



République Algérienne Démocratique et Populaire

Université Abou Bakr Belkaid- Tlemcen

Faculté des Sciences

Département d'Informatique

Mémoire de fin d'études

pour l'obtention du diplôme de Master en Informatique

*Option: Modèle Intelligent et Décision(M.I.D)*

*Thème*

**Classifieur Neuro-Flou  
Évolutionnaire du Cancer du Sein**

**Réalisé par :**

- M<sup>r</sup> Mostafa EL HABIB DAHO

*Présenté le 02 Juillet 2011 devant la commission composé de MM.*

*Président :* - Mr BENAMAR A.

*Encadreur :* - Mr CHIKH M. A.

*Examineurs :* - Mr BENZIAN Y.

- Mr BENAZOUZ M.

- Mr LEHSAINI M.

- Mr SMAHI I.

- Mr HADJILA F.

Année universitaire : 2010-2011

## REMERCIEMENTS

AVANT TOUT, JE REMERCIE DIEU LE TRÈS HAUT QUI M'A DONNÉ LE COURAGE ET LA VOLONTÉ DE RÉALISER CE MODESTE TRAVAIL.

Mes remerciements s'adresse à MR CHIKH M.A, MAÎTRE DE CONFÉRENCES À L'UNIVERSITÉ DE TLEMCEM, qui a encadré ce travail. Sa confiance ne m'a jamais fait défaut, Sa gentillesse, ses connaissances et sa pertinence m'ont énormément facilité la tâche. En dépit d'un emploi du temps fort chargé, j'ai conscience des efforts qu'il a du fournir pour se rendre disponible. Et je lui suis extrêmement reconnaissant. Je le remercie aussi de m'avoir accueilli dans son équipe de recherche CREDOM du laboratoire G.B.M qui m'a permet d'enrichir mes connaissances.

Qu'il me soit également permis de remercier chaleureusement les membres de la commission MID pour l'honneur qu'ils m'ont fait d'avoir expertisé ce travail. Mes vifs remerciements à tous mes enseignants qui m'ont fait apprécier et aimer ma spécialité tout au long de ma formation. Qu'ils reçoivent ici tout ma gratitude.

Enfin, je tiens à exprimer mes remerciements à tous ceux qui ont participé au travail de recherche présenté dans ce document, spécialement Melle DIB Ibtissem, SETTOUTI Nesma et BELOUFA Fayçal, ainsi qu'à tous mes collègues et amis du laboratoire GBM et de la promo.

## DÉDICACE

À mes Chers Parents pour tant d'innombrables sacrifices dans ma formation humaine et intellectuelle,

À ma petite famille: Lila, Mounia, Meryem, Radia et Mohammed,

À tous mes cousins et cousines, oncles et tantes,

À mes amis,

À tous ceux et celles qui me sont chers.

Je dédie ce travail.

Mostafa.

# Table des matières

Introduction Générale	1
<b>1 Le Cancer du Sein</b>	<b>9</b>
1.1 Introduction	9
1.2 Anatomie du sein	10
1.3 Le cancer du sein	10
1.3.1 Définition	10
1.3.2 Les symptômes	11
1.3.3 Épidémiologie	11
1.3.4 Les facteurs de risques	12
1.3.5 Différents types de cancer du sein	12
1.3.6 Classification des cancers du sein	13
1.3.7 Différentes détection du cancer	13
1.3.8 Traitement du cancer du sein	15
1.4 Conclusion	16
<b>2 Revue de la Littérature</b>	<b>17</b>
2.1 Introduction	17
2.2 Approches probabilistes et approches statistiques	17
2.3 Approches déductives et symboliques	20
2.4 Approches évolutionnaires	22
2.5 Approches connexionnistes	26
2.6 Contribution et Conclusion	31
<b>3 Notions sur les Techniques de l'Intelligence Artificielle</b>	<b>32</b>
<b>I Les Réseaux de Neurones</b>	<b>33</b>
3.1 Introduction	34
3.2 Les réseaux de neurones	34
3.3 Architecture des Réseaux de Neurones	34
3.3.1 Réseaux de neurones non bouclés (feed-forward)	35
3.3.2 Réseaux de neurones bouclés (récurrents)	36
3.3.3 Le choix de la bonne architecture	37
3.3.4 Domaine d'utilisation des réseaux de neurones	37
3.4 L'Apprentissage	37
3.4.1 L'apprentissage non supervisé	38
3.4.2 L'apprentissage supervisé	38
3.4.3 L'apprentissage semi-supervisé (par renforcement)	39

3.4.4	L'apprentissage des perceptrons multicouches . . . . .	39
3.4.5	Sous-Apprentissage, généralisation et sur-apprentissage . . . . .	40
3.5	Les étapes de mise en oeuvre d'un PMC . . . . .	42
3.6	Conclusion . . . . .	42
<b>II</b>	<b>La Logique Floue</b>	<b>43</b>
3.7	Introduction . . . . .	44
3.8	Etat de l'art . . . . .	44
3.9	Notion d'ensemble et de sous ensemble flou . . . . .	44
3.9.1	Appartenance graduelle à un ensemble . . . . .	45
3.10	Les variables linguistiques . . . . .	45
3.11	Description des variables linguistiques par les fonctions d'appartenances . . . . .	45
3.12	Règles linguistiques . . . . .	46
3.13	Extension des opérations usuelles . . . . .	46
3.14	Structure générale d'un système flou . . . . .	47
3.15	Principe de fonctionnement du système flou . . . . .	47
3.15.1	Fuzzification ou quantification floue . . . . .	47
3.15.2	L'inférence floue . . . . .	48
3.15.3	Défuzzification . . . . .	49
3.16	Conclusion . . . . .	50
<b>III</b>	<b>Les Systèmes Neuro-Flous</b>	<b>51</b>
3.17	Introduction . . . . .	52
3.18	ANFIS : <i>Adaptative Neuro-Fuzzy Inference System</i> . . . . .	52
3.19	NEFCON : <i>Neuro-Fuzzy Control</i> . . . . .	53
3.20	FALCON : <i>Fuzzy Adaptive Learning Control Network</i> . . . . .	53
3.21	FUN : <i>Fuzzy Net</i> . . . . .	54
3.22	NEFPROX : <i>Neuro-Fuzzy Function Approximation</i> . . . . .	55
3.23	NEFCLASS : <i>Neuro-Fuzzy Classification</i> . . . . .	55
3.23.1	La structure du NEFCLASS . . . . .	56
3.23.2	L'apprentissage de la base de règles ( <i>structurel</i> ) . . . . .	57
3.23.3	L'apprentissage des sous-ensembles flous ( <i>paramétrique</i> ) . . . . .	60
<b>IV</b>	<b>Optimisation par Essaim Particulaires</b>	<b>62</b>
3.24	Introduction . . . . .	63
3.25	Intérêt de l'algorithme PSO . . . . .	63
3.26	Les éléments de la P.S.O. . . . .	64
3.27	Notion de voisinage . . . . .	64
3.27.1	Les voisinages géographiques . . . . .	64
3.27.2	Les voisinages sociaux . . . . .	65
3.28	Coefficients de confiance . . . . .	66
3.29	Vitesse maximale . . . . .	66
3.30	Facteur d'inertie . . . . .	66
3.31	Initialisation de l'essaim . . . . .	66
3.32	Critères d'arrêt . . . . .	66
3.33	La P.S.O. pour l'optimisation continue . . . . .	67

3.33.1 Principe fondamental . . . . .	67
3.34 Conclusion . . . . .	68
<b>4 Expérimentations, Résultats et Interprétations</b>	<b>69</b>
4.1 Introduction . . . . .	69
4.2 Description de la base de données . . . . .	70
4.2.1 Historique de la base de données . . . . .	70
4.2.2 Les descripteurs . . . . .	70
4.3 La structure utilisée pour l'optimisation . . . . .	70
4.3.1 Configuration de la méthode . . . . .	71
4.4 La procédure de classification . . . . .	73
4.4.1 Phase de test (d'évaluation) . . . . .	73
4.5 Les résultats du Nefclass . . . . .	75
4.5.1 Interprétation des résultats . . . . .	75
4.6 Les résultats de l'Opt-Nefclass . . . . .	77
4.6.1 Interprétation des résultats . . . . .	78
4.7 Etude comparative . . . . .	79
4.8 Conclusion . . . . .	79
<b>Conclusion Générale et Perspectives</b>	<b>80</b>
<b>Références Bibliographiques</b>	<b>82</b>
<b>Annexes</b>	<b>86</b>
<b>A Modélisation de l'application</b>	<b>87</b>
A.1 Définition des besoins des utilisateurs . . . . .	88
A.2 Spécification des cas d'utilisations . . . . .	88
A.3 Les diagrammes de classes . . . . .	89
A.4 Environnement de développement . . . . .	92
<b>B Manuel d'utilisation</b>	<b>93</b>
B.1 Extraction des règles . . . . .	93
B.2 Apprentissage des sous-ensembles flous (Nefclass) . . . . .	94
B.3 Optimisation (Opt-Nefclass) . . . . .	94
B.4 Classification . . . . .	95

# Table des figures

1.1	Incidence annuelle pour 100000 habitants. . . . .	9
1.2	Anatomie d'un sein et Glandes Mammaires . . . . .	10
1.3	Résultat d'un Électromammogramme . . . . .	15
2.1	Architecture du système expert multiple [10] . . . . .	21
2.2	Architecture du système d'aide au diagnostic [8] . . . . .	26
2.3	Organigramme de l'approche proposée [23] . . . . .	27
2.4	Architecture du réseau neuro-flou [24] . . . . .	29
3.1	Réseau de neurones multi-couhes . . . . .	36
3.2	Réseaux de neurones Récurrent . . . . .	37
3.3	Différence des capacités de généralisation entre le PMC (a) et les RFR (b). . . . .	41
3.4	Complexité dun réseau de neurones . . . . .	42
3.5	Variables et Valeurs Linguistiques . . . . .	45
3.6	Structure dun système flou . . . . .	47
3.7	Les formes de la fonction d'appartenance . . . . .	48
3.8	Le modèle ANFIS . . . . .	52
3.9	Architecture de NEFCON . . . . .	53
3.10	Architecture de FALCON . . . . .	54
3.11	Architecture du modèle FUN montant . . . . .	54
3.12	Architecture du modèle NEFPROX . . . . .	55
3.13	Une vue de réseau de neurones d'un classifieur neuro-flou . . . . .	56
3.14	Larchitecture du modèle NEFCLASS . . . . .	57
3.15	Classification avec NEFCLASS après l'apprentissage des règles . . . . .	59
3.16	Adaptation dun sous-ensemble . . . . .	61
3.17	Classification après l'apprentissage des sous-ensembles flous . . . . .	61
3.18	Volée d'Anser en formation en V . . . . .	63
3.19	Un voisinage géographique [52] . . . . .	65
3.20	Un voisinage en cercle [52] . . . . .	65
3.21	Organigramme de la méthode des essaims particuliers . . . . .	68
4.1	Les trois points du sous-ensemble flou . . . . .	72
4.2	Création d'une population . . . . .	72
4.3	SEF avant et après apprentissage . . . . .	73
4.4	Histogramme des taux derreurs . . . . .	75
4.5	Histogramme des taux derreurs . . . . .	78
A.1	Schéma global du processus de modélisation UML . . . . .	87
A.2	Le diagramme des cas d'utilisations. . . . .	88
A.3	Diagramme de classe du cas d'utilisation Lancer apprentissage . . . . .	89

A.4	Diagramme de classe du cas d'utilisation Lancer Test . . . . .	90
A.5	Diagramme de classe global . . . . .	91
B.1	Génération des règles . . . . .	93
B.2	Apprentissage du réseau . . . . .	94
B.3	Angle de l'optimisation . . . . .	95
B.4	Configuration du PSO . . . . .	95
B.5	Classification d'un exemple . . . . .	96
B.6	Classification d'une base . . . . .	96



# Liste des tableaux

1.1	Taux de détection du cancer par rapport à l'âge . . . . .	11
1.2	Tableau T.N.M . . . . .	13
2.1	Taux d'erreur de classification des SVM (base de données breast cancer) .	18
2.2	Taux de reconnaissance des trois algorithmes conventionnels et de l'ap- proche MBBC proposée par Madden . . . . .	19
2.3	Taux de reconnaissance de chaque sous-système et du système global . . .	21
2.4	Classification basée sur les algorithmes d'apprentissage et de sélection de variables . . . . .	23
2.5	Taux de reconnaissance de l'algorithme proposé (Ant_Miner3) par rapport à l'Ant_Miner conventionnel . . . . .	24
2.6	Résultats de classification de la base de données Brest Cancer . . . . .	25
2.7	Taux de classification du FNN . . . . .	28
2.8	Performance de l'approche MPNN . . . . .	30
2.9	Performance de l'approche HFNN . . . . .	30
3.1	Vue comparative entre les RNA et les SIF . . . . .	52
4.1	Attributs de la base de diagnostic . . . . .	70
4.2	Modélisation d'une particule . . . . .	72
4.3	les résultats de la classification avec Nefclass . . . . .	75
4.4	Résultats de la 10-Fold Cross Validation . . . . .	77
4.5	Résultats de la classification avec Opt-Nefclass . . . . .	77
4.6	Résultats de la 10-Fold Cross Validation . . . . .	79
4.7	Etude comparative . . . . .	79

# Introduction Générale

La médecine (du latin *medicus*, " qui guérit ") est la science qui étudie l'organisation du corps humain (anatomie), son fonctionnement normal (physiologie) et cherche à restaurer la santé par le traitement (thérapie) et la prévention (prophylaxie) des pathologies. La médecine contemporaine utilise les soins de santé, la recherche et les technologies biomédicales pour diagnostiquer et traiter les blessures et les maladies, habituellement à travers la prescription de médicaments, la chirurgie ou d'autres formes de thérapies. Depuis plusieurs décennies, le soulagement de la souffrance s'est également imposé comme un objectif médical à travers des solutions chimiques mais également par la relation médecin-patient.

Parmi les maladies tuantes que la médecine moderne défait, on trouve les cancers. La lutte contre les cancers commence par comprendre les mécanismes qui mènent à cette maladie afin de trouver des meilleures solutions pour la prévenir et la traiter.

Le cancer du sein est le cancer le plus diagnostiqué chez les femmes à travers le monde. Il apparaît le plus souvent après la ménopause. Il n'est cependant pas le plus mortel : le

cancer du poumon demeure la première cause de décès chez les femmes. Une femme sur 9 risque d'être atteinte d'un cancer du sein. Les hommes aussi peuvent en être touchés, mais rarement. Moins de 1 % des cas de cancer du sein touchent les hommes. Le nombre de personnes atteintes a progressé légèrement, mais régulièrement, au cours des 3 dernières décennies. Par contre, le taux de mortalité a continuellement diminué au cours de la même période, grâce aux progrès réalisés en matière de dépistage, de diagnostic et de traitement.

Le progrès quotidien de la médecine, les nouvelles maladies découvertes et les nouveaux traitements introduits génèrent des diagnostics très variés.

Un diagnostic médical est le résultat du raisonnement d'un médecin avec une décision très souvent prise à partir d'informations incertaines et/ou incomplètes qui sont le résultat d'un interrogatoire et d'un examen clinique. Le médecin puise dans ses connaissances pour trouver la maladie en cause.

Il existe des difficultés qui peuvent fausser un diagnostic médical et qui peuvent être liées aux praticiens et/ou aux malades. Parmi les problèmes qui peuvent influencer le bon diagnostic du médecin nous citons :

- Le médecin peut être fatigué ou avoir un souci diminuant ses capacités de concentration.
- Le cerveau humain n'est pas fait pour garder un trop grand nombre de données, présentes à l'esprit en même temps.
- De nombreuses études sur le raisonnement médical montrent que, généralement, le médecin élabore un petit nombre d'hypothèses très tôt en phase d'examen, sur un faible nombre de signes caractéristiques avec un bon degré de fiabilité.

A ce titre plusieurs travaux sont effectués afin de développer des outils d'aide au diagnostic et de classification des maladies. De nombreuses techniques d'intelligence artificielle

ont été appliquées pour essayer de résoudre cette problématique.

L'introduction de l'intelligence artificielle en médecine a permis non seulement de faciliter la tâche aux experts du domaine mais aussi d'automatiser la prise de décision. L'utilisation des ordinateurs pour aider les professionnels de la santé dans leurs activités a été étudiée depuis les années cinquante.

La médecine moderne a besoin d'une aide informatisée pour le dépistage, le diagnostic et la classification de certaines maladies dans un temps très réduit d'où la nécessité d'un système de classification.

Les techniques telles que les réseaux de neurones, la logique floue et d'autres architectures hybrides qui intègrent ces outils, sont les plus couramment utilisées dans les systèmes déjà développés par de nombreuses équipes de recherches.

Dans ce mémoire nous avons développé le système de classification neuro-flou " NEF-CLASS", et en suite nous l'avons optimisé avec les essais particuliers. Notre travail est reparti sur quatre chapitres organisés comme suite :

- Chapitre I : nous donnons un bref aperçu sur la maladie du cancer du sein.
- Chapitre II : Un état de l'art sur les travaux les plus récents et les techniques utilisées pour le diagnostic du cancer du sein.
- Chapitre III : Est composé de quatre parties à savoir l'essentiel des réseaux de neurones, la logique floue, l'hybridation des deux méthodes et enfin l'optimisation par essaim particulier.
- Chapitre IV : Dans ce dernier chapitre nous présentons nos résultats et une étude comparative avec les travaux existants dans la littérature.
- Annexe A : Contient la modélisation de notre application.
- Annexe B : Présente le manuel d'utilisation.

Notre application porte le nom : OPT-NEFCLASS (OPTimized NEuro-Fuzzy Classification)

# Chapitre 1

## Le Cancer du Sein

### 1.1 Introduction

La maladie du cancer a coûté 312 milliards de dollars en 2010. C'est, en tout cas, ce qu'a révélé une étude présentée lors de la Conférence internationale sur le cancer, organisée par la Fondation Lance Armstrong, à Dubai [1]. Selon cette même étude il devrait y avoir pour cette année 12,9 millions de nouveaux cas de cancer à travers le monde. Ce chiffre jugé alarmant pourrait fortement augmenter pour atteindre 16,8 millions en 2020 et 27 millions en 2030. Le cancer est devenu l'une des causes essentielles des décès dans le monde. Les statistiques montrent que la majorité des personnes atteintes de cette maladie vivent dans les pays en voie de développement dont l'Algérie [2]. Environ 35.000 nouveaux cas de cancer sont diagnostiqués chaque année en Algérie avec une augmentation de plus de 50% du nombre de cas depuis une décennie.

Les cancers les plus fréquents ne sont pas les mêmes chez l'homme et chez la femme comme le montre la figure suivant :

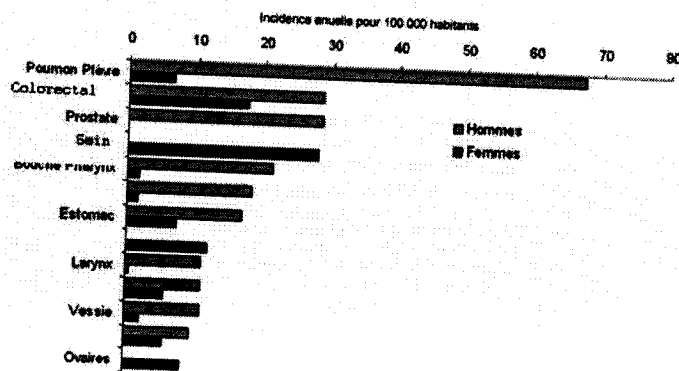


FIGURE 1.1 – Incidence annuelle pour 100000 habitants.

## 1.2 Anatomie du sein

Le sein est un organe de nature glandulaire dont la fonction est de produire du lait. Chaque sein se divise en 15 à 20 secteurs, appelés lobes. Chacun de ces lobes se divise en nombreux lobules, plus petits, qui s'achèvent en douzaines de minuscules bulbes sécrétant le lait. Les lobes, lobules et bulbes sont reliés entre eux par les canaux galactophores. Ceux-ci aboutissent au mamelon, situé au centre d'une zone pigmentée, l'aréole. La plus grande partie du sein est constituée de tissu adipeux qui comble l'espace situé entre les différentes structures du sein. Des muscles recouvrant les côtes sous-tendent le sein, mais n'en font pas partie.

Les figures suivantes illustrent l'anatomie d'un sein.

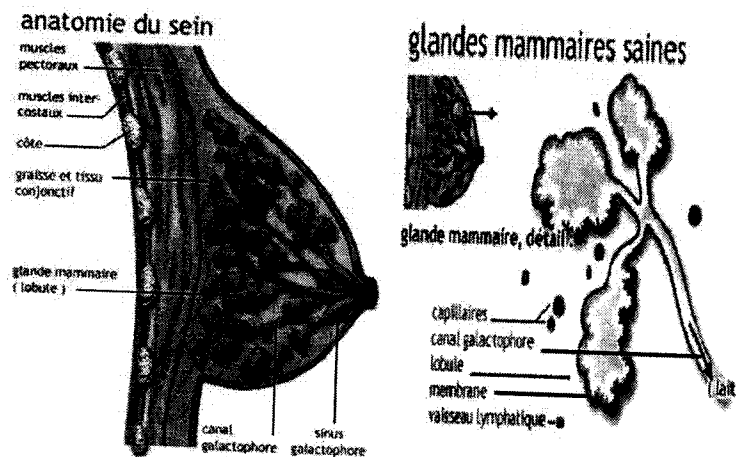


FIGURE 1.2 – Anatomie d'un sein et Glandes Mammaires

Chez l'homme, les seins demeurent immatures et ne contiennent pas de glandes mammaires. Ils n'ont aucun rôle.

## 1.3 Le cancer du sein

### 1.3.1 Définition

Le cancer est une maladie qui a pour mécanisme une prolifération cellulaire anarchique, incontrôlée et incessante.

Le terme cancer recouvre un vaste ensemble de maladies, cataloguées selon les cellules et les tissus à partir duquel les cancers se forment. La tumeur développée dans un organe (tumeur primitive) va se greffer à distance sur d'autres organes (cerveau, poumon, foie etc.) en passant par les voies lymphatiques ou sanguines. Ces tumeurs secondaires, qui reproduisent la structure de la tumeur mère, s'appellent des métastases.

Le cancer du sein est une tumeur maligne dans le sein touchant le plus souvent les femmes. Les hommes aussi peuvent être atteints de ce cancer, mais ceci est beaucoup plus rare. Le cancer touche la glande mammaire de la femme, principalement sous la forme d'un adénocarcinome (cancer du tissu glandulaire) parfois sous la forme d'un sarcome (cancer du tissu conjonctif).

### 1.3.2 Les symptômes

L'induration ou la nodosité palpable est comme "fixée" à la peau, constitue généralement le symptôme le plus fréquent du cancer du sein. Elle est parfois accompagnée de douleurs ou d'une sensation de tension ou de lourdeur. Les modifications palpables du sein ne sont pas toutes des cancers mais doivent toujours être examinées par un médecin, seule personne apte à poser un diagnostic.

D'autres signes peuvent apparaître, par exemple une rougeur ou une rétraction de la peau, une lésion du mamelon ou (rarement) un écoulement sanguinolent.

La majorité des femmes atteintes d'un cancer du sein ne présentent aucun signe d'altération de l'état général et se sentent "en bonne santé" au moment du diagnostic.

Au stade avancé, les symptômes du cancer du sein sont très variables et dépendent principalement de la localisation et de l'étendue de la tumeur.

Ils peuvent se manifester sous la forme de gros ganglions dans le creux axillaire du côté atteint, de douleurs dorsales (métastases osseuses), d'une toux d'irritation accompagnée d'une gêne respiratoire (métastases pulmonaires), d'une perte de l'appétit, d'une perte pondérale, d'un ictère (jaunisse, métastases hépatiques), etc.

Tous ces symptômes doivent faire immédiatement consulter un médecin, même s'ils ne sont pas forcément le signe d'un cancer du sein.

### 1.3.3 Épidémiologie

Le cancer du sein est de loin le cancer le plus fréquent chez la femme (il en représente 1/4). Il touchera une femme sur dix, soit 7.000 nouveaux cas en Algérie tandis qu'en France se sera une femme sur treize, soit l'apparition chaque année de 25.000 nouveaux cas.

Depuis les années 50, nous constatons une augmentation régulière de l'incidence (nombre de nouveau cas recenser pour 100.000 personnes et par ans). Cette augmentation est de l'ordre de 2% par an.

Le cancer du sein est la première cause de décès chez la femme dans les pays industriels, soit 8.000 décès par an en France. Alors que la recherche avance, l'incidence augmente et les victimes sont de plus en plus jeunes. Dans 30% des cas, elles ont moins de 50 ans, certaines n'ont qu'une vingtaine d'année, ce qui était extrêmement rare il y a quelques années. L'âge moyen du diagnostic est d'environ 50 ans.

Le cancer du sein est rare avant 30 ans et son incidence augmente en suite jusqu'à l'âge de 75 ans. Nonante-cinq pourcents des cancers du sein surviennent après 40 ans. Le tableau suivant nous donne plus de détaille :

Taux	Tranche d'âge
±5%	Moins de 40 ans
±17%	entre 40 et 49 ans
±22%	entre 50 et 59 ans
±22%	entre 60 et 69 ans
±22%	entre 70 et 79 ans
±12%	au-delà de 80 ans

TABLE 1.1 – Taux de détection du cancer par rapport à l'âge

### 1.3.4 Les facteurs de risques

Il existe un certain nombre de facteurs de risques : Les antécédents familiaux de cancer du sein. Le risque est multiplié par 14 s'il y a eu un cancer du sein dans les deux lignées parentales. C'est dire l'intérêt d'une surveillance clinique et para clinique importante chez ces femmes à haut risque. Le cancer du sein est un cancer hormonodépendant ; des récepteurs strogéniques et/ou progestéroniques sont souvent présents. Cela explique la particularité fréquente des cancers du sein chez les femmes ayant présenté :

- Une puberté précoce.
- Une ménopause tardive (plus longue exposition aux oestrogènes).
- Un traitement oestrogénique insuffisant compensé par les progestatifs pilules séquentielles par exemple (mais il n'y pas encore été prouvé qu'elle avait un rôle dans le cancer).
- L'âge est très important. A partir de 50 ans, le risque de développer un cancer du sein augmente sensiblement.
- Consommation d'alcool et obésité peuvent augmenter le risque de développer un cancer du sein.

### 1.3.5 Différents types de cancer du sein

Il existe différentes formes de cancer du sein, chacune ayant une évolution qui lui est propre. On peut les diviser en deux grandes catégories : le cancer non invasif ou in situ, qui touche uniquement les canaux galactophores, et le cancer invasif ou infiltrant, qui se propage dans le tissu gras du sein. (Attention, le fait que le cancer soit invasif ne signifie pas qu'il forme des métastases, mais qu'il envahit les tissus autour des canaux).

#### Cancer non invasif

- Carcinome canalaire in situ : C'est le type le plus fréquent de cancer du sein non invasif chez la femme. C'est une forme de cancer très " jeune ", au tout début de sa formation. Comme son nom l'indique, il se forme à l'intérieur des canaux de lactation du sein. Il ne se dissémine pas. On diagnostique beaucoup plus fréquemment ce type de cancer depuis l'utilisation plus répandue de la mammographie. Le traitement de ce cancer mène à la guérison dans presque tous les cas. Sans traitement, il poursuit sa croissance et peut alors devenir " infiltrant ", donc se propager à l'extérieur des canaux galactophores.

#### Cancers invasifs

- Carcinome canalaire : Il se forme dans les canaux galactophores. Les cellules cancéreuses traversent la paroi des canaux.
- Carcinome lobulaire : Les cellules cancéreuses apparaissent dans les lobules, puis traversent la paroi des lobules et se disséminent dans les tissus environnants.
- Carcinome inflammatoire : Un cancer rare qui se caractérise principalement par un sein qui peut devenir rouge, enflé et chaud. La peau du sein peut aussi prendre l'aspect d'une peau d'orange. Ce type de cancer progresse plus rapidement et est plus difficile à traiter.
- Autres carcinomes (médullaires, collodes ou mucineux, tubulaires, papillaires) : Ces types de cancer du sein sont plus rares. Les principales différences entre ces types de cancer reposent sur le type de cellules touchées.
- Maladie de Paget : Un cancer rare qui se manifeste par une petite plaie au mamelon qui ne guérit pas.

**Remarque 1** *Le carcinome lobulaire in situ (ou néoplasie lobulaire) n'est pas cancéreux. Il est cependant un facteur de risque important de cancer du sein qui exige un suivi serré.*

### 1.3.6 Classification des cancers du sein

Il existe différentes possibilités de classification des tumeurs : selon l'organe, selon le tissu dont elles sont issues ou bien selon leur extension. La classification internationale se fait selon, le système T.N.M. dont voici le tableau :

Stade 0	Cancer in situ (localisé)	/
Stade 1	Petite tumeur sans envahissement ganglionnaire	T1 N0 M0
Stade 2	Tumeur plus volumineuse et envahissement ganglionnaire minime	T2 N1 M0
Stade 3	Tumeur développée au-delà de l'organe atteint avec envahissement ganglionnaire important	T3 N2 M0
Stade 4	Tumeur extensive avec grosses adénopathies ou métastases	T4 N3 M1

TABLE 1.2 – Tableau T.N.M

Lorsque le diagnostic du cancer est établi, on devra pratiquer un bilan complet à l'issu duquel il sera possible de classer ce cancer.

- T = tumeur (tumor).
- N = atteinte ganglionnaire (nodes).
- M = métastase (metastasis).

### 1.3.7 Différentes détection du cancer

#### Individuel (auto-évaluation)

##### - Examiner les seins

Se mettre face à un miroir, les bras le long du corps et observer ses seins : la symétrie, une modification du volume ou de la forme, un gonflement, un pli, une fossette ou une déformation du mamelon.

Observer aussi la peau : décoloration, rougeur, rétraction, " peau d'orange ". Refaire la même démarche avec les bras levés au-dessus de la tête, puis de profil, en ayant les bras le long du corps puis levés au-dessus de la tête.

##### - Palper les seins

Se mettre en position allongée, un oreiller sous l'épaule droite pour l'examen de sein droit, la main droite sous la tête. Avec la main gauche, les doigts à plat, exercer de légères pressions sur le sein droit.

Suivre un mouvement de rotation dans le sens des aiguilles d'une montre, en commençant dans la partie supérieure, c'est-à-dire à 12 heures. Rechercher d'éventuelles indurations, épaisissements ou grosseurs.

Effectuer plusieurs cercles en se déplaçant de 2,5 cm vers le mamelon à chaque nouveau cercle, de façon à examiner tout le sein. Ne pas oublier de palper le mamelon et aisselle.

Terminer l'examen en pressant le mamelon entre le pouce et l'index afin de détecter tout



écoulement clair ou sanguin. Suite à l'auto-examen, signaler toute anomalie au médecin. Le diagnostic d'une éventuelle tumeur sera précisé par une mammographie, mais souvent, il s'agit de tumeurs bénignes c'est-à-dire non-cancéreuses.

### **Examen diagnostique**

Souvent, la femme consulte son médecin parce qu'elle a une " boule au sein " appelée nodule. Ce nodule n'est pas nécessairement malin mais pour avoir la certitude il faut effectuer plusieurs examens.

#### **- Examen clinique complet**

Le médecin va procéder à une investigation détaillée des antécédents personnels et familiaux de la patiente. Il pourra également recourir à un ou plusieurs examens complémentaires.

#### **- La palpation**

Elle devra être pratiquée en position couchée et debout et explorer tout le sein sans oublier le prolongement axillaire. Cette palpation sera terminée par la recherche de ganglions axillaires. Elle s'effectuera la main ouverte posée à plat sur le sein en faisant rouler la glande sur les côtés.

### **La mammographie**

La radiothérapie des seins, permet de mieux juger de l'aspect d'une tumeur. Cet examen doit être pratiqué de façon bilatérale et comparative. La période idéale se situe dans la période post menstruelle. Les appareillages modernes permettent de réduire les quantités de Rx délivrés tout en donnant des clichés lisibles, d'où l'intérêt de les effectuer dans des centres spécialisés. Il faut noter que cette technique est moins performante chez la femme jeune (aux seins très denses) que chez la femme plus âgée. Des études récentes effectuées en Suède semblent apporter la preuve qu'un dépistage par la pratique qu'un dépistage systématique d'une mammographie après 45 ans permettait de nombreux diagnostics précoces et réduirait le pourcentage de mortalité liée à cette maladie. Par ailleurs, en Belgique, le dépistage gratuit bisannuel vient d'être instauré pour les femmes entre 50 et 69 ans.

### **L'échographie**

Cette méthode récente présente l'avantage de ne pas utiliser de Rx et peut donc être renouvelée sans danger. Cet examen apporte des renseignements précis dans le diagnostic des kystes du sein. Il permet d'obtenir des images plus intéressantes chez la femme jeune aux seins denses. C'est un examen complémentaire de la mammographie qui permet donc de préciser sur la malignité ou la bénignité de la tumeur et de guider la ponction d'une masse mammaire.

### **La ponction**

Cet examen consiste à ponctionner la tumeur à l'aide d'une fine aiguille. Les quelques cellules ramenées sont ensuite analysées au laboratoire à la recherche de cellules anormales. Elle se pratique sans anesthésie si la lésion est superficielle et sous anesthésie locale si elle est profonde. Elle intervient souvent après repérage par échographie ou scanner et se déroule sous contrôle échographique. La patiente doit être à jeun. Une fois la ponction réalisée, l'injection d'air ou de liquide radio opaque dans la cavité sert à préciser les contours du kyste, lors d'une mammographie.

## Biopsie

C'est le seul examen capable d'apporter la certitude absolue en matière de bénignité ou malignité. Elle sera pratiquée chaque fois que les examens précédents laissent planer un doute. Le plus souvent, on pratiquera un examen extemporané c'est-à-dire que la tumeur suspecte sera retirée au cours d'une intervention chirurgicale. Elle sera examinée en salle d'opération par un anatomopathologiste. Cet examen peut être effectué en dehors de l'intervention, lorsque l'on fait une ponction, l'on peut analyser la partie ponctionnée pour voir si la masse au niveau du sein est cancéreuse ou pas.

## La galactographie

Il s'agit d'une radio du sein après opacification d'un canal galactophore. Cette méthode est indiquée dans l'écoulement de sang par un seul port du mamelon.

## L'électromammogramme

Cette technique, ni irradiante, ni traumatisante, consiste à localiser les potentiels électriques anormaux. En mesurant les différentes zones saines et zones cancéreuses ; 90% des tumeurs sont ainsi décelables. Le seul inconvénient est que l'appareil signale une fois sur deux une anomalie qui n'est pas tumorale ; ce qui provoque souvent chez la femme une angoisse bien compréhensible mais inutile, toutefois, si la sensibilité et la fiabilité de l'examen s'améliorent, il pourrait devenir l'élément principal d'un premier dépistage.

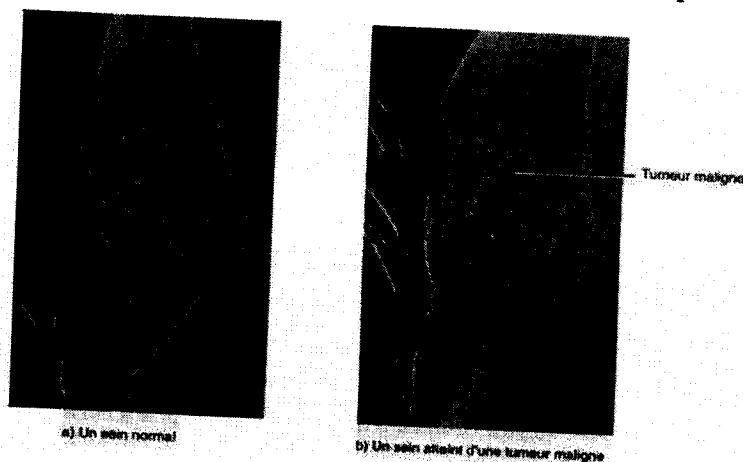


FIGURE 1.3 - Résultat d'un Électromammogramme

### 1.3.8 Traitement du cancer du sein

Il existe trois types de traitements du cancer du sein :

1. La radiothérapie.
2. La chimiothérapie.
3. L'hormonothérapie.

Il existe encore d'autres types de traitement qui sont encore en essais clinique, il s'agit de :

- La thérapie biologique qui consiste à utiliser le système immunologique de défense de l'organisme pour combattre le cancer.
- La transplantation de la moelle épinière.

### **La radiothérapie**

La radiothérapie est utilisée en association avec la chimiothérapie ou la chirurgie. Le principe de radiothérapie est l'utilisation de radiations ionisantes (photons ou électrons) qui agissent sur l'ADN cellulaire. Une différence de radiosensibilité entre les tissus sains et les tissus cancéreux fait qu'elle est moins toxique pour le tissu sain. En effet, le tissu sain se restaure alors que le tissu cancéreux qui se répare difficilement, est détruit.

Il existe deux types de radiothérapie :

- Radiothérapie externe : on irradie à distance la zone à traiter.
- Radiothérapie interne ou curiethérapie : on place directement dans le sein des aiguilles d'une substance radioactive (Iridium 192 ou Césium 137).

### **La chimiothérapie**

La chimiothérapie est un traitement général, c'est-à-dire que l'on administre des substances dans l'organisme soit par voie orale soit par voie intraveineuse ou encore par voie intramusculaire. De cette manière les substances médicamenteuses peuvent agir sur les cellules cancéreuses de tout l'organisme et pas uniquement au niveau du sein.

Ces médicaments agissent soit par blocage des mitoses, par destruction des cellules, ou par blocage du métabolisme des cellules, ce qui empêche celles-ci de se reproduire. Malheureusement, ils n'agissent pas en sélectionnant les cellules cancéreuses, et peuvent détruire des cellules saines, tout spécialement les cellules à reproduction rapide : cheveux, muqueuses, éléments figurés du sang.

### **L'hormonothérapie**

Si les tests montrent que les cellules cancéreuses ont des récepteurs d'oestrogènes et de progestérone, un traitement hormonal pourrait être donné. Le traitement hormonal est utilisé pour changer le mode d'action des hormones sur la croissance du cancer. Cela se fait soit en utilisant des médicaments ou en enlevant des organes qui fabriquent ces hormones comme les ovaires.

## **1.4 Conclusion**

La prévention vise à diminuer le risque du cancer du sein à court et à long terme. Il existe des consignes qui réduisent probablement le risque d'avoir cette maladie ; il est recommandé de cesser de fumer, avoir une alimentation pauvre en graisses. L'OMS en partie conseille de manger des fibres, des fruits et des légumes. Ces aliments ont un rôle protecteur. Il est également recommandé de pratiquer régulièrement du sport.

Aujourd'hui, le cancer du sein se guérit quand il est diagnostiqué tôt, et est considéré comme une maladie chronique quand il est diagnostiqué tardivement. C'est une maladie avec laquelle on peut vivre, travailler...

# Chapitre 2

## Revue de la Littérature

### 2.1 Introduction

Beaucoup de recherches liées à la détection du cancer du sein se sont multipliées durant la dernière décennie. Plusieurs de ces travaux se sont dirigés vers la détection de la présence de tissus cancéreux dans le sein et la classification de tumeurs. Les approches utilisées proviennent de plusieurs domaines : probabilités et statistiques, connexionnisme, ainsi que d'autres outils issus de l'intelligence artificielle et des sciences cognitives. À partir de là, une taxonomie des approches récentes de classification dans le cadre du dépistage du cancer du sein a été établie.

### 2.2 Approches probabilistes et approches statistiques

Dans cette première partie de la revue de littérature, nous nous intéressons aux approches probabilistes et statistiques employées en tant que classificateurs pour la détection du cancer du sein. Des méthodes statistiques et probabilistes reviennent dans la littérature, proposant souvent des versions améliorées des approches classiques telles que les réseaux bayésiens et la règle des  $k$  plus proches voisins.

Dans le travail de Subhash et al. en 2003 [3], les auteurs proposent une approche qui est basée sur une généralisation de la règle des  $k$  plus proches voisins pour la reconnaissance du cancer du sein. C'est une méthode qui concerne un classifieur non paramétrique mais dont la performance dépend des vecteurs de distributions moyennes et des matrices de covariance. Si, ces distributions sont de nature gaussienne, la performance de cette approche devient plus intéressante. L'approche a été implémentée et testée sur deux bases de données portant sur le cancer du sein, WDBC (Wisconsin Diagnosis Breast Cancer) et WBC (Wisconsin Breast Cancer).

La première base de données a été partitionnée et la classification a été effectuée sur chacune de ses partitions. Les résultats obtenus à travers les expérimentations ont été plus performants par rapport à ceux obtenus en utilisant la règle conventionnelle des  $k$  plus proches voisins.

À travers la classification des différentes partitions, les auteurs dans [3] montrent que la méthode utilisée fait preuve de robustesse et offre une meilleure performance que la règle classique des  $k$  plus proches voisins. Le meilleur taux de reconnaissance obtenu est de 98.1% pour la base de données WDBC et de 97% avec WBC.

On note que la méthode proposée dans [3] représente un classifieur à la fois simple à implémenter et d'application générale car non paramétrique, mais dont la performance

dépend des vecteurs de distributions moyennes et des matrices de covariance. Par ailleurs, les résultats de classification obtenus sont intéressants dans la mesure où il s'agit d'une classification binaire (classe bénigne/classe maligne). Or, dans les problèmes réels relatifs au cancer du sein, on a souvent affaire à un plus grand nombre de classes de tumeurs.

Dans Fei et al. en 2003 [4], les auteurs proposent une méthode de séparation à vastes marges pour effectuer la classification dans le cadre du dépistage du cancer du sein, soit, une machine à vecteurs de support (SVM pour Support Vector Machine). L'approche utilise des mises à jour multiplicatives "multiplicative updates" pour résoudre le problème de programmation quadratique non-négative dans les machines à vecteurs de support. Pour la mise en oeuvre des SVMs, les auteurs ont pris en considération un cas simple où, en projetant les données d'entrée dans un espace à grande dimension, les classes deviennent linéairement séparables et l'hyperplan de séparation est contraint de traverser l'origine.

Kernel	Polynomial		Radial		
	$k = 4$	$k = 6$	$\sigma = 0.3$	$\sigma = 1.0$	$\sigma = 3.0$
Data					
Sonar	9.6%	9.6%	7.6%	6.7%	10.6%
Breast cancer	5.1%	3.6%	4.4%	4.4%	4.4%

TABLE 2.1 – Taux d'erreur de classification des SVM (base de données breast cancer)

Les résultats de classification sont illustrés dans la table (2.1). Les expérimentations ont été faites avec une fonction noyau polynomiale et une fonction à base radiale. Les auteurs ont attribué à la fonction polynomiale les degrés  $k = 4$  et  $k = 6$ , et pour la fonction à base radiale des variances  $\sigma$  de 1.0 et de 3.0. Les taux d'erreur de la classification obtenus pour la base de données "cancer du sein" varient entre 3.6% et 5.1%. Il est à noter que l'approche proposée dans [4] aide à pallier aux problèmes relatifs à la programmation quadratique non-négative dans les machines à vecteurs de support. Selon les taux de rappel obtenus à la classification, la performance des SVMs en utilisant cette approche s'avère intéressante. Toutefois, pour la conception de la méthode, les auteurs ont pris en considération un cas simple où, dans un espace à grande dimension, les classes sont linéairement séparables et l'hyperplan séparateur est contraint de traverser l'origine. Cela entraîne une simplification du problème de départ et permet de remettre en question l'applicabilité de l'approche proposée dans des problèmes réels de cas de cancer du sein. Les classes de tumeurs ne sont effectivement pas toujours linéairement séparables. La machine à vecteur de support dans [4] constitue un apprentissage hors-ligne. D'autres travaux comprennent des classifieurs à vastes marges, tel celui proposé dans Baback et al. en 2002 [5], où il est question d'une approche qui utilise des fonctions "noyaux dyadiques" stimulées, ceci étant dans un but de maximiser la marge en offrant la possibilité d'un apprentissage en ligne.

Dans le travail de Huang en 2004 [6], les auteurs ont construit un classifieur à partir de la méthode de machines à probabilités *Minimax* (MPM pour Minimax Probability Machine). Cette méthode permet de générer un effet pire-cas sur la probabilité de mauvaise classification des données, en se basant sur des estimations fiables des matrices de moyennes et de covariance des classes d'apprentissage. En ce qui concerne les MPMs classiques, on y suppose de manière générale que le poids de chaque classe est non biaisé. Or, dans le cas de l'approche proposée par les auteurs dans [6], cette hypothèse est ignorée et

la méthode est appelée en conséquence : machine à probabilité *Minimax biaisée* (BMPM pour **B**aised **M**inimax **P**robability **M**achine). Le modèle a été transformé secondairement en une sorte de problème pseudo-concave avec un minimum local qui est aussi un maximum global. Delà, le classifieur conçu a été donc applicable à des tâches de classification biaisées et a été évalué avec des données pour le diagnostic du cancer du sein.

La classification a été effectuée sur une base de données non identifiées avec l'approche classique MPM et l'approche proposée par les auteurs Huang [6], BMPM, en vue de mettre en avant la performance de celle-ci. Plus de 94% de précision à la classification a été atteinte avec la méthode proposée.

Il est important de noter que le classifieur proposé dans [6] est biaisé et binaire. Le but de ce type de classifieur est de favoriser la classification de la catégorie la plus importante par rapport à la classification globale, cela tout en essayant de maintenir la classification de la catégorie de moindre importance à un taux de précision acceptable. Selon les auteurs, le modèle proposé représente la première méthode quantitative pour contrôler la position de l'hyperplan en faveur de la classification pour la classe la plus importante. Cela met en avant le fait que cette méthode n'est applicable que pour les problèmes de classification binaire.

Dans l'article de Madden en 2002 [7], l'auteur propose une étude empirique pour l'évaluation d'un classifieur probabiliste construit à partir de l'algorithme Bayésien Markov-Blanket. Ce classifieur a été testé avec la base de données "cancer du sein" et les taux de reconnaissance résultants ont été comparés à ceux obtenus avec trois autres classifieurs Bayésiens soient : Bayes naïf, Bayes naïf avec arbre, et un réseau Bayésien conventionnel. Les réseaux Bayésiens représentent graphiquement la distribution des probabilités de jonctions pour un ensemble de variables aléatoires. La construction d'un réseau Bayésien pour les besoins d'une classification pourrait être computationnellement inefficace dans le sens où la structure globale du réseau ne serait pas cohérente pour la classification [7]. Plus précisément, la classification n'est pas influencée par les parties de la structure qui se situent en dehors du noeud de classification Markov-Blanket. Pour un noeud  $x$ , son Markov-Blanket est l'union de ses parents directs, de ses fils directs, ainsi que tous les parents directs de ses fils directs. Le noeud  $x$  n'est pas affecté par les noeuds qui sont en dehors de l'intervalle où son Markov-Blanket est définit.

L'approche que l'auteur a utilisée a pour but de construire un Markov-Blanket approximatif autour du noeud classifieur. La performance des trois modèles conventionnels par rapport au modèle proposé est illustrée à la table (2.2). On s'intéresse particulièrement à la performance relative à la base de données " cancer du sein" (WBCD).

	Naïve	TAN	K2	MBBC
Chess	87.63 ± 1.61	91.68 ± 1.09	94.03 ± 0.87	<b>97.03 ± 0.54</b>
WBCD	<b>97.81 ± 0.53</b>	<b>97.47 ± 0.68</b>	<b>97.17 ± 1.05</b>	<b>97.30 ± 1.01</b>
LED-24	<b>73.28 ± 0.70</b>	<b>73.18 ± 0.63</b>	<b>73.14 ± 0.73</b>	<b>73.14 ± 0.73</b>
DNA	94.80 ± 0.44	94.75 ± 0.42	<b>96.22 ± 0.64</b>	<b>95.99 ± 0.42</b>
Lymph.	83.60 ± 9.82	<b>85.47 ± 9.49</b>	81.47 ± 10.4	83.47 ± 9.45
Nursery	90.48 ± 0.41	<b>94.16 ± 0.33</b>	92.63 ± 0.67	<b>94.16 ± 0.33</b>

TABLE 2.2 – Taux de reconnaissance des trois algorithmes conventionnels et de l'approche MBBC proposée par Madden

D'après les résultats que l'auteur a obtenu avec la méthode MBBC (Markov Blanket Bayesian Classifier) proposée, on pourrait remarquer qu'avec la base de données "cancer du sein" le classifieur Bayésien Markov-Blanket a fait preuve d'une performance comparable à celles des algorithmes Bayes naïf, Bayes naïf avec arbre (TAN), et réseau bayésien conventionnel (K2) [7]. Toutefois, la difficulté à l'implémentation des réseaux bayésiens demeure dans l'estimation de leurs paramètres. Les travaux présentés ci-haut montrent l'applicabilité de certaines méthodes issues du domaine des probabilités et statistiques dans le cadre du dépistage du cancer du sein. Les taux de classification obtenus à travers les différentes expériences sont intéressants mais les modèles proposés présentent tout de

Dans un autre travail réalisé par De Santo et al. en 2003 [10], les auteurs proposent une approche basée sur un système expert multiple pour faire de la reconnaissance de microcalcifications à partir d'une base d'image mammographiques. Un tel système rassemble plusieurs sous-systèmes experts, certains d'entre eux sont consacrés à classifier les microcalcifications en tant que telles et d'autres se chargent de faire la classification d'une partition de la mammographie dans sa globalité. La sortie du système est le résultat d'une combinaison des deux groupes de sous-systèmes. L'architecture du système global est illustrée à la figure (2.1).

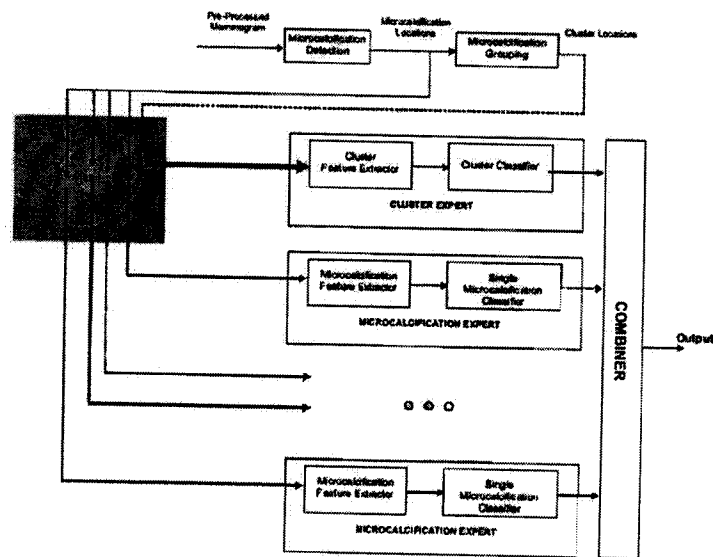


FIGURE 2.1 – Architecture du système expert multiple [10]

Un traitement préliminaire est effectué sur les données avant qu'elles ne soient affectées à l'entrée du classifieur. Ce prétraitement est réalisé dans le but de réduire la taille des images en les convertissant au format 8bits/pixel. La base de données contient 40 mammographies, 102 clusters et 2123 micro calcifications (1792 malignes et 331 bénignes). Les résultats de classification obtenus sont illustrés dans la table (2.3).

	Cluster Expert	$\mu$ C-Expert ( $A_b, A_m$ )	MES
Malignant clusters	60.0%	75.7%	75.7%
Benign clusters	70.0%	67.5%	73.5%

TABLE 2.3 – Taux de reconnaissance de chaque sous-système et du système global

À travers les résultats rapportés dans la table (2.3), on pourrait noter que lorsque le système expert global MES (Mutti-Expert System) fait preuve de bonne performance à la reconnaissance des clusters bénins, sa performance est moindre à la reconnaissance des clusters malins. Ceci est dû à l'aptitude du système à exploiter les capacités de complémentarité sur les deux sous-systèmes experts pour la reconnaissance des différents types de clusters.

La performance du système expert multiple est intéressante mais les taux de reconnaissance restent moindres par rapport à ceux obtenus dans d'autres travaux (Grohman et

Dhawan en 2001 [11], Karnan et al. en 2008 [12]). Mise à part la limitation due à la classification binaire, l'approche présente l'inconvénient majeur rencontré dans la conception de tout système expert et qui est décuplé dans ce cas de figure : le volume d'expertises à capturer, formaliser, et exploiter par les différents sous-systèmes experts ainsi que la mise en oeuvre de l'expertise globale à travers le MES constituent des tâches complexes et nécessitent un effort cognitif appréciable. Au final, il y a eu capture d'une connaissance présente dans l'image et représentée par l'interprétation d'un expert et non pas la capture d'une expertise en soi.

Les travaux présentés ci-haut montrent l'applicabilité de certaines méthodes déductives et symboliques dans le cadre du dépistage du cancer du sein. Les arbres de décisions présentés fournissent des taux de classification intéressants. Néanmoins, ces modèles présentent des limites ; en particulier, leur performance est tributaire de la structure de l'arbre de décision qui doit être représentative des données [8]. On pourrait donc obtenir des structures différentes si on fait un moindre changement sur les données ou sur le choix de variables. Cette dernière propriété peut être une source importante d'erreur et de variance dans les résultats. De plus, ces approches ont une faible capacité de généralisation. Le système expert multiple quant à lui, offre une aide dans le diagnostic final du cancer du sein par le fait qu'il s'intéresse non seulement à la classification des microcalcifications en tant que telles mais aussi à la classification de clusters de microcalcifications. Ceci permet de travailler avec différentes granularités de représentation de connaissance, ce qui est souvent considéré comme irréalisable avec d'autres approches [10].

## 2.4 Approches évolutives

Dans la troisième partie de cette revue de littérature, on s'intéresse aux approches évolutives employées dans le cadre de la classification du cancer du sein. Dans la littérature, on retrouve plusieurs méthodes basées notamment sur les algorithmes génétiques, la programmation génétique, ou sur des modèles issus de l'intelligence collective tels que les algorithmes de colonies de fourmis et les algorithmes à essaims particuliers. Dans Karnan et al. en 2008 [12], les auteurs visent à dépister le cancer du sein en employant deux approches issues de l'intelligence collective, soient, les algorithmes de colonies de fourmis (ACO pour Ant Colony Optimization) et les essaims particuliers (PSO pour Particular Swarm Optimization). Pour leur besoins expérimentaux, un prétraitement des données est effectué. Il s'agit de faire de l'extraction de variables à partir d'une base d'images mammographiques, avec trois méthodes évolutives différentes : un algorithme génétique, un algorithme de colonies de fourmis et une approche à essaims particuliers. Les données prétraitées ont ensuite été affectées à l'entrée de deux systèmes hybrides pour la phase de classification. Le premier système comprend un PMC à rétropropagation d'erreur et un algorithme d'optimisation à colonie de fourmis. Quant au deuxième système, il consiste en un PMC à rétropropagation d'erreur et un algorithme d'optimisation à essaims particuliers. Dans le premier système hybride, le réseau de neurones utilise une topologie à trois couches. La première couche comprend un nombre de neurones égal à la taille du vecteur d'entrée, la deuxième est une couche cachée qui comprend autant de neurones que la couche d'entrée et la dernière couche contient une unité de sortie. Les poids initiaux du réseau sont générés par l'algorithme de colonie de fourmis. Ce dernier produit  $N$  nombres aléatoires avec  $d$  digits, où  $N$  correspond aux nombre



total de neurones dans le PMC. Les poids synaptiques sont ensuite extraits à partir d'une population de nombre aléatoires pour déterminer les valeurs de la fonction fitness. La valeur fitness optimale est sélectionnée en utilisant l'algorithme d'optimisation à colonie de fourmis. Les poids sont mis-à-jour en fonction de la valeur fitness actuelle et le processus d'apprentissage est exécuté de nouveau. La procédure est répétée jusqu'à ce que l'erreur générée par le réseau soit inférieure à la valeur de tolérance fixée dans l'algorithme ACO.

Le deuxième système hybride, PMC-PSO, est un algorithme d'optimisation qui combine l'algorithme à essais particuliers au PMC. L'idée fondamentale de cette hybridation est qu'au début du processus de recherche de l'optimum, le PSO est employé pour accélérer l'apprentissage. Quand la valeur de la fonction fitness ne change plus pour certaines générations, ou que cette valeur est inférieure à la valeur prédéfinie, le processus de recherche est réalisé en appliquant la règle de descente du gradient. Cet algorithme effectue en premier lieu une initialisation d'un groupe aléatoire de particules, et procède secondairement à la mise-à-jour de ces valeurs jusqu'à ce qu'une nouvelle génération de particules soit générée. Les nouvelles particules procèdent à la recherche de la meilleure position globale dans l'espace de solution. L'algorithme à rétropropagation est utilisé dans le but d'évaluer cet optimum global. C'est de cette façon que l'approche hybride proposée peut trouver un optimum avec une certaine rapidité.

Method	GA	ACO	PSO
Jack-Knif Method	0.856	0.948	0.952
Round Robin Method	0.861	0.947	0.950
Ten-Fold Validation Method	0.849	0.938	0.949

TABLE 2.4 – Classification basée sur les algorithmes d'apprentissage et de sélection de variables

Le feedback de la phase de classification a permis aux auteurs de procéder à l'évaluation, de la performance des trois méthodes de sélection de variables (AG, ACO, PSO). Cette évaluation a été effectuée en générant des courbes ROC (Receiver Operating Characteristic Curve) en fonction de trois méthodes de validation : Jack Knife, Round Robin, et Ten-Fold. Les résultats de classification obtenus à travers les trois approches évolutionnaires sont illustrés à la table (2.4). Ils montrent que l'approche par essais particuliers procure une meilleure performance à la classification que l'algorithme génétique conventionnel ou l'algorithme de colonie de fourmis.

La performance des deux modèles hybrides proposés est intéressante, en particulier celle de l'algorithme PSO-PMC. La méthode PSO est une approche globale ayant une grande capacité à trouver l'optimum global. Ceci dit, son inconvénient dans le cadre du tra-

*vail présenté en [12] est que la recherche de cet optimum est lente et coûteuse. Par ailleurs, l'algorithme de rétropropagation a une grande capacité à trouver l'optimum local mais présente une certaine limitation quant à trouver l'optimum global [12]. Par conséquent, l'hybridation PSO-PMC permet de palier aux faiblesses des deux modèles d'une manière complémentaire. On note que la performance de l'algorithme PSO dépend considérablement du choix de la fonction fitness, et celle de l'algorithme ACO dépend du choix de la valeur de tolérance. Une bonne estimation de ces paramètres demande une expertise dans le domaine afin de pouvoir trouver un compromis entre la qualité du modèle et la plausibilité de ses résultats.*

Dans Liu et al. en 2004 [13], les auteurs proposent un algorithme basé sur l'approche des colonies de fourmis (ACO) pour faire de la classification dans le cadre du dépistage du cancer du sein. Il s'agit d'une version améliorée d'un algorithme connu dans la littérature sous le nom d'Ant\_Miner. Cet algorithme est détaillé dans la papier de Parepinelli et al. en 2002 [14]. La méthode proposée par les auteurs emploie un élément stochastique pour la construction des règles, ce qui l'amène à procurer un équilibre entre l'exploitation et l'exploration des données. Cette propriété permet de mieux modéliser le comportement des colonies de fourmis réelles, et aussi d'avoir un plus grand nombre de choix de chemins d'exploration à emprunter. Elle propose de plus une nouvelle stratégie pour contrôler l'influence des valeurs de la phéromone. Cette stratégie représente une règle pour effectuer la mise-à-jour de la phéromone qui amène les futures particules (fourmis) à converger vers de meilleures décisions. Cela permet de favoriser la qualité de la règle ainsi que la performance globale de la série de règles.

Run Number	Breast	Cancer
	Ant_Miner1	Ant_Miner3
1	92.05	94.32
2	93.15	93.15
3	91.67	91.67
4	95.59	97.06
5	88.41	92.75
6	94.20	95.65
7	90.77	93.84
8	96.55	96.55
9	91.04	92.54
10	92.86	95.71

TABLE 2.5 – Taux de reconnaissance de l'algorithme proposé (Ant\_Miner3) par rapport à l'Ant\_Miner conventionnel

La base de données " cancer du sein" est divisée en 10 partitions, chaque algorithme est exécuté 10 fois en utilisant une partition différente de celles consacrées à la phase de test ; les 9 partitions restantes sont consacrées à la phase d'apprentissage. Les résultats de performance de l'algorithme proposé (Ant\_Miner3) pour les 10 exécutions sont illustrés dans la table (2.5). Nous remarquons que plus de 97% de précision à la reconnaissance a été atteinte pour la classification des données en deux classes, bénigne ou maligne. La fonction principale de l'algorithme Ant\_Miner est d'extraire des règles à partir de données [13]. Les règles extraites avec Ant\_Miner3 procurent de meilleurs taux de classification que ceux de la méthode classique. Néanmoins, Ant\_Miner3 nécessite un nombre plus grand de fourmis pour trouver la solution, ce qui rend l'algorithme avantageux en termes de performance à la classification mais pas forcément en termes de temps d'exécution. De plus, cet algorithme comprend un nombre de paramètres dont les auteurs ignorent les effets [13], ainsi, des expériences détaillées seraient requises et profitables pour pouvoir les déterminer d'une manière appropriée.

Une autre variante de l'algorithme Ant\_Miner a été proposée dans Jaganathan et al. en 2007 [15] pour effectuer une classification sur une base de données d'images mam-

mographiques. Le nouvel algorithme est une combinaison de l'algorithme Ant\_Miner et d'une version améliorée de l'algorithme quickreduct. Ce dernier a pour but de calculer un sous-ensemble minimal sans pour autant générer d'une manière exhaustive tous les sous-ensembles possibles ; une manière rapide de réduire la dimensionnalité de l'espace d'entrée de Ant\_Miner.

Run Number	Ant_Miner	Ant_Miner with Improved Quick Reduct
1	70.28	70.11
2	70.40	71.34
3	69.47	70.84
4	70.98	72.00
5	69.42	70.23
6	72.19	74.20
7	73.65	75.43
8	63.55	65.63
9	71.84	73.84
10	69.54	68.96

TABLE 2.6 – Résultats de classification de la base de données Brest Cancer

Les expérimentations ont été effectuées de la même manière que dans [13] (10 exécutions sur 10 partitions de données), les résultats obtenus avec l'algorithme Ant\_Miner classique et l'approche proposée sont illustrés à la table (2.6). On peut remarquer à travers les taux de reconnaissance obtenus, que la performance de l'algorithme Ant\_Miner est meilleure s'il est utilisé avec la version améliorée de l'algorithme Quickreduct.

Un autre travail a été fait par El Hamdi et al. en 2010 [16], les auteurs ont proposé une approche évolutive de la conception d'un modèle flou type TSK (TFM pour TSK-Type Fuzzy Model) pour résoudre le problème de diagnostic du cancer du sein. Dans cette méthode, le nombre de règles floues et les paramètres réglables dans le TFM sont conçus en même temps que la combinaison de l'algorithme génétique compact (CGA pour Compact Genetic Algorithm) et l'algorithme de l'état d'équilibre génétique (SSGA pour Steady-State Genetic Algorithm). Les résultats obtenus montrent que l'approche présentée peut obtenir une meilleure généralisation que certaines méthodes existantes en utilisant la base de données du cancer de sein (WBCD).

Les travaux présentés ci-haut montrent l'applicabilité de certaines méthodes évolutionnaires dans le cadre du dépistage du cancer du sein. Toutefois, il est à considérer que les approches citées comprennent certaines limites. Celles-ci résident essentiellement dans la façon de déterminer leurs paramètres. Aucune règle n'est établie pour les définir. Par ailleurs, les hybridations proposées sont intéressantes en termes de complémentarités des systèmes qu'elles contiennent, mais semblent toutefois complexes à la réalisation, ceci est notamment dû au fait que le système global résultant soit très paramétré. De plus, il s'agit dans les travaux cités de classification binaire (classe bénigne et classe maligne), ce qui remet en question l'applicabilité des méthodes proposées dans les problèmes réels, en particulier dans le cadre du dépistage du cancer du sein où on a plusieurs catégories de tumeurs à identifier.

## 2.5 Approches connexionnistes

Dans la dernière partie de cette revue de la littérature, nous nous intéressons aux approches connexionnistes employées dans le domaine du dépistage du cancer du sein. à cet effet, on retrouve dans la littérature plusieurs modèles de RNAs : des réseaux de neurones à rétropropagation [17],[18],[19], des réseaux utilisant des fonctions à base radiale [20],[21], mais aussi des modèles neuronaux hybrides, où le réseau de neurones est combiné à une approche autre que connexionniste. Nous présentons certains travaux plus en détail dans ce qui suit. Le premier travail de cette revue de littérature sur les approches connexionnistes procède à une analyse de méthodes de classification dont deux modèles de réseaux neurones artificiels [8]. Comme dans certains travaux cités précédemment, avant d'affecter les données aux classifieurs proposés, une étape de prétraitement des données (images mammographiques) est réalisée par un processus de sélection de variables. La démarche suivie pour effectuer la tâche de classification est illustrée à la figure (2.2).

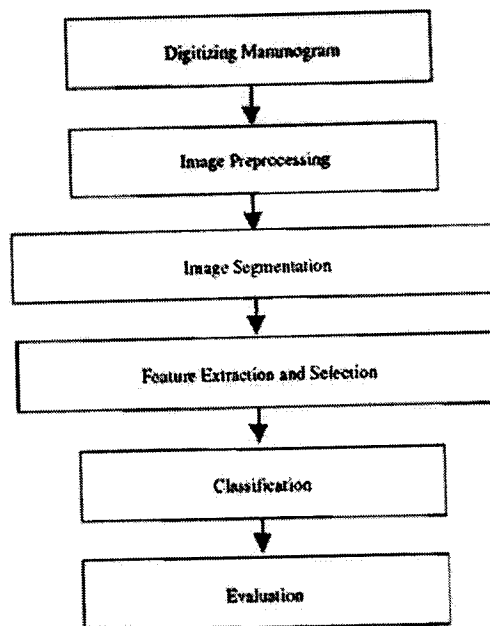


FIGURE 2.2 – Architecture du système d'aide au diagnostic [8]

Une fois que les variables relatives aux masses sont extraites et sélectionnées, elles sont affectées aux entrées des classifieurs. Les masses obtenues à travers les étapes de prétraitement doivent être classées comme étant des tissus normaux, des masses bénignes, ou des masses malignes. Les deux réseaux connexionnistes discutés dans [8] ont été implémenté dans [20]. Le premier est un réseau à rétropropagation 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 65%.

Dans le papier Guo et Nandi en 2006 [22], les auteurs ont proposé un perceptron multicouche (PMC) en tant que classifieur pour diagnostiquer le cancer du sein. La base de données WDBC (Wisconsin Diagnosis Breast Cancer) a été utilisée, qui contient 569 cas de tumeurs dont 357 de cas bénins et 212 de cas malins. Dans un premier temps, une étape de sélection de variables est effectuée sur les données en utilisant la programmation

génétique (PG). Les auteurs ont mis en oeuvre une variante améliorée de l'analyse discriminante linéaire de Fisher (MFLDA pour **M**odified **F**isher **L**inear **D**iscriminant **A**nalysis) pour servir la méthode de programmation génétique dans l'optimisation de la sélection des variables. Les variables sont ensuite affectées à l'entrée du PMC pour évaluer la performance de classification.

L'architecture du PMC consiste en une couche d'entrée au réseau, une couche cachée comportant 2 à 14 neurones, et une couche de sortie. Une fonction d'activation logistique est utilisée pour les neurones cachés, et une fonction linéaire pour la couche de sortie. Un algorithme d'apprentissage par rétropropagation d'erreur avec moment et adaptation a été employé.

Le réseau de neurones a convergé avec un taux moyen de classification de 96.21 %. Le meilleur pourcentage obtenu durant les expérimentations est de 97.34%. Les résultats obtenus par les auteurs ont mis de l'avant la capacité de la méthode de PG à transformer l'information en réduisant la dimensionnalité de l'espace de variables, et de définir la relation entre les données d'une manière automatique. Ces deux dernières propriétés aident à améliorer la précision de la classification [22]. Les taux de reconnaissance obtenus sont intéressants. Cela dit, le classifieur était de type binaire avec une classe de cas malins et une classe de cas bénins.

Dans un autre travail de Cheng et Cui en 2004 [23], les auteurs ont proposé une approche neuro-floue pour la détection de masses cancéreuses. Le réseau consistait en 4 couches. La couche d'entrée et la deuxième contiennent 4 neurones chacune. La troisième couche comporte N neurones flous. Le nombre N est déterminé durant la phase d'apprentissage et varie selon les paramètres du réseau et la distribution des données. Enfin, la quatrième couche contenait un maximum de 2 unités floues et une unité compétitive floue. L'architecture du système est telle qu'illustrée à la figure 2.5. Tout comme dans le travail proposé dans [22], un prétraitement des données a été effectué dans le but de faire une sélection de variables avant de procéder à l'étape de classification.

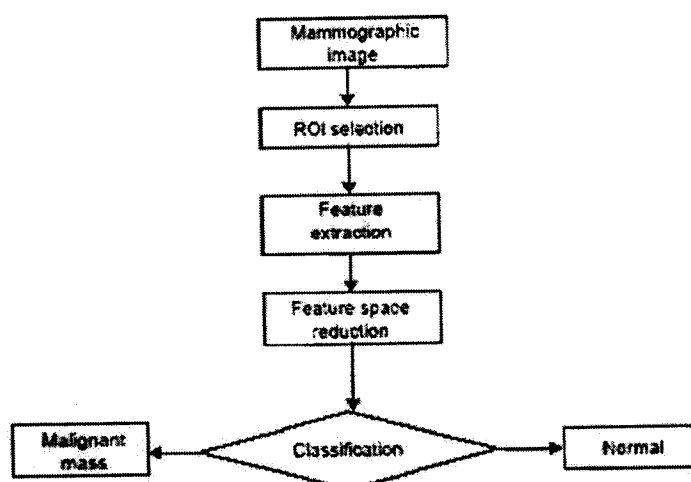


FIGURE 2.3 – Organigramme de l'approche proposée [23]

La base de données utilisée est une série d'images mammographiques qui a été divisée en deux sous-ensembles pour les besoins expérimentaux, un ensemble d'apprentissage et un ensemble de test.

Alpha ( $\alpha$ )	Beta ( $\beta$ )	Number of neurons in the third layer	Recognition rate (%)
0.1	0	65	94
0.1	0.1	51	97
0.1	0.2	47	96
0.1	0.3	49	93
0.1	0.4	54	93
0.1	0.5	49	92
0.1	0.6	52	91
0.1	0.7	55	91
0.1	0.8	57	90
0.1	0.9	58	90

TABLE 2.7 – Taux de classification du FNN

Les données sont affectées à l'entrée du réseau neuro-flou, et sont flouifiées au niveau de la deuxième couche. Ensuite, tout au long du processus d'apprentissage, le réseau procède à une adaptation et un ajustement autonome. L'adaptation consiste en l'ajout ou non d'un neurone dans l'architecture du réseau ; un neurone est ajouté à la troisième couche du réseau si une donnée ne peut être classée dans les classes déjà existantes, la couche reste telle qu'elle était durant le dernier cycle d'apprentissage sinon. Quant à l'ajustement, il se situe dans la mise à jour d'une valeur qui représente le seuil de décision pour choisir la classe à laquelle doit être affectée la donnée en entrée.

Les résultats de classification sont résumés dans la table (2.7),  $\alpha$  est le paramètre d'apprentissage du réseau et  $\beta$  est le paramètre de flouification. Comme on peut le remarquer, jusqu'à 97% de précision à la classification a été atteinte. Les auteurs rapportent que le taux de rappel a été influencé par le paramètre  $\beta$  ; il est donc important de bien le définir, et ceci se fait en général d'une manière empirique.

Un autre système neuro-flou a été proposé dans l'article de Grohman et Dhawan en 2001 [24], où les auteurs visent à effectuer la classification en identifiant des sous-ensembles convexes de données dans l'espace de variables. Cette méthode permet de définir l'espace de variables et de situer sa contribution dans la tâche de classification. Le résultat de l'apprentissage est un classifieur comprenant un ensemble de fonctions de flouification qui reflètent le degré d'appartenance du pattern d'entrée à un nombre de sous-ensembles convexes dans l'espace de définition de variables. Ces sous-ensembles convexes sont identifiés durant le processus d'apprentissage du système neuro-flou. Le fonctionnement du processus d'apprentissage du réseau tente en quelque sorte de trouver une dichotomie optimale de l'espace d'entrée dans ces ensembles convexes.

La structure du classifieur neuro-flou est illustrée à la figure (2.4). Dans l'approche proposée, une fonction d'appartenance  $M_f$  est conçue pour chaque sous-ensemble convexe ( $f = 1, 2, 3, \dots, K$ ). C'est à la couche de sortie que la décision de classification est effectuée, cette couche applique le principe de l'algorithme *the winner takes all*. La catégorie résultante est la catégorie de l'ensemble convexe du pattern d'entrée avec la valeur de la fonction d'appartenance la plus grande.

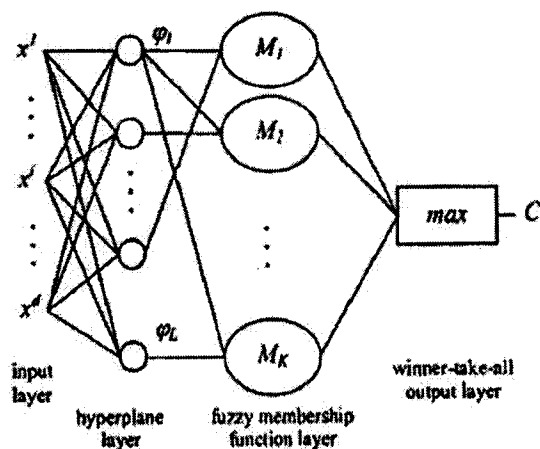


FIGURE 2.4 – Architecture du réseau neuro-flou [24]

Le fonctionnement de la méthode neuro-floue proposée comprend 4 étapes : la création d'un ensemble convexe, le placement d'un hyperplan séparateur (au niveau de la deuxième couche du réseau), la construction de la fonction d'appartenance pour chaque ensemble convexe généré (au niveau de la troisième couche du réseau), et la sélection d'une classe en fonction de la plus grande valeur de la fonction d'appartenance issue de la troisième couche (au niveau de la quatrième couche).

Dans le but d'établir une comparaison, les auteurs ont employé un réseau à rétropropagation d'erreur pour effectuer la même tâche de classification. Ensuite, à travers un système ROC (Receiver Operating Characteristic), la performance des deux classifieurs a été évaluée selon les taux de classification qu'ils ont générés respectivement à la phase de rappel. Les résultats de l'approche proposée par les auteurs [24] sont meilleurs que ceux obtenus avec un réseau à rétropropagation classique. Néanmoins, le système conçu semble comporter une hybridation qui pourrait être complexe, il est en effet nécessaire d'avoir des connaissances à la fois sur les systèmes flous, les réseaux connexionnistes à couches et les algorithmes compétitifs pour pouvoir élaborer une telle hybridation. De plus, il y a certains paramètres du système neuro-flou qui ne peuvent être fixés que d'une manière empirique tel que le choix de la fonction d'appartenance.

Un autre travail de recherche dans le cadre du dépistage du cancer du sein a été proposé par Abbas en 2002 [25]. Il s'agit d'une approche neuro-évolutive basée sur l'algorithme Pareto-différentiel avec une recherche locale. Cette hybridation a été réalisée dans le but de remédier au problème de l'apprentissage lent que connaissent les approches neuro-évolutives classiques.

Le système a une architecture multicouche. Dans ce type de réseaux, il est souvent problématique de décider du nombre d'unités cachées à allouer au réseau. Ce nombre est fixé par l'expérience. Pour pallier à cette limite, l'auteur a incorporé une méthode d'optimisation, qui a pour but de déterminer une bonne topologie pour le réseau hybride. Il s'agit d'une méthode d'optimisation multi-objective ayant pour buts de minimiser l'erreur de reconnaissance et donc la performance du système mais aussi de minimiser le nombre d'unités de traitement à la couche cachée du réseau neuro-évolutive.

La base de données "cancer du sein" de Wisconsin est utilisée en vue de faire la classification du cancer du sein. Les probabilités de croisement et de mutation de l'algorithme

Pareto ont été variées entre 0 et 1 avec un incrément de 0.1, et la taille de la population a été fixée à 12.

Le paramètre d'apprentissage de l'algorithme de rétropropagation a été fixé à 0.03, et le nombre d'unités cachées à un maximum de 10. L'auteur note que la taille d'une population doit être plus grande que le nombre des unités de traitement cachées car le nombre maximum des solutions que génère l'algorithme Pareto dans ce travail est le nombre maximum des unités cachées (par définition de l'optimalité de Pareto). Les résultats de classification sont illustrés à la table (2.8).

Method	Error	Number of epochs
MPANN	$0.981 \pm 0.005$	5100
Fogel et <i>al.</i>	$0.981 \pm 0.464$	200000
Abbas et <i>al.</i>	$0.975 \pm 1.800$	10000

TABLE 2.8 – Performance de l'approche MPNN

L'auteur a mis en avant les résultats obtenus avec l'approche qu'il a proposée d'une part ses taux de reconnaissance mais aussi d'autre part sa capacité de généralisation, en les comparant à ceux obtenus dans deux travaux antérieurs, l'un utilisant une approche évolutionnaire uniquement et l'autre un réseau de neurones à rétropropagation d'erreur. Nous remarquons selon les résultats illustrés dans la table (2.8) que l'approche MPANN (Mimetic Pareto Neural Network) proposée a abouti à de meilleurs taux de reconnaissance en un temps de rappel moindre par rapport aux deux autres approches.

Parmi les travaux les plus récents sur les systèmes neuro-flou, nous trouvons celui de Seyedeh et al. en 2010 [26] ; une nouvelle approche pour la détection du cancer de sein basée sur les réseaux neuro-flous hiérarchiques (HFNN pour Hierarchical Fuzzy Neural Network). Ces types de réseaux se composent de multiples réseaux neuro-flous réunis dans différents niveaux ou sur une architecture en cascade. Dans ce travail un réseau neuro-flou hiérarchique évolué a été utilisé pour la détection du cancer de sein. Le système a donné de bons résultats par rapport à un système neuro-flou classique (FNN pour Fuzzy Neural Network).

Method	Sensitivity %	Specificity %	Accuracy %
FNN	94.5	99.5	98.5
	Var = 0.0000	Var = 0.0000	Var = 0.0000
HFNN	98.18	99.29	99.04
	Var = 0.0000	Var = 0.0026	Var = 0.0020

TABLE 2.9 – Performance de l'approche HFNN

Un autre résultat de la même année a été obtenu par Ashraf et al. [27] , les auteurs présentent une nouvelle approche pour le diagnostic du cancer de sein en utilisant une combinaison d'un réseau adaptatif en fonction du système d'inférence floue (ANFIS) et la méthode de gain de l'information. Dans cette approche l'ANFIS a pour tâche de construire une cartographie des entrées-sorties en utilisant les connaissances de l'expert et la capacité d'apprentissage de la machine ; la méthode de gain de l'information est utilisé pour réduire le nombre d'entrées pour l'ANFIS. Un résultat expérimental montre



98,23% de précision qui souligne la capacité de l'algorithme proposé.

Les travaux présentés ci-haut montrent l'applicabilité de certains types de réseaux de neurones artificiels dans le domaine du dépistage du cancer du sein. Les taux de classification obtenus à travers les différentes expériences sont intéressants. Notons par ailleurs que les conceptions proposées sont pour la plupart une combinaison de deux systèmes ou plus, ceci engendre une certaine complexité à la réalisation. De plus, dans l'utilisation des modèles connexionnistes cités, il était question d'effectuer une classification binaire ou à trois classes. Il s'agit d'une caractéristique qui n'est pas avantageuse pour l'applicabilité des modèles aux problèmes réels où il est souvent question d'une classification du cancer du sein multi classes.

## 2.6 Contribution et Conclusion

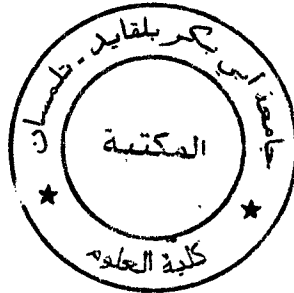
La revue de littérature présentée dans ce chapitre a permis de mettre en évidence plusieurs avantages et certaines limites des approches de classification employées dans le cadre du dépistage du cancer du sein. Il faut noter cependant que certains auteurs omettent de donner la référence exacte à la base de données "cancer du sein" utilisée durant leurs expérimentations. Cela risque d'engendrer des comparaisons incohérentes lorsque l'on veut situer la performance d'un travail par rapport à un autre.

Notons par ailleurs que plusieurs des conceptions proposées sont une combinaison de deux systèmes ou plus, ceci cause une certaine complexité à la réalisation. De plus, ces hybridations améliorent quelques fois le temps de convergence du système mais la performance de celui-ci dépend amplement du choix des paramètres des différentes méthodes qu'il comprend. Aussi, dans la majorité des travaux, la classification des données devait se faire selon 2 ou 3 classes au plus. Ceci étant, pour les problèmes réels dans le domaine du dépistage du cancer du sein, on a souvent affaire à plusieurs cas de tumeurs, et donc un plus grand nombre de classes. Il est tout aussi à considérer que les différents systèmes de classification existants peuvent être remis en question de par leur applicabilité pour l'aide au diagnostic du cancer du sein. En particulier, quelques recherches se sont consacrées à la conception de modèles de réseaux neuronaux artificiels pour tenter de remédier aux limites que connaissent des modèles déjà existants. Il serait intéressant de pouvoir éliminer les inconvénients cités afin de favoriser l'applicabilité des RNAs dans le domaine de l'aide au diagnostic du cancer du sein et les combinés avec d'autres systèmes tels que la logique floue et les méthodes évolutionnaire pour obtenir un classifieur performant et interprétable. C'est l'objectif de ce mémoire.

Nous concevons dans ce travail deux systèmes neuro-flou basé sur l'architecture du Nefclass (pour Neuro Fuzzy Classification) dans le but d'évaluer leur performance à la classification de cas de tumeurs en utilisant la même base de données, le premier est le système Nefclass standard et le deuxième est une optimisation de la version original en utilisant les essaims particuliers. Deuxièmement, nous proposons une étude comparative des deux modèles en essayant de mettre en avant quelques points discutés dans cette revue de littérature.

## Chapitre 3

# Notions sur les Techniques de l'Intelligence Artificielle



**Première partie**  
**Les Réseaux de Neurones**

## 3.1 Introduction

Le rêve de créer une machine, dotée d'une forme d'intelligence, est présent depuis fort longtemps dans l'imagination humaine. Des recherches menées par des scientifiques ont abouties à plusieurs approches dont l'une d'entre elles est l'étude des réseaux de neurones formels. Le réseau de neurones (**Artificial Neural Network**) est un modèle de calcul dont la conception est très schématiquement inspirée du fonctionnement de neurones biologiques. Les réseaux de neurones sont généralement optimisés par des méthodes d'apprentissage de type statistique, si bien qu'ils sont placés d'une part dans la famille des applications statistiques, qu'ils enrichissent avec un ensemble de paradigmes permettant de générer de vastes espaces fonctionnels souples et partiellement structurés, et d'autre part dans la famille des méthodes de l'intelligence artificielle qu'ils enrichissent en permettant de prendre des décisions s'appuyant davantage sur la perception que sur le raisonnement logique formel. Les réseaux de neurones sont utilisés dans de nombreux domaines tels que la classification, l'évaluation, ou la prédiction. Une des principales propriétés des RNA est la capacité d'apprendre à partir des données. L'apprentissage d'un réseau de neurones consiste à changer son comportement de façon à lui permettre de se rapprocher d'un but bien défini.

## 3.2 Les réseaux de neurones

Un réseau de neurone artificiel est un système de traitement de l'information qui a certaines caractéristiques en commun avec les réseaux de neurones biologiques. La première modélisation d'un neurone découle des résultats des travaux de Mac Culloch et Pitts. Les réseaux de neurones artificiels ont été développés comme des généralisations des modèles mathématiques de la cognition humaine ou de la biologie des neurones en se basant sur les hypothèses suivantes [28] :

- Traitement des informations se fait dans des éléments simples appelés neurones.
- Les signaux sont transmis entre les neurones par des liaisons de connexion.
- Chaque liaison de connexion a un poids associé.
- Chaque neurone applique une fonction d'activation à son entrée pour déterminer son signal de sortie.

Les RNAs ont la capacité de stocker la connaissance empirique et de la rendre disponible à l'usage. Les habiletés de traitement (et donc la connaissance) du réseau vont être entreposées dans les poids synaptiques, obtenus par des processus d'adaptation ou d'apprentissage [29]. En ce sens, les RNAs ressemblent donc au cerveau car non seulement, la connaissance est acquise au travers d'un apprentissage mais en plus stockée dans les connexions entre les entités (poids synaptiques).

## 3.3 Architecture des Réseaux de Neurones

On distingue deux types de réseaux de neurones : les réseaux non bouclés et les réseaux bouclés.

### 3.3.1 Réseaux de neurones non bouclés (feed-forward)

Les réseaux de neurones feed-forward sont des réseaux où les neurones ne sont connectés que dans un sens, le sens orienté de l'entrée vers la sortie.

#### Les Perceptrons

**Le Perceptron Monocouche** C'est l'un des premiers réseaux de neurones, conçu en 1958 par Rosenblatt. Il est linéaire et muni d'une seule couche. Il est inspiré du système visuel. La première couche (d'entrée) représente la rétine. Les neurones de la couche suivante sont les cellules d'association, et la couche finale les cellules de décision. Les sorties des neurones ne peuvent prendre que deux états (-1 et 1 ou 0 et 1). Seuls les poids des liaisons entre la couche d'association et la couche finale peuvent être modifiés. La règle de modification des poids utilisée est la règle de Widrow-Hoff [29] : si la sortie (celle d'une cellule de décision) est égale à la sortie désirée, le poids de la connexion entre ce neurone et le neurone d'association qui lui est connecté n'est pas modifié. Dans le cas contraire le poids est modifié en fonction de l'entrée :

$$W_i \leq W_i + k(d - s)$$

Avec :

- $k$  : constante positive
- $s$  : sortie
- $d$  : sortie désirée

En 1969, Papert et Minsky (un des pères des systèmes experts) [30] démontrent les limites du perceptron classique, incapable, par exemple de simuler la fonction ou exclusif (xor).

**Les Perceptrons Multicouches (PMC)** Le perceptron multicouche est sans doute le plus simple et le plus connu des réseaux de neurones. La structure est relativement simple : une couche d'entrée, une couche de sortie et une ou plusieurs couches cachées. Chaque neurone n'est relié qu'aux neurones des couches précédentes, mais à tous les neurones de la couche précédente. Les PMCs sont une amélioration du perceptron monocouche comprenant une ou plusieurs couches intermédiaires dites couches cachées. Ils utilisent, pour modifier leurs poids, un algorithme de rétro-propagation du gradient, qui est une généralisation de la règle de Widrow-Hoff [29]. Il s'agit toujours de minimiser l'erreur quadratique, ce qui est assez simple quand on utilise une fonction  $f$  dérivable (la sigmoïde par exemple). On propage la modification des poids de la couche de sortie jusqu'à la couche d'entrée. Les PMCs agissent comme un séparateur non linéaire et peuvent être utilisés pour la classification, le traitement de l'image ou l'aide à la décision.

- **Les noeuds d'entrée** : La première couche est appelée couche d'entrée. Elle recevra les données source que l'on veut utiliser pour l'analyse. Dans le cas de l'aide au diagnostic médical, cette couche recevra les caractéristiques des symptômes. Sa taille est donc directement déterminée par le nombre de variables d'entrées.

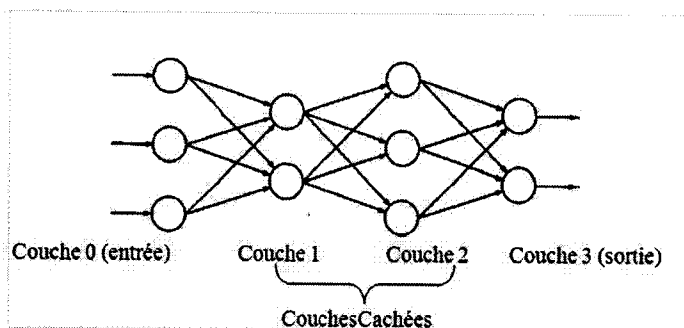


FIGURE 3.1 – Réseau de neurones multi-couches

- **Les noeuds cachés** : La seconde couche est une couche cachée, en ce sens qu'elle n'a qu'une utilité essentielle pour le réseau de neurones et n'a pas de contact direct avec l'extérieur. Les fonctions d'activations sont en général non linéaires sur cette couche mais il n'y a pas de règle à respecter. Le choix de sa taille n'est pas implicite et doit être ajusté. En général, on peut commencer par une taille moyenne des couches d'entrée et de sortie mais ce n'est pas toujours le meilleur choix. Il sera souvent préférable pour obtenir de bons résultats, d'essayer le plus de tailles possibles.
- **Les noeuds de sortie** : La troisième couche est appelée couche de sortie. Elle donne le résultat obtenu après traitement des données d'entrée de la première couche. Dans le cas de l'aide au diagnostic médical, cette couche donne le diagnostic. Sa taille est directement déterminée par le nombre de classe ou de types de maladies.

Un noeud reçoit des valeurs en entrée et renvoie 0 à n valeurs en sortie. Toutes ces valeurs sont normalisées pour être comprises entre 0 et 1 (ou parfois entre -1 et 1, selon les bornes de la fonction de transfert).

Une fonction de combinaison calcule une première valeur à partir des noeuds connectés en entrée et du poids des connexions [31] Ainsi, dans les réseaux les plus courants, les perceptrons en général, il s'agit de la somme pondérée  $\sum_i n_i p_i$  des valeurs  $n_i$  des noeuds d'entrée. Afin de déterminer une valeur à sa sortie une seconde fonction, appelée fonction d'activation est appliquée à cette valeur. Les noeuds de la couche d'entrée sont triviaux, dans la mesure où ils ne combinent rien, et ne font que transmettre la valeur de la variable qui leur correspond et leurs sorties seront injecté dans l'entrée des neurones de la couche suivante.

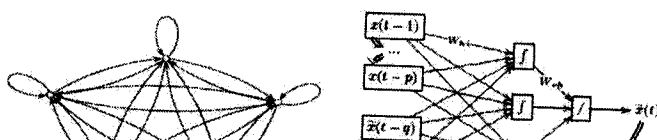
### Les réseaux à fonction radiale

Ce sont les réseaux que l'on nomme aussi RBF ("Radial Basic Functions"). L'architecture est la même que pour les PMC cependant, les fonctions de base utilisées ici sont des fonctions Gaussiennes.

### 3.3.2 Réseaux de neurones bouclés (récurrents)

Les réseaux de neurones récurrents sont des réseaux où les neurones de sortie par exemple peuvent voir leur sortie utilisée comme entrée d'un neurone d'une couche précédente ou de la même couche. Par conséquent, instinctivement il est évident que ce modèle est

bien plus compliqué. Il existe beaucoup de ces réseaux, les plus connus sont les cartes auto-organisatrices de Kohonen, les réseaux Hopfield et les réseaux ART (Adaptive Resonance Theory).



L'apprentissage des réseaux de neurones est la procédure qui consiste à estimer les paramètres des neurones du réseau, afin que celui-ci remplisse au mieux la tâche qui lui est affectée. On peut distinguer trois types d'apprentissages : l'apprentissage "supervisé", "non supervisé" et "semi-supervisé".

### 3.4.1 L'apprentissage non supervisé

Dans la réalité, l'homme a souvent beaucoup de mal à mémoriser, de façon individualisée, un ensemble d'objets, surtout quand ils sont en très grand nombre. L'homme préfère généralement catégoriser ces objets en classes en fonction de certaines propriétés communes ou en fonction d'un critère. Ce que nous retenons sont les caractéristiques de chaque classe, non les individus qui la composent [33]. En apprentissage non supervisé, la structuration des objets en classe homogènes est réalisée de façon automatique au moyen de méthodes implantées sur l'ordinateur. Le souhait du spécialiste est que ce dispositif automatique puisse lui proposer des regroupements ayant un sens pratique. Nous parlons d'apprentissage non supervisé, car l'utilisateur ne sait pas, a priori, quelles classes, groupes ou catégories il va obtenir. Ce mode d'apprentissage est également appelé "apprentissage sans professeur".

Les techniques employées sont appelées "méthodes de classification automatique" (clustering analysis), on s'intéressera aux méthodes basées sur les réseaux de neurones. Bien que cela ne soit pas leur seule caractéristique, ces techniques utilisent une structure de réseaux de neurones artificiels ; Un réseau de neurones non bouclé peut également être utilisé dans un but de visualisation ou d'analyse de données : On dispose d'un ensemble de données, représentées par un vecteur de grande dimension, et l'on cherche à les regrouper, selon des critères de ressemblance qui sont inconnus a priori. Ce type de tâches est connu en statistique sous le nom de méthodes "d'agrégation" (clustering) [34]. On peut utiliser les réseaux de neurones non bouclés pour réaliser une tâche assez voisine : à partir de données, décrites par des vecteurs de grande dimension, trouver une représentation de ces données dans un espace de dimension beaucoup plus faible (typiquement de dimension 2) tout en conservant les "proximités" ou "ressemblances" entre ces données. Il n'y a donc pas de "professeur", puisque c'est au réseau de découvrir les ressemblances entre les éléments de la base de données, et de les traduire par une proximité dans la "carte" de dimension 2 qu'il doit produire. On pourra résumer cela en disant que le processus d'apprentissage est incrémental, c'est à dire que les objets sont affectés séquentiellement à des groupes en fonction de leur proximité. Nous retrouvons, dans cette catégorie de méthodes, les cartes topologiques de Kohonen, et les techniques basées sur la résonance adaptative.

### 3.4.2 L'apprentissage supervisé

Dans cette catégorie d'apprentissage, nous avons une notion de valeurs ciblées, définie par les états d'une variable particulière, appelée "variable endogène", fixée a priori [33]. L'objectif est alors de mettre au point un processus permettant de calculer des valeurs de la variable endogène de façon automatique à partir d'autres informations. Quand la variable endogène est discrète, c.à.d. qu'elle prend un nombre restreint de valeurs, on parle alors de classe. D'où le travail d'apprentissage consiste à analyser les ressemblances entre les formes d'une même classe et les dissemblances entre les formes de classes différentes pour

en déduire les meilleures séparations possibles entre les classes.

### 3.4.3 L'apprentissage semi-supervisé (par renforcement)

L'apprentissage par renforcement est en fait une sorte d'apprentissage supervisé et certains auteurs le classent d'ailleurs, dans la catégorie des modes supervisés. Dans cette approche le réseau doit apprendre la corrélation entrée/sortie via une estimation de son erreur, c'est-à-dire du rapport échec/succès. Le réseau va donc tendre à maximiser un index de performance qui lui est fourni, appelé signal de renforcement. Le système étant capable ici, de savoir si la réponse qu'il fournit est correcte ou non, mais il ne connaît pas la bonne réponse.

### 3.4.4 L'apprentissage des perceptrons multicouches

L'apprentissage consiste en un entraînement du réseau. On présente au réseau des entrées et on lui demande de modifier sa pondération de telle sorte que l'on retrouve la sortie correspondante.

L'algorithme consiste dans un premier temps à propager vers l'avant les entrées jusqu'à obtenir une entrée calculée par le réseau. La seconde étape compare la sortie calculée à la sortie réelle connue. L'algorithme modifie alors les poids de telle sorte qu'à la prochaine itération, l'erreur commise entre la sortie calculée et connue soit minimisée. Malgré tout, il ne faut pas oublier que l'on a des couches cachées. On rétro-propage alors l'erreur commise vers l'arrière jusqu'à la couche d'entrée tout en modifiant la pondération.

On répète ce processus sur tous les exemples jusqu'au temps que l'on obtienne une erreur de sortie considérée comme négligeable.

#### Algorithme d'apprentissage de la rétro-propagation

L'algorithme d'apprentissage par rétro-propagation du gradient de l'erreur est un algorithme itératif quia pour objectif de trouver les poids des connexions minimisant l'erreur quadratique moyenne commise par le réseau sur l'ensemble d'apprentissage. Cette minimisation par une méthode du gradient conduit à l'algorithme d'apprentissage de rétro-propagation [35]. La fonction d'activation d'un neurone du PMC est souvent une fonction de type sigmoïde,

$$f(x) = \frac{1}{1 + e^{-x}} \text{ sa dérivée est : } f'(x) = (1 - f(x)) * f(x)$$

#### Algorithme

- **Initialiser les paramètres** : Initialiser tous les poids et les seuils par des nombres aléatoires.
- **Activation** :
  - Couche cachée :

$$y_i(P) = \text{sigmoïde} \left[ \sum_{i=1}^n x_i(P)w_{ij}(P) - \theta_j \right]$$

Avec n est le nombre d'entrée du neurone j.



- Couche de sortie :

$$y_k(P) = \text{signoide} \left[ \sum_{j=1}^n x_{jk}(P)w_{jk}(P) - \theta_k \right]$$

- Ajustement des poids :

- Calcul du gradient de l'erreur des neurones de la couche de sortie :

$$\begin{aligned} \delta_k(P) &= [1 - y_k(P)] \cdot y_k(P) \cdot e_k(P) \\ \Delta w_{ij}(P) &= \alpha x_j(P) \cdot \delta_k(P) \\ w_{ij}(P+1) &= w_{ij}(P) + \Delta w_{ij}(P) \end{aligned}$$

- Calcul du gradient de l'erreur des neurones de la couche cachée :

$$\begin{aligned} \delta_j(P) &= y_j(P) \cdot [1 - y_j(P)] \cdot \sum_{k=1}^l \delta_k(P) w_{jk}(P) \\ \Delta w_{ij}(P) &= \alpha x_j(P) \cdot \delta_j(P) \\ w_{ij}(P+1) &= w_{ij}(P) + \Delta w_{ij}(P) \end{aligned}$$

- **Itération suivante** : Passer à l'itération  $P + 1$  et retourner à l'étape d'activation ; répéter le processus jusqu'à satisfaction du critère d'arrêt.
- **Critères d'arrêt** : Si un nombre maximal d'itération a été atteint, ou si la variation des poids est inférieure à une limite fixée, ou si la variation du module du gradient est inférieure à une limite fixée (l'algorithme n'évolue pas), ou encore si le module du gradient de la fonction cout est inférieur à une valeur fixée (on est arrivé à un minimum ou dans un voisinage très proche), arrêter la procédure ; sinon, reprendre le calcul du gradient et effectuer de nouvelle itération.

**Remarque 2** Cependant, il existe plusieurs algorithmes d'apprentissages que l'on peut séparer en trois catégories distinctes. Une catégorie constituée d'algorithmes non applicable à des problèmes réels vue leur extrême lenteur (Descente du Gradient et Descente du gradient avec Momentum). La deuxième catégorie regroupe des algorithmes rapides et qui utilisent une technique heuristique pour modifier le "learning rate" et le momentum (variable learning rate et resilient back propagation). Enfin, la dernière catégorie rassemble des algorithmes rapides et qui font appel à deux optimisations particulières : l'optimisation de la Descente du Gradient Conjugué et la Méthode de Newton. La méthode développée par Levenberg-Marquardt est astucieuse car elle fait justement appel aux avantages de deux algorithmes : l'optimisation de Newton de la 3eme catégorie et la Descente du Gradient de la 1ere catégorie.

### 3.4.5 Sous-Apprentissage, généralisation et sur-apprentissage

La capacité de généralisation est une raison de base qui motive l'étude et le développement des réseaux de neurones artificiels. Elle peut être définie par la capacité d'élargir les connaissances acquises après apprentissage à des données nouvellement rencontrées par le réseau de neurones. C'est de cette façon que les réseaux de neurones sont capables d'approximer une fonction uniquement à partir d'une partie des données, ou encore d'associer un vecteur d'entrée qui n'a pas fait l'objet d'un apprentissage, à une classe. On

peut distinguer deux types de généralisation : locale et globale.

Dans l'approche locale, chaque neurone est associé à une région d'activation (région d'influence) localisée dans l'espace des données. Seule une partie des neurones participe donc à la réponse du réseau. Deux types d'architectures neuronales possèdent cette particularité : les Réseaux à base de Fonctions Radiales et la carte de Kohonen.

Par contre, dans l'approche globale, l'ensemble des neurones du réseau participe à l'élaboration de la sortie du réseau. L'information est donc distribuée dans le réseau tout entier. C'est le cas des réseaux de neurones de type Perceptron Multi Couches ou le modèle de Hopfield.

L'approche globale est supposée plus robuste aux pannes éventuelles de quelques neurones isolés. Par ailleurs, lors d'un apprentissage incrémental, des problèmes d'interférences catastrophiques peuvent apparaître : la modification des paramètres d'un neurone a des répercussions sur l'ensemble de la fonction modélisée par le réseau.

Des régions de l'espace des données éloignées de la région de la donnée à mémoriser risquent d'être perturbées. En d'autres termes, tandis que le réseau apprend dans une région de l'espace des données, le modèle peut oublier ce qu'il a appris dans d'autres régions.

La figure ci-dessous illustre la différence de généralisation entre les deux architectures neuronales caractérisées par les deux fonctions d'activation (locale pour le RFR et globale pour le PMC) :

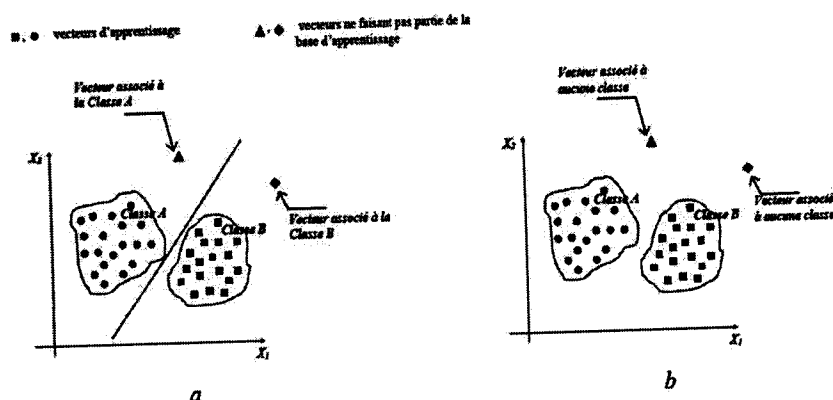


FIGURE 3.3 – Différence des capacités de généralisation entre le PMC (a) et les RFR (b).

La capacité de généralisation est très liée à la notion de sur-apprentissage (Over-fitting en anglais). Ces deux caractéristiques sont complètement antagonistes. On parle de sur-apprentissage quand le réseau a trop parfaitement appris les exemples proposés. Il sera donc incapable de généraliser.

Un indicateur utilisé pour étudier ce phénomène est la mesure de complexité  $k$  du système d'apprentissage (indicateur lié à la taille du réseau de neurones). En pratique, on calcule alors la moyenne des erreurs quadratiques - appelée 'erreur base apprentissage' - sur l'ensemble A de données d'apprentissage, et - 'erreur base test' - sur l'ensemble T de données de test (différent de celui d'apprentissage). Plus on agrandit l'ensemble A, plus l'erreur base apprentissage diminue, plus l'erreur base test augmente, ce qui implique la perte des capacités de généralisation. La figure suivante illustre clairement ce compromis entre sur-apprentissage, sous-apprentissage et bonne généralisation.

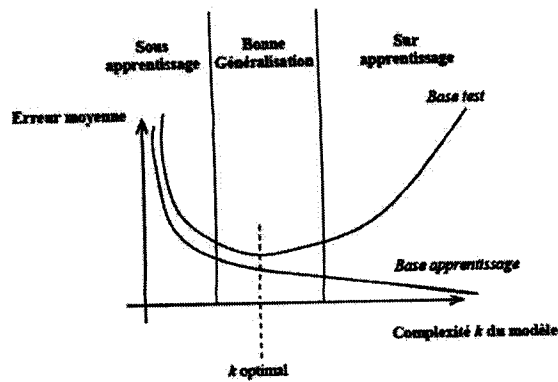


FIGURE 3.4 – Complexité dun réseau de neurones

### 3.5 Les étapes de mise en oeuvre d'un PMC

- Déterminer les entrées pertinentes, c.à.d. les grandeurs qui ont une influence significative sur le phénomène que l'on cherche à modéliser.
- Normalisation des données.
- Trouver le nombre de neurones cachés nécessaire pour obtenir une approximation satisfaisante.
- Estimer les valeurs des paramètres correspondant à un minimum de la fonction cout, c.à.d. d'effectuer un apprentissage.
- Evaluer les performances du réseau de neurones à l'issue de l'apprentissage.
- L'application du modèle généré par l'apprentissage.
- La dé-normalisation des données en sortie.

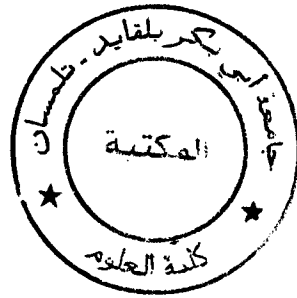
En fonction des résultats obtenus, il peut être nécessaire de lancer plusieurs itérations de tout ou partie de cette procédure.

### 3.6 Conclusion

A partir d'un modèle simple des neurones biologiques, on a construit un modèle plus complexe, celui des perceptrons multicouches. Ces outils permettent de calculer des fonctions vectorielles, adaptables à un ensemble d'exemples par le biais d'algorithmes d'optimisation utilisant la technique de la rétro-propagation.

De cette façon, on peut espérer faire apprendre une fonction complexe à un PMC, contenant assez peu de neurones. On obtiendra ainsi une modélisation analytique compacte d'une fonction obtenue expérimentalement.

Deuxième partie  
La Logique Floue



### 3.7 Introduction

L'homme perçoit, raisonne, imagine et décide à partir de modèles ou de représentations. Sa pensée n'est pas binaire. L'idée de la logique floue est de capturer l'imprécision de la pensée humaine et de l'exprimer avec des outils mathématiques appropriés. La résolution d'un problème passe par la recherche d'un modèle qui est le plus objectif et le plus certain possible. Ce modèle peut être défini en langage naturel ou par les mathématiques. Vu que notre environnement est incertain, dans certaines situations nous n'arriverons pas à définir qui représente la réalité avec certitude.

### 3.8 Etat de l'art

Les points de repères historiques suivants permettent de situer dans le temps le développement de la logique floue et ses applications [36] :

- En 1973 Lotfi propose d'appliquer la logique floue aux problèmes de réglage. En 1974, première application du réglage par la logique floue appliquée sur une turbine à vapeur.
- En 1975 E.H.Mandani expérimente un régulateur flou qu'il perfectionne au cours des années suivantes.
- En 1985 M.Sugeno décrit des applications industrielles possibles en régulation floue, tel que premières produits industriels (JAPON) utilisant le principe de la logique floue appliqué à des problèmes de réglage et de commande.
- En 1995 J.S.R.Jang élargit la logique floue aux systèmes à réseaux de neurones et l'intelligence artificielle.

### 3.9 Notion d'ensemble et de sous ensemble flou

Une des notions fondamentales dans les mathématiques est la notion d'ensemble. La théorie des ensembles a été créée par le mathématicien allemand Georg Cantor. Il a défini les ensembles comme des collections d'objets, de notre intuition ou de notre intellect, bien spécifiés et tous différents [37].

Dans son article " Fuzzy Sets " [38], L.A. Zadeh, le fondateur de la théorie des ensembles flous, commence par la définition de collection d'objets. Il donne comme exemple la collection des "nombres réels beaucoup plus grands que 1". Ce n'est pas un ensemble au même sens du terme qu'un ensemble de Cantor. Cette collection est définie imprécisément et un ensemble flou peut très bien exprimer cette imprécision. Zadeh définit un ensemble flou comme : "une collection telle que l'appartenance d'un élément quelconque à cette collection peut prendre toutes les valeurs entre 0 et 1" [38].

Par analogie avec les ensembles usuels, les ensembles flous permettent de définir des conditions (ou prédicats) graduelles pouvant intervenir dans des requêtes. Par nature même, un prédicat flou induit de la gradualité puisqu'il permet une transition souple entre la pleine satisfaction et le rejet total. Par suite, chacun des résultats délivrés par un tel prédicat est gradué et ordonné par rapport aux autres (sur l'échelle  $[0, 1]$ ). Les prédicats flous élémentaires correspondent aux adjectifs de la langue naturelle tels que "grand", "moyen" ou bien "jeune" qu'on appelle des variables linguistiques, et sont modélisées par une fonction (de forme trapézoïdale, triangulaire ou gaussienne...) d'un ensemble de domaines, cette fonction s'appelle fonction d'appartenance.

Nous pouvons résumer tous cela en disant que : les ensembles flous sont la mesure qui permettent de représenter et de manipuler des données imprécises et incertaines, et donc de rendre compte de concept de la vie courante, ou reflétant notre mode de perception, qui n'étaient pas pris en compte jusque-là par les approches classiques. En logique classique, le fait d'appartenir à une classe exclut en général le fait d'appartenir à une classe différente. L'intérêt de l'approche "floue" vient du fait qu'un élément peut appartenir simultanément à plusieurs classes, à des degrés divers.

### 3.9.1 Appartenance graduelle à un ensemble

La logique floue permet de caractériser une appartenance graduelle à un sous ensemble, appelé sous ensemble flou. On peut dire en cela que la théorie des sous-ensembles flous et la logique floue sont une généralisation de la théorie des ensembles classiques et de la logique booléenne, mais fondées sur un nouvel opérateur d'appartenance qui, au lieu d'être binaire, est multi-valué.

## 3.10 Les variables linguistiques

La description d'une certaine situation, d'un phénomène ou d'un procédé contient en général des expressions floues comme :

- Quelque, beaucoup, souvent,
- Chaud, froid, rapide, lent
- Grand, petit, etc.

Les expressions de ce genre forment les valeurs des variables linguistiques de la logique floue. D'où une variable linguistique représente un état dans le système à régler ou une variable de réglage dans un contrôleur flou. Sa valeur est définie dans des termes linguistiques d'un langage naturel ou artificiel.

### 3.11 Description des variables linguistiques par les fonctions d'appartenances

Chaque sous ensemble flou peut être représenté par sa fonction d'appartenance. Si l'ensemble de référence est un ensemble fini, les valeurs d'appartenances sont des valeurs discrètes définies dans  $[0, 1]$ . Si l'ensemble de référence est fini, on peut représenter ces valeurs comme des fonctions d'appartenance continues.

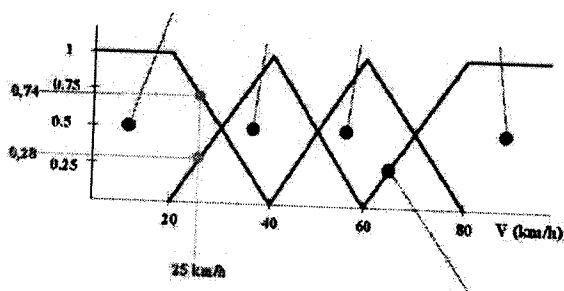


FIGURE 3.5 - Variables et Valeurs Linguistiques

### 3.12 Règles linguistiques

L'idée principale des systèmes basés sur la logique floue est d'exprimer la connaissance humaine sous la forme de règles linguistiques de forme Si... Alors... Chaque règle a deux parties :

- Partie antécédente (prémisse ou condition), exprimée par Si...
- Partie conséquente (conclusion), exprimée par Alors...

La partie antécédente est la description de l'état du système. La partie conséquente exprime l'action que l'opérateur qui contrôle le système doit exécuter. Chaque règle floue est basée sur l'implication floue.

Il y a plusieurs formes de règles **Si ... Alors ...**

La forme générale est :

*Si (un ensemble de conditions est satisfait) alors (un ensemble de conséquences peut être exécuté).*

Zadeh a été le premier à introduire la notion de règle floue sous la forme [39] :

*Si x est A, alors y est B.*

Mamdani [40] a proposé une forme de règle où les conclusions, comme les prémisses, sont des propositions floues :

*Si x est A et y est B, alors Z est C*

Takagi et Sugeno [41] ont proposé une forme de règle où les sous-ensembles flous figurent seulement dans la partie antécédente de la règle et où la partie conséquente est une équation précise de la variable d'entrée. La forme générale de cette règle est :

*Si x est A, alors y est f(x).*

### 3.13 Extension des opérations usuelles

Les opérations de l'algèbre relationnelle peuvent être étendues de façon mutuelle aux relations floues d'une part en les considérant comme des ensembles flous (pour ce qui est des opérations ensemblistes), d'autre part en introduisant des prédicats graduels au lieu des seuls prédicats booléens.

Il existe de nombreuses variantes dans ces opérateurs. Cependant, les plus répandues sont ceux dites "de Zadeh" décrites ci-dessous. En considérant deux relations A et B définies sur le même ensemble d'attributs X, et le degré de vérité d'une proposition A sera noté  $\mu(A)$  :

- |                                     |                                                |
|-------------------------------------|------------------------------------------------|
| - Union :                           | $\mu_{A \cup B}(X) = \max(\mu_A(X), \mu_B(X))$ |
| - Intersection :                    | $\mu_{A \cap B}(X) = \min(\mu_A(X), \mu_B(X))$ |
| - Complémentaire :                  | $\mu_{\bar{A}} = 1 - \mu_A$                    |
| - Différence : on part de l'égalité | $A - B = A \cap \bar{B}$                       |
| Et donc :                           | $\mu_{A-B}(X) = \min(\mu_A(X), 1 - \mu_B(X))$  |

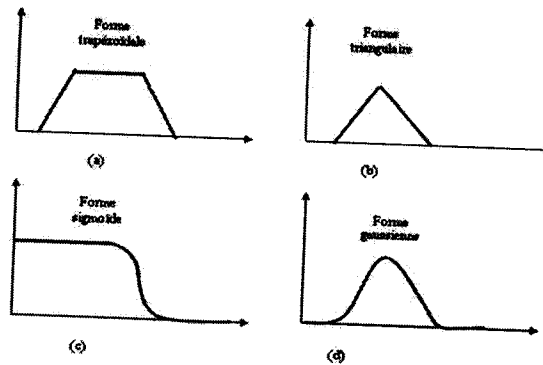


FIGURE 3.7 – Les formes de la fonction d'appartenance

### 3.15.2 L'inférence floue

Les degrés d'appartenance de chaque variable à chaque état permettent d'appliquer les règles floues qui ont été préalablement définies. Le degré d'appartenance des variables de sortie à chaque état est ainsi obtenu.

La stratégie de réglage dépend essentiellement des interférences adoptées. Les interférences lient les grandeurs mesurées et les variables de sortie par des règles linguistiques. Ces règles sont combinées en utilisant les connexions ET et OU. Il y a plusieurs sortes d'interférences floues, on s'intéressera aux deux méthodes suivantes :

#### Méthode de Takagi-Sugeno

Chaque fonction d'appartenance de la sortie est une combinaison linéaire de valeurs d'entrée. La sortie précise est la moyenne pondérée des poids d'activation et des sorties de fonctions d'appartenance. La méthode de Sugeno, où les fonctions d'appartenance sont des valeurs précises, constitue un cas particulier de cette méthode [37].

On parle parfois de méthode de Takagi-Sugeno simplifiée ou d'ordre zéro quand la conclusion est une constante. Pour un vecteur d'entrée  $x = (x_1, x_2, \dots, x_n)^i$  la sortie inférée est obtenue par :

- Calcul du degré d'appartenance de chaque entrée aux différents sous ensemble flous,

$$\mu_{A_j}^i(x_i)$$

- Calcul de la valeur de vérité de chaque règle,

$$\alpha_i(x) = ET(\mu_{A_1}^1(x_i), \dots, \mu_{A_n}^i(x_i))$$

- Calcul de la sortie du SIF,

$$y = \frac{\sum_i^r \alpha_i f(x_1, x_2, \dots, x_n)}{\sum_i^r \alpha_i}$$

La sortie obtenue n'est pas floue, ce qui supprime une étape dans l'inférence. Les SIF de type Takagi-Sugeno permettent donc le passage aisé d'une expression symbolique (la base de règle) à sa traduction numérique.



## Méthode de Mamdani

L'identification du modèle flou Mamdani s'effectue soit d'une manière analytique comme le cas du modèle Takagi-Sugeno soit en intégrant une connaissance à priori. Par exemple, un expert peut définir la partition des entrées et un certain nombre de règles. Le formalisme explicite du modèle permet d'unifier les règles définies à partir des données et celles formulées à partir d'une connaissance.

C'est un avantage quand les mesures, ou l'expertise, ou les deux sont incomplètes. Cependant, pour approcher le comportement global d'un système par un modèle flou Mamdani, un grand nombre de règles est nécessaire et le nombre de paramètres à identifier augmente. Malgré cet inconvénient, si un ensemble de règles linguistiques complet et compact est identifiable (toutes les situations sont traduites dans la base de règles), la précision du modèle flou Mamdani est correcte voire supérieure à celle des modèles Takagi-Sugeno. Pour un vecteur d'entrée  $x = (x_1, x_2, \dots, x_n)^i$  la sortie inférée est obtenue par :

- Calcul du degré d'appartenance de chaque entrée aux différents sous ensemble flous,

$$\mu_{A_j}^i(x_i)$$

- Calcul de la valeur de vérité de chaque règle,

$$\alpha_i(x) = ET(\mu_{A_1}^1(x_i), \dots, \mu_{A_n}^i(x_i))$$

- Calcul de la sortie du SIF,

$$y = OU(\mu_{A_1}^1(x_i), \dots, \mu_{A_n}^i(x_i))$$

En conclusion, les modèles flous Takagi-Sugeno et Mamdani sont deux représentations complémentaires et non concurrentes. Le choix d'une structure dépend de l'objectif recherché avec une précision fixée à l'avance qui peut être soit au niveau de la simulation du modèle, soit la connaissance explicite des relations liant les entrées et les sorties du système.

### 3.15.3 Défuzzification

En sortie, le système flou ne peut pas communiquer des valeurs floues qu'il peut seul exploiter. Il lui est donc nécessaire de fournir des valeurs précises, c'est le rôle de la défuzzification. Cette étape s'effectue toujours à l'aide des fonctions d'appartenance. A partir des degrés d'appartenance, on obtient autant de valeurs qu'il y a d'états. Pour déterminer la valeur précise à utiliser, il y a plusieurs méthodes proposées :

#### Méthode du centre de gravité

C'est une des méthodes les plus utilisées. Lorsque la fonction  $\mu(Z)$  est discrétisé, le centre de gravité est donnée par :

$$Z = \frac{\sum_{i=1}^n \mu_i Z_i}{\sum_{i=1}^n \mu_i}$$

Où  $n$  est le nombre de niveaux de quantification,  $Z_i$  la valeur de sortie pour le niveau  $i$  et  $\mu_i$  sa valeur d'appartenance [43].

### Méthode de moyenne de maximum

Cette méthode génère une commande précise en calculant la moyenne des valeurs pour lesquelles l'appartenance est maximale. Si la fonction est discrétisée la valeur de défuzzification est donnée par :

$$Z = \sum_{i=1}^l \frac{r_i}{l}$$

Où  $l$  est le nombre de valeurs quantifiées  $r$  pour lesquelles l'appartenance est maximale [37].

### Méthode de Tsukamoto

Si l'on utilise les fonctions d'appartenance monotones et non décroissantes pour la sortie, la valeur précise de la commande est calculée comme suit :

$$Z = \frac{\sum_{i=1}^n W_i Z_i}{\sum_{i=1}^n W_i}$$

Où  $n$  est le nombre de règles activées avec le poids  $W_i > 0$  et  $Z_i$  la valeur de la fonction de sortie pour la règle  $i$  [44].

### Méthode de moyenne pondérée

Cette méthode est utilisée lorsque les sorties sont définies comme fonction linéairement dépendantes d'entrées [37]. Si  $W_i$  est le poids d'activation de la règle  $i$ , la valeur précise de la commande est :

$$Z = \frac{\sum_{i=1}^n W_i f(x_i, y_i)}{\sum_{i=1}^n W_i}$$

Où  $n$  est le nombre de règles activées.

## 3.16 Conclusion

Le point le plus intéressant des systèmes flous réside dans l'absence de méthode standard pour la transformation de la connaissance ou de l'expérience humaine vers la base de règles linguistiques d'un système flou ; il n'y a pas de procédure générale pour choisir un nombre optimal de règles. De plus, lorsqu'un expert humain est disponible, la connaissance est plutôt incomplète que systématique. Des lors, il est indispensable de trouver une méthode automatique pour l'ajustement des fonctions d'appartenance, qui minimise une mesure d'erreur de la sortie du système ou un index de performance préétabli.

En résumé, il serait utile d'automatiser le processus de conception des systèmes flous. On pourrait diviser cette tâche en deux parties [37] :

- Identification de la structure du système (recherche du nombre idéal de règles),
- Identification des paramètres (ajustement des fonctions d'appartenance).

Ces deux raisons conduisent vers l'idée de fusionner l'approche des réseaux de neurones avec la technique floue. Nous allons aborder ce problème dans la partie qui suit.

**Troisième partie**  
**Les Systèmes Neuro-Flous**

### 3.17 Introduction

Les systèmes hybrides qui combinent la logique floue, les réseaux neurones, les algorithmes génétiques, et les systèmes experts prouvent leur efficacité dans une variété de problèmes de monde réel et dans l'industrie. Chaque technique intelligente a des propriétés particulières (par exemple capacité d'apprentissage, explication de décisions). Chaque technique convient à résoudre certains problèmes particuliers [45].

En effet, les réseaux neuronaux sont par exemple utilisés pour la reconnaissance des modèles. Cependant, ils sont incapables d'expliquer comment ils atteignent leurs décisions. Aussi pour les Systèmes flous qui raisonnent avec l'information imprécise et expliquer leurs décisions mais ne peuvent cependant pas acquérir automatiquement les règles qu'ils l'utilisent pour prendre ces décisions. Ces limites ont été une raison derrière la création de systèmes hybrides intelligents où ; deux ou plus de techniques sont combinées afin de vaincre les limitations d'une seule technique.

Le tableau ci-dessous regroupe une vue comparative entre les deux techniques.

Réseau de neurones	Système d'inférence flou
- La base de règle ne peut être utilisée	- La base de règle peut être utilisée
- L'apprentissage peut être démarré de zéro	- Pas d'apprentissage (utilise la connaissance linguistique)
- Boite noire	- Interprétable (la règle Si - Alors)
- Complexion des algorithmes d'apprentissage	- Universel et implémentation simple
- Difficulté pour extraire la connaissance	- La connaissance doit être disponible

TABLE 3.1 – Vue comparative entre les RNA et les SIF

Ces systèmes d'inférences flous (SIFs) possèdent deux points forts par rapport aux RNs. D'une part, ils sont généralement construits à partir de la connaissance humaine, d'autre part, ils ont une capacité descriptive élevée due à l'utilisation de variables linguistiques. Il est donc apparu naturel de construire des systèmes hybrides qui combinent les concepts des systèmes d'inférence flous et des RN. Ainsi l'approche Neuro-floue est apparue [46].

Plusieurs auteurs ont proposé des modèles flous comme : Nefcon, Falcon, Fun, Nefprox, Anfis. Dans le cadre de notre mémoire nous utilisons le modèle Nefclass.

### 3.18 ANFIS : *Adaptative Neuro-Fuzzy Inference System*

Le modèle ANFIS est un réseau neuro-flou proposé par Jang en 1993 [47]. Ce réseau est composé de cinq couches représentées dans la figure ci-dessous avec deux entrées X et Y. Les règles floues utilisées sont du type Takagi-Sugeno.

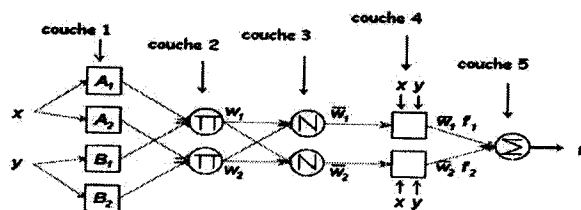


FIGURE 3.8 – Le modèle ANFIS

La première couche permet de convertir les variables d'entrée dans l'espace flou à l'aide des fonctions d'appartenance. Dans la deuxième couche chaque neurone est associé à une règle floue et permet de définir le poids de cette règle à l'aide d'une t-norme. La troisième couche a pour but de normaliser les différents poids et la quatrième permet d'obtenir une sortie définie par une fonction pondérée du premier ordre des entrées (approche de Takagi-Sugeno). La sortie Y dans la cinquième couche est définie par la somme des sorties de la couche précédente.

L'ajustement des paramètres de l'ANFIS est réalisé lors de la phase d'apprentissage. Pour cela un ensemble de données associant une séquence d'entrées et de sorties est nécessaire. Pour la réalisation de cette phase, l'algorithme d'apprentissage hybride est utilisé. L'algorithme d'apprentissage hybride est une association de la méthode de descente de gradient et de la méthode d'estimation des moindres carrés. La méthode de descente de gradient permet d'ajuster les prémisses (règles floues) en fixant les paramètres conséquents alors que la méthode LSM (Least Square Method) ajuste les paramètres conséquents en fixant les prémisses (conclusions) [48].

### 3.19 NEFCON : *Neuro-Fuzzy Control*

NEFCON vise à mettre en oeuvre le SIF de type Mamdani. Il est composé de trois couches : une couche d'entrée incluant les noeuds d'entrées et les sous ensemble flous d'antécédent, une couche cachée formé par des règles et un neurone de sortie pour les sous-ensembles flous des conséquences. NEFCON peut être employé pour apprendre des règles initiales, lorsque aucune connaissance du système n'est disponible, ou tout simplement pour optimiser une base de règle définie.

L'architecture de ce modèle est illustrée dans la figure suivante :

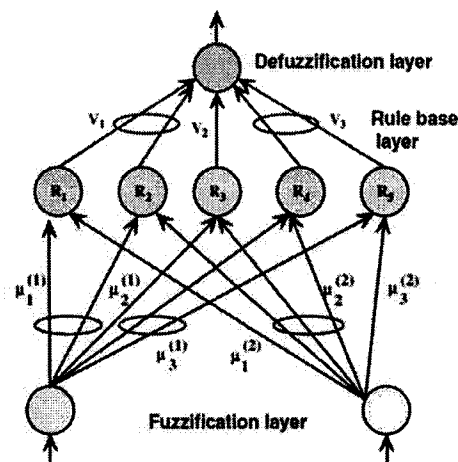


FIGURE 3.9 – Architecture de NEFCON

### 3.20 FALCON : *Fuzzy Adaptive Learning Control Network*

Comme nous remarquons dans la figure ci-dessous, Falcon est un réseau comportant cinq couches. Il possède deux noeuds linguistiques, une pour la sortie désiré et l'autre pour la sortie du Falcon.

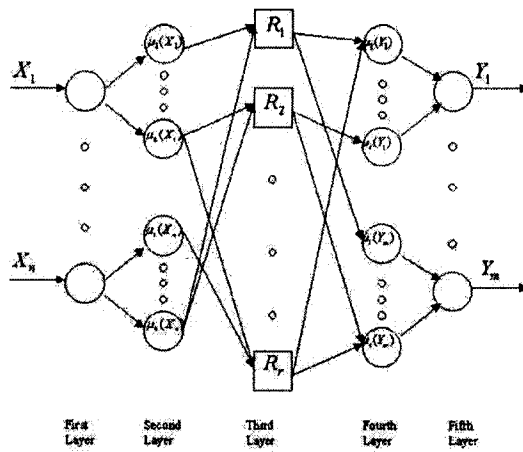


FIGURE 3.10 – Architecture de FALCON

La première couche cachée est responsable de la fuzzification des variables d'entrées tandis que la deuxième couche définit les parties antécédentes des règles floues suivie par les parties conséquences des règles dans la troisième couche cachée. FALCON utilise un apprentissage hybride pour l'ajustement des paramètres des fonctions d'appartenance afin de générer les sorties désirées.

### 3.21 FUN : *Fuzzy Net*

Pour FUN, les neurones dans la première couche cachée contiennent les fonctions d'appartenance pour appliquer la fuzzification des valeurs d'entrée. Dans la deuxième couche cachée, les conjonctions (fuzzy-AND) sont calculées. Les fonctions d'appartenance des variables de sortie sont stockées dans la troisième couche cachée. Leur fonction d'activation est un (fuzzy-OR).

A la fin, le neurone de sortie effectue la défuzzification. Le réseau est initialisé avec une base de règles floues et les fonctions d'appartenance correspondant. Il utilise une technique d'apprentissage qui change les paramètres des fonctions d'appartenance.

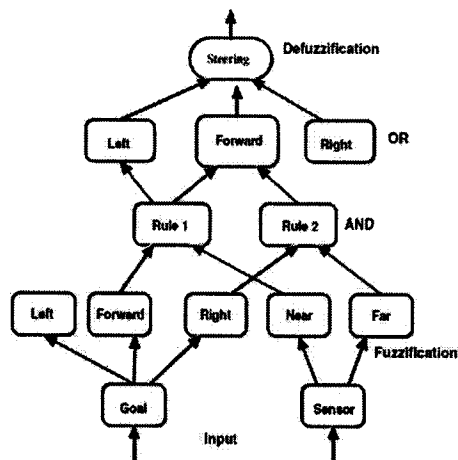


FIGURE 3.11 – Architecture du modèle FUN montant

### 3.22 NEFPROX : *Neuro-Fuzzy Function Approximation*

Le système NEFPROX est basé sur l'apprentissage supervisé et il est utilisé pour l'approximation des fonctions [49]. Il s'agit d'une version modifiée du modèle NEFCON sans l'apprentissage par renforcement. NEFPROX est très similaire à NEFCON et NEFCLASS à l'exception du fait que NEFCON a un noeud de sortie unique et le NEFCLASS n'utilise pas les fonctions d'appartenance dans la partie conclusion. Nous pouvons initialiser le système NEFPROX si nous avons déjà des règles appropriées sinon il est capable d'apprendre progressivement toutes les règles.

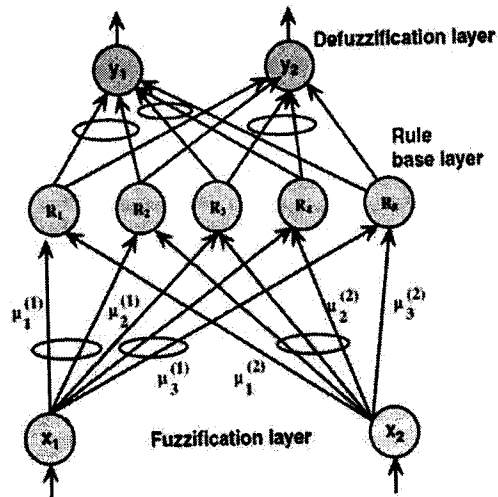


FIGURE 3.12 – Architecture du modèle NEFPROX

Notons que ANFIS est basé sur les modèles de Takagi-Sugeno uniquement, par contre NEFPROX peut adopter les deux modèles (Takagi-Sugeno et Mamdani). On outre NEFPROX donne des résultats rapidement en le comparant à ANFIS.

### 3.23 NEFCLASS : *Neuro-Fuzzy Classification*

Dans cette partie, nous parlons du modèle neuro-flou NEFCLASS. Ce modèle est utilisé généralement en classification, à partir d'un ensemble de données, il peut générer des règles floues et adapter les sous-ensembles flous après un apprentissage supervisé, basée sur la rétro-propagation du gradient.

L'objectif principal du NEFCLASS est la création d'un classificateur interprétable qui fournit également une précision acceptable. Toutefois, l'utilisateur doit être conscient, que l'interprétabilité et la précision ne vont pas ensemble à cent pour cent.

Un système flou interprétable doit comporter les caractéristiques suivantes :

- Peu de règles constructives avec peu de variables dans leurs antécédents,
- Le minimum de sous-ensembles flous pour chaque variable,
- Il n'existe pas de règles pondérées
- Des termes linguistiques identiques sont représentés par des ensembles flous identiques,

Ces caractéristiques constituent des contraintes lors de l'apprentissage du modèle flou. si nous basons lors de l'apprentissage des modèles flous uniquement sur la précision, nous risquons d'obtenir des modèles de type boîte noire. Un utilisateur doit donc choisir entre :

### la précision ou l'interprétabilité.

NEFCLASS fournit des moyens pour assurer une lisibilité complète de la solution en donnant à l'utilisateur la main pour contrôler le processus d'apprentissage. Il convient également de souligner que la solution interprétable peut généralement ne pas être obtenue sans la coopération de l'utilisateur. L'utilisateur doit décider si la lisibilité de la solution est suffisante ou non, et il doit être prêt à intervenir dans le processus d'apprentissage lorsque cela est nécessaire.

Le modèle NEFCLASS doit être considéré comme un outil qui aide les utilisateurs à construire des classifieurs neuro-flou interprétables. Ce n'est pas un concepteur automatique de classificateur où les données sont présentées à l'intérieur et une solution est retournée. Il est nécessaire que l'utilisateur utilise cet outil d'une manière interactive [50]. Pour cette raison les stratégies d'apprentissage rapides sont utilisées afin de donner à l'utilisateur la possibilité d'interagir avec l'outil.

### 3.23.1 La structure du NEFCLASS

Il est possible de voir un système neuro-flou comme un réseau de neurones à trois couches où :

- La première couche représente les variables d'entrée (les attributs de l'exemple),
- La couche cachée représente l'ensemble des règles floues,
- La troisième couche représente sorties, chaque sortie représente une classe,
- Les neurones utilisent des t-normes et des t-conormes comme fonctions d'activation,
- Les ensembles flous sont codés comme des poids de connexions floues.

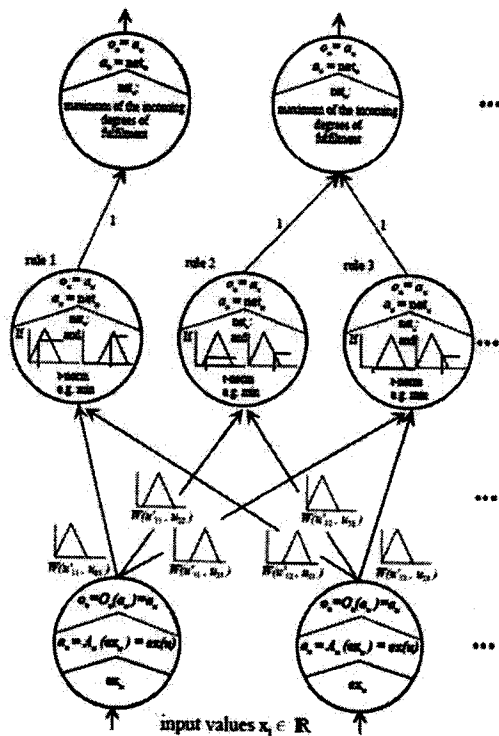


FIGURE 3.13 – Une vue de réseau de neurones d'un classifieur neuro-flou



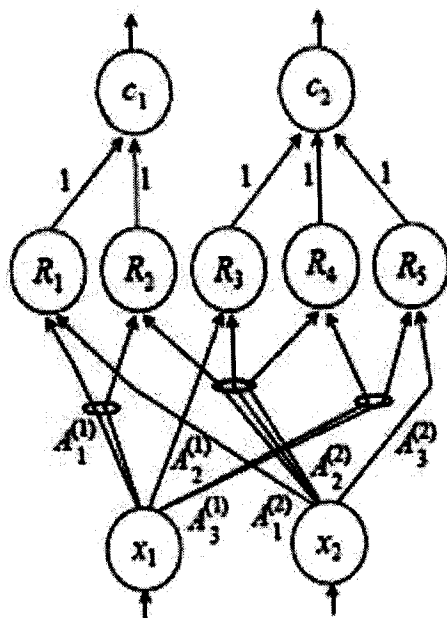


FIGURE 3.14 – L'architecture du modèle NEFCLASS

La figure (3.13) et la figure (3.14) présentent la structure du NEFCLASS. La figure (3.14) est une version simplifiée de la figure (3.13). Les  $R_k$  représentent les règles et  $A_j^{(i)}$  les poids  $W(x_i, R_k)$ , où l'indice  $j$  sélectionne les sous-ensembles flous de la partition. NEFCLASS utilise des poids partagés sur certaines connexions (représentés par des cercles sur des connexions, voir la figure). Afin de garder la sémantique des règles, chaque ensemble flous maintient son sens pour une variable d'entrée donnée.

$W(R_k, c_m)$  est la connexion de la règle  $R_k$  à l'unité de sortie  $c_m$ . Pour des raisons sémantiques, c'est à dire d'éviter les règles pondérées, ces connexions sont fixées à 0 (pas de connexion) ou 1 (connexion existante). Chaque règle correspond à une seule sortie [50]. L'activation de la sortie est calculée par une l'opération maximum au lieu de la somme pondérée. Nous expliquons par la suite les deux modes d'apprentissages adopter (paramétrique et structurel) afin de garder un compromis entre la précision et l'interprétabilité.

### 3.23.2 L'apprentissage de la base de règles (*structurel*)

Un système NEFCLASS peut être construit à partir des connaissances partielles sur les exemples, et peut être affiné par l'apprentissage, ou il peut commencer par une base de règles vide complétée par la création de règles à partir des données d'apprentissage. Pour chaque variable d'entrée, l'utilisateur doit décider du nombre des partitions floues utilisées. Généralement, l'utilisateur peut choisir deux partitions floues ou plus et au même temps il peut sélectionner le nombre maximum de règles. Au minimum chaque classe possède une règle [50].

Dans notre travail, nous avons choisi des fonctions d'appartenances triangulaires décrites par trois paramètres :

$$\mu : \mathbb{R}^n \rightarrow \mu(x) = \begin{cases} \frac{x-a}{b-a} & \text{si } x \in [a, b[ \\ \frac{c-x}{c-b} & \text{si } x \in [b, c] \\ 0 & \text{sinon} \end{cases}$$

On considère un système NEFCLASS avec :

- $n$  unités d'entrées  $x_1, \dots, x_n$ ,
- $k \leq k_{max}$  règles initiales  $R_1, \dots, R_k$  (connaissance a priori,  $k = 0$  veut dire pas de connaissance donnée),
- $m$  unités de sorties  $c_1, \dots, c_m$ ,
- Un ensemble d'apprentissage  $\tilde{L} = (p_1, t_1), \dots, (p_s, t_s)$  de  $s$  patterns, chaque couple est une entrée  $P \in R^n$ , et une classe  $t \in 0, 1^m$ .

Supposons que NEFCLASS est initialisé avec  $k \leq k_{max}$  règles floues. La base de règles du NEFCLASS est complétée en donnant pour chaque pattern  $p$  une combinaison des ensembles flous qui donne le plus haut degré d'appartenance pour chaque valeur de  $p_i$ . Si un antécédent d'une règle n'existe pas, il est ajouté dans une liste. La somme des degrés de vérité dans l'antécédant, nous permet de fixer la classe correspondante (choix de la somme maximale).

Après le traitement de chaque exemple d'apprentissage, nous obtenons une base de  $k'$  règles. Si  $k' > k_{max}$ , seulement les meilleur  $k_{max}$  règles ("Best Rule Learning"), ou les meilleures ( $k_{max}/m$ ) règles pour chaque classe ("Best per Class") sont conservées, toutes les autres règles sont supprimées de la base de règles.

Les meilleures règles sont déterminées par le calcul de leurs de performance. Si une règle classe correctement un exemple, son degré d'activation (degree of fulfillment) est ajouté à sa valeur de performance, si non, le degré d'activation est soustrait.

#### Algorithme : le pseudo code de l'algorithme d'apprentissage

```

For each pattern  $(p, t)$  of  $\tilde{L}$  do
  begin
    For each input feature do
      find  $\mu_{ji}^{(i)}$  such that  $\mu_{ji}^{(i)}(p_i) = \max_{j \in \{1, \dots, q_j\}} \{ \mu_j^{(i)}(p_i) \}$ 
      Create antecedent  $A = (\mu_{j1}^{(1)}, \dots, \mu_{jn}^{(n)})$ 
    If  $A$  is not in list of antecedents
      then add antecedent  $A$  to list of antecedents;
    end;
  For each pattern  $(p, t)$  of  $\tilde{L}$  do
    For each antecedent  $A_j$  do
      begin
         $c$  = class index of  $p$  given by  $t$ ;
         $C_j(c) = C_j(c) + A_j(p)$  (* add degree of fulfillment *)
      end;
    For each antecedent  $A_j$  do
      Begin
         $c = \underset{i \in \{1, \dots, m\}}{\operatorname{argmax}} \{ C_j(i) \}$ ;
        create rule  $R_j$  with antecedent  $A_j$  and consequent  $c$ ;
        add  $R_j$  to the list of rule base candidates;
         $performance_j = C_j(c) - \sum_{i \in \{1, \dots, m\}, i \neq c} C_j(i)$ 
      end;
  
```

```

If "Best RuleLearning"
  then For i = 1 to k_max do
    begin
       $R = \underset{R_j}{\operatorname{argmax}} \{performance_j\};$ 
      add R to rule base;
      delete R from list of rule candidates;
    end;
else If "Best Per Class Rule Learning"
  then For each class c do
    For i = 1 to  $\frac{k_{max}}{m}$  do
      begin
         $R = \underset{R_j, \text{consequent}_j=c}{\operatorname{argmax}} \{performance_j\};$ 
        add R to rule base;
        delete R from list of rule candidates;
      end;

```

L'algorithme d'apprentissage peut être visualisé via un exemple dans une structure de grille. La figure (3.15) montre comment les règles sont sélectionnées à partir des exemples. Dans cet exemple, le système est autorisé à créer trois règles.

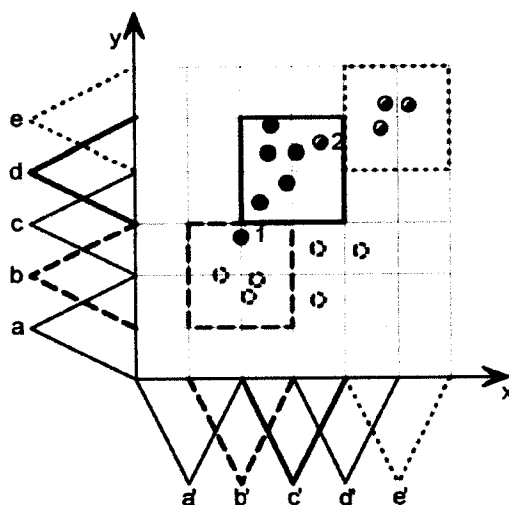


FIGURE 3.15 – Classification avec NEFCLASS après l'apprentissage des règles

Nous remarquons que le résultat de classification n'est pas mauvais, mais des améliorations sont souhaitées. Les exemples 1 et 2 sont mal classés et il y a aussi trois autres qui ne sont pas classés. Le déplacement et la modification des ensembles flous vont améliorer la classification :

L'exemple 1 pourrait être correctement classé si nous modifions l'emplacement des ensembles flous  $b$ ,  $d$  et  $c'$ . De même l'exemple 2 pourrait être correctement classé si nous modifions l'emplacement des ensembles flous  $e'$  et  $c'$ , et les exemples non classés pourraient être correctement classés si l'ensemble flou  $b'$  est un peu plus large vers la droite.

### 3.23.3 L'apprentissage des sous-ensembles flous (*paramétrique*)

Le NEFCLASS utilise un algorithme d'apprentissage supervisé et itératif pour adapter ses sous-ensembles flous jusqu'à ce qu'un critère donné est vérifié, par exemple si un taux d'erreur minimal de classification admissible est atteint.

L'attribut responsable du degré d'activation est identifiée et son sous-ensemble flou sera adapté en conséquence. Un ensemble flou est modifié si et seulement s'il ne viole pas les contraintes spécifiées par l'utilisateur. Par exemple on peut avoir les contraintes suivantes :

- Les ensembles flous doivent se chevaucher dans une mesure fixe (par exemple le sous-ensemble flou petit ne doit pas englober le sous-ensemble moyen).
- Un ensemble flou ne doit pas dépassé un autre ensemble (c.-à-d. échanger ler positions relatif).
- Ect.

Les utilisateurs peuvent sélectionner une ou plusieurs contraintes en fonction de leurs besoins. Les contraintes de ce genre aident à obtenir une base de règles interprétables, mais peut entraîner une perte de performance en classification [50].

#### Algorithme : le pseudo code de l'algorithme d'apprentissage

```
Repeat Propagate the next pattern ( $p, t$ )
  For each output unit  $c_i$  do
     $ec_i = t_i - activation(c_i)$ 
  For each rule unit  $R$  with activation ( $R$ ) > 0 do
    Begin
       $e_R = activation(R) \cdot (1 - activation(R)) \cdot \sum_{c_i} (W(R, c) \cdot ec_i)$ 
       $j = \arg \min_{i \in \{1, \dots, n\}} \{(W(x_i, R) \cdot (p_i))\}$ ;
       $\mu = W(x_i, R)$ 
      (*  $\mu_a, \mu_b, \mu_c$  are the parameters of the fuzzy set  $\mu$  *)
       $\delta_b = \sigma \cdot e_R \cdot (c_\mu - a_\mu) \cdot sgn(p_i - b_\mu)$ ;
       $\delta_a = -\sigma \cdot e_R \cdot (c_\mu - a_\mu) + \delta_b$ ;
       $\delta_c = \sigma \cdot e_R \cdot (c_\mu - a_\mu) + \delta_b$ ;
      Modify  $\mu$  with  $\delta_a, \delta_b, \delta_c$  without violating the constraints for  $\mu$ ;
    End;
  Until end criterion;
```

La procédure d'apprentissage pour les sous-ensembles flous est une heuristique simple. Il en résulte dans la modification des supports des fonctions d'appartenance (les rendre moins ou plus large). En changeant seulement l'ensemble flou qui a donné le plus petit degré d'appartenance pour l'exemple actuel, on assure que les changements sont aussi réduits que possible.

La somme dans le calcul de  $e_R$  (ligne 7 de l'algorithme) n'est pas vraiment nécessaire, parce que chaque règle est liée à une seule unité de sortie (i.e. il y a un seul  $W(R, c) \neq 0$ ). Mais il rend le modèle plus souple, car il serait également possible d'utiliser des poids de règles. Bien que la mise en oeuvre du modèle NEFCLASS permet d'utiliser des poids de règle, il est recommandé de ne pas les utiliser afin de conserver la sémantique du système NEFCLASS.

La procédure d'apprentissage ne peut pas atteindre une valeur d'erreur de zéro, et donc la variation de l'erreur est généralement utilisée comme un critère d'arrêt pour l'algorithme d'apprentissage.

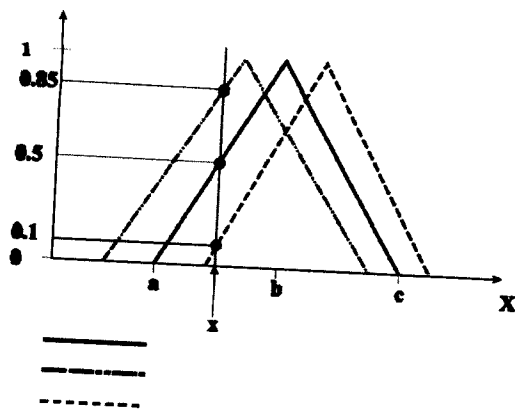


FIGURE 3.16 – Adaptation dun sous-ensemble

L'adaptation d'un ensemble flou est effectuée en changeant simplement les paramètres de sa fonction d'appartenance suivant le changement du degré d'appartenance de la valeur caractéristique actuelle.

La figure (3.17) montre la situation après l'apprentissage des sous-ensembles flous qui a mené à améliorer le résultat de la classification. Ici, aucune contrainte n'a été utilisée pour restreindre le processus d'apprentissage. Comme on le voit, les nouvelles partitions floues ne sont plus très bien interprétables.

Un tel résultat indique qu'il faut répéter le processus d'apprentissage avec d'autres paramètres. Dans ce cas, il aurait été préférable de permettre au système de créer quatre règles et d'utiliser des contraintes pour la création des fonctions d'appartenance. Cet exemple montre qu'il est important pour l'utilisateur de travailler de manière interactive avec des approches telles que NEFCLASS pour obtenir des solutions interprétables [50].

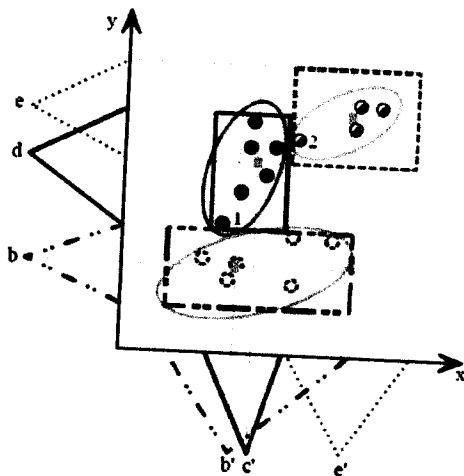


FIGURE 3.17 – Classification après l'apprentissage des sous-ensembles flous

Du point de vue de l'architecture NEFCLASS et le flux de données, les ensembles flous sont formés par l'algorithme de rétro-propagation. Ce modèle est facile à mettre en application, il évite l'étape de défuzzification, tout en étant précis dans le résultat final, avec une rapidité bien supérieure aux autres architectures.

**Quatrième partie**  
**Optimisation par Essaim**  
**Particulaires**

### 3.24 Introduction

L'optimisation est une branche des mathématiques qui permet de résoudre des problèmes en déterminant le meilleur élément d'un ensemble selon certains critères prédéfinis. De ce fait, l'optimisation est omniprésente dans tous les domaines et évolue sans cesse depuis Euclide.

En 1995, Russel Eberhart, ingénieur en électricité, et James Kennedy, socio-psychologue, s'inspirent du monde du vivant pour mettre en place une méta-heuristique : l'optimisation par essaim particulaire. Cette méthode se base sur la collaboration des individus entre eux : chaque particule se déplace et à chaque itération, la plus proche de l'optimum communique aux autres sa position pour qu'elles modifient leur trajectoire. Cette idée veut qu'un groupe d'individus peu intelligents puisse posséder une organisation globale complexe.

De part sa récence, de nombreuses recherches sont faites sur la P.S.O., nous avons donc, afin de comprendre l'efficacité et les limites de cette approche, appliqué cette méthode au système NEFCLASS, un classificateur neuro-flou, qui se base sur la descente de gradient pour faire l'apprentissage de son réseau de neurone. La mise en oeuvre et les résultats seront détaillés dans le dernier chapitre.

Tout d'abord, nous présentons l'optimisation par essaim particulaire.

### 3.25 Intérêt de l'algorithme PSO

Pour de nombreux problèmes, il n'existe pas de solution déterministe qui donne le résultat en un temps raisonnable, et ceci malgré la création d'ordinateurs de plus en plus performants. Pour pallier à ce problème, on a recours à des méthodes dites heuristiques, c'est-à-dire des méthodes qui fournissent une solution approchée. Toutefois, il faut reproduire le processus sur plusieurs itérations pour tendre vers une solution acceptable.

On retrouve parmi ces heuristiques, certains algorithmes qui possèdent un principe générique adaptable et qui s'applique donc à plusieurs problèmes d'optimisation. On les appelle des méta-heuristiques.

La plus courante est la descente stochastique : on part d'une solution initiale, on la compare à tous ses voisins en conservant à chaque fois le meilleur résultat.

L'optimisation par essaim particulaire, qui dérive de la descente stochastique, entre dans cette famille d'algorithmes. Elle s'inspire fortement des relations grégaires des oiseaux migrateurs qui doivent parcourir des longues distances et qui doivent donc optimiser leurs déplacements en termes d'énergie dépensée, comme par exemple la formation en V.

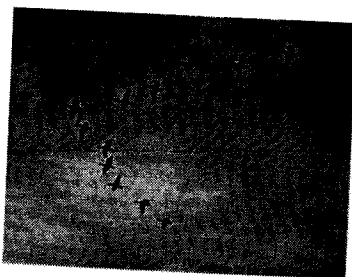


FIGURE 3.18 – Volée d'Anser en formation en V

### 3.26 Les éléments de la P.S.O.

L'algorithme de base de la P.S.O. travaille sur une population appelée essaim de solutions possibles, elles-mêmes appelées particules. Ces particules sont placées aléatoirement dans l'espace de recherche de la fonction objectif.

Pour appliquer la PSO il faut définir un espace de recherche constitué de particules et une fonction objective à optimiser. Le principe de l'algorithme est de déplacer ces particules afin qu'elles trouvent l'optimum.

Chacune de ces particules est dotée :

- D'une position, c'est-à-dire ses coordonnées dans l'ensemble de définition.
- D'une vitesse qui permet à la particule de se déplacer. De cette façon, au cours des itérations, chaque particule change de position. Elle évolue en fonction de son meilleur voisin, de sa meilleure position, et de sa position précédente. C'est cette évolution qui permet de tomber sur une particule optimale.
- D'un voisinage, c'est-à-dire un ensemble de particules qui interagissent directement sur la particule, en particulier celle qui a le meilleur critère.

A tout instant, chaque particule connaît :

- Sa meilleure position visitée. On retient essentiellement la valeur du critère calculée ainsi que ses coordonnées.
- La position du meilleur voisin de l'essaim qui correspond à l'ordonnement optimal.
- La valeur qu'elle donne à la fonction objectif car à chaque itération il faut une comparaison entre la valeur du critère donnée par la particule courante et la valeur optimale.

On se rend compte, en accord avec Maurice Clerc et Patrick Siarry [51], que l'évolution d'une particule est finalement une combinaison de trois types de comportements : égoïste (suivre sa voie suivant sa vitesse actuelle), conservateur (revenir en arrière en prenant en compte sa meilleure performance) et panurgien (suivre aveuglement le meilleur de tous en considérant sa performance).

On voit alors que la bio-inspiration à l'origine de l'optimisation par essaim particulaire ressort dans l'algorithme sous la forme d'une intelligence collective : coordination du groupe, instinct individuel et interaction locale entre les individus (grognements, phéromones...). On observe donc un compromis psycho-social entre confiance en soi et influence des relations sociales.

### 3.27 Notion de voisinage

Le voisinage d'une particule est le sous-ensemble de particules de l'essaim avec lequel il a une communication directe. Ce réseau de rapports entre toutes les particules est connu comme la sociométrie, ou la topologie de l'essaim.

Il existe deux principaux types de voisinage, géographique et social.

#### 3.27.1 Les voisinages géographiques

Les voisins sont considérés comme les particules les plus proches. Cependant, à chaque itération, les nouveaux voisins doivent être recalculés à partir d'une distance prédéfinie dans l'espace de recherche. C'est donc un voisinage dynamique [52].



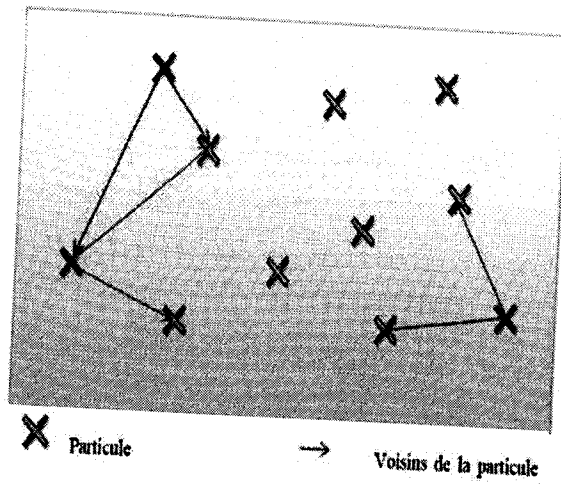


FIGURE 3.19 – Un voisinage géographique [52]

Dans cet exemple, le voisinage de la particule est composé des deux particules les plus proches.

### 3.27.2 Les voisinages sociaux

Les voisins sont définis à l'initialisation et ne sont pas modifiés ensuite. C'est le voisinage le plus utilisé, pour plusieurs raisons :

- Il est plus simple à programmer.
- Il est moins coûteux en temps de calcul.
- En cas de convergence, un voisinage social tend à devenir un voisinage géographique.

Pour ce faire, on dispose (virtuellement) les particules en cercle puis, pour la particule étudiée, on inclut progressivement dans ses informatrices, d'abord elle-même, puis les plus proches à sa droite et à sa gauche, jusqu'à atteindre la taille voulue. On peut aussi choisir les informatrices au hasard [52].

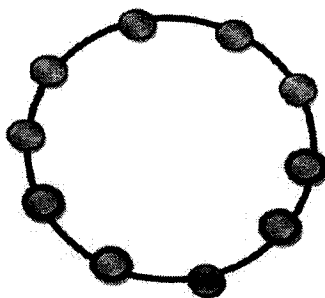


FIGURE 3.20 – Un voisinage en cercle [52]

Dans cet exemple, la particule principale est en bas et ses informatrices correspondent au deux particules directement à sa droite et à sa gauche.

### 3.28 Coefficients de confiance

Les variables de confiance pondèrent les tendances de la particule à vouloir suivre son instinct de conservation ou son panurgisme. Les variables aléatoires  $r_1$  et  $r_2$  peuvent être définis de la façon suivante :

$$p_1 = r_1 c_2$$
$$p_2 = r_2 c_3$$

Où  $r_1$  et  $r_2$  suivent une loi uniforme sur  $[0..1]$  et  $c_2$  et  $c_3$  sont des constantes positives déterminées de façon empirique et suivant la relation  $c_2 + c_3 \leq 4$ .

### 3.29 Vitesse maximale

Afin d'éviter que les particules ne se déplacent trop rapidement dans l'espace de recherche, passant éventuellement à côté de l'optimum, il peut être nécessaire de fixer une vitesse maximale (notée  $V_{max}$ ) pour améliorer la convergence de l'algorithme.

### 3.30 Facteur d'inertie

Le facteur d'inertie  $C_1$  - introduit par SHI et EBERHART - permet de définir la capacité d'exploration de chaque particule en vue d'améliorer la convergence de la méthode. Une grande valeur de  $C_1 (> 1)$  est synonyme d'une grande amplitude de mouvement et donc, une exploration globale. A contrario, une faible valeur de  $C_1 (< 1)$  est synonyme de faible amplitude de mouvement et donc, une exploration locale. Fixer ce facteur, revient donc à trouver un compromis entre l'exploration locale et l'exploration globale. La taille du facteur d'inertie influence directement la taille de l'hyper-espace exploré et aucune valeur de  $C_1$  ne peut garantir la convergence vers la solution optimale.

### 3.31 Initialisation de l'essaim

La position des particules ainsi que leur vitesse initiale doivent être initialisés aléatoirement selon une loi uniforme sur  $[0..1]$ . Cependant, en ce qui concerne la position des particules, il est préférable d'utiliser un générateur de séquence de SOBOL qui est plus pertinent dans la disposition homogène des particules dans un espace de dimension  $n$ .

### 3.32 Critères d'arrêt

Comme indiqué précédemment, la convergence vers la solution optimale globale n'est pas garantie dans tous les cas de figure même si les expériences dénotent la grande performance de la méthode. De ce fait, il est fortement conseillé de doter l'algorithme d'une porte de sortie en définissant un nombre maximum d'itération.

L'algorithme doit alors s'exécuter tant que l'un des critères de convergence suivant n'a pas été atteint :

- le nombre maximum d'itération a été atteint ;
- la variation de la vitesse est proche de 0 ;
- le fitness de la solution est suffisant.

# Chapitre 4

## Expérimentations, Résultats et Interprétations

### 4.1 Introduction

La classification des pathologies dans le domaine médical a pour but de rassembler en classes (groupes) les cas qui ont des similitudes biologiques fondamentales et qui sont susceptibles de partager certains facteurs pathologiques. Il est important de faire la reconnaissance de ces classes car cela permet non seulement de comprendre le processus de la maladie mais aussi d'instaurer l'approche thérapeutique adéquate. Après l'étape de reconnaissance, il est intéressant de pouvoir faire l'identification de cas en les attribuant à la classe qui leur convient.

Tel que présenté dans le chapitre 2, plusieurs méthodes de classification comprenant des approches issues de multiples domaines ont été utilisées dans l'aide au diagnostic du cancer du sein. Les approches neuronales présentées dans la revue de littérature montrent l'applicabilité de plusieurs types de réseaux de neurones artificiels. Les taux de classification obtenus à travers les différents travaux sont très intéressants. Cependant, un prétraitement des données était requis dans plusieurs cas d'application, et les systèmes proposés sont pour la plupart des systèmes hybrides qui comprennent un réseau de neurones et un autre ou plusieurs autres paradigmes de l'intelligence artificielle. Cela engendre une certaine complexité à la réalisation. De plus, dans tous les travaux présentés, il était question d'une classification de données selon 2 ou 3 classes tout au plus ; pour les problèmes réels dans le domaine du dépistage du cancer du sein, on a souvent affaire à plusieurs cas de diagnostic, et donc un plus grand nombre de classes.

Nous avons pour but dans ce travail de tester l'efficacité du système neuro-flou optimisé (OptNefclass) par rapport au Nefclass classique, et ce, pour faire la classification de cas de tumeur dans le cadre du cancer du sein. Dans ce contexte, nous avons exploité une base de données en premier lieu sur notre système. Puis nous avons fait une étude comparative avec les résultats déjà existant dans la littérature.

tant donné que les paramètres d'un réseau neuro-flou, entre autres son architecture, dépendent beaucoup des données, on commence par définir celles-ci.

## 4.2 Description de la base de données

### 4.2.1 Historique de la base de données

Depuis 1988, l'Université du Wisconsin alimente une base avec des données médicales (WBCD pour Wisconsin Breast Cancer Data) concernant des prélèvements cytologiques (épaisseur, uniformité de la taille et de la forme des cellules, etc.) avec leurs diagnostics : bénin ou malin [54].

Ces échantillons arrivent sur plusieurs groupes :

- 1er groupe : 367 cas (janvier 1989).
- 2ème groupe : 70 cas (octobre 1989).
- 3ème groupe : 31 cas (février 1990).
- 4ème groupe : 17 cas (avril 1990).
- 5ème groupe : 48 cas (avril 1990).
- 6ème groupe : 49 cas (janvier 1991).
- 7ème groupe : 31 cas (juin 1991).
- 8ème groupe : 86 cas (novembre 1991).

### 4.2.2 Les descripteurs

La base comporte environ 683 exemples qui sont caractérisés par neuf attributs décrit dans le tableau suivant :

Nombre d'attributs		9	
Attributs	Index	Nom de l'attribut	Moyenne
	A1	Épaisseur de l'échantillon	4.44
	A2	Uniformité de la taille	3.15
	A3	Uniformité de la forme	3.22
	A4	Adhésion marginale	2.83
	A5	Taille cellule épithéliale	3.23
	A6	Noyaux	3.54
	A7	Chromatine terne	3.45
	A8	Nucleoli normal	2.87
	A9	Mitose	1.60
Valeur Min		1.0	
Valeur Max		10.0	
Étiquette des classes		Bénin, Malin	
Nombre de patterns		683	

TABLE 4.1 - Attributs de la base de diagnostic

La base contient 239 cas malin (codé avec le chiffre '4') et 444 cas bénin (codé avec le chiffre '2').

## 4.3 La structure utilisée pour l'optimisation

Pour l'optimisation du système Nefclass classique, nous avons utilisé une méta-heuristique pour l'apprentissage du réseau de neurone.

Le Nefclass utilise une heuristique plus simple que la descente de gradient pour l'apprentissage de son réseau de neurone (décrite dans la troisième partie du troisième chapitre), mais cette approche tombe dans le problème du minimum local. Pour remédier à ce problème nous proposons une autre méthode basée sur les essais particuliers connus par leur performance à trouver le minimum global d'une fonction donnée.

### 4.3.1 Configuration de la méthode

A première vue, il semblerait que de nombreux paramètres sont à prendre en compte pour l'application de l'optimisation par essaim particulière sur le système Nefclass. Toutefois, la plupart d'entre eux peuvent être fixés, d'autres au contraire ne peuvent être définis qu'empiriquement.

C'est le cas, par exemple de la taille de l'essaim. La quantité de particules allouées dépend essentiellement de deux paramètres : la taille de l'espace de définition, et le rapport entre la capacité de calcul de la machine et le temps maximum de recherche. Le meilleur moyen d'affiner ce coefficient est donc de faire de nombreux essais afin de se doter de l'expérience nécessaire. Il faut aussi d'autre part considérer l'initialisation de l'essaim ; elle est généralement faite aléatoirement suivant une loi uniforme sur  $[0, 1]$ .

Deux autres paramètres importants sont les coefficients de confiance que l'on nomme précédemment dans la quatrième partie du troisième chapitre  $c_2$  et  $c_3$ . Ils permettent de pondérer les tendances des particules à suivre leur instinct de conservation ou leur pargisme. De manière générale, ces variables aléatoires sont évaluées à chaque itération suivant une loi uniforme sur le domaine de définition.

De même un paramètre important à prendre en compte est le coefficient d'inertie appelé  $c_1$  dans la formule vue auparavant. Il permet de définir la capacité d'exploration de chaque particule en vue d'améliorer la convergence de la méthode. Fixer ce paramètre revient à trouver un compromis entre une exploration globale ( $c_1 > 1$ ) et une exploration locale ( $c_1 < 1$ ). Il représente l'instinct aventureux de la particule.

Enfin, il reste à configurer le critère d'arrêt. En effet, la convergence vers la solution optimale globale n'est pas garantie dans tous les cas. Il est donc important de doter l'algorithme d'une porte de sortie en définissant un nombre maximum d'itérations. Le programme s'arrête alors si et seulement si le nombre maximum d'itérations est atteint ou que la valeur du critère obtenue est acceptable pour l'utilisateur.

Le Nefclass, comme nous l'avons présenté, utilise le système Mamdani, donc l'agrégation Min-Max. Le Min entre les degrés d'appartenance de chaque attributs pour une règle donnée et le Max entre les sorties de toutes les règles.

La règle qui a le plus grand degré d'activation (Max) est la responsable de la classification du pattern propagé.

L'idée est, donc, de prendre le sous-ensemble flou, de la règle Max, qui a le plus petit degré d'appartenance parmi les attributs (Min) et de le modifier afin de diminuer le taux d'erreur en respectant les contraintes sur les sous-ensembles flous. Une population est créée où chaque particule représente le sous ensemble flou Min. La particule est modélisée par un vecteur de trois dimensions qui contient les trois points du sous ensemble flou (point-min, point-moyen, point-max).

On résumé nous avons une population de N particules de dimension 3. L'initialisation de l'essaim se fait aléatoirement dans un intervalle prédéfini (pour éviter de violé les contraintes).

L'intervalle de chaque point change selon le sous-ensemble flou, si :

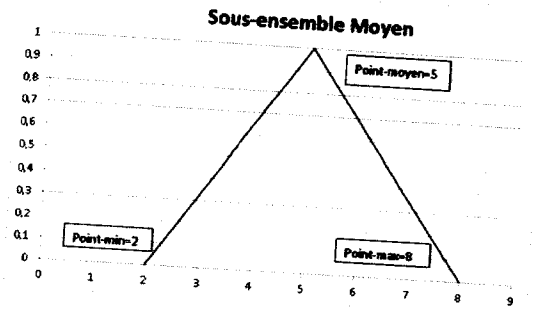


FIGURE 4.1 – Les trois points du sous-ensemble flou

P-min	P-moy	P-max
-------	-------	-------

TABLE 4.2 – Modélisation d'une particule

L'ensemble flou **petit** :

- p.min = 0
- P.moy [0, ..., 4.9]
- p.max [2.1, 7.9]

L'ensemble flou **moyen** :

- p.min [0, ... , 4.9]
- P.moy [0, ... , 4.9]
- p.max [5.1, , 10]

L'ensemble flou **grand** :

- p.min [2.1, 7.9]
- P.moy [5.1, , 10]
- p.max = 10

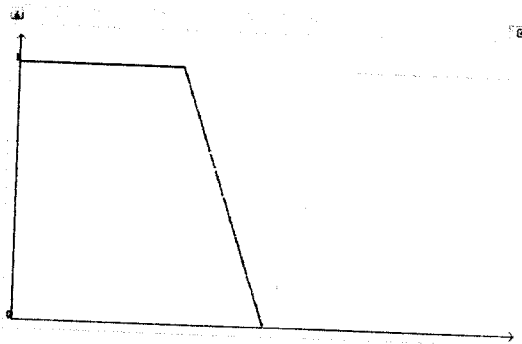


FIGURE 4.2 – Création d'une population

à chaque itération de l'algorithme, le système calcule l'erreur avec chaque sous-ensemble flou crée (particule), le minimum est donc le " global best ", la position de cette particule est sauvegardé et le reste de la population approche de cette solution en calculant leurs nouvelles positions.

Si l'algorithme atteint le nombre d'itération prédéfini ou bien un taux d'erreur qui est nul, il récupère la " best particule ", et avec ses paramètres, le sous-ensemble flou est remplacé dans la règle. Dans les figures ci-dessous, nous pouvons voir le résultat de l'apprentissage du sous-ensemble flou Moyen :

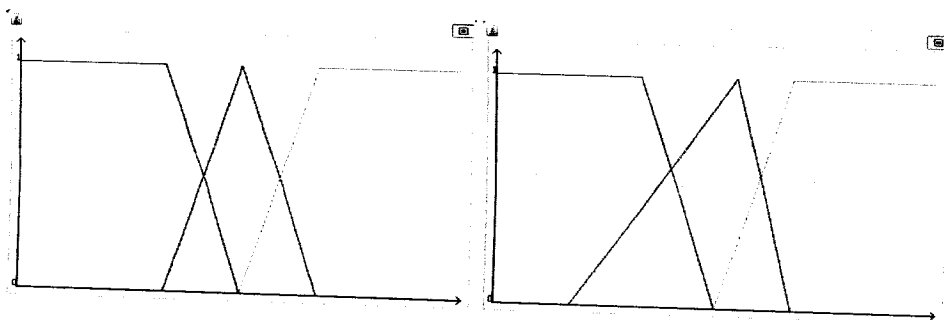


FIGURE 4.3 – SEF avant et après apprentissage

L'algorithme donne de meilleurs résultats par rapport au Nefclass classique.

## 4.4 La procédure de classification

Un système d'apprentissage a induit une procédure de classification à partir d'un échantillon ; nous souhaitons estimer la qualité de la procédure induite en estimant l'erreur réelle.

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. Donc 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, de même, la précision de l'estimation augmente avec la taille de l'ensemble test. 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.

### 4.4.1 Phase de test (d'évaluation)

Cette phase permet d'affecter un objet nouveau à une classe au moyen d'une règle de décision obtenu durant la phase d'apprentissage. Pour évaluer le classifieur on utilise le critère classique comme

$$\text{Taux de classification} = 100 * \frac{N_{correct}}{N_{total}}$$

$$\text{Taux d'erreur} = 100 * \frac{N_{erreur}}{N_{total}}$$

Erreur de classification : E(d) est la probabilité qu'un élément de P ayant d pour description soit mal classé. D'autre critère comme la spécificité est la sensibilité apportent ainsi des informations utiles :

**Sensibilité :** on appelle sensibilité ( $S_e$ ) 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 ( $S_p$ ) 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.

$$S_p = \frac{VN}{VN + FP}$$

Avec :

- VP : Vrai Positif : nombre de positifs classés positifs.
- VN : Vrai Négatif : nombre de positifs classés négatifs.
- FP : Faux Positif : nombre de négatifs classés positifs.
- FN : Faux Négatif : nombre de négatifs classés négatifs.

Mais il importe de confronter la spécificité et la sensibilité du test aux caractéristiques de la maladie dans la population où l'on pratique le dépistage. Il faut calculer les valeurs prédictives positives et négatives du test.

- La valeur prédictive positive (VPP) d'un test est le risque que le sujet soit effectivement " malade " si son test est positif, c'est le rapport des vrais positifs sur la somme de tous les individus dont le test est revenu positif.
- La valeur prédictive négative (VPN) est la chance que le sujet ne soit pas " malade " si le test est négatif, c'est le rapport des vrais négatifs sur le total de tous les négatifs.

	Sujets malades	sujets seins
Test positif	Vrais positifs (VP)	Faux positifs (FP)
Test négatif	Faux négatifs (FN)	Vrais négatifs (VN)

Valeur prédictive positive = Vrais positifs / Tous les positifs  
 Valeur prédictive négative = Vrais négatifs / Tous les négatifs

Donc le Taux de classification :

$$TC = \frac{VP + VN}{VP + VN + FP + FN} * 100\%$$

## 4.5 Les résultats du Nefclass

Dans ce qui suit, nous allons présenter les résultats obtenus selon trois modes de partitionnement de la base WBCD et avec des paramètres différents. Premièrement, nous utilisons la base complète pour l'apprentissage et la même base pour le test. Deuxièmement, nous consacrons deux tiers des échantillons de la base de données à la phase d'apprentissage, et le tiers restant à la phase de test. Finalement, nous utiliserons une méthode, qui est devenu, très utilisé dans la dernière décennie, c'est la 10-fold cross validation.

	K-max	Taille b.App	Taille b.Test	Nbr. epochs	Pas d'App	Performance	Taux	Spéci.	Sensi.
1	8	100%	100%						



Procédure d'apprentissage : **Best Per Class.**

Contraintes sur la génération des règles :

- Garder l'ordre des SEF.
- Ne pas dépasser la valeur Min et Max de chaque SEF.
- Chevauchement autorisé.

Poids des règles : **Non utilisé.**

Pas d'apprentissage : **0.15.**

Nombre d'époques : **200.**

Les règles créées :

**R1** : Si **A1** est Petit **A2** est Petit **A3** est Petit **A4** est Petit **A5** est Petit **A6** est Petit **A7** est Petit **A8** est Petit **A9** est Petit alors la classe est **2.0 (Bénin).**

**R2** : Si **A1** est Moyen **A2** est Petit **A3** est Petit **A4** est Petit **A5** est Petit **A6** est Petit **A7** est Petit **A8** est Petit **A9** est Petit alors la classe est **2.0(Bénin).**

**R3** : Si **A1** est Moyen **A2** est Petit **A3** est Petit **A4** est Petit **A5** est Petit **A6** est Petit **A7** est Moyen **A8** est Petit **A9** est Petit alors la classe est **2.0(Bénin).**

**R4** : Si **A1** est Moyen **A2** est Petit **A3** est Petit **A4** est Petit **A5** est Moyen **A6** est Petit **A7** est Petit **A8** est Petit **A9** est Petit alors la classe est **2.0(Bénin).**

**R5** : Si **A1** est Grand **A2** est Moyen **A3** est Moyen **A4** est Grand **A5** est Moyen **A6** est Grand **A7** est Moyen **A8** est Moyen **A9** est Grand alors la classe est **4.0(Malin).**

**R6** : Si **A1** est Grand **A2** est Grand **A3** est Grand **A4** est Grand **A5** est Moyen **A6** est Grand **A7** est Moyen **A8** est Grand **A9** est Moyen alors la classe est **4.0(Malin).**

**R7** : Si **A1** est Grand **A2** est Grand **A3** est Grand **A4** est Moyen **A5** est Moyen **A6** est Grand **A7** est Moyen **A8** est Grand **A9** est Moyen alors la classe est **4.0(Malin).**

**R8** : Si **A1** est Grand **A2** est Grand **A3** est Grand **A4** est Moyen **A5** est Grand **A6** est Grand **A7** est Moyen **A8** est Grand **A9** est Grand alors la classe est **4.0(Malin).**

**Cas n :1**

Nous utilisons 100% de la base pour l'apprentissage et 100% pour le test. Les résultats obtenus sont résumés dans le tableau suivant :

	Malin			Bénin			Non classé		Somme	
Malin	VP	225	32.94%	FN	14	2.05%	0	0%	239	34.99%
Bénin	FP	12	1.75%	VN	432	63.26%	0	0%	444	65.01%
									683	100%

**Taux : 96.19 % Spécificité : 97.29 % Sensibilité : 94.14 %**

**Cas n :3**

Nous utilisons cette fois 2/3 de la base pour l'apprentissage et 1/3 pour le test. Les résultats obtenus sont résumés dans le tableau ci-dessous :

	Malin			Bénin			Non classé		Somme	
Malin	VP	49	21.59%	FN	3	1.32%	0	0%	52	22.91%
Bénin	FP	5	2.20%	VN	170	74.29%	0	0%	175	77.09%
									227	100%

**Taux : 96.47 % Spécificité : 97.14 % Sensibilité : 94.23 %**

### Cas n :7

Les meilleurs résultats obtenus avec la 10-Fold Cross Validation. Dans cette méthode la base est divisée en 10 ensembles. Nous utiliserons à chaque fois 9 ensembles (614 patterns) pour apprentissage et un ensemble (22 patterns) pour test, et à chaque fois on change les ensembles. Le tableau suivant illustre ces résultats :

	Taux
1	95.65 %
2	98.55 %
3	97.10 %
4	98.52 %
5	95.58 %
6	95.58 %
7	97.05 %
8	97.05 %
9	97.05 %
10	95.58 %
Moyenne	96.74

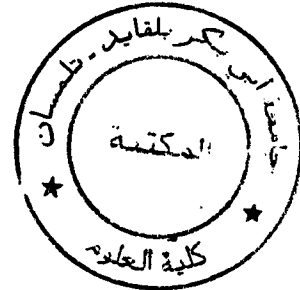


TABLE 4.4 – Résultats de la 10-Fold Cross Validation

- Si nous revenons au tableau des résultats, nous pouvons remarquer que :
- La deuxième expérience montre que le système a obtenu 94.85% d'exemples bien classés. Par rapport aux trois expériences citées auparavant, l'erreur est plus grande, donc nous constatons que le système a fait un sur-apprentissage.
  - Pour le cinquième cas, le système n'a pas bien appris la base (peu d'itérations et un pas réduit), c'était naturel d'avoir un taux qui dépasse pas les 89.97%.
  - Dans le quatrième cas nous avons utilisé la configuration qui a donné les meilleurs résultats mais avec une autre stratégie de choix des k-max règle (Best Rule Learning). Les résultats étaient moins performants à cause de cette stratégie car sur 8 règles nous avons obtenu 7 règles de la même classe (Malin), ce qui provoque une pénalisation de la classe bénigne.

## 4.6 Les résultats de l'Opt-Nefclass

Nous avons fait beaucoup de combinaisons afin d'avoir une population qui converge vers l'optimum globale. Les meilleurs résultats sont groupés dans le tableau suivant :

	K-max	Taille b.App	Taille b.Test	Taille Essaim	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	Nbr. Epochs	Taux	Spéci.	Sensi.
1	8	100%	100%	50	0.72	1.8	1.8	100	97.21%	97.97%	95.81%
2	8	100%	100%	15	0.72	1.8	1.8	100	97.21%	97.97%	95.81%
3	8	2/3	1/3	15	0.72	1.8	1.8	100	97.78%	98.28%	96.15%
4	8	2/3	1/3	15	0.1	2	2	100	96.92%	97.74%	95.39%
5	8	100%	100%	5	0.72	1.8	1.8	100	96.48%	97.29%	94.97%
6	8	10-Fold	Cross Val	15	0.72	1.8	1.8	100	98.24%	99.01%	96.98%

TABLE 4.5 – Résultats de la classification avec Opt-Nefclass

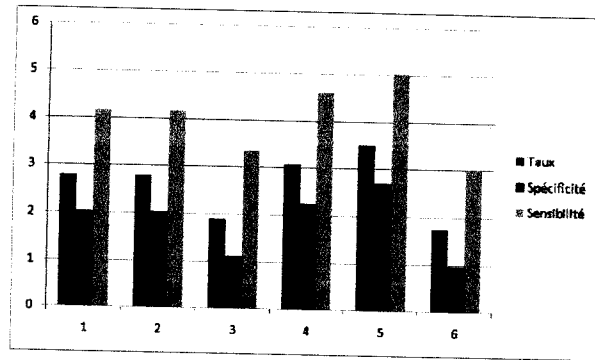


FIGURE 4.5 – Histogramme des taux de performances

#### 4.6.1 Interprétation des résultats

En analysant le tableau des résultats, nous remarquons que le système donne des résultats très prometteurs. Notons que la base de règle ne change pas vu que le PSO est utilisé juste dans l'étape d'apprentissage des sous-ensembles flous.

– Le premier et le deuxième cas donnent le même résultat avec un nombre de particule différent (50 dans le premier et 15 dans le deuxième), par contre dans la cinquième expérience avec un essaim de 5 particules, le taux est plus faible. Cela signifie que nous devons avoir un minimum de 15 particules pour assurer la convergence de l'algorithme.

– Les meilleurs résultats sont obtenus avec un facteur d'inertie  $C_1 = 0.72$ ,  $C_2 = C_3 = 1,8$ , ( $C_2 + C_3 \leq 4$ ).

Nous allons présenter les meilleurs cas :

##### Cas n :1

Nous utilisons 100% de la base pour l'apprentissage et 100% pour le test. Les résultats obtenus sont résumés dans le tableau suivant :

	Malin			Bénin			Non classé		Somme	
Malin	VP	229	33.53%	FN	10	1.46%	0	0%	239	34.99%
Bénin	FP	9	1.33%	VN	435	63.68%	0	0%	444	65.01%
									683	100%

**Taux : 97.21 % Spécificité : 97.97 % Sensibilité : 95.81 %**

##### Cas n :3

Nous utilisons cette fois 2/3 de la base pour l'apprentissage et 1/3 pour le test. Les résultats obtenus sont résumés dans le tableau ci-dessous :

	Malin			Bénin			Non classé		Somme	
Malin	VP	50	22.02%	FN	2	0.89%	0	0%	52	22.91%
Bénin	FP	3	1.32%	VN	172	75.77%	0	0%	175	77.09%
									227	100%

**Taux : 97.78 % Spécificité : 98.28 % Sensibilité : 96.15 %**

## Cas n :6

	Taux
1	97.07 %
2	99.17 %
3	98.68 %
4	99.12 %
5	97.16 %
6	97.21 %
7	98.92 %
8	98.34 %
9	98.71 %
10	98.02 %
Moyenne	<b>98.24 %</b>

TABLE 4.6 – Résultats de la 10-Fold Cross Validation

## 4.7 Etude comparative

Auteur	Technique	Taux de reconnaissance
Subhash et al., 2003 [3]	K-NN	98.1 %
Fei et al., 2003 [4]	SVM	96.4 %
Madan, 2002 [7]	Naïve Bayes	97.81 %
Seydeh et al.,2010 [26]	HFNN	<b>99.04 %</b>
El-Hamdi et al.,2010 [16]	Neuro-flou + AG	97.87 %
Pena et al.,1999 [55]	Fuzzy Genetic	97.80 %
Arulampalan et al., 2001 [56]	NN	<b>100 %</b>
Ben Kadour,2010 [57]	PMC	<b>99.56 %</b>
Notre Travail	NEFCLASS	96.74 %
Notre Travail	OPT-NEFCLASS	<b>98.24 %</b>

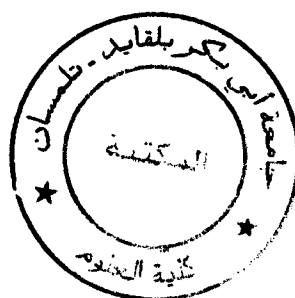
TABLE 4.7 – Etude comparative

Dans cette partie nous avons fait la comparaison entre les différents travaux existants dans la littérature et les résultats obtenus avec Nefclass et Opt-Nefclass, et cela sur la même base de donnée du cancer de sein.

## 4.8 Conclusion

Nous remarquons à travers les travaux présentés dans ce chapitre, notamment pour le cas des approches utilisant les systèmes hybrides, que les taux de reconnaissance que l'on a obtenu avec les deux modèles réalisés, sont compétitifs voir meilleurs que ceux obtenus dans plusieurs travaux de recherche. Par contre en terme de performance, les réseaux de neurones ont obtenus les meilleurs résultats vu leur capacité d'apprentissage, mais ils

sont des modèle de type boîte noir. Ce dernier constitue un inconvénient majeur dans le domaine médical. Il est toutefois important de noter qu'une comparaison objective de la performance des différents modèles de classification est très difficile et parfois irréalisable à cause de l'utilisation de bases de données différentes.



## Conclusion Générale et Perspectives

La classification dans le domaine du dépistage du cancer du sein a pour but d'identifier les catégories de cas qui ont des similitudes biologiques fondamentales et qui sont susceptibles de partager certains facteurs étiopathologiques. Il est important de faire la reconnaissance de ces classes dans le but de permettre non seulement de comprendre le processus de la maladie mais aussi d'instaurer l'approche thérapeutique adéquate. Après l'étape de reconnaissance de ces classes, il est tout aussi important de pouvoir faire l'identification de cas de cancer du sein en les attribuant à la classe qui leur convient. Dans cette optique, les méthodes de classification automatiques offrent de nos jours une multitude d'outils profitables pour faire l'identification de différentes données et étudier leurs propriétés. En particulier, dans le cadre de la classification, les réseaux neuro-flous ont largement fait leurs preuves au sein de la communauté scientifique et industrielle.

Chaque modèle neuro-flou diffère d'un autre par ses paramètres expérimentaux, sa complexité d'implémentation et son domaine d'applicabilité. Ainsi, chaque modèle possède ses avantages et ses inconvénients faisant que son utilisation soit appropriée ou non à un problème donné, selon la nature de celui-ci, ses buts, mais aussi la nature des données. Nous nous sommes intéressés dans ce travail à étudier la performance d'un modèle neuro-flou appelé NEFCLASS sur une tâche de classification reliée à la détection du cancer du sein. à cet effet, nous avons procédé en premier temps à une analyse de travaux effectués dans le domaine qui a permis de mettre en évidence plusieurs avantages ainsi que certaines limites des approches de classification employées dans le cadre du dépistage du cancer du sein. Nous avons constaté en conséquence que les classifieurs proposés dans le domaine font preuve d'une certaine complexité à la réalisation par les hybridations que comportent la plupart sans pour autant que leurs performances soit irréprochables, tant dit que les systèmes moins complexes n'offrent pas une interprétabilité des résultats. Le modèle neuro-flou que nous avons réalisé se caractérise de par sa simplicité mais aussi de par son interprétabilité sa performance.

Nous avons proposé et mis en oeuvre un autre modèle neuro-flou, à savoir, le système NEFCLASS hybridé avec la PSO(OPT-NEFCLASS) pour effectuer la même tâche de classification. Nous avons procédé en second lieu à une étude comparative à travers laquelle nous mettons en avant l'apport fait par l'OPT-NEFCLASS par rapport aux autres classifieurs existants. La performance de l'OPT-NEFCLASS a souvent été irréprochable comparativement au NEFCLASS.

Nous considérons qu'à travers les différentes réalisations citées, il a été possible de parler à certaines limites que connaissent les classifieurs de cancer du sein déjà existants dans la littérature. Ces palliatifs favorisent en particulier l'applicabilité des réseaux neuro-flous pour l'aide au diagnostic du cancer du sein. Toutefois, nous estimons pouvoir élargir l'ex-

exploitation de cette application, et il serait dans cette optique intéressant de mettre plus d'investigations pour expérimenter d'autres modèles de réseaux neuro-floous au service de l'aide au diagnostic du cancer du sein.

Nous notons par ailleurs que nous disposions dans le cadre de ce travail d'un nombre relativement petit de données, à savoir, 683 échantillons. On envisage de ce fait des travaux futurs dans cet axe de recherche en exploitant une base de données plus volumineuse. Ceci serait en effet intéressant pour des fins d'analyser le comportement de réseaux neuro-floous face à un plus grand nombre de cas de tumeurs et un plus grand nombre de patterns pour chaque cas. Nous pensons à la réalisation d'une base de données locale, riche et variée au sein du laboratoire GBM.

Notre approche constitue un premier pas dans le développement d'un système d'aide au diagnostic médical intelligent qui peut aider les experts dans le domaine du cancer de sein à renforcer leurs diagnostics. Nous souhaitons dans le futur que ce premier modèle peut être intégré dans un système d'aide au diagnostic avec un système de gestion d'un cabinet médical pour garder la traçabilité des données d'un malade vu que le test du cancer de sein est périodique pour les femmes plus de 40 ans. Nous souhaitons aussi appliqué sur le même système NEFCLASS d'autres méthodes évolutionnaires tel que les algorithmes génétiques ou bien les colonies de fourmis pour tester leurs efficacités sur une architecture qui reste la plus simple des réseaux neuro-floous.

# Bibliographie

- [1] Conférence internationale sur le cancer. Fondation Lance Amstrong, Dubai. 8 septembre 2010.
- [2] L'Organisation Mondiale de la Santé (OMS). <http://www.who.int/cancer/en/>. Novembre 2010.
- [3] Subhash c., Bagui, Sikha Bagui, Kuhu Pal, Nikhil R. Breast cancer detection using rank nearest neighbor classification rules . Elsevier Pattern recognition, vol 36, pp.25–34. 2003.
- [4] Fei Sha, Lawrence K., Saul, Daniel D., Lee. Multiplicative updates for non-negative quadratic programming in support vector machines . Advances in Neural Information Processing Systems 15, Sebastian and K. Obermayer, Eds. Cambridge, MA : MIT Press, 2003.
- [5] Moghaddam Baback, et Shakhnarovich Gregory. Boosted dyadic kernel discriminants. Mitsubishi electric research laboratories . Advances in Neural information processing systems (NIPS 15). Vancouver B.C, Canada. 2002.
- [6] Huang Kaizhu, Yang Haiqin, King Irwin, Lyu Michael R, Chan Laiwan. Biased minimax probability machine for medical diagnosis . The 8th International Symposium on Artificial Intelligence and Mathematics, pp.4 – 6. 2004.
- [7] Madden Michael G. Evaluation of the Performance of the Markov Blanket Bayesian Classifier Algorithm. CoRR, cs. *LG/02/1003*. 2002.
- [8] Cheng RD., Shi X.J, Min R, Hu l.m., X.P., Cai, RN. Du. Approaches for automated detection and classification of masses in mammograms . Pattern recognition, vol 39, pp. 646 – 668. 2006.
- [9] Esmeir Saher, Markovitch Shaul. Lookahead-based Algorithms for anytime induction of decision trees . 21 th International conference on machine learning, vol169, pp. 33. 2004.
- [10] De Santo Massimo, Molinara Mario, Tortorella Francesco, Vento Mario. Automatic classification of clustered microcalcifications by a multiple expert system . Elsevier Pattern Recognition, 36, pp. 1467 – 1477. 2003.
- [11] Wojciech M., Grohman, Atam P., Dhawan; Fuzzy convex set-based pattern classification for analysis of mammographie microcalcifications . Elsevier Pattern recognition, vol 34, pp. 1469 – 1482. 2001.
- [12] Karnan M., Thangavel K., Ezhilarasu P. Ant colony optimization and a new particle swarm optirnizaion algorithm for classification of microcalcifications in mammograms . The 6th International Conference on Advanced Computing and Communication. 2008.
- [13] Liu Bo, Abbass Hussein A, McKay Bob. Classification mIe discovery with ant colony optimization . IEEE Computational intelligence bulletin, Vol.3 No. 1. 2004.



- [14] Parepinelli R. S., Lopes H. S., Freitas A. An Ant Colony Algorithm for Classification Rule Discovery . Data Mining : Heuristic Approach : Idea Group Publishing, H. A a. R. S. a. C. Newton Edition. 2002.
- [15] Jaganathan P., Thangavel K., Pethalakshmi A, Karnan M. Classification rule discovery with ant colony optimization and improved quick reduct algorithm . IAENG International journal of computer science, 33 – 1, *IJCS<sub>3</sub>L9*. 2007.
- [16] R. El Hamdi, M. Njah, M. Chtourou. An Evolutionary Neuro-Fuzzy Approach to Breast Cancer Diagnosis . IEEE, 978-1-4244-6588-0/10/, P. 142-146. 2010.
- [17] Cheng S.N.C., Chan H.P., Helvie 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 . 1. Imaging Sci, Technol, 38. 1994.
- [18] Baeg S., Kehtarnavaz N. Texture based classification of mass abnormalities in mammograms . Proceedings of the IEEE Symposium on Computer-based Medical Systems, pp. 163 – 168. 2000.
- [19] Floyd C.E., Lo J.Y., Yun A.J., Sullivan D.C., Kornguth P.J. Prediction of breast cancer malignancy using an artificial neural network . Cancer 74, pp. 2944 – 2948. 1994.
- [20] Bovis K., Singh S., Fieldsend J., Pinder c. Identification of masses in digital mammograms with MLP and RBF nets . Proceedings of the IEEE-INNSENS International Joint Conference on Neural Networks Com, pp. 342 – 347. 2000.
- [21] Christoyianni I., Dermatas E., Kokkinakis G. Fast detection of masses in computer-aided mammography . IEEE Signal Process, pp. 54 – 64. 2000.
- [22] Asoke Hong G., Nandi K. Breast cancer diagnosis using genetic programming generated feature . Elsevier Pattern recognition, vol 39, pp. 980987. 2006.
- [23] Cheng H. D., Cui Muyi Mass lesion detection with a fuzzy neural network . Elsevier Pattern recognition, vol 37, pp. 1189 – 1200. 2004.
- [24] Wojciech M., Grohman, Atam P., Dhawan. Fuzzy convex set-based pattern classification for analysis of mammographic microcalcifications . Elsevier Pattern recognition, vol 34, pp. 1469 – 1482. 2001.
- [25] Abbass Hussein A. An evolutionary artificial neural networks approach for breast cancer diagnosis . Artificial intelligence in medicine, vol 25, pp. 265 – 281. 2002.
- [26] S. N. Seyedeh, M. Teshnehlab, M. A. Shoorehdeli. Breast Cancer Detection by using Hierarchical Fuzzy Neural System with EKF Trainer . Proceedings of the 17th Iranian Conference of Biomedical Engineering (ICBME2010). 3-4 November 2010.
- [27] M. Ashraf, Kim Le, Xu Huang. Information Gain and Adaptive Neuro-Fuzzy Inference System for Breast Cancer Diagnoses . IEEE, P. 911-915. 2010.
- [28] L. Personnaz, J. Rivals. Réseaux de neurones formels pour la modélisation, la commande et la classification . CNRS Editions, Paris. 2003
- [29] McCulloch, PITTS W. A logical Calculus of Ideas Immanent in Nervous Activity . Research Studies Press, Taunton. 1997.
- [30] Marc Parizeau . Réseaux De Neurones . Université de LAVAL. P39,51. Automne 2004.
- [31] Stéphane Tufféry. Data Mining et statistique décisionnelle : L'intelligence des données . Editions Ophrys. P 148, 157, 160. 2007.

- [32] Setiono R, Liu H. Symbolic representation of neural networks . IEEE Computer, 717. 1996.
- [33] Rosenblatt F. Principles of Neurodynamics : Perceptrons and the Theory of Brain Mechanisms . Spartan, Washington. 1962.
- [34] Alain Faure. Classification et commande par réseaux de neurones. Edition Lavoisier. P 149-153. 2006.
- [35] Kiyant. and Yildirim T. Breast Cancer Diagnosis Using Statistical Neural Networks. Journal of electrical and electronics engineering, pp. 1149-1153. 2004.
- [36] Claude Rosental. Histoire de la logique floue. Une approche sociologique des pratiques de démonstration , Revue de Synthèse. vol. 4, 4, pp. 575 – 602. octobre-décembre 1998.
- [37] Hopfield J.J. Neural Networks and Physical Systems with Emergent Collective Computational Abilities . Proc. Of the Nalt. Acad. Sci., vol. 79, p. 2554 – 2558. 1982.
- [38] L.A. Zadeh. Fuzzy-Sets . Information and Control, 8 :338-353. 1965.
- [39] Djamel A. Zighed, Ricco Rakotomalala. Graphes d'induction (Apprentissage et Data Mining) . Edition hermès, P. 25 – 27. Année 2000.
- [40] Rumelhart, D. E., Hinton, G. E., et Williams, R. J. Learning internal representations by error propagation . In D. E. Rumelhart et J. L. McClelland(Eds.). Parallel distributed processing : Explorations in the microstructure of cognition (Vol. 1, pp. 318 – 362). Cambridge, MA : MIT Press. 1986.
- [41] Hagan, M. T., et Menhaj, M. Training feed forward networks with the Marquardt algorithm . IEEE Transactions on Neural Networks, 5, 989 – 993. 1994.
- [42] N. SETTOUTI, T. FANDI. Reconnaissance par Neuro-Floue du Diabète . PFE. juin 2009.
- [43] Julien Beaudry. Nicolas Vaillancourt-Veilleux. Système neuronal pour le positionnement et la classification d'éléments visuels d'un terrain de soccer robotisé . Laboratoire de mécatronique, DG, PM. Montréal, Québec, Canada.
- [44] Alain Faure. Classification et commande par réseaux de neurones. Edition Lavoisier .P 149 – 153. 2006.
- [45] R.Mahdaoui. H.L Mouss. Diagnostic Industriel par Neuro-Floue, Application à Un Système de Production . Laboratoire d'Automatique et Productique (LAP) Université de Batna.
- [46] Y. Nakoula, Apprentissage des modèles linguistiques flous, par jeu de règles pondérées , Thèse de doctorat, Université de Savoie. 1997.
- [47] R. Jang, ANFIS : Adaptive-Network-Based Fuzzy Inference System , IEEE Trans.on Systems, Man and Cybernetics, J.S. 1993.
- [48] T. Takagi and M. Sugeno. Derivation of fuzzy control rules from human operators control actions , Proc. Of the IFAC Symp. On Fuzzy Information, Knowledge Representation and Decision analysis, P 555 – 60. July 1983.
- [49] D. Nauck and R. Kruse. NEFCLASS - A Neuro-Fuzzy Approach for the Classification of Data .In K. George, J.H.Carrol, E. Deaton, D. Oppenheim and J. Heightower, eds : Applied Computing 1995. Proc. 1995 ACM Symposium on Applied Computing, Nashville, Feb. 26-28, pp. 461-465. ACM Press, New York. 1995.

- [50] D. Nauck and R. Kruse. Obtaining Interpretable Fuzzy Classification Rules From Medical Data. In Artificial Intelligence in Medicine. 1998.
- [51] CLERC, SIARRY. Une nouvelle métaheuristique pour l'optimisation difficile : la méthode des essaims particulaires . 2003
- [52] EUDES F, M. RIOLAND. Optimisation par essaim particulaire pour un problème d'ordonnement et d'affectation de ressources . 2007
- [53] GOURGAND, KEMMOE. Particle Swarm Optimization : A study of particle displacement for solving continuous and combinatorial optimization problems . 2009
- [54] Juan-Manuel TORRES-MORENO. Apprentissage et généralisation par des réseaux de neurones : étude de nouveaux algorithmes constructifs . THSE Sciences Cognitives. FRANCE 1997.
- [55] Pena-Reyes CA, Sipper M. cA fuzzy-genetic approach to breast cancer diagnosis . Artificial Intelligence in Medicine 17;131155. 1999.
- [56] G. Arulampalam and A. Bouzerdoum. Application of Shunting Inhibitory Artificial Neural Networks to Medical Diagnosis . In Seventh Australian and New Zealand Intelligent Information Systems Conference, pp. 89 -94. 2001.
- [57] F. BEKADOUR, La conception d'un Système d'Aide au Diagnostic Médical , PFE. 2010.
- [58] OMG. Unified Modeling Language Specification. Version 1.5. 2003.
- [59] Rocques, P. UML Modéliser un site e-commerce, les Cahiers du Programmeur . p.152 , Ed. Eyrolles. 2003.

# Annexe A

## Modélisation de l'application

Notre approche de conception suit le modèle UML (UnifiedModellingLanguage). Plusieurs raisons nous ont conduits à ce choix :

- La première est sa normalisation par l'OMG [58]. L'historique a montré que la profusion des notations est, à moyen terme, préjudiciable aux entreprises comme à leurs fournisseurs. Toute norme doit donc être considérée avec le plus grand sérieux, en particulier les normes du domaine public, comme c'est le cas pour l'UML.
- La deuxième raison est l'intérêt montré par les informaticiens pour ce langage de modélisation.
- La troisième raison est la possibilité d'utiliser le même atelier de génie logiciel, depuis l'expression des besoins jusqu'à la génération de tout ou partie de l'application.
- La dernière raison, mais non la moindre, est d'utiliser les principes et concepts objet pour enrichir la démarche de conception de systèmes d'aide à la décision. On en attend des améliorations dans le sens tout à la fois de la richesse, d'une modularité, d'une cohérence et d'une rigueur accrues.

UML s'articule autour de plusieurs types de diagrammes, chacun d'eux étant dédié à la représentation des concepts particuliers d'un système logiciel. Le schéma suivant montre comment, en partant des besoins utilisateurs formalisés par des cas d'utilisation et une maquette, et avec l'apport du modèle du domaine, on peut aboutir à des diagrammes de conception qui permettent de dériver du code assez directement [59].

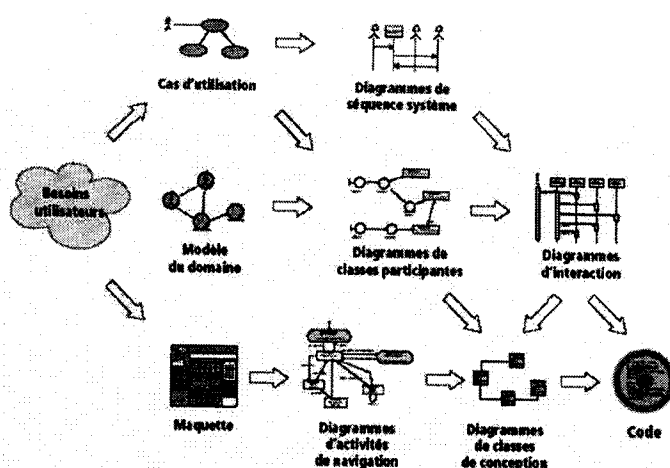


FIGURE A.1 – Schéma global du processus de modélisation UML

## A.1 Définition des besoins des utilisateurs

Lors de l'étude des besoins des utilisateurs, nous avons spécifié les exigences fondamentales d'un outil d'aide au diagnostic moderne :

- Utilisation simple.
- Capacité d'interfaçage avec les outils de gestion (gestion d'un cabinet médical, ).
- La possibilité ultérieure de prendre en compte les nouvelles expériences (si possible en ligne).
- Interprétabilité du système.
- possibilité d'interfaçage avec des IHM - clients légers (PDA, ordinateurs portables) (Perspective).
- intégration de l'outil dans un système embarqué. (perspective).

## A.2 Spécification des cas d'utilisations

La première étape nécessaire à l'établissement des cas d'utilisation de l'outil est constituée par la définition des acteurs. Dans notre cas, les acteurs humains sont le médecin et le responsable de la maintenance (l'administrateur ou bien l'informaticien).

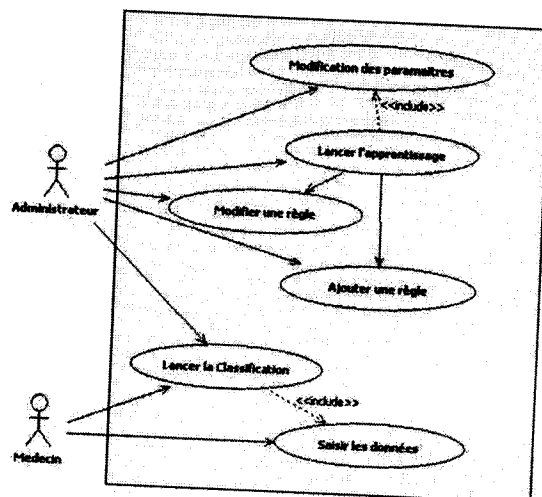


FIGURE A.2 - Le diagramme des cas d'utilisations.

En ce qui concerne le médecin, ses cas d'utilisation principaux ont été mis en évidence par l'expression des besoins, à savoir :

- Saisir la donnée d'un malade.
- Lancer la classification des données insérées.
- Notifier à l'administrateur qu'il doit modifier ou bien ajouter une règle de classification car un médecin est un expert du domaine médical.

L'administrateur, en tant que informaticien, peut faire la modification et l'apprentissage du système ainsi que les tâches du médecin (avec son accord dans le cas d'ajout ou de la modification d'une règle).

### A.3 Les diagrammes de classes

Après avoir effectué une modélisation par les cas d'utilisation qui correspond à une description fonctionnelle, il est nécessaire de passer à une modélisation structurale, statique du système à réaliser, sous forme d'ensemble de classes logicielles, éventuellement regroupées en paquetages.

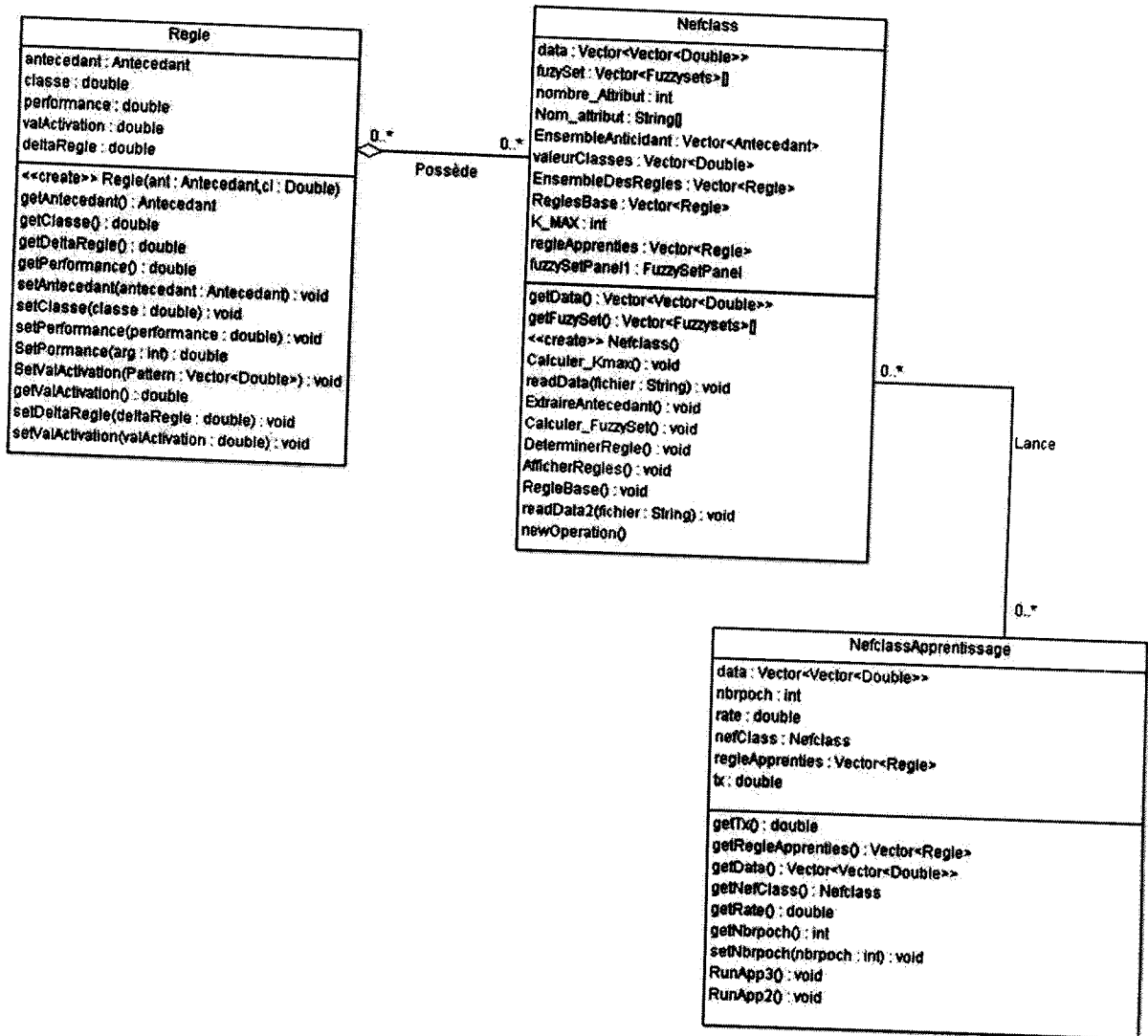


FIGURE A.3 – Diagramme de classe du cas d'utilisation Lancer apprentissage

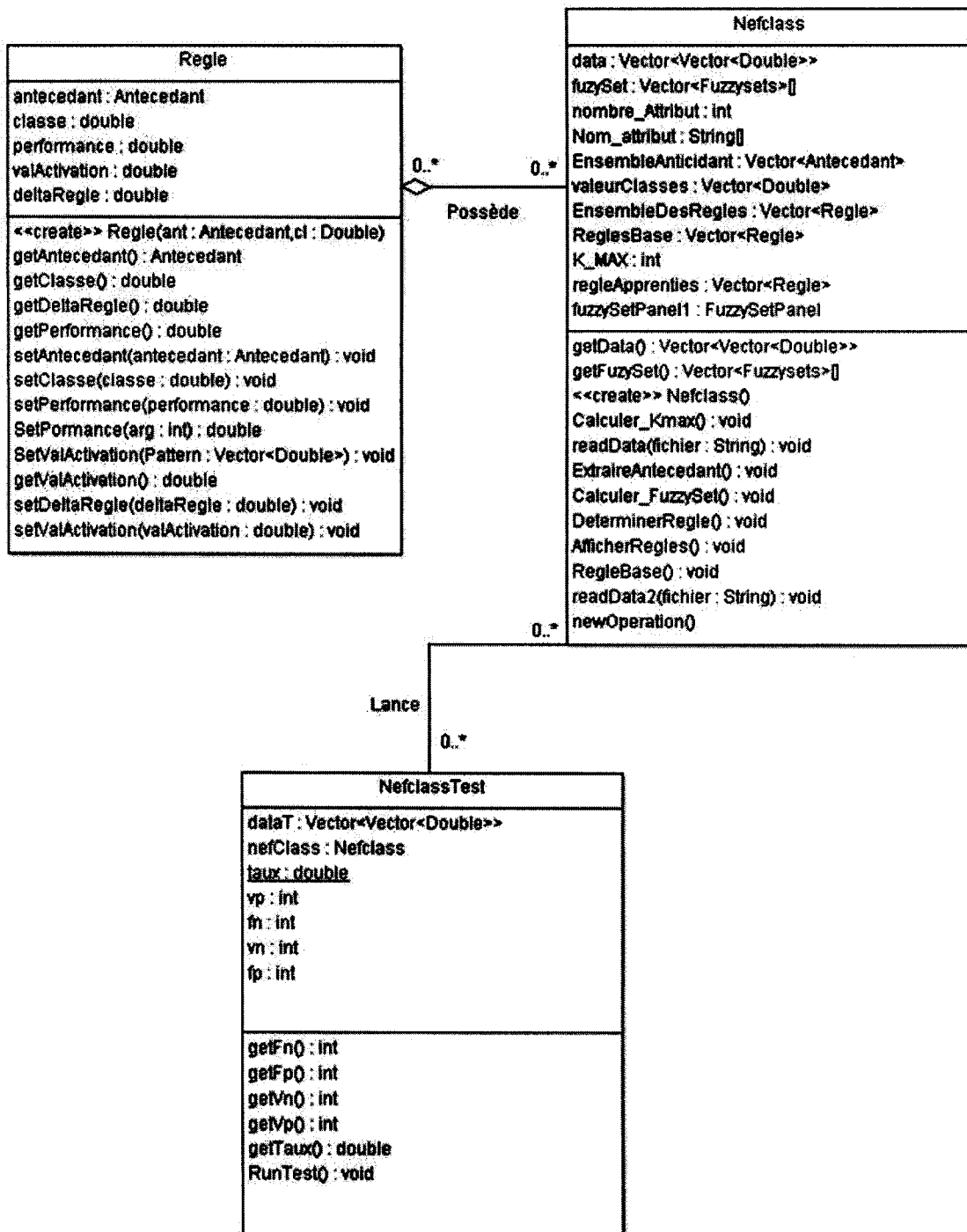


FIGURE A.4 – Diagramme de classe du cas utilisation Lancer Test





# Annexe B

## Manuel d'utilisation

Opt-Nefclass a une interface très facile à utiliser. Elle contient des onglets modulaires, chaque onglet est responsable d'une tâche bien définie.

### B.1 Extraction des règles

1. Sélection de la base d'apprentissage
2. Traitement de la base (perspective)
3. Choix du nombre maximal de règles
4. Choix de la méthode de performance (Best per Class ou Best Rule Learning)
5. Extraction des règles
6. Zone d'affichage des règles

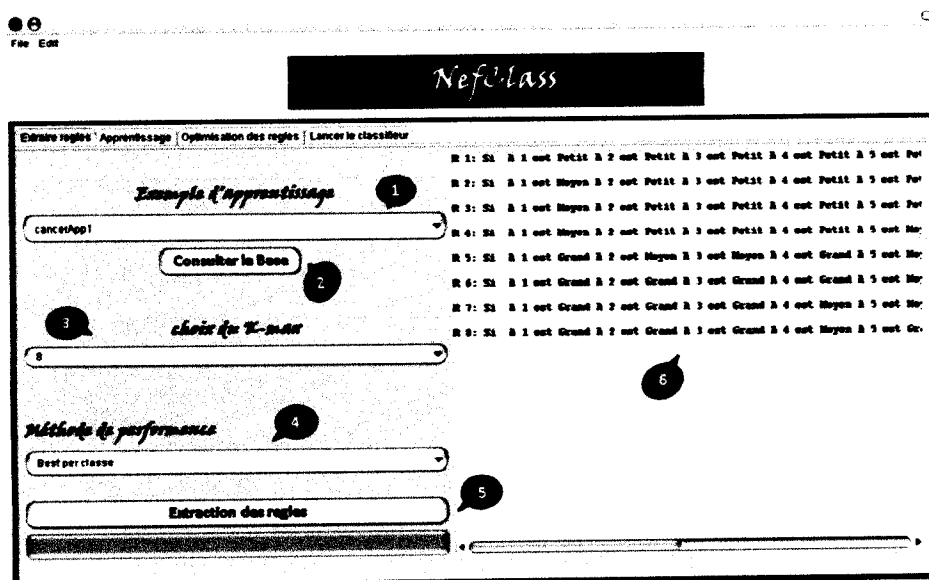


FIGURE B.1 – Génération des règles

## B.2 Apprentissage des sous-ensembles flous (Nefclass)

1. Saisie des paramètres (nombre d'itérations et le pas d'apprentissage)
2. Création du réseau de neurone
3. Affichage de la base de règles
4. Visualisation des sous-ensembles flous
5. Lancement de l'apprentissage
6. Panneau d'affichage des sous-ensembles flous

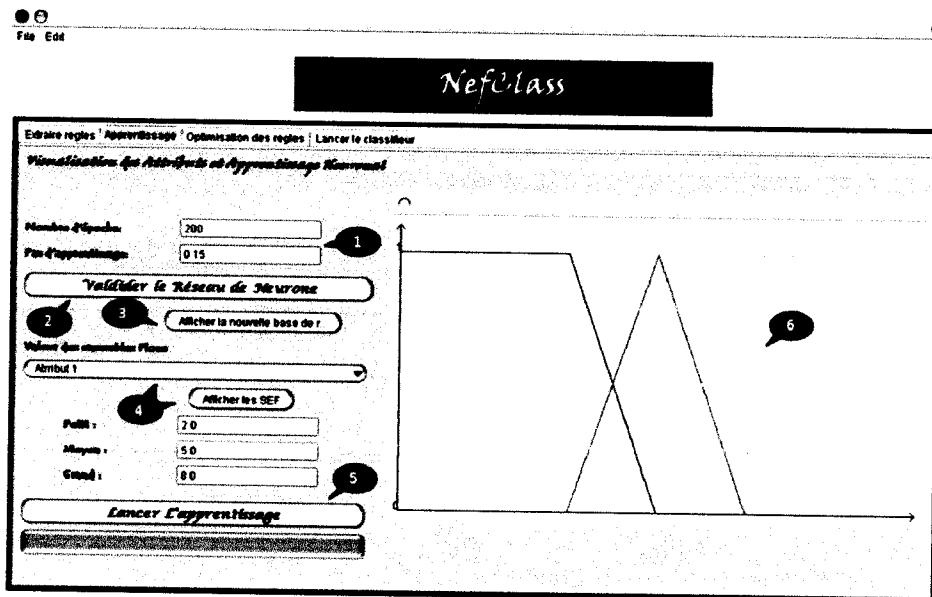


FIGURE B.2 – Apprentissage du réseau

## B.3 Optimisation (Opt-Nefclass)

1. Choix du type d'optimisation ; soit l'application du PSO sur le réseau de neurone (objectif de ce mémoire), soit l'application du PSO sur les règles flous (perspective).
2. Sélection de l'algorithme d'optimisation ; soit la PSO (déjà fait), soit les AG (perspective).
3. Configuration du PSO (une autre fenêtre s'affiche pour la modification des paramètres).
4. Lancement de l'optimisation.
5. Choix de la taille de l'essaim.
6. Choix de la taille du voisinage.
7. Choix du facteur de l'inertie et des coefficients.
8. Limitation de la vitesse des particules.
9. Initialisation de l'essaim.

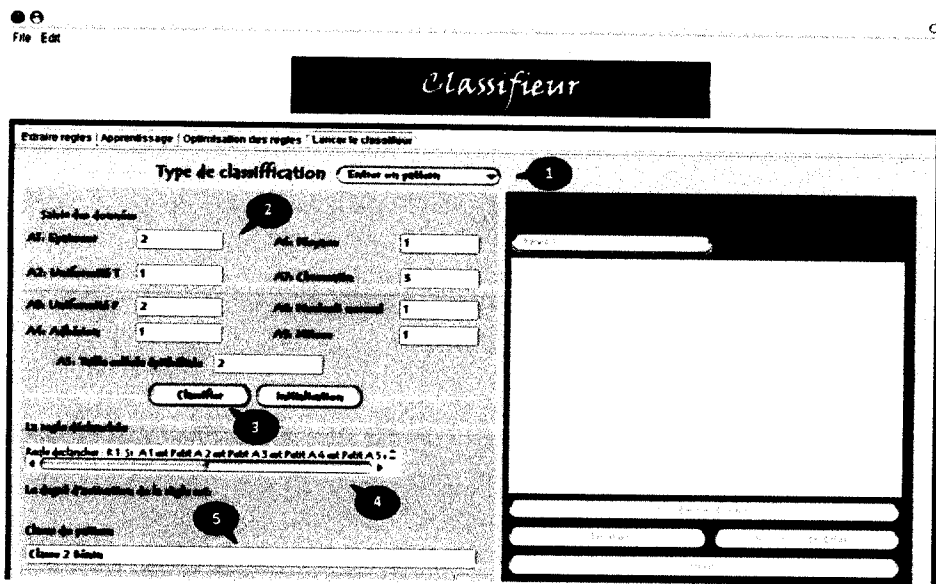


FIGURE B.5 – Classification d'un exemple

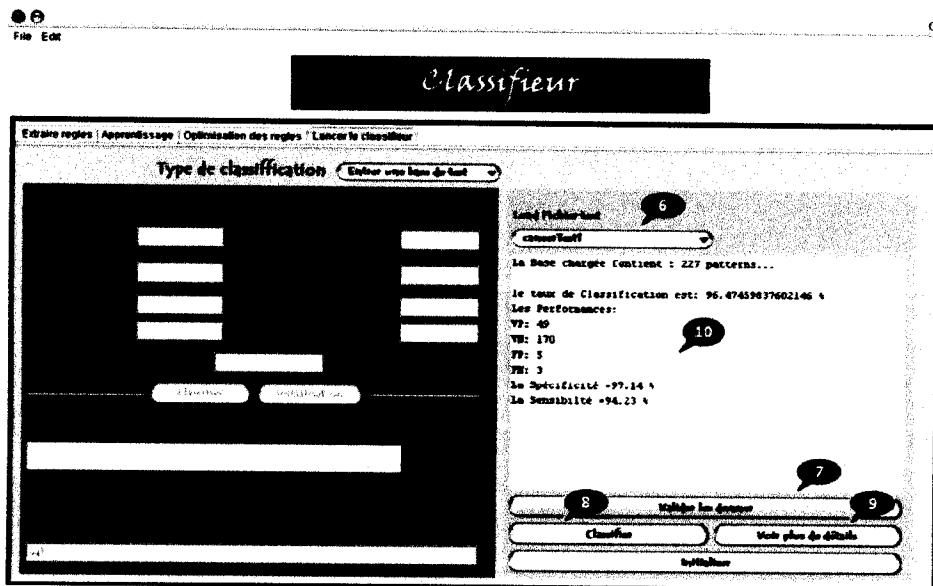


FIGURE B.6 – Classification d'une base