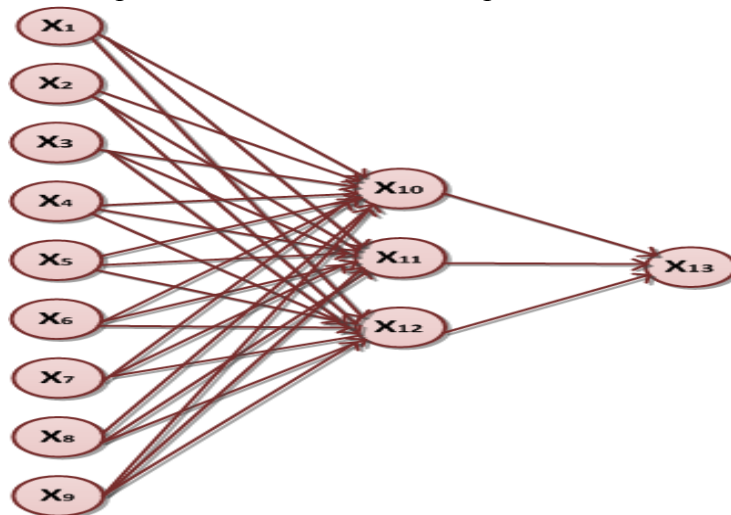


Figure IV.22 : Architecture du réseau RNMC

- ✓ **Apprentissage paramétrique**

- **Choix de l'algorithme d'apprentissage**

L'algorithme d'apprentissage utilisé c'est l'algorithme de rétro propagation par la méthode de Levenberg-Marquardt (LM)

Cet algorithme est utilisé dans les réseaux de type feed forward, ce sont des réseaux de neurones à couches, ayant une couche d'entrée, une couche de sortie, et au moins une couche cachée. Il n'y a pas de récursivité dans les connexions, et pas de connexions entre neurones de la même couche. Le principe de la rétro propagation consiste à présenter au réseau un vecteur d'entrées, de procéder au calcul de la sortie par propagation à travers les couches, de la couche d'entrée vers la couche de sortie en passant par les couches cachées. Cette sortie obtenue est comparée à la sortie désirée, une erreur est alors obtenue. A partir de cette erreur, est calculé la sortie qui est à son tour propagé de la couche de sortie vers la couche d'entrée, d'où le terme de rétro propagation. Cela permet la modification des poids du réseau et donc l'apprentissage. L'opération est réitérée pour chaque vecteur d'entrée et cela jusqu'à ce que le critère d'arrêt soit vérifié.

### 2.1.2 Classifieur neuro-génétique

Pour ce classifieur, nous avons débuté par la même architecture que le RNMC

#### ✓ Apprentissage paramétrique

En premier lieu une population initiale est créée (taille de la population fixée expérimentalement, en ce qui nous concerne nous avons utilisé  $N=100$ ) avec des poids aléatoires compris entre  $-1,0$  à  $+1,0$ .

une fois le jeu de poids généré il sera représenté sous la forme d'un chromosome

Sachant que le chromosome est une collection de gènes (voir chapitre 3); et qu'en ce qui nous concerne nous avons gardé la structure suivante (9-3-1) avec 30 connections (liens pondérés entre un neurone et un autre), le chromosome sera alors représenté par 30 gènes.

Maintenant, nous évaluons les performances de chaque chromosome en définissant la fonction fitness ; dans notre cas nous avons utilisé le calcul de l'erreur quadratique

Selon les résultats obtenus, les meilleurs chromosomes (ayant l'erreur quadratique la plus petite) sont sélectionnés (Stochastique Uniform), mutés (Gaussian) et croisés ( $P_c=0.8$ ) formant ainsi une nouvelle population

La procédure se répète jusqu'à atteindre le nombre de génération ( $N=100$ )

1	2	3	4	5	6	7	8	9	10
-0.6041	-0.7186	-0.4667	0.3974	-0.6138	-0.8826	1.5129	0.1252	3.2231	-0.4317
11	12	13	14	15	16	17	18	19	20
2.5582	1.2547	4.1567	-0.6688	2.1089	2.1227	1.4714	2.1566	-0.5217	0.4696
21	22	23	24	25	26	27	28	29	30
1.5524	-0.7317	-0.5079	0.7260	-0.5946	0.9116	-2.3084	-0.3190	2.0591	0.5714

Figure IV.23 : chromosome des poids synaptiques

✓ *Apprentissage structurel* il suit le même principe que pour l'apprentissage paramétrique, sauf que nous avons :

- 1) remplacer le poids par une connexion binaire : 1 s'il y a une connexion entre un neurone et un autre sinon 0.
- b) dans ce cas nous avons gardé la structure (9-4-1) avec 40 connexions, (Figure IV.24)

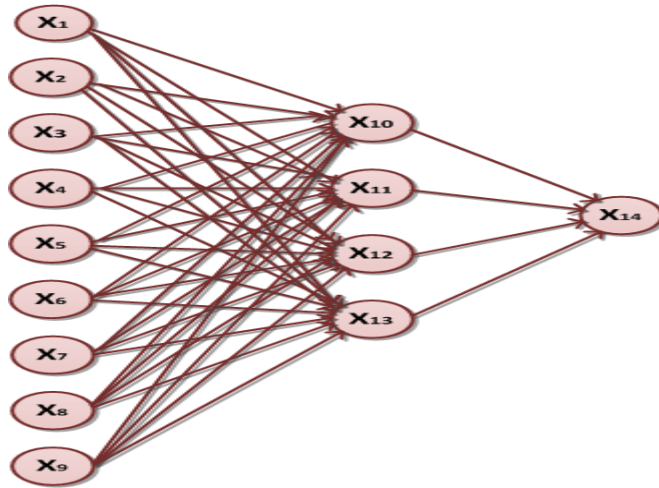


Figure IV.24 architecture du réseau neuro génétique avant apprentissage

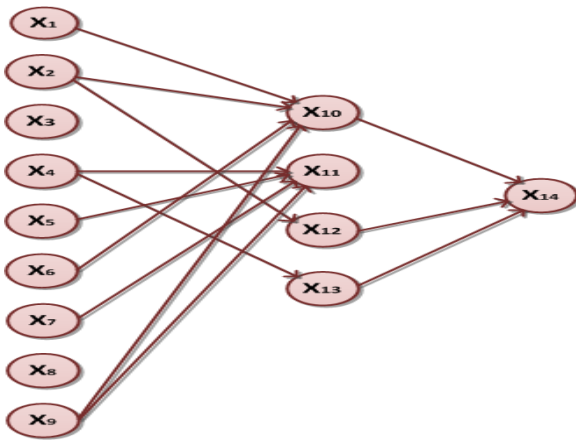


Figure IV.25: Architecture du réseau neuro génétique après apprentissage

1	2	3	4	5	6	7	8	9	10
1	1	0	0	0	1	0	0	0	0
11	12	13	14	15	16	17	18	19	20
0	1	1	0	1	0	1	0	0	1
21	22	23	24	25	26	27	28	29	30
0	0	0	0	0	0	0	0	0	0
31	32	33	34	35	36	37	38	39	40
1	0	0	0	0	0	1	0	1	1

Figure IV.26: Chromosome des connexions

### 2.1.3 Résultats

Pour effectuer la répartition des 683 patientes : nous avons gardé 2/3(456) pour la phase d'apprentissage et 1/3 (227) pour la phase de test

#### ➤ Résultats obtenus par le classifieur neuronal

Erreur atteinte du réseau	Nombre de neurones cachés	nombre d'itération	CC	Se	Sp	VP	VN	FP	FN
0.01	15	150	98,65%	96%	99,42%	48	172	1	2

Tableau IV. 1 : Résultats du classifieur neuronal (CRN)

#### ➤ Résultats obtenus par le classifieur neuro-génétique

##### ✓ Apprentissage paramétrique

Population	génération	Nombres de neurones cachés	CC	Se	Sp	VP	VN	FP	FN
100	100	3	97,23%	97,07%	98,28%	51	172	3	1

Tableau IV.2 : Résultats du classifieur neuro-génétique paramétrique(CNGP)

##### ✓ Apprentissage structurel

Population	génération	Nombre de neurones cachés	CC	Se	Sp	VP	VN	FP	FN
100	100	4	86.34%	100%	82.28%	52	142	31	0

Tableau IV.3 : Résultat du classifieur neuro-génétique structurel(CNGS)

#### ➤ Interprétation des résultats

Les classifieurs neuro-génétiques ont donné des meilleures performances particulièrement dans la reconnaissance des cas malins (voir tableaux)

Cette performance est due essentiellement à l'optimisation de la structure et à la modification des poids en plus pour l'apprentissage structurel (2 entrées X3 et X8) ont été éliminées, ces deux dernières représentent respectivement : Uniformity of cell shape et Normal Nucleoli

Du point de vue médical, ils sont les moins significatifs (confirmé par les experts du domaine)

## 2.2 Approche multi agents

Nous avons adopté l'approche distribuée (SMA) pour profiter des contributions des trois modèles (CNC, CNGP, CNGS) représentant chacun un agent auquel nous avons intégré un quatrième agent que nous avons dénommé « AgentContrôleur ».

Au préalable, ces différents agents ont été dénommés comme suit :

1-*Agent Classifieur* pour les réseaux de neurones CNC

2-*AgentClassifieur1* pour les neuro-génétique paramétrique CNGP

3-*AgentClassifieur2* pour les neuro-génétique structurel CNGS

L'AgentContrôleur récupère les résultats des trois agents ensuite il calcule ses propres performances afin d'établir le diagnostic final les résultats obtenus sont présentés dans le tableau suivant :

AgentContrôleur		
CC	Se	Sp
97,5%	98%	97%

Tableau IV.4 : Résultats AgentContrôleur

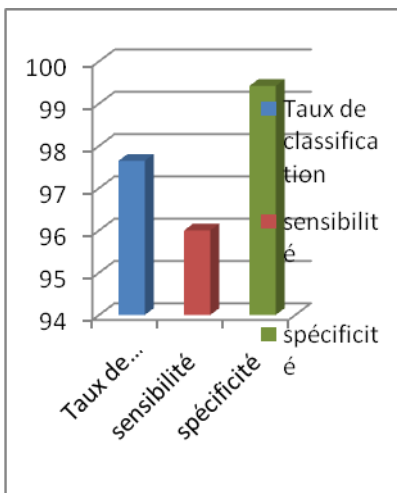


Figure IV.27 : Histogramme Des performances du CNC

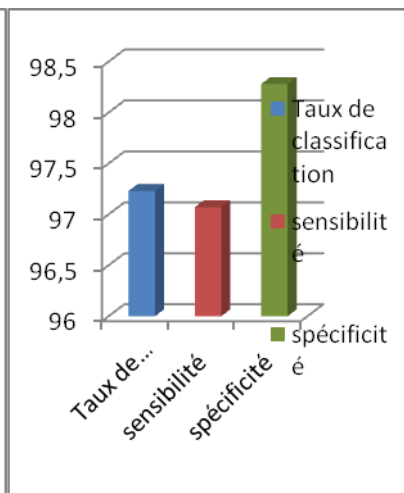


Figure IV.28 : Histogramme des performances du CNGP

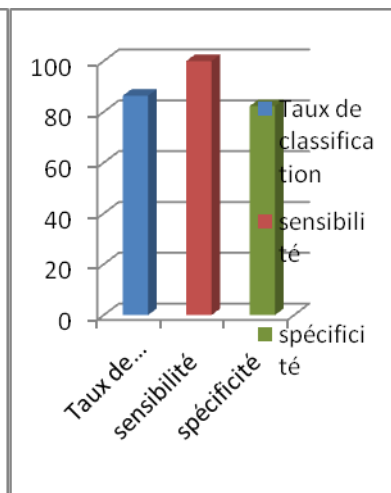


Figure IV.29 Histogramme des performances du CNGS

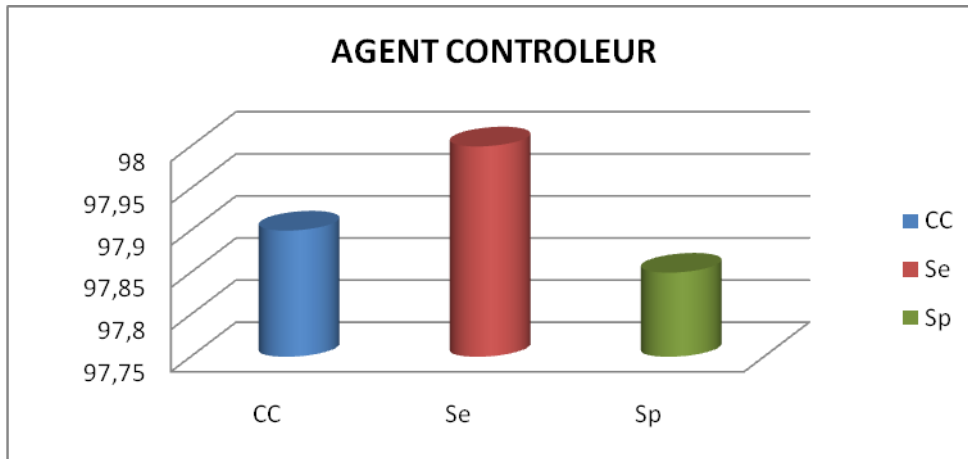


Figure IV.30 :Histogramme des performances de l'AgentControleur

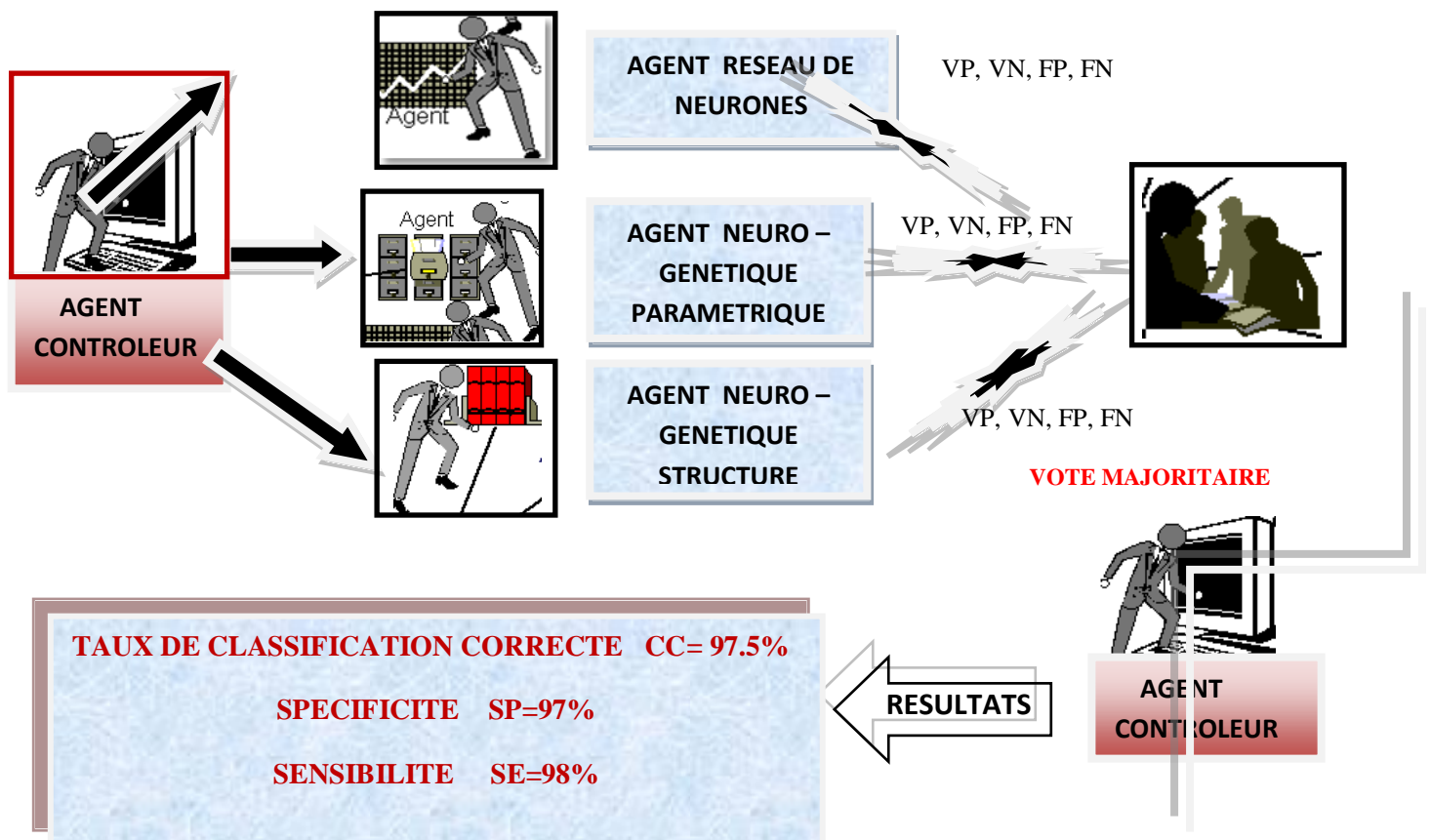


Figure IV.31 : processus de relation agent contrôleur - agents classifieurs

### IV.2.3 Conclusion et interprétation :

Nous remarquons que l'AgentContrôleur a bénéficié des meilleurs résultats obtenus par les trois agents (vote majoritaire) pour calculer ses propres performances qui ont augmenté visiblement ses résultats.

### 2.4 Menu d'utilisation de notre application :



Figure IV.32: Interface crée

Lancement de la plateforme Jade (*launch plateforme*) de *launch RMA* apparition de l'interface



```

Output - JanchGraphicq (run)
18 juin 2011 23:56:49 jade.core.BaseService init
INFO: Service jade.core.event.Notification initialized
18 juin 2011 23:56:49 jade.core.messaging.MessagingService clearCachedSlice
INFO: Clearing cache
18 juin 2011 23:56:50 jade.mtp.http.HTTPServer <init>
INFO: HTTP-MTP Using XML parser com.sun.org.apache.xerces.internal.jaxp.SAXParserImpl$JAXPSAXParser
18 juin 2011 23:56:50 jade.core.messaging.MessagingService boot
INFO: MTP addresses:
http://pc:7778/acc
18 juin 2011 23:56:50 jade.core.AgentContainerImpl joinPlatform
INFO: -----
Agent container Main-Container@192.168.1.5 is ready.
-----

```

Figure IV.33 :Extrait d'un exemple d'exécution

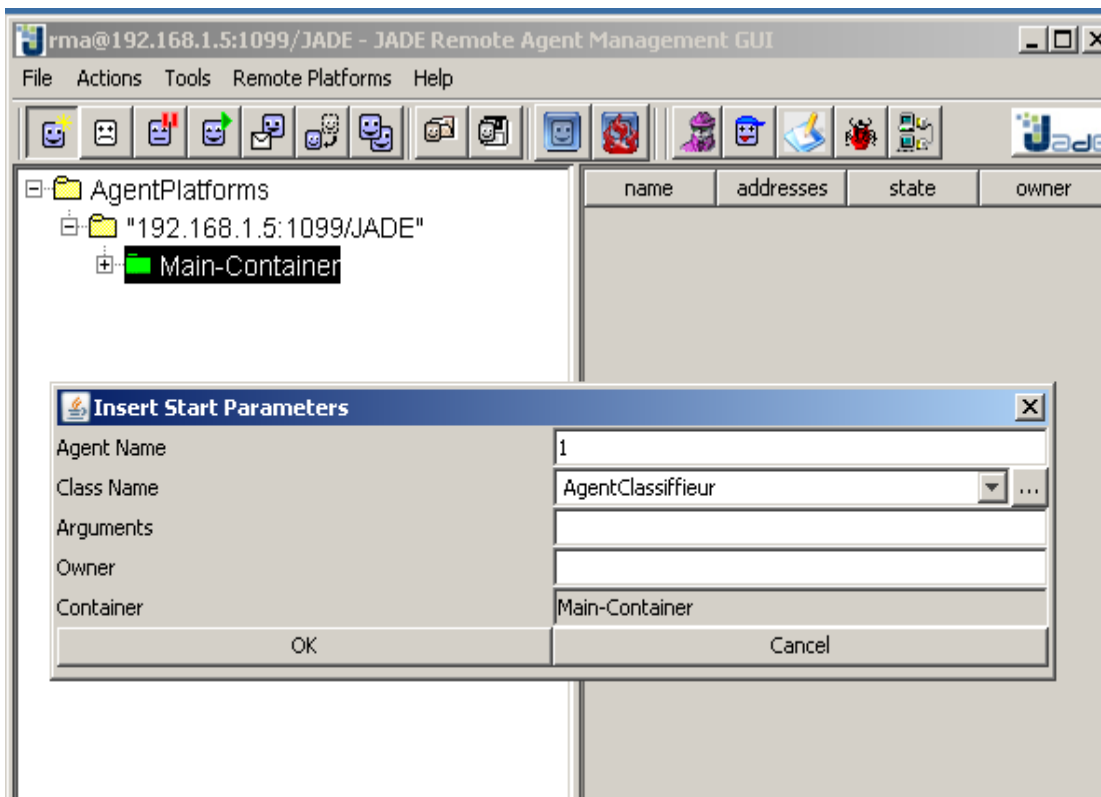


Fig IV.34 : plateforme Jade Création des 3 agents classifieurs

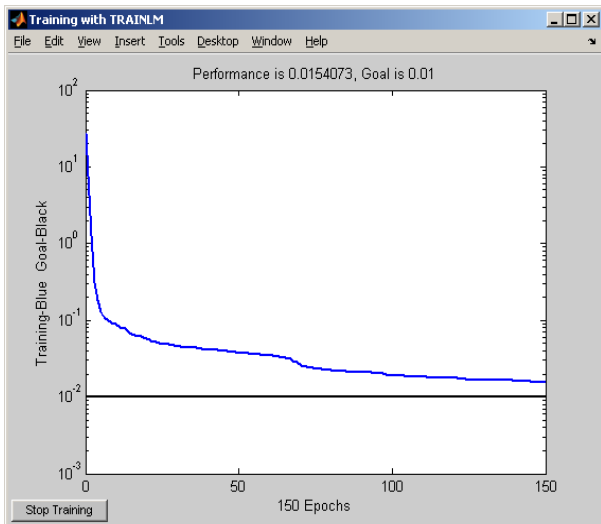


Figure IV.35 :Apprentissage du CNC

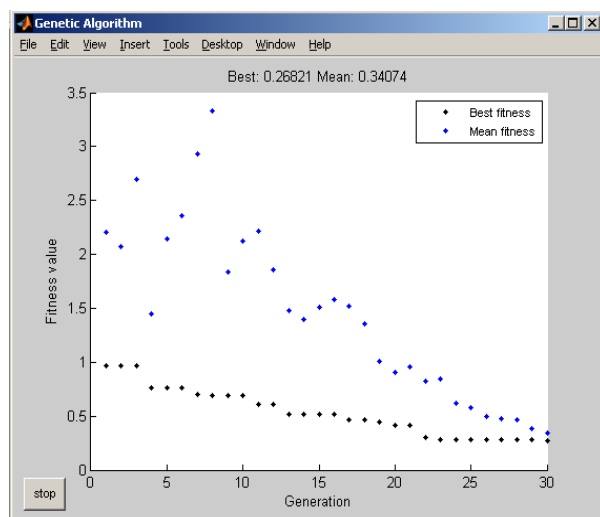


Figure IV.36 :Apprentissage du CNGP

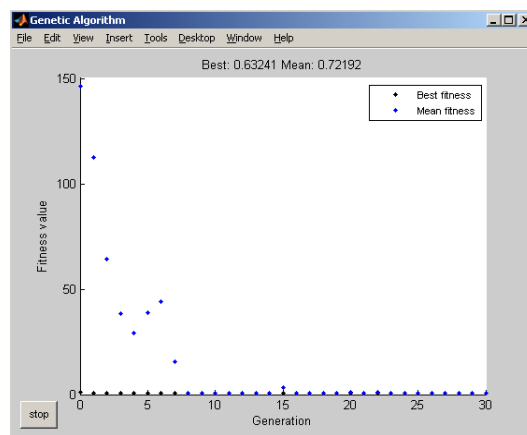


Figure IV.37 : Apprentissage du CNGS

➤ Retour sur interface jade pour la création de l'agent Contrôleur : Apparition des résultats (Tableau IV.4)

NB :une autre démarche de raccourcie est possible ,elle consiste à lancer au niveau de l'interface(figure IV.32) :

1-La plateforme

2-Agents Classifieurs

3-Agent Contrôl

## CONCLUSION GENERALE

Notre approche multi agents pour la reconnaissance du cancer vise à renforcer le diagnostic d'une manière distribué.

Pour cela nous avons utilisé la base de données universelle WBCD pour évaluer notre modèle.

Nous avons en premier lieu conçu des classifieurs mono agents en utilisant les réseaux de neurones multicouches, leurs hybridations avec les algorithmes génétiques

Nous avons évalué et testé les performances de chaque agent en terme de sensibilité (Se),Spécifité(Sp) et le taux de classification correcte (CC)

En dernier lieu nous avons utilisé un modèle SMA composé de quatre agents : les trois agents conçus auparavant, plus un agent contrôleur.  
Les résultats de classifications des données par l'agent ont été très prometteurs.

Notre approche peut se développer davantage par :

- L'élargissement de la base de données
- La multiplication du nombre d'agents (classifieurs)
- L'augmentation de la communication entre différents agents

Nous souhaiterons intégrer la notion d'interprétabilité chez les agents par l'hybridation de leurs techniques de classification avec l'approche floue.

### **Références Bibliographiques :**

- [1] Bartels PH, Weber JE,. Expert systems in histopathology, Introduction and overview. Anal. Quant. Cytol. Histology, Vol. 11, pp. 1-7, 1989.
- [2] Jelonek J, Krawiec K, Slowinski R, Neural networks and rough sets-comparison and combination for classification of histological pictures, In: Proc. of the first National Conference Neural Networks and their Applications, Kule, Czqstochawa, pp. 268-73, 1994.
- [3] Sushmita M. Fuzzy MLP based expert system for medical diagnosis, Fuzzy Sets and Systems, Vol. 65, pp. 285-96, 1994.
- [4] Verma, B, Novel network architecture and learning algorithm for the classification of mass abnormalities in digitized mammograms, In: Artificial Intelligence in Medicine, pp. 4267-79. 2008.
- [5] Cheng S.N.C, Chan H.P, He1vie M.A, Goodsitt M.M, Adler D.D, St. Clair D.C, Classification of mass and non-mass regions on mammograms using artificial neural networks, Imaging Sei, Technol, pp. 38, 1994.
- [6] Zhangn P, Verma B, Kumar K, A neural-genetic algorithm for feature selection and breast abnormality classification in digital mammography,2004.
- [7] Sekkal M, Chikh Ma, Neuro – Genetic Approach To Classification Of Cardiac Arrhythmias, Research student, Laboratory of Biomedical Engineering ,Department of Biomedical Electronics, UABB Tlemcen, Algérie,2011.
- [8] Iantovics 2007, 2008, 2009, 2010, IEEE, Springer.

[9] Mangasarian O. L, Wolberg W. H, Cancer diagnosis via linear programming, SIAM News, Vol. 23, pp 1- 18, 1991.

[10] Shibakali G, Shiladitya P, IEEE, 2009.

[11] Fougères A.J, Une architecture cognitive d'agents communicants dans des systèmes d'information complexes. Laboratoire M3M, Université de Technologie de Belfort-Montbéliard, France ,2008.

[12]Asoke Hong .G, Nandi K, Breast cancer diagnosis using genetic programming generated feature , Elsevier Pattern recognition, Vol 39, pp 980-987, 2006.

[13] Sekkal M, Chikh Ma ,Settoui N, Evolving neural networks using a genetic algorithm for heartbeat Classification, Journal of Medical Engineering & Technology,Biomedical Engineering Laboratory, pp 1-9 , Tlemcen, Algeria ,2011.