

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

APPROCHE NEURO-MIMÉTIQUE
AU SERVICE DU DÉPISTAGE
DU CANCER DU SEIN

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN INFORMATIQUE

PAR
HANAË NAOUM

OCTOBRE 2009

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

*À ma chère mère C. Bensaoula et à mon père A. Naoum.
À mon frère S. Naoum.
À toute ma famille, mes proches ainsi qu'à ceux qui pourront bénéficier de ce travail.*

REMERCIEMENTS

Mes remerciements vont au Pr. Mounir Boukadoum, qu'il trouve dans cet ouvrage un témoignage de ma profonde reconnaissance.

Je présente ma gratitude à mes collègues, au personnel du département d'informatique de l'UQÀM, au personnel du département d'informatique de l'UDL (Université Djillali Liabes), ainsi qu'à tous ceux qui ont pu contribuer de près ou de loin à l'aboutissement de ce travail.

TABLE DES MATIÈRES

LISTE DES FIGURES.....	vi
LISTE DES TABLES.....	viii
RÉSUMÉ.....	ix
CHAPITRE I	
INTRODUCTION ET PROBLÉMATIQUE.....	01
CHAPITRE II	
REVUE DE LA LITTÉRATURE.....	04
2.1 Introduction.....	04
2.2 Approches probabilistes et approches statistiques.....	04
2.3 Approches déductives et symboliques.....	12
2.4 Approches évolutionnaires.....	16
2.5 Approches connexionnistes	23
2.6 Conclusion.....	31
CHAPITRE III	
NOTIONS THÉORIQUES.....	33
3.1 Introduction.....	33
3.2 Notions fondamentales sur les mémoires associatives bidirectionnelles.....	34
3.3 Notion de chaos et de système dynamique : Application aux BAMs.....	38
3.3.1 Points fixes et points périodiques.....	38
3.3.2 Les cartes multidimensionnelles et cartes non linéaires.....	42
3.3.3 Le chaos et les attracteurs.....	44
3.3.4 Le contrôle du chaos.....	47

3.3.5 Les cartes chaotiques.....	48
3.4 Notions fondamentales sur le Perceptron multicouche.....	62
3.5 Conclusion.....	66
CHAPITRE IV	
ÉTUDES EXPÉRIMENTALES.....	67
4.1 Introduction.....	67
4.2 Descriptions et précisions sur les données utilisées.....	69
4.3 Expérimentation du réseau BAM.....	73
4.4.1 Architecture du BAM	74
4.4.2 Fonction de sortie et règle d'apprentissage.....	75
4.4 Expérimentation du réseau CNN-BAM	80
4.4.1 Architecture du CNN-BAM.....	80
4.4.2 Fonctions de sortie testées et règle d'apprentissage.....	80
4.5 Expérimentation du PMC.....	83
4.5.1 Architecture du Perceptron multicouches.....	83
4.5.2 Fonction de sortie et règle d'apprentissage.....	86
4.6 Étude comparative.....	91
4.6.1 Architectures.....	91
4.6.2 Règle d'apprentissage.....	92
4.6.3 Fonction de sortie.....	92
CONCLUSION.....	96
BILIOGRAPHIE.....	99

LISTE DES FIGURES

Figure 1. Architecture du système expert multiple	14
Figure 2. Architecture du système d'aide au diagnostic	24
Figure 3. Organigramme de l'approche proposée	26
Figure 4. Architecture du réseau neuro-flou.....	28
Figure 5. Le fonctionnement du BAM : (a) Direction avant ; (b) Direction arrière.....	34
Figure 6. Une orbite convergeant vers un attracteur 2-période.....	40
Figure 7. Les dynamiques locales proches d'un point fixe.....	43
Figure.8. Diagramme de bifurcation, carte logistique	49
Figure.9. Diagramme de bifurcation, carte logistique1	49
Figure.10. Diagramme de bifurcation, carte logistique3	49
Figure.11. Diagramme de bifurcation, carte logistique2	49
Figure.12. Diagramme de bifurcation, carte logis-cubic.....	50
Figure.13. Diagramme de bifurcation, carte Mira	51
Figure.14. Diagramme de bifurcation, carte Tent.....	53
Figure.15. Diagramme de bifurcation, carte Tent1	54
Figure.16. Diagramme de bifurcation, carte Tent2.....	54
Figure.17. Diagramme de bifurcation, carte Tailed_Tent.....	55
Figure 18. Le bassin d'attraction de la carte Hénon avec $\beta = -0.3$	56
Figure.19. Diagramme de bifurcation, carte Bernoulli.....	57
Figure.20. Diagramme de bifurcation, carte PWAM1.....	57
Figure.21. Diagramme de bifurcation, carte PWAM 2.....	58

Figure.22. Diagramme de bifurcation, carte PWAM 3.....	58
Figure.23. Diagramme de bifurcation, carte PWAM 4.....	59
Figure 24. Bassin d'attraction de la carte Ikeda ($C1 = 0.4$, $C2 = 0.9$, $R = 1$, $C3=6$).....	61
Figure 25. Illustration du l'apprentissage à minimisation de l'erreur.....	64
Figure 26. Distribution des neurones sur la carte SOM [20 20].....	72
Figure 27. Architecture du BAM	75
Figure 28. Courbe d'erreur au rappel en fonction du nombre de cycles d'apprentissage.....	78
Figure 29. Performance du BAM chaotique avec les différentes cartes	82
Figure 30. Architecture du Perceptron multicouche (10-10-1)	86
Figure 31. Architecture du Perceptron multicouche (10-10-10-1).....	87
Figure 32. Taux de reconnaissance du PMC avec une couche cachée.....	89
Figure 33. Taux de reconnaissance du PMC avec deux couches cachées.....	90

LISTE DES TABLES

Table 1. Résultats de reconnaissance pour la 10 ^{ième} partition de WDBC.....	05
Table 2. Taux d'erreur de classification des SVM (base de données <i>breast cancer</i>).....	07
Table 3. Taux de rappel à la classification avec la base de données « cancer du sein »	09
Table 4. Taux de reconnaissance des trois algorithmes conventionnels et de l'approche MBBC proposée dans	11
Table 5. Taux de reconnaissance de chaque sous-système et du système glob.....	15
Table 6. Classification basée sur les algorithmes d'apprentissage et de sélection de variables	18
Table 7. Taux de reconnaissance de l'algorithme proposé (<i>Ant_Miner3</i>) par rapport à l' <i>Ant_Miner</i> conventionnel	20
Table 8. Résultats de classification de la base de données <i>BrestCancer</i>	22
Table 9. Taux de classification du FNN	27
Table 10. Performance de l'approche MPNN	30
Table 11. Les différentes orbites de $g(x) = 3.3x(1-x)$	40
Table 12. Classification des cartes chaotiques selon le taux de succès à la reconnaissance.....	81

RÉSUMÉ

Les méthodes de classification automatiques sont un outil important de nos jours pour identifier différentes données et étudier leurs propriétés. On retrouve aujourd'hui l'utilisation des réseaux de neurones en tant que classificateurs dans plusieurs domaines d'application, notamment, le dépistage du cancer du sein. En effet, dans cet axe de recherche, beaucoup de travaux ont été réalisés en employant une approche neuronale pour la reconnaissance de formes et ont abouti à des résultats intéressants. Ceci étant, chaque modèle de réseau de neurones diffère d'un autre de par ses paramètres expérimentaux, sa complexité d'implantation et son domaine d'applicabilité. Ainsi, chaque modèle possède ses avantages et ses inconvénients, faisant qu'il soit plus approprié pour une tâche donnée qu'un autre.

Dans (Freeman 1987, Freeman et al., 1988) des expérimentations électrophysiologiques ont été réalisées sur des lapins prouvant l'existence de la dynamique chaotique dans les neurones réels. Depuis les années quatre-vingts, les réseaux de neurones chaotiques ont été le centre d'intérêt de plusieurs travaux en raison de leur dynamique complexe et leur application potentielle dans la dynamique associative, l'optimisation, et le traitement de l'information (Adachi et Aihara, 1997 ; Aihara et al., 1990 ; Tokuda et al., 1997). Dans cette optique, on s'intéresse dans ce travail à étudier la performance d'un modèle de réseau de neurones chaotique sur une tâche de classification reliée à la détection du cancer du sein. À cet effet, on procède dans un premier temps à la revue critique de certains travaux portant sur le sujet et qui ont été rapportés dans la littérature, de manière à situer le travail par rapport à ces travaux et mettre en évidence les contributions qu'il permet. Une étude comparative est ensuite établie où l'on met en avant les avantages d'utiliser une mémoire associative bidirectionnelle chaotique pour le travail de classification, relativement à deux modèles de réseau de neurones classiques, à savoir, une mémoire associative bidirectionnelle standard et un réseau multicouche à rétropropagation d'erreur.

CHAPITRE I

INTRODUCTION ET PROBLÉMATIQUE

Les méthodes de classification automatiques sont un outil important de nos jours pour identifier différentes données et étudier leurs propriétés. Dans le cadre de la classification, les réseaux de neurones artificiels (RNAs) ont largement fait leurs preuves au sein de la communauté scientifique et industrielle. À l'origine, ils constituent une tentative de modélisation formelle du cerveau humain. Ils représentent aussi une technique d'apprentissage machine qui a connu une évolution constante depuis les travaux de *Mac Culloch et Pitts* en 1943 (McCulloch 1943). Aujourd'hui, on retrouve l'utilisation des réseaux de neurones comme classifieurs avec plusieurs modèles et dans plusieurs domaines, notamment le contrôle des systèmes, la robotique et le domaine biomédical. En effet, beaucoup de travaux ont été réalisés en employant une approche neuronale pour la reconnaissance de formes et ont abouti à des résultats intéressants. Cela dit, chaque modèle de réseau de neurones 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 qu'il soit plus approprié pour une tâche donnée qu'un autre.

L'existence du chaos dans le cerveau a fait l'objet de plusieurs recherches durant les vingt dernières années. Le comportement chaotique a été découvert autant au niveau

microscopique (neuronal) que macroscopique du cerveau. Dès 1983, Guevara et ses collègues ont fait l'hypothèse qu'un comportement chaotique peut être responsable de certaines maladies telles que la schizophrénie, l'insomnie, l'épilepsie, et d'autres désordres dans le cerveau (Guevara et al., 1983 ; Babloyantz et al., 1985). L'équipe du Dr. Freeman (Freeman 1987, Freeman et al., 1988) a réalisé dans des expérimentations électrophysiologiques sur des lapins et a prouvé que la dynamique chaotique existe dans les neurones réels. Depuis les années 80s, les réseaux de neurones chaotiques ont été le centre d'intérêt de plusieurs travaux en raison de leur dynamique complexe et leur application potentielle dans la dynamique associative, l'optimisation, et le traitement de l'information (Adachi et Aihara, 1997 ; Aihara et al., 1990 ; Tokuda et al., 1997).

Par ailleurs, le cancer du sein est l'une des causes éminentes de mortalité chez les femmes dans les pays développés. Ainsi, l'étude du dépistage du cancer du sein représente un sujet de recherche d'actualité dans l'aide au diagnostic. La classification des pathologies dans le domaine médical a pour but de rassembler en classes (catégories) les 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 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 de ces classes, il est nécessaire de pouvoir faire l'identification de cas en les attribuant à la classe qui leur convient.

Le présent travail a pour but d'étudier la performance d'un réseau de neurones de type chaotique sur une tâche de classification reliée à la détection du cancer du sein. À cet effet, on procède dans un premier temps à la revue critique de certains travaux portant sur le sujet et qui ont été rapportés dans la littérature, de manière à situer le travail par rapport à ces travaux et mettre en évidence les contributions qu'il amène.

Secondairement, à travers une étude comparative, on met en avant les avantages d'utiliser une mémoire associative bidirectionnelle à fonction de sortie chaotique (*CNN-BAM* pour *Chaotic Neural Network-Bidirectional Associative Memory*) pour le travail de classification, relativement à deux modèles de réseau de neurones classiques, à savoir, une mémoire associative bidirectionnelle standard et un réseau multicouche à rétropropagation d'erreur .

Le mémoire comporte quatre chapitres dont le premier est une introduction générale. On présente dans le deuxième chapitre une revue de la littérature sur les techniques de reconnaissance de formes utilisées dans le cadre du dépistage du cancer du sein. Une présentation sommaire des notions fondamentales sur les trois modèles de réseaux de neurones étudiés est proposée dans le chapitre 3; on fournit en l'occurrence des concepts sur le chaos et ses fonctionnalités. Dans le chapitre 4, on présente les trois modélisations des RNAs implémentés ainsi que les résultats obtenus durant les différentes expérimentations respectives à chaque modèle. Le chapitre se termine par une étude comparative concernant les trois modèles et une discussion subséquente. Une conclusion fait le point sur les contributions de ce travail et met en avant les perspectives envisagées.

CHAPITRE II

REVUE DE LA LITTÉRATURE

2.1 Introduction

Les recherches liées à la détection du cancer du sein se sont multipliées durant la dernière décennie. Beaucoup de 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, on s'intéresse aux approches probabilistes et statistiques employées en tant que classifieurs 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 (Subhash et al., 2003), les auteurs proposent une approche qui est basée sur une généralisation de la règle des k plus proches voisins pour faire la classification dans le cadre du dépistage du cancer du sein. C'est une méthode qui représente un classifieur non paramétrique mais dont la performance dépend des vecteurs de distributions moyennes et des matrices de covariance. Si, de plus, ces distributions sont de nature gaussienne, la performance de cette approche devient 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. La table 1 illustre un exemple de résultat de classification de la dixième partition de cette série de données.

k	k -RNN rule				k -NN rule			
	Confusion matrix		Prob. of false positive false negative	Avg. error rate	Confusion matrix		Prob. of false positive false negative	Avg. error rate
1	335	9	0.026	0.041	337	7	0.020	0.033
	11	128	0.079		9	130	0.065	
2	334	10	0.029	0.039	338	6	0.017	0.048
	9	130	0.065		17	122	0.122	
3	334	10	0.029	0.037	335	9	0.026	0.037
	8	131	0.056		9	130	0.065	
4	334	10	0.029	0.037	338	6	0.017	0.041
	8	131	0.056		14	125	0.101	
5	334	10	0.029	0.037	337	7	0.020	0.035
	8	131	0.056		10	129	0.072	
6	334	10	0.029	0.039	338	6	0.017	0.041
	9	130	0.065		14	125	0.101	

**Table 1 Résultats de reconnaissance pour la 10^{ième} partition de WDBC
(Subhash et al., 2003)**

Les résultats obtenus à travers les expérimentations ont été mis en avant 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 (Subhash et al., 2003) 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 (Subhash et al., 2003) 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., 2003), 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 œuvre 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. Alors, l'hyperplan à marge maximale entre les classes est obtenu en minimisant la fonction de perte :

$$L(\alpha) = - \sum_i \alpha_i + \frac{1}{2} \sum_{ij} \alpha_i \alpha_j y_i y_j K(x_i, x_j);$$

où :

$\{(x_i, y_i)\}_{i=1}^N$ représente les paires d'entrées-sorties à deux classes telles que $y_i = \pm 1$, $K(x_i, x_j)$ est une fonction noyau appliquée aux entrées et $\{\alpha_i\}_{i=1}^N$ représente des valeurs de mise-à-jour multiplicatives.

En respectant la contrainte de non-négativité $\alpha_i \geq 0$. Le vecteur de l'hyperplan à marge maximal est $w = \sum_i \alpha_i^* y_i x_i$ et satisfait les contraintes de marge $y_i K(w, x_i) \geq 1$ pour toutes les données d'apprentissage ; α^* dénote la localisation du minimum de la fonction de perte.

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%

Table2. Taux d'erreur de classification des SVM (base de données *breast cancer*) (Fei et al., 2003)

Les résultats de classification sont illustrés dans la table 2. 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 σ de 1.0 et de 3.0. Les coefficients α_i ont été initialisés à une valeur de 1 d'une manière uniforme dans chaque expérimentation. Les taux d'erreur à 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 (Fei et al., 2003) aide à palier 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 (Fei et al., 2003) constitue un apprentissage hors-ligne. D'autres travaux comprennent des classifieurs à vastes marges, tel celui proposé dans (Baback et Shakhnarovich, 2002), où il est question d'une approche qui utilise des fonctions noyau dyadiques stimulées, ceci étant dans un but de maximiser la marge en offrant la possibilité d'un apprentissage en-ligne.

Dans un autre travail (Huang 2004), 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 (Huang 2004), cette hypothèse est ignorée et la méthode est appelée en conséquence : machine à probabilité Minimax biaisée (*BMPM* pour *Biased Minimax Probability Machine*). Le modèle a été transformé secondairement en une sorte de problème pseudo-concave avec un minimum local qui est aussi un maximum global. De là, 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.

Kernel	BMPM					MPM			
	α		Accuracy			α	Accuracy		
	α_x	α_y	TSA _x	TSA _y	TSA	α	TSA _x	TSA _y	TSA
Linear(%)	90.0 ± 0.3†	50.0 ± 0.0	99.9 ± 0.1†	92.0 ± 0.2	94.9 ± 0.2	84.2 ± 0.3	96.9 ± 0.4	97.1 ± 0.5	96.9 ± 0.3
Gaussian(%)	97.6 ± 0.3†	50.0 ± 0.0	100.0 ± 0.0†	88.9 ± 0.2	92.8 ± 0.2	90.1 ± 0.3	96.6 ± 0.2	97.1 ± 0.3	96.8 ± 0.2

Table 3. Taux de rappel à la classification avec la base de données « cancer du sein » (Huang 2004).

La classification a été effectuée sur une base de données non identifiée¹ avec l'approche classique *MPM* et l'approche proposée par les auteurs (Huang 2004), *BMPM*, en vue de mettre en avant la performance de celle-ci. Les simulations ont été faites avec deux types de fonctions, l'une étant linéaire et l'autre non, soit une fonction gaussienne. Les résultats obtenus avec les deux approches sont illustrés à la table 3. Les colonnes TSA_x, TSA_y, TSA (*Test-Set Accuracy*) représentent les taux de classification à la phase de test pour la classe x, la classe y, et la performance globale respectivement. Le terme α représente un indicateur qualitatif de performance. 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 (Huang 2004) 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. Or, pour l'aide au diagnostic médical dans le cadre du cancer du sein, les méthodes de

¹ Malheureusement, ce cas n'était pas unique durant la revue de la littérature.

classification multi-classes seraient plus appropriées. De plus, le modèle proposé repose essentiellement sur des estimations fiables des matrices de moyennes et de covariances. Toutefois, les méthodes d'estimation s'avèrent empiriques et difficiles à mettre en œuvre.

Dans (Madden 2002), 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 (Madden 2002). Plus précisément, la classification n'est pas influencée par les parties de la structure qui se situent en dehors du nœud de classification *Markov-Blanket*. Pour un nœud 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 nœud x n'est pas affecté par les nœuds qui sont en dehors de l'intervalle où son *Markov-Blanket* est défini. L'approche que l'auteur a utilisée a pour but de construire un *Markov-Blanket* approximatif autour du nœud classifieur. La performance des trois modèles conventionnels par rapport au modèle proposé est illustrée à la table 4. 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	97.03 ± 0.54
WBCD	97.81 ± 0.51	97.47 ± 0.68	97.17 ± 1.05	97.30 ± 1.01
LED-24	73.28 ± 0.70	73.18 ± 0.63	73.14 ± 0.73	73.14 ± 0.73
DNA	94.80 ± 0.44	94.75 ± 0.42	96.22 ± 0.64	95.99 ± 0.42
Lymph.	83.60 ± 9.32	85.47 ± 9.49	81.47 ± 10.4	83.47 ± 9.45
Nursery	90.48 ± 0.41	94.16 ± 0.33	92.63 ± 0.67	94.16 ± 0.33
SPECT	71.70 ± 6.56	81.25 ± 4.78	80.19 ± 4.66	80.75 ± 4.97
TTT	70.69 ± 1.94	75.08 ± 1.86	74.04 ± 3.51	77.37 ± 4.37

Table 4. Taux de reconnaissance des trois algorithmes conventionnels et de l'approche MBBC proposée dans (Madden 2002)

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) (Madden 2002). 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 mêmes des limites considérables. Dans plusieurs cas, on a eu des conceptions de méthodes qui reposent sur des paramètres dont l'estimation affecte sensiblement la performance globale du classifieur. Ce qui est notamment le cas pour les approches bayésiennes (Cheng 2006) et les méthodes utilisant les matrices de moyennes et les matrices de covariances. L'estimation des paramètres doit donc être faite d'une manière correcte,

et étant donné que cela se fait généralement d'une manière empirique, ceci peut rendre ce type de classifieur complexe à réaliser. De plus, dans quelques travaux cités, des hypothèses sont établies à la construction du classifieur pour faire en sorte de simplifier le problème de départ en le rendant dans certains cas linéairement séparable. Ceci remet en question la plausibilité des résultats obtenus et l'applicabilité du modèle proposé à des problèmes de classification réels. En dernier lieu, les classifieurs présentés sont pour la plupart binaires, donc applicables aux problèmes de classification à deux catégories, ce qui est loin d'être le cas des problèmes réels, en particulier dans le domaine du diagnostic du cancer du sein où l'on peut avoir affaire à plusieurs catégories de tumeurs.

2.3 Approches déductives et symboliques

Dans cette deuxième partie de la revue de littérature, on s'intéresse aux approches déductives et symboliques employées dans des travaux de recherche dans le cadre de la détection du cancer du sein.

Dans (Esmeir 2004), les auteurs proposent deux algorithmes pour l'induction d'arbres de décision dans le but de palier aux problèmes qu'on retrouve dans les algorithmes d'apprentissage pour les arbres de décision conventionnels. En effet, la majorité des algorithmes existants pour les arbres de décisions sont gloutons, le parcours de l'arbre se fait de haut en bas en calculant des décisions optimales locales au niveau de chaque nœud ; de plus, dans la plupart des cas l'arbre construit n'est pas globalement optimal. De là, les auteurs proposent deux algorithmes dans le but de réaliser un compromis entre la qualité de l'arbre et le temps d'apprentissage.

Le premier algorithme suggéré est inspiré de l'algorithme classique ID3. Il fait un parcours de K nœuds en profondeur dont le nombre de nœuds K dépend du temps d'apprentissage alloué, plus celui-ci est grand plus le paramètre K l'est en conséquence. Cet algorithme est appelé ID3-K.

Le second algorithme est appelé LSID3 (*Lookahead Stochastic ID3*), son fonctionnement est basé sur une recherche stochastique permettant d'évaluer des nœuds candidats à supprimer de l'arbre, une estimation de la taille de l'arbre est effectuée à chaque fois qu'une suppression de nœud est produite. L'approche fonctionne de manière à minimiser la taille de l'arbre escomptée.

Une série d'expérimentations a été effectuée pour évaluer la performance de classification des deux algorithmes ID3-K et LSID3, ainsi que leur capacité à minimiser la taille de l'arbre de décision. La moyenne des taux de reconnaissance obtenus avec le premier algorithme est de 65.1% avec une déviation à travers les différentes expériences de $\pm 8.4\%$. Les résultats avec l'algorithme employant la recherche stochastique ont été meilleurs avec une moyenne de 67% et une déviation de 9.3% à travers les différents tests.

On note que les algorithmes proposés sont intéressants pour minimiser les problèmes des algorithmes gloutons utilisés pour l'induction des arbres de décision, et ce, en essayant de réaliser un compromis dans l'optimisation de la taille de l'arbre et de sa qualité. En particulier dans (Esmeir 2004), l'algorithme LSID3 a procuré un meilleur taux de classification avec la recherche stochastique que l'algorithme ID3-K. Cependant, les résultats obtenus sont bien moindres que les taux de reconnaissance qui reviennent dans la littérature pour la base de données « cancer du sein ».

Dans un autre travail (De Santo et al., 2003), 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 1.

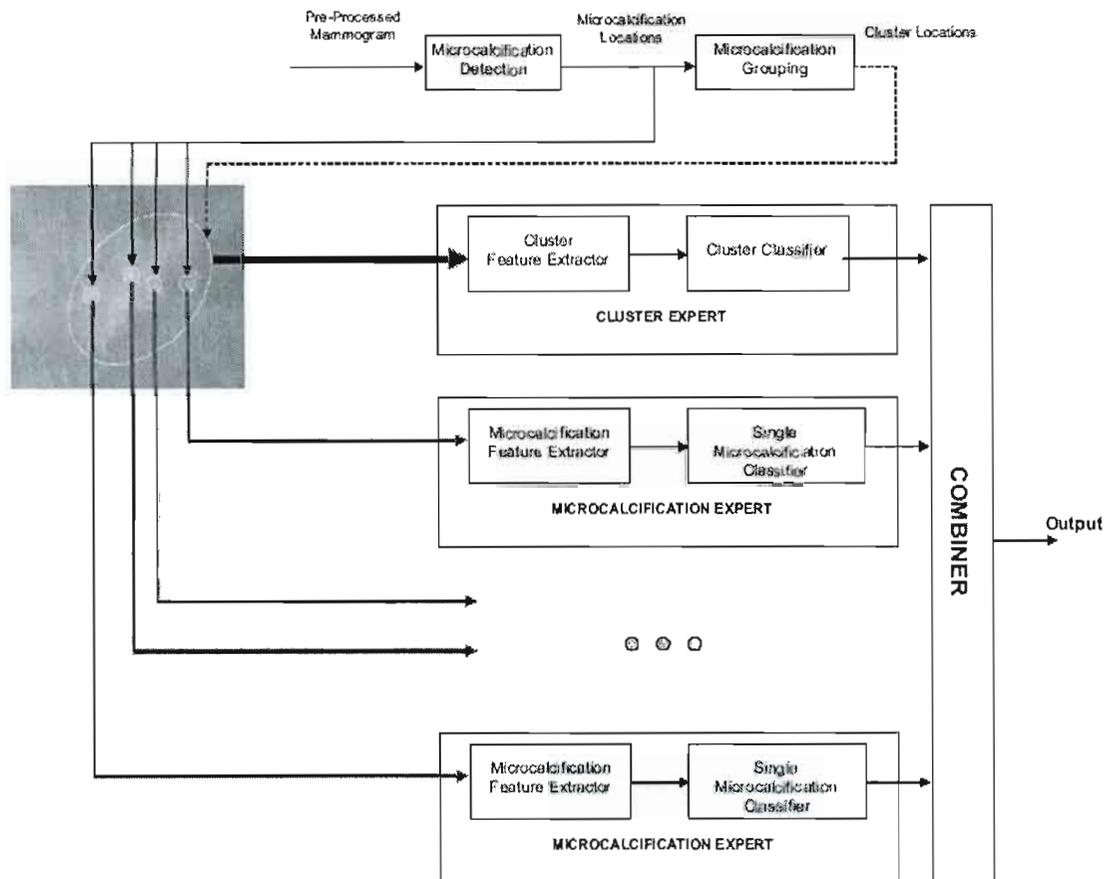


Figure 1. Architecture du système expert multiple (De Santo et al., 2003)

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 microcalcifications (1792 malignes et 331 bénignes). Les résultats de classification obtenus sont illustrés à la table 5.

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

**Table 5. Taux de reconnaissance de chaque sous-système et du système global
(De Santo et al., 2003)**

À travers les résultats rapportés dans la table 5, on pourrait noter que lorsque le système expert global MES (*Multi-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., 2001 ; Karnan et al., 2008). 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 œuvre 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 (Cheng et al., 2006). 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 (De Santo et al., 2003).

2.4 Approches évolutionnaires

Dans la troisième partie de cette revue de littérature, on s'intéresse aux approches évolutionnaires 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 essais particuliers.

Dans (Karnan et al., 2008), 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 essais 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 évolutionnaires différentes : un algorithme génétique, un

algorithme de colonies de fourmis et une approche à essais particulières. 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 essais particulières.

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 au 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*. Un poids est défini selon l'équation suivante :

$$w_k = \{c * [x_{kd+2} 10^{d-2} + x_{kd+3} 10^{d-3} + \dots + x_{(k+1)d}]\} / 10^{d-2}$$

où le paramètre k représente la population. Les poids sont extraits pour chaque chaîne dans la population. Les valeurs de la fonction *fitness* sont calculées comme suit :

$$F = 1 / E, \text{ où } E = \text{sqrt} ([E_1 + E_2 + \dots + E_m] / m)$$

où :

m représente le nombre total de patrons d'apprentissage;

E_1, E_2, \dots, E_m sont les erreurs respectives pour chacun des patrons.

La valeur fitness est ainsi calculée pour chaque population générée.

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-Knife 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 6. Classification basée sur les algorithmes d'apprentissage et de sélection de variables (Karnan et al., 2008)

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 6. 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 travail présenté en (Karnan et al., 2008) 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 (Karnan et al., 2008). 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., 2004), 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 (Parepinelli et al., 2002). 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 7. Taux de reconnaissance de l'algorithme proposé (*Ant_Miner3*) par rapport à l'*Ant_Miner* conventionnel (Liu et al., 2004)

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 à la table 7. On remarque 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 (Liu et al., 2004). 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 (Liu et al., 2004), ainsi, des expériences détaillées seraient requises et profitables pour pourvoir les déterminer d'une manière appropriée.

Une autre variante de l'algorithme *Ant_Miner* a été proposée dans (Jaganathan et al., 2007) pour effectuer une classification sur une base de données d'images mammographiques. 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 8. Résultats de classification de la base de données *BrestCancer* (Jaganathan et al., 2007)

Les expérimentations ont été effectuées de la même manière que dans (Liu et al., 2004) (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 8. 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*.

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, on s'intéresse 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 (Cheng et al., 1994; Baeg et Kehtarnavaz, 2000; Floyd et al., 1994), des réseaux utilisant des fonctions à base radiale (Bovis et al., 2000; Christoyianni et al., 2000), mais aussi des modèles neuronaux hybrides, où le réseau de neurones est combiné à une approche autre que connexionniste. On présente 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 (Cheng et al., 2006). 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.

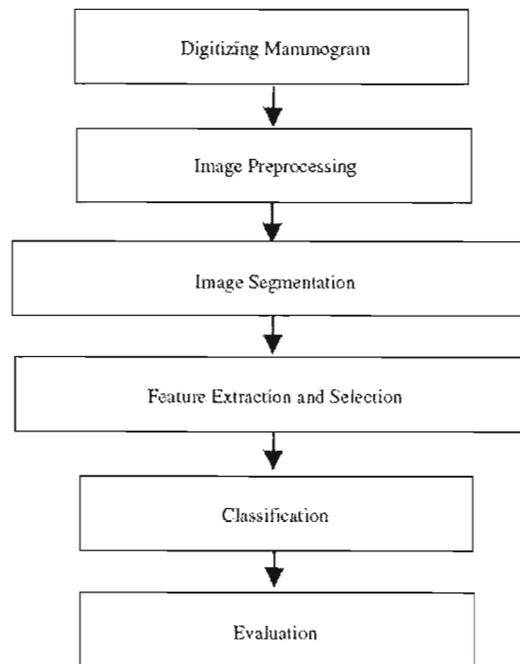


Figure 2. Architecture du système d'aide au diagnostic (Cheng et al., 2006)

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 (Cheng et al., 2006) ont été implémentés dans (Bovis et al., 2000). 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 de 65%.

Selon les résultats obtenus à travers les expérimentations ainsi que l'analyse qu'a fait l'auteur, on pourrait noter que bien qu'intéressants, les taux de classification obtenus sont moindres par rapport à ceux obtenus dans d'autres travaux cités précédemment. Aussi, étant donné qu'aucune règle n'existe pour fixer la taille de l'architecture de ce type de réseau de neurones, ce paramètre doit être fixé empiriquement. L'auteur

rapporte que l'apprentissage de ces modèles a été long et qu'il y avait surapprentissage.

Dans (Guo et Nandi, 2006), 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 œuvre une variante améliorée de l'analyse discriminante linéaire de Fisher (*MFLDA* pour *Modified Fisher Linear Discriminant Analysis*) 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 (Guo et Nandi, 2006). 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 (Cheng et Cui, 2004), 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 3. Tout comme dans le travail proposé dans (Guo et Nandi, 2006), un prétraitement des données a été effectué dans le but de faire une sélection de variables avant de procéder à l'étape de classification.

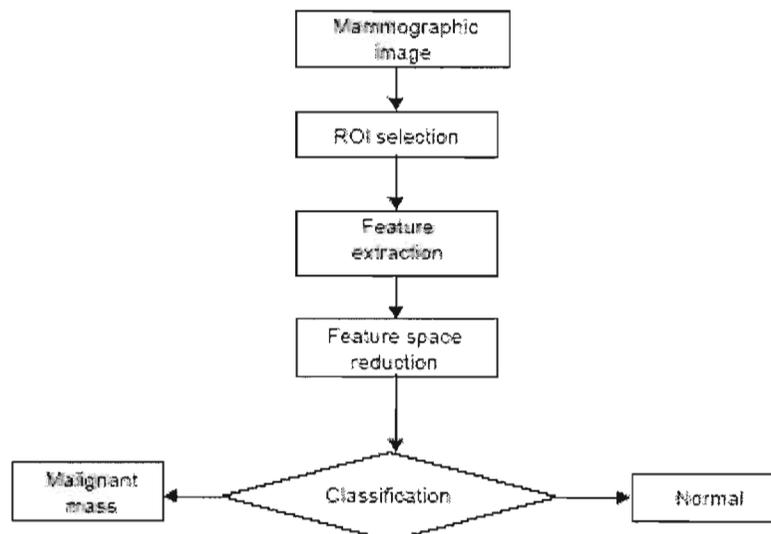


Figure 3. Organigramme de l'approche proposée (Cheng et Cui, 2004)

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

Alpha (α)	Beta (β)	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 9. Taux de classification du FNN (Cheng et Cui, 2004)

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 9, α est le paramètre d'apprentissage du réseau et β 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 β ; 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 (Grohman et Dhawan, 2001) 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 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.

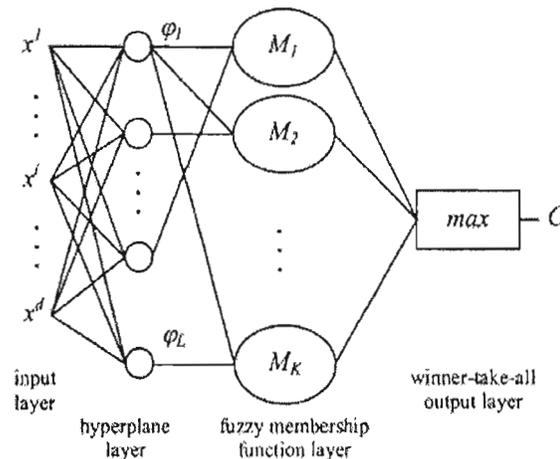


Figure 4. Architecture du réseau neuro-flou (Grohman et Dhawan, 2001)

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 (Grohman et Dhawan, 2001) 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 dans (Abbas 2002). Il s'agit d'une approche neuro-évolutionnaire 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-évolutionnaires 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 palier 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-évolutionnaire.

La base de données « cancer du sein » de Wisconsin est utilisée en vue de faire la classification du cancer du sein. Le nombre de classes est de 2 : classe de cas bénins et classe de cas malins. 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 10.

Method	Error	Number of epochs
MPANN	0.981 ± 0.005	5100
Fogel et al.	0.981 ± 0.464	200000
Abbass et al.	0.975 ± 1.800	10000

Table 10. Performance de l'approche MPNN (Abbas 2002)

L'auteur a mis en avant les résultats obtenus avec l'approche qu'il a proposée de part ses taux de reconnaissance mais aussi de 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. On remarque selon les résultats illustrés dans la table 10 que l'approche

MPANN (*Mimetic Pareto Neural Network*) proposée a aboutit à de meilleurs taux de reconnaissance en un temps de rappel moindre par rapport aux deux autres approches. Il est cependant à noter qu'il s'agit dans ce travail d'une classification binaire. Or, pour les problèmes réels dans le domaine du dépistage du cancer du sein, on a souvent affaire à un plus grand nombre de classes. Aussi, l'hybridation que comprend l'approche proposée amène à avoir un système global à paramètres multiples ayant un impact sur la performance.

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 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. C'est l'objectif de ce mémoire.

On conçoit dans ce travail trois différents modèles de réseaux de neurones artificiels dans le but d'évaluer leur performance à la classification de cas de tumeurs en utilisant la même base de données. Secondairement, on propose une étude comparative des trois modèles en essayant de mettre en avant quelques points discutés dans cette revue de littérature.

CHAPITRE III

NOTIONS THÉORIQUES

3.1 Introduction

Dans le chapitre précédent, on a passé en revue différents travaux portant sur les techniques de classification dans le cadre du dépistage du cancer du sein. On a mis de l'avant certaines limites des différentes méthodes auxquelles on tente de remédier à travers les expérimentations proposées dans le chapitre 4. Dans le présent chapitre, on présente les notions fondamentales sur lesquelles est basée la conception des trois modèles de réseaux de neurones proposés et testés dans le chapitre expérimental. On commence par citer les notions de base sur les mémoires associatives bidirectionnelles, puis on présente la règle d'apprentissage utilisée dans les deux modèles à mémoire que l'on propose, l'un d'eux étant un réseau de neurones chaotique. À cet effet, dans la deuxième section de ce chapitre on présente certaines notions sur le chaos, les systèmes dynamiques et leurs utilisations dans les mémoires associatives. Enfin, à la troisième section, on introduit le réseau à couches, soit le Perceptron multicouche ainsi que la règle d'apprentissage que l'on utilise dans ce modèle.

3.2 Notions fondamentales sur les mémoires associatives bidirectionnelles

La mémoire associative bidirectionnelle a été proposée pour la première fois par Bart Kosko (Negnevitsky 2005). Sa fonction primaire est d'associer les patrons d'une série de données A à ceux d'une autre série B , et vice versa. Le modèle de réseau BAM (*Bidirectional Associative Memory*) peut faire de la généralisation en plus de sa caractéristique de robustesse puisqu'il peut générer des sorties correctes malgré des données incomplètes ou bruitées.

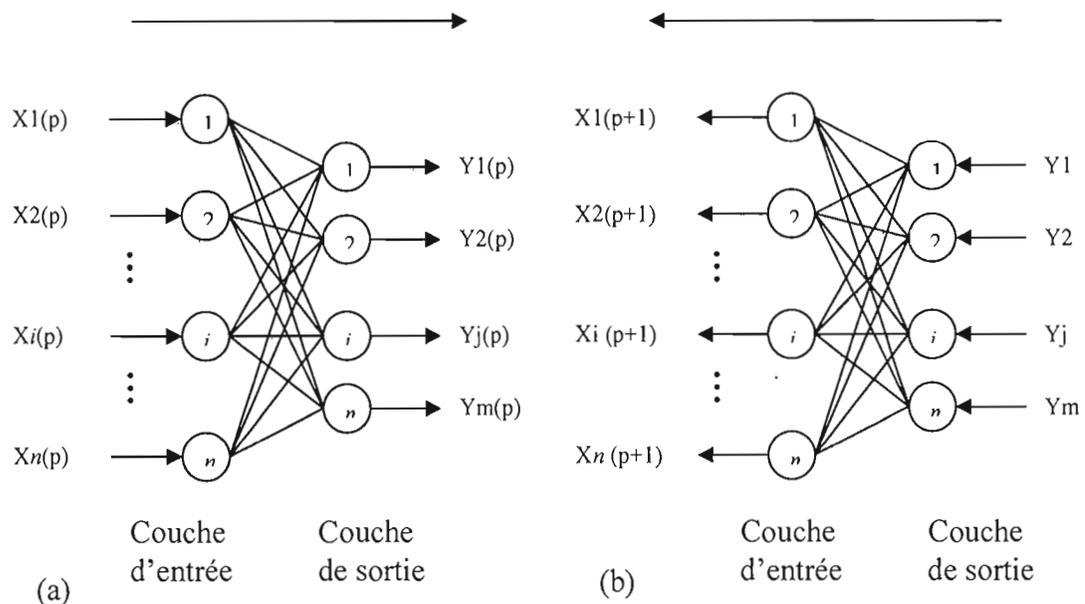


Figure 5. Le fonctionnement du BAM : (a) Direction avant ; (b) Direction arrière (Negnevitsky 2005)

L'architecture BAM de base est illustrée à la figure 5. Ce réseau de neurones est constitué de deux couches entièrement connectées, une couche d'entrée et une couche de sortie (Negnevitsky 2005).

Règle d'apprentissage

Un vecteur d'entrée $X(p)$ est affecté à la matrice transposée des poids W^T pour produire un vecteur de sortie $Y(p)$, tel qu'illustré à la figure 5 (a). Ensuite, le vecteur de sortie $Y(p)$ est affecté à la matrice de poids W pour produire un nouveau vecteur d'entrée $X(p+1)$, tel qu'illustré à la figure 5 (b). Ce processus est répété jusqu'à ce que le vecteur d'entrée et le vecteur de sortie ne changent plus, en d'autres termes, jusqu'à ce que le BAM atteigne un état stable (Negnevitsky 2005).

Pour réaliser un réseau BAM, on doit créer une matrice de corrélation pour chaque paire de patrons que l'on veut mémoriser. Cette matrice est le produit du vecteur d'entrée X et du transposé du vecteur de sortie Y^T . La matrice de poids du réseau est la somme de toutes les matrices de corrélation, soit :

$$W = \sum_{m=1}^M X_m Y_m^T$$

où M est le nombre des paires de patterns à mettre en mémoire par le BAM (Negnevitsky 2005). Dans ce modèle, l'apprentissage se fait donc hors-ligne.

Dans le cadre du présent travail, on utilise une règle d'apprentissage différente de celle présentée précédemment. Pour les expérimentations sur le BAM et le réseau BAM chaotique, on a employé une règle inspirée de celle utilisée dans (Chartier et Boukadoum, 2006). Elle fait intervenir des matrices de poids différentes pour les deux directions entre les couches x et y et inclut la fonction de sortie lors de l'évaluation des matrices de poids. De ce fait, elle tente de résoudre les deux contraintes non-linéaires suivantes :

$$X = f(VY)$$

$$Y = f(W^T X)$$

où V et W sont les deux matrices de poids dans les deux directions. Ainsi, la forme de ces deux contraintes et la récurrence du réseau font que le processus d'apprentissage est exécuté en-ligne. La règle d'apprentissage que l'on a utilisée sur les deux modèles à mémoire est dérivée de l'approche hebbienne/anti-hebbienne (Chartier et Boukadoum, 2006) :

$$W_{[k+1]} = W_{[k]} + \eta (y_{[0]} x_{[0]}^T + y_{[0]} x_{[t]}^T - y_{[t]} x_{[0]}^T - y_{[t]} x_{[t]}^T)$$

$$V_{[k+1]} = V_{[k]} + \eta (x_{[0]} y_{[0]}^T + x_{[0]} y_{[t]}^T - x_{[t]} y_{[0]}^T - x_{[t]} y_{[t]}^T)$$

où :

W et V représentent les matrices de poids synaptiques du réseau dans les deux directions,

$x_{[0]}$ et $y_{[0]}$ sont les entrées initiales à associer,

η est le paramètre d'apprentissage,

et k est le numéro de l'itération courante.

On peut noter que cette règle d'apprentissage inclut un retour à partir de la fonction de sortie non-linéaire via $x_{[t]}$ et $y_{[t]}$. Ceci permet au réseau d'apprendre en-ligne et à contribuer ainsi à la convergence des poids synaptiques. Pour plus de détails, voir (Chartier et Boukadoum, 2006).

Il est important de préciser que les deux modèles à mémoire proposés dans le chapitre 4 sont des mémoires associatives utilisant la règle d'apprentissage dérivée

de l'approche hebbienne/anti-hebbienne. Toutefois, l'un de ces deux modèles a pour fonction de sortie une carte chaotique. Notons par ailleurs, que plusieurs cartes chaotiques ont été exploitées dans ce modèle. Dans la section qui suit, on présente quelques notions de base sur les systèmes dynamiques ainsi que certaines spécifications sur les différentes cartes chaotiques utilisées sur le CNN-BAM.

3.3 Notion de chaos et de système dynamique : Application aux *BAMs*

Dans (Alligood et al., 1996), les auteurs définissent un système dynamique avec un ensemble d'états possibles et doté d'une règle qui détermine l'état courant en fonction des états précédents. La règle est déterministe, i.e. on ne peut déterminer l'état présent qu'à partir des états précédents, et il n'y a pas de hasard dans un tel système. L'évolution du processus dynamique est déterminée par la composition d'une fonction f telle que :

$$f^k(x) = f(\dots\dots f(x))$$

est le résultat de l'application de la fonction f à un état initial k fois.

3.3.1 Points fixes et points périodiques

Le réseau de neurones chaotique que l'on présente dans le chapitre expérimental est une mémoire associative bidirectionnelle qui a comme fonction de sortie une fonction chaotique, qui est en fait une carte chaotique. On définit une carte (*map*) comme étant une fonction dont le domaine est un espace qui est le même que l'espace de sortie (Alligood et al., 1996).

Une orbite de x sous la fonction chaotique f est définie par l'ensemble de points $\{x, f(x), f^2(x), \dots\}$; où le 1^{er} point est appelé : valeur initiale de l'orbite (Alligood et al., 1996). La carte chaotique f possède un point fixe p si $f(p) = p$. Les points fixes d'une carte f sont obtenus soit en calculant les solutions de l'équation $f(x)=x$, ou bien, d'une manière graphique, en localisant les intersections de la carte f avec la droite $y = x$, des exemples sont disponibles dans (Alligood et al., 1996).

Le voisinage d'un point fixe p , $N_\epsilon(p)$, est défini par l'ensemble de points x tels que pour une distance réelle ϵ , on a (Alligood et al., 1996):

$$N_\varepsilon(p) = \{ x \in \mathbb{R} : |x-p| < \varepsilon \}, \varepsilon > 0$$

Un point fixe p est attracteur si p peut rattraper tous les points qui lui sont proches tel que (Alligood et al., 1996):

$$\text{Si } \exists \varepsilon > 0, \text{ tel que } \forall x \in N_\varepsilon(p), \text{ alors } \lim_{k \rightarrow \infty} f^k(x) = p.$$

Un point fixe p est repousseur si p peut repousser tous les points qui lui sont proches, tel que (Alligood et al., 1996):

$$\text{Si } \exists N_\varepsilon(p), \forall x \in N_\varepsilon(p) \ x \neq p$$

En d'autres termes, x est projeté hors du voisinage $N_\varepsilon(p)$.

La valeur absolue de la dérivée d'une carte f au point fixe p permet de mesurer l'effet de f sur la distance entre p et un point au voisinage de p . Si $|f'(p)| > 1$, p est un attracteur si $|f'(p)| < 1$.

Une carte chaotique peut aussi avoir des points périodiques (Alligood et al., 1996). On peut illustrer ceci à travers l'exemple de la carte logistique $g_a(x) = ax(1-x)$. L'orbite avec une condition initiale 0.2 est montrée à la figure 6. Cette figure montre une orbite convergeant vers un attracteur $\{p_1, p_2\}$. Cette orbite est alternativement rattrapée vers p_1 à une itération, et vers p_2 à l'autre. On note aussi que l'orbite reste dans un patron de valeurs alternatives $p_1 = 0.4794$ et $p_2 = 0.8236$ quand $a = 3.3$. Un exemple de ce type d'orbite en partant de trois états initiaux différents est illustré dans la table 11.

n	$g^n(x)$	$g^n(x)$	$g^n(x)$
0	0.2000	0.5000	0.9500
1	0.5280	0.8250	0.1568
2	0.8224	0.4764	0.4362
3	0.4820	0.8232	0.8116
4	0.8239	0.4804	0.5047
5	0.4787	0.8237	0.8249
6	0.8235	0.4792	0.4766
7	0.4796	0.8236	0.8232
8	0.8236	0.4795	0.4803
9	0.4794	0.8236	0.8237
10	0.8236	0.4794	0.4792
11	0.4794	0.8236	0.8236
12	0.8236	0.4794	0.4795
13	0.4794	0.8236	0.8236
14	0.8236	0.4794	0.4794

Table 11. Les différentes orbites de $g(x) = 3.3x(1-x)$ (Alligood et al., 1996)

Si on suppose qu'une carte f a un point p qui est un point k -périodique, et que p est un attracteur de la carte f^k , alors l'orbite k -périodique de p est un attracteur périodique. Aussi, l'orbite k -périodique de p est une source périodique, si p est une source de la carte f^k .

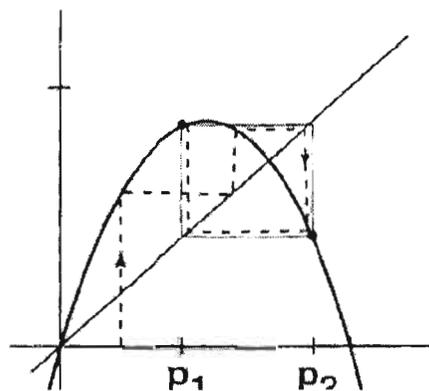


Figure 6. Une orbite convergent vers un attracteur 2-période (Alligood et al., 1996)

Si on reprend l'exemple de la carte logistique $g_a(x) = ax(1-x)$, celle-ci a en général le comportement suivant (Alligood et al., 1996) :

- Si $0 \leq a < 1$, la carte possède un attracteur à $x = 0$. Dans ce cas toutes les conditions initiales entre 0 et 1 sont rattrapées par cet attracteur.

- Si $1 < a < 3$, alors on a un attracteur à $x = (a-1/a)$ car $|g'_a(x)| < 1$.

- Si $a \geq 3$, alors le point fixe $x = (a-1/a)$ est instable car $|g'_a(x)| > 1$; par contre on a un attracteur de période 2 qui se génère (voir la table 11 pour $a = 3.3$).

On note que pour certaines valeurs de a , l'orbite prend un comportement aléatoire, par exemple pour $a = 3.86$, l'attracteur est chaotique (Alligood et al., 1996).

Soit $\{p_1, \dots, p_k\}$ une orbite k -période de la carte f , f étant une carte une-dimension.

On a :

$$\begin{aligned} (f_k)'(p_1) &= (f(f_{k-1}))'(p_1) \\ &= f'(f_{k-1}(p_1))(f_{k-1})'(p_1) \\ &= f'(f_{k-1}(p_1)) f'(f_{k-2}(p_1)) \dots f'(p_1) \\ &= f'(p_k) f'(p_{k-1}) \dots f'(p_1) \end{aligned}$$

On détermine la stabilité des orbites périodiques telle qu'une orbite périodique $\{p_1, \dots, p_k\}$ est un attracteur (*sink*) si :

$$|f'(p_k) \dots f'(p_1)| < 1$$

et un repousseur (source) si :

$$|f'(p_k) \dots f'(p_1)| > 1$$

Cela indique que la $k^{\text{ième}}$ dérivée f^k de f à un point de l'orbite k -période est le produit des dérivées de f aux points k de l'orbite. En particulier, la stabilité est une propriété collective de l'orbite périodique, tel que :

$$(f^k)'(p_i) = (f^k)'(p_j), \text{ pour tout } i \text{ et } j.$$

3.3.2 Les cartes multidimensionnelles et les cartes non linéaires

Soit f une carte définie dans \mathbb{R}^m et $p \in \mathbb{R}^m$ tel que p est un point fixe de f , i.e. $f(p) = p$ (Alligood et al., 1996):

- ⬇ Si $\exists \varepsilon > 0$ tel que $\forall x \in N_\varepsilon(p)$, alors $\lim_{k \rightarrow \infty} f^k(x) = p$. En d'autres termes, si tous les points suffisamment proches de p sont rattrapés par p , alors p est un point fixe attracteur.
- ⬇ Si $\exists N_\varepsilon(p)$, tel que $\forall x \in N_\varepsilon(p) \ x \neq p$, alors x sera mappé à l'extérieur du voisinage $N_\varepsilon(p)$. En d'autres termes, si tous les points suffisamment proches de p sont repoussés, alors p est un point fixe repousseur.

La figure 7 montre le point attracteur et le point repousseur d'une carte f à deux dimensions ainsi que leurs images sous f . Le schéma 7(a) représente l'origine, 7(b) un attracteur, et 7(c) illustre un type de point fixe qui ne peut exister dans une carte à une dimension. Ce type de point fixe a au moins une direction d'attraction et une autre direction de répulsion (Alligood et al., 1996).

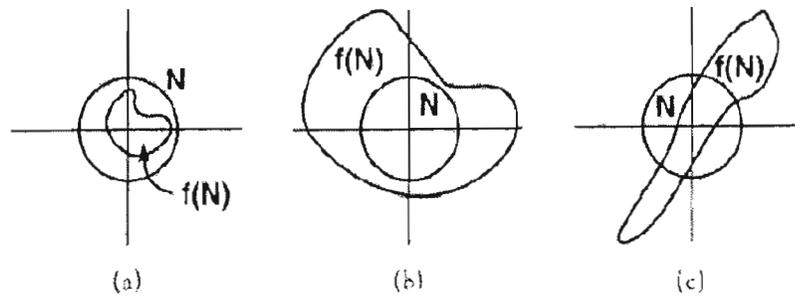


Figure 7. Les dynamiques locales proches d'un point fixe (Alligood et al., 1996)

Dans (Alligood et al., 1996), il est montré que la clé décisive de la stabilité d'un point fixe est la dérivée de la carte en ce point. La dynamique au voisinage d'un point fixe est déterminée par la meilleure approximation linéaire de la carte en ce point. Cette meilleure approximation linéaire est donnée par la matrice Jacobienne et, pour déterminer la stabilité d'une carte non linéaire de dimension supérieure à 1, on doit calculer la matrice Jacobienne (Alligood et al., 1996) :

Soient $f = (f_1, \dots, f_m)$ une carte dans \mathbb{R}^m , et un point $p \in \mathbb{R}^m$; alors la matrice Jacobienne de f au point p est :

$$Df(p) = \begin{bmatrix} \sigma f_1 / \sigma x_1 (p) & \dots & \sigma f_m / \sigma x_m (p) \\ \vdots & & \vdots \\ \sigma f_m / \sigma x_1 (p) & \dots & \sigma f_m / \sigma x_m (p) \end{bmatrix}$$

où $\sigma f_i / \sigma x_j (p)$ représente la dérivée partielle de f évaluée au point p .

Par ailleurs, pour ce qui est des cartes linéaires, la matrice Jacobienne est égale à la carte même (Alligood et al., 1996).

Dans ce qui suit, on présente quelques bases sur la théorie du chaos.

3.3.3 *Le chaos et les attracteurs*

La description des attracteurs représente un aspect important pour expliquer le phénomène dynamique. Deux types de mouvement d'attraction ont été définis par Newton: la pomme qui reste sur terre en équilibre et les planètes dans le système solaire qui sont périodiques ou plus précisément quasi périodiques. Durant les 300 ans qui suivirent, ces deux types de mouvement sont restés les seuls mouvements connus pour les systèmes dynamiques simples. Maxwell et Poincaré étaient parmi la minorité de scientifiques qui n'étaient pas d'accord avec ce point de vue. Au dernier quart du 20^{ème} siècle, un troisième type de mouvement devient reconnu : le Chaos (Alligood et al., 1996).

Contrairement à ce que pensait Newton, tel que rapporté par James Gleick dans (Gleick 1999), on sait maintenant que tous les systèmes dynamiques ne sont pas identiques. On distingue deux types de systèmes, les systèmes stables et les systèmes instables. Parmi les systèmes instables, il y a une classe particulièrement intéressante, qui est associée au chaos déterministe. Dans le chaos déterministe, les lois microscopiques sont déterministes mais les trajectoires prennent un aspect aléatoire. Cela provient de la " sensibilité aux conditions initiales " : la moindre modification des conditions initiales entraîne des divergences exponentielles. Dans une seconde classe de systèmes, l'instabilité va jusqu'à détruire les trajectoires. Une particule n'a plus de trajectoire unique, mais différentes trajectoires sont possibles, auxquelles s'attache chaque fois une probabilité.

Techniquement, le terme " *chaos* " définit un état particulier d'un système caractérisé par le comportement suivant (Alligood et al., 1996; Gleick 1999):

- Ne se répète jamais et semble erratique,

- A une dépendance sensible par rapport aux conditions initiales : des différences extrêmement faibles dans les valeurs des paramètres peuvent aboutir à des résultats largement divergents,

- N'en est pas moins ordonné et caractérisé par un déterminisme imprévisible. Le déterminisme imprévisible signifie que même un modèle parfait de système chaotique (équations de mouvement identiques et mêmes conditions initiales) débouche sur des résultats imprévisibles.

Les systèmes en état de chaos sont donc « légitimes, ordonnés, déterministes et imprévisibles » (Gleick 1999).

Exposant Lyapunov

On a vu précédemment dans cette section que la stabilité des systèmes dynamiques est fortement sensible à la dérivée des cartes définissant leurs points fixes. Par exemple, dans un système discret si x_1 est un point fixe de la carte f , et $f'(x_1) = \delta > 1$, alors l'orbite de chaque point x proche de x_1 serait séparé de x_1 d'une distance qui vaut δ à chaque itération jusqu'à ce que l'orbite se déplace suffisamment de x_1 . De la même façon, si x_1 est un point k -périodique la distance moyenne de séparation à chaque itération serait $\delta^{1/k}$ [34]. À cet effet, le nombre Lyapunov est introduit pour quantifier la distance moyenne de séparation des points x très proches des points périodiques ou fixes tout le long d'une orbite.

L'exposant de Lyapunov est un algorithme naturel pour trouver le nombre de Lyapunov. Tel qu'évoqué auparavant, le chaos est caractérisé par un exposant de Lyapunov supérieur à 0. Une orbite chaotique se définit par le comportement instable proche d'un point repousseur (source) ou bien *saddle* i.e. proche d'un état d'équilibre non stable (*unstable steady state*) (Alligood et al., 1996).

Dans (Alligood et al., 1996), on définit le nombre Lyapounov tel que pour une carte continue f qui est dérivable dans \mathbb{R} , le nombre Lyapounov $L(x_1)$ de l'orbite $\{x_1, \dots, x_n\}$ est:

$$L(x_1) = \lim_{n \rightarrow \infty} (|(f)'(x_1)| \dots |(f)'(x_n)|)^{1/n}$$

Dans le cas où cette limite existe alors l'exposant Lyapounov $h(x_1)$ est défini par :

$$h(x_1) = \lim_{n \rightarrow \infty} (1/n) [\ln|(f)'(x_1)| + \dots + \ln|(f)'(x_n)|],$$

Si cette limite existe. On note que h existe si et seulement si L existe, qu'il est différent de la valeur 0, et que $\ln L = h$.

Selon (Alligood et al., 1996), pour une carte f dans \mathbb{R} , l'orbite bornée $\{x_1, x_2, \dots\}$ de f est chaotique si :

1. $\{x_1, x_2, \dots\}$ n'est pas asymptotiquement périodique, et
2. l'exposant Lyapounov $h(x_1) > 0$.

Les bassins d'attraction

Soient f une carte définie dans \mathbb{R}^n , et p est un point fixe attracteur ou point fixe périodique de f . Le bassin d'attraction de p est l'ensemble de points x tel que [34] :

$$|f^k(x) - f^k(p)| \rightarrow 0, k \rightarrow \infty.$$

Les attracteurs Chaotiques

On sait qu'une orbite chaotique n'est ni périodique ni asymptotiquement périodique, et a au moins un exposant Lyapunov positif. Le chaos introduit un mouvement « *Twist* ». Les orbites chaotiques peuvent être rattrapés; si l'ensemble de limites directes d'une telle orbite chaotique contient l'orbite elle-même, alors l'attracteur est chaotique. On peut noter par conséquent qu'un attracteur chaotique est un ensemble de limites directes « *Forward Limit set* » (Alligood et al., 1996).

3.3.4 Le contrôle du chaos

Dans (Adachi et Aihara, 1997), le comportement associatif dynamique d'un réseau de neurones chaotique a été étudié en détails. Le réseau conçu réussit à retrouver les patrons en mémoire ; toutefois, ces patrons surviennent d'une manière non-périodique car le réseau opère dans un mode chaotique. Il est ainsi difficile de distinguer les différents patrons mémorisés dans un réseau de neurones chaotique (Kobori et al., 1996), ce qui peut limiter l'usage de la fonction de mémoire associative dans certaines applications. Par conséquent, bien qu'il ait été montré que la dynamique chaotique peut être une technique prometteuse pour le traitement de l'information, les sorties du réseau neuronal chaotique arpentent autour des patrons mémorisés et ne peuvent être stabilisés à un de ces patrons (Guoguang et al., 2003 ; Guoguang et al., 2007 ; Kushibe et al., 1996). Cela dit, les problèmes de convergence dans la dynamique chaotique n'ont pas été tout à fait résolus. Dès lors, on fait face à certaines difficultés dans l'application de la dynamique chaotique dans le domaine du traitement de l'information. À cet effet, pour réaliser un traitement de l'information tel que la reconnaissance de formes avec les réseaux de neurones chaotiques, il y est nécessaire de procéder au contrôle du chaos (Watanabe et al., 2000 ; Guoguang et al., 2003 ; Guoguang et al., 2004 ; Guoguang et al., 2007 ; Guoguang et al., 2007 ; Shin et Takuya, 1998).

Depuis le travail de Ott, Grebogi et York (OGY) dans (Ott et al., 1990), plusieurs travaux de recherche se sont orientés vers l'étude du contrôle du chaos. On contrôle le chaos en général afin d'atteindre les trois objectifs (Guoguang et al., 2007) :

- Pour changer l'état chaotique en un état stable ou périodique.
- Pour permettre à un état stable ou périodique de se transformer en un état chaotique.
- Pour basculer d'un état chaotique à un autre.

Dans la sous-section qui suit, on présente les différentes cartes chaotiques que l'on utilise par la suite dans les expérimentations sur le BAM chaotique. On détermine les paramètres de chaque fonction, notamment les paramètres de contrôle du chaos.

3.3.5 Les cartes chaotiques

On décrit dans cette section les différentes cartes chaotiques que l'on a utilisées durant les expérimentations sur le réseau BAM chaotique. Les résultats des simulations sont détaillés dans le chapitre 4. On détermine les valeurs des différents paramètres de chacune des fonctions de manière à ce qu'elles opèrent en mode chaotique. Le diagramme de bifurcation est illustré pour les fonctions dont on a fixé les paramètres. Pour définir les paramètres de certaines de ces cartes, on s'est principalement inspiré d'expériences faites dans d'autres travaux.

Les cartes logistiques:

Elles sont définies dans (Alligood et al., 1996 ; W-E ; Chartier et Boukadoum, 2006) comme suit:

- La carte logistique :

$$f(x) = \delta x(1-x), \text{ avec } \delta=4.0.$$

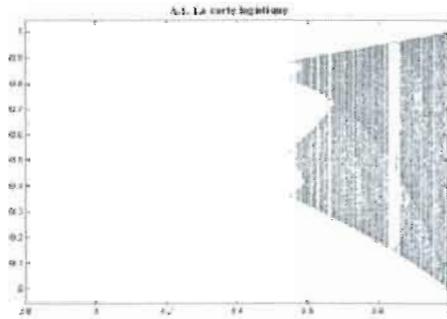


Figure.8. Diagramme de bifurcation, carte logistique.

- La carte logistique1 :

$$f(x) = 2\delta x(1-x)(1-2x)^2, \text{ avec } \delta=4.$$

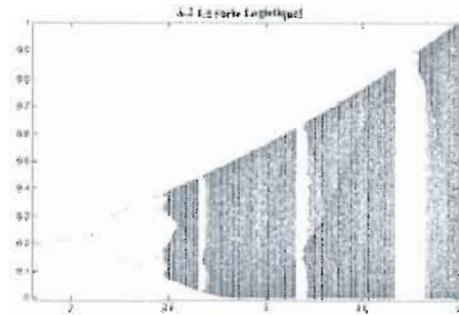


Figure.9. Diagramme de bifurcation, carte logistique1.

- La carte logistique3 :

$$f(x) = 1.25\delta x^2 - 5x^4 + 4x^4, \text{ avec } \delta=4.$$

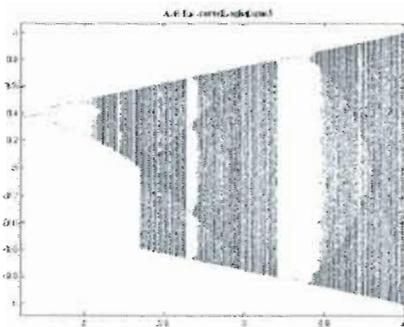


Figure.10. Diagramme de bifurcation, carte logistique3.

- La carte logistique2 :

$$f(x) = \delta x(0.75-x)^2, \text{ avec } \delta=4.$$

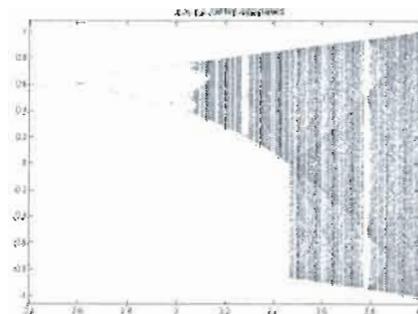


Figure.11. Diagramme de bifurcation, carte logistique2.

- La carte Logis-Cubic (Chartier et Boukadoum, 2006):

$$f(x) = (\delta+1)x^3 - \delta x^3, \text{ avec } \delta=0.1 \text{ dans le BAM et } \delta=1.4 \text{ dans le CNN-BAM.}$$

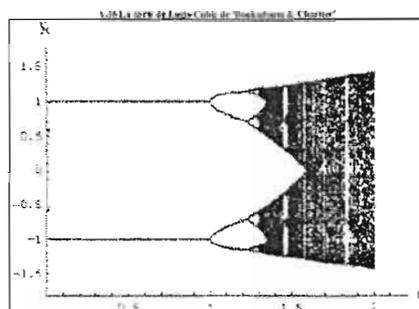


Figure.12. Diagramme de bifurcation, carte logis-cubic (Chartier et Boukadoum, 2006)

où :

x représente les conditions initiales i.e les entrées du système,

$f(x)$ représente les sorties du système, et

δ est une constante déterminant la proportionnalité dans la fonction logistique.

On utilise la carte logistique cubique dans le BAM où elle opère en mode non chaotique, et dans le modèle CNN-BAM de manière à ce qu'elle opère en mode chaotique. C'est le paramètre δ qui détermine le mode de fonctionnement de cette carte. En effet, si l'on observe le diagramme de bifurcation de cette fonction à la figure 12, on remarque que pour certaines valeurs de ce paramètre, cette carte génère des attracteurs fixes (pour $\delta \leq 1$), des attracteurs périodiques (tel que pour $\delta = 1.1$), ou des attracteurs chaotiques (tel que pour $\delta = 1.4$).

Les cartes Mira

La carte Mira à une dimension est définie par (Gumowski et Mira, 1980) :

$$x[n] = y[n-1],$$

$$y[n] = y[n-1] + \begin{cases} -ax[n-1], & \text{si } x[n-1] < 6 \\ bx[n-1] - 6(a+b), & \text{autrement} \end{cases}$$

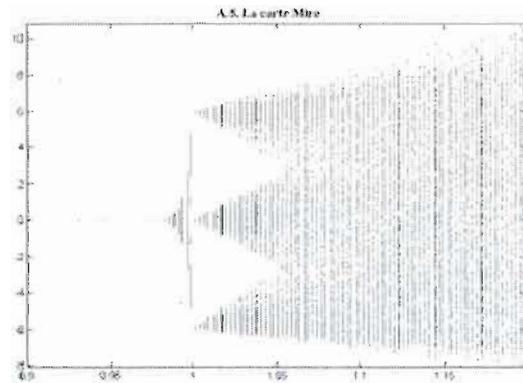


Figure.13. Diagramme de bifurcation, carte Mira.

où x et y représentent les conditions initiales pour les trajectoires de la carte.

Les paramètres de contrôle utilisés durant les expérimentations sont : $a=1.05$; $b=2$.

La carte Mira modifiée :

C'est une carte à deux dimensions utilisée pour imiter le régime impulsionnel (*spiking*) et le régime de silence des neurones biologiques (England et al., 2004).

Elle est définie par :

$$f(x) = y,$$

$$f(y) = ax + bx^2 + y^2$$

On a utilisé cette carte avec les paramètres de contrôle suivants:

- Mira_map1 avec $a = 0.8$, $b = 1$, où la carte a un point fixe avec deux valeurs propres positives.
- Mira_map2 avec $a = -0.8$, $b = 0.2$, où la carte a un ensemble stable.

La carte Gumowski-Mira a une particularité qui réside dans le fait qu'elle est *multistable* et non inversible; ceci joue un rôle très important dans la structure des bassins d'attraction avec une coexistence des attracteurs qui peut constituer des régions déconnectées. Généralement, la connaissance de la structure des bassins d'attraction aide à comprendre l'évolution d'un système à long terme (Gumowski et Mira, 1980 ; England et al., 2004). Cette carte comprend une récurrence qui est définie comme suit :

$$x_{[n+1]} = y_{[n]} + F(x_{[n]})$$

$$y_{[n+1]} = -bx_{[n]} + F(x_{[n+1]}),$$

$$F(x) = ax + 2(1-a)(x^2 / (1+x^2))$$

où :

x et y sont des valeurs réelles qui représentent l'entrée et la sortie du système respectivement.

F représente une fonction non linéaire arbitraire qui admet un certain nombre de dérivées continues et des séries Mc Laurin qui convergent d'une manière asymptotique (Gumowski et Mira, 1980).

On a utilisé cette carte avec les paramètres de contrôle: $a=0.7$; $b=0.9998$.

Les cartes Tent

La carte Tent est assez similaire à la carte logistique, elle possède des orbites chaotiques et elle est définie dans (Alligood et al., 1996 ; Hauptmann et al., 2003 ; Yoshida et al., 1983 ; Hasler et Yu, 1997) comme suit:

$$x[n] = y[n-1]$$

$$y[n] = \delta x [n-1] (1 - |1-2x[n-1]|),$$

avec $\delta=0.99$.

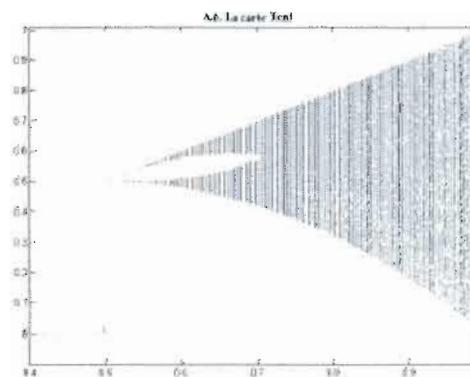


Figure.14. Diagramme de bifurcation, carte Tent.

où x et y représentent l'entrée et la sortie du système respectivement, δ un paramètre de contrôle. Dans le cas où $0 < \delta < 1$, la carte a un point fixe unique à zéro qui est attracteur. Les trois cartes qui suivent sont dérivées de la carte Tent.

La carte Tent1 :

$$x[n] = y[n-1],$$

$$y[n] =$$

$$\begin{cases} \delta x[n-1] & , \text{if } (0 \leq x[n-1]) \text{ et } (x[n-1] < 1/3) \\ \delta(2/3 - x[n-1]) & , \text{if } (1/3 \leq x[n-1]) \text{ et } (x[n-1] < 2/3), \\ \delta(-2/3 + x[n-1]) & , \text{autrement} \end{cases}$$

avec $\delta=3$.

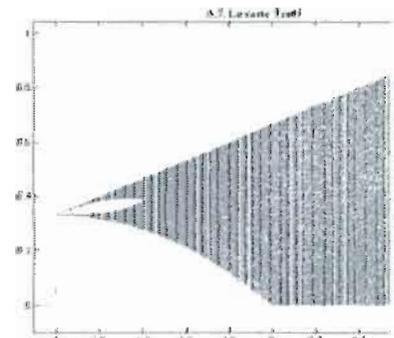


Figure.15. Diagramme de bifurcation, carte Tent1.

La carte Tent2 :

$$x[n] = y[n-1],$$

$$y[n] =$$

$$\begin{cases} \delta x[n-1] & , \text{si } (0 \leq x[n-1]) \text{ et } (x[n-1] < 1/4) \\ \sqrt{(\delta)} \cdot (1 - 2 * x[n-1]) & , \text{si } (1/4 \leq x[n-1]) \text{ et } (x[n-1] < 1/2) \\ \sqrt{(\delta)} \cdot (2 * x[n-1] - 1) & , \text{si } (1/2 \leq x[n-1]) \text{ et } (x[n-1] < 3/4) \\ \delta \cdot (1 - x[n-1]) & , \text{autrement} \end{cases}$$

avec $\delta=3.99$.

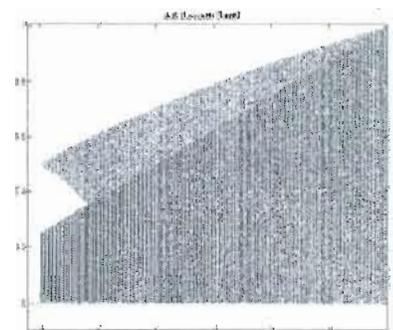


Figure.16. Diagramme de bifurcation, carte Tent2.

La carte *Tailed_Tent* :

$$x[n] = y[n-1] ,$$

$$y[n] = 1 - 2 \cdot (|[n-1] - (1-\delta)/2|) + \max(x[n-1] - 1 + \delta, 0),$$

avec $\delta=0.1$.

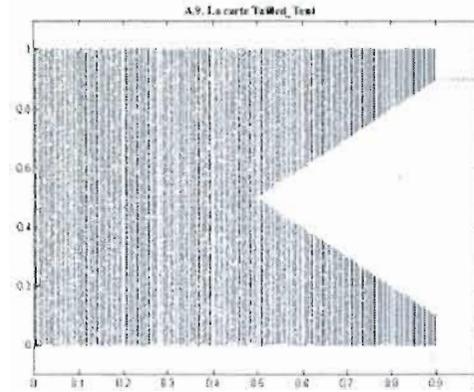


Figure.17. Diagramme de bifurcation, carte *Tailed_Tent*.

La carte Hénon

C'est une famille de carte qui diffère de celles précédemment évoquées, elle est définie dans (Alligood et al., 1996) comme suit :

$$f(x, y) = (\alpha - x^2 + \beta \cdot y, x).$$

On note que cette carte a deux variables d'entrée, soient, x et y, et deux variables de sortie, soient, les nouveaux x et y.

À l'utilisation des valeurs $\alpha = 1.28$, $\beta = -0.3$ autant que paramètres de contrôle. En utilisant les valeurs initiales $(x, y) = (0,0)$, on remarque que l'orbite converge vers un attracteur 2-période.

À la figure 18.(a) sont illustrés les résultats obtenus avec la valeur $\alpha = 1.28$, et à la figure 18.(b) ceux de $\alpha = 1.4$ où l'attracteur est fractal.

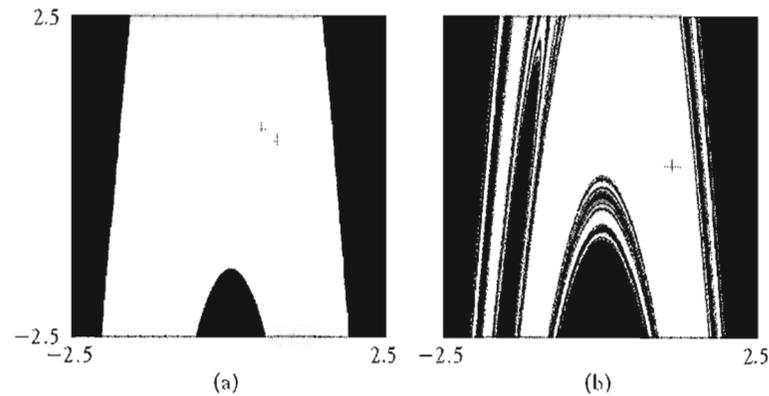


Figure 18. Le bassin d'attraction de la carte Hénon avec $\beta = -0.3$ (Alligood et al., 1996)

La figure 18. (a) illustre les résultats d'une itération avec des valeurs initiales générales. En noir sont illustrées les valeurs initiales dont les trajectoires divergent vers l'infini. Les valeurs initiales représentées par les zones blanches sont tirées par l'attracteur 2-périodes. La limite du bassin est une courbe continue qui se déplace de l'intérieur vers l'extérieur du rectangle des valeurs initiales.

La carte Bernoulli

C'est une carte chaotique à une dimension, une analyse détaillée sur cette carte est présentée dans (Alligood et al., 1996 ; Baek et Ott, 2004), et est définie comme suit :

$$x[n] = y[n-1],$$

$$y[n] = \text{mod}(\delta \cdot x[n-1], 1),$$

avec $\delta=1.99$ comme paramètre de contrôle

où x et y représentent l'entrée et la sortie du système respectivement.

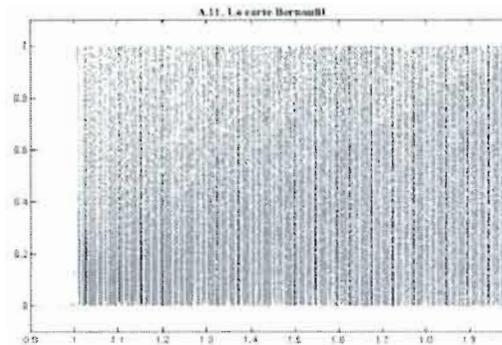


Figure.19. Diagramme de bifurcation, carte Bernoulli.

La carte « Piece Wise Affine Markov » (PWAM)

C'est une carte linéaire, avec cette famille de cartes, presque toutes les conditions initiales ont des orbites qui convergent vers un attracteur chaotique (Alligood et al., 1996 ; Chen et al., 2001 ; Tomoki 1994 ; Giovaiznrdi et Mazziizi, 2001 ; Rovatti et al., 1998). Des exemples de membres de cette famille suivent :

La carte PWAM1 :

$$x[n] = y[n-1],$$

$$y[n] = \begin{cases} \delta \cdot (1 - \text{abs}(x[n-1])), & \text{si } \text{abs}(x[n-1]) \leq 1 \\ \delta \cdot (\text{abs}(x[n-1]) - 2), & \text{autrement} \end{cases},$$

avec $\delta=3$.

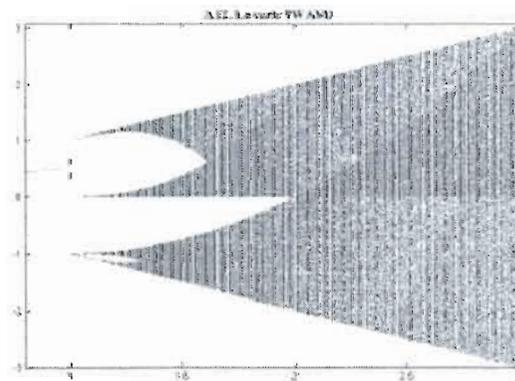


Figure.20. Diagramme de bifurcation, carte PWAM1.

La carte PWAM2 :

$$x[n] = y[n-1]$$

$$y[n] =$$

$$\begin{cases} -\delta \cdot (1 - \text{abs}(x[n-1])), & \text{si } \text{abs}(x[n-1]) \leq 1 \\ -\delta \cdot (\text{abs}(x[n-1]) - 2), & \text{sinon} \end{cases},$$

avec $\delta=3$

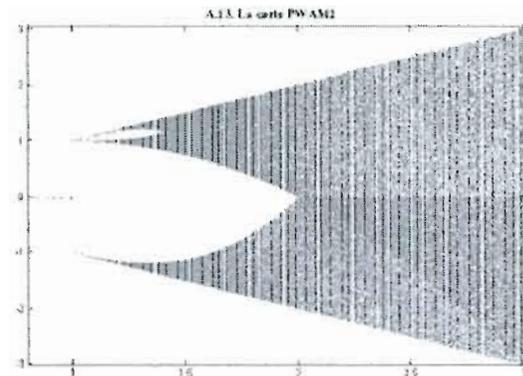


Figure.21. Diagramme de bifurcation, carte PWAM 2.

La carte PWAM3 :

$$x[n] = y[n-1],$$

$$y[n] =$$

$$\begin{cases} -\delta \cdot x[n-1], & \text{si } \text{abs}(x[n-1]) \leq 1 \\ -\delta \cdot (x[n-1] - 2 \cdot \text{sign}(x[n-1])), & \text{sinon} \end{cases},$$

avec $\delta=3$.

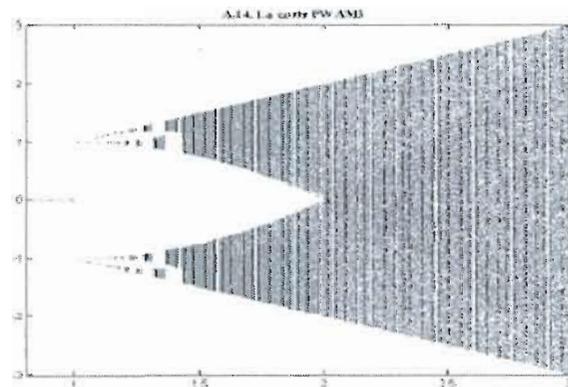


Figure.22. Diagramme de bifurcation, carte PWAM 3.

La carte PWAM4 :

$$x[n] = y[n-1],$$

$$y[n] =$$

$$\begin{cases} \delta \cdot (x[n-1] + \text{sign}(x[n-1])), & \text{si } \text{abs}(x[n-1]) \leq 1 \\ \delta \cdot (x[n-1] - 2 \cdot \text{sign}(x[n-1])), & \text{autrement} \end{cases},$$

avec $\delta=2.1$.

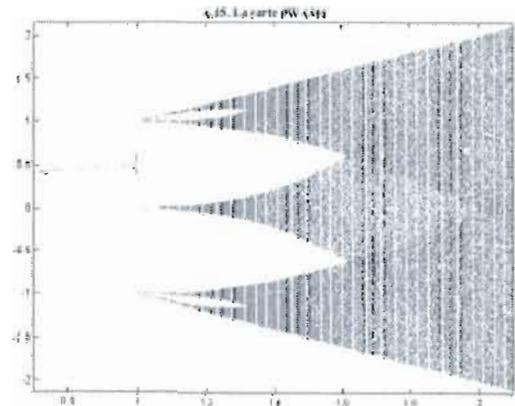


Figure.23. Diagramme de bifurcation, carte PWAM 4.

La carte impulsionnelle « Spiking »

L'origine du chaos a été étudiée dans (Shilnikov et Rulkov, 2003) avec un modèle de carte à deux dimensions. Le système est caractérisé par deux types d'attracteurs : un point fixe stable répliquant le silence dans un neurone biologique et un deuxième attracteur de type « *superstable limit cycle* » émettant des impulsions (*spikes*). En couplant ce système avec un sous-système lent, on peut générer un comportement '*spiking-bursting*' chaotique. En d'autres termes, ce type de carte non linéaire est capable de générer une grande variété de régimes dynamiques, ce qui permet de comprendre et d'imiter les dynamiques d'oscillations observées dans un neurone biologique, et de montrer une restructuration correcte du comportement collectif. Cette carte est définie dans (Shilnikov et Rulkov, 2003) comme suit:

$$x[n] = f(x_{[n-1]}, y_{[n-1]}),$$

$$y[n] = y_{[n-1]} - \mu (x_{[n-1]} + 1) + \mu \sigma,$$

$$f(x, y) = \begin{cases} \alpha(1-x) + y, & \text{si } x \leq 0 \\ \alpha + y & , \text{si } 0 < x < \alpha + y, \\ -1 & , \text{si } x \geq \alpha + y \end{cases}$$

où :

$x_{[n]}$ est la variable dynamique rapide (*fast dynamical variable*),

μ est une constante fixée à 0.001.

$y_{[n]}$ est la variable dynamique lente (*slow dynamical variable*), le fait que l'évolution de cette variable soit lente est dû à la petite valeur du paramètre $\mu = 0.001$.

α et σ sont des paramètres de contrôle de la carte.

Pour que cette carte opère en mode chaotique, on l'a utilisé avec des paramètres repris à partir des expériences faites dans (Shilnikov et Rulkov, 2003), soient:

- *Spiking_Map1*: $\mu=0.001$; $\alpha = 5.6$ et $\sigma= 0.322$.

- *Spiking_Map2*: $\mu=0.001$; $\alpha = 4.6$ et $\sigma= 0.16$.

- *Spiking_Map3*: $\mu=0.001$; $\alpha = 4.6$ et $\sigma= 0.225$.

La carte Ikeda

Cette carte a été proposée comme modèle du type de cellule qui peut être utilisé dans les ordinateurs optiques.

Cette carte est définie par l'équation suivante (Alligood et al., 1996) :

$$f(x,y) = (R + C_2 \cdot (x \cos(\tau) - y \sin(\tau)), C_2 \cdot (x \sin(\tau) + y \cos(\tau)))$$

$$\tau = C_1 - C_3 / (1 + x^2 + y^2)$$

où :

R , C_1 , C_2 et C_3 qui sont des paramètres de contrôle.

x et y sont les entrées du système.



**Figure 24. Basin d'attraction de la carte Ikeda ($C_1 = 0.4$, $C_2 = 0.9$, $R = 1$, $C_3 = 6$)
(Alligood et al., 1996)**

La figure 24 montre le bassin d'attraction de cette carte. La carte Ikeda est inversible avec le déterminant Jacobien C_2^2 pour chaque conditions initiales (x, y) , d'où $L_1 L_2 = C_2^2$. Pour certains paramètres, elle admet deux points fixes attracteurs. Avec les

paramètres : $C_1 = 0.4$, $C_2 = 0.9$, $R = 1$, $C_3=6$, l'un des deux points fixes converge vers un attracteur chaotique avec les nombre Lyapunov : $L_1 = 1.66$ et $L_2 = 0.487$ (exposants Lyapunov : 0.51 et -0.72). Tel qu'illustré à la figure 24, les orbites de la plupart des points initiaux choisis dans le voisinage de l'attracteur convergent vers le même ensemble limite.

3.4 Notions fondamentales sur le Perceptron multicouche

À l'instar de tout réseau de neurones artificiel, un Perceptron multicouches (PMC) est déterminé par les connexions entre ses neurones (son architecture), la fonction d'activation utilisée par ses neurones, et la règle d'apprentissage qui désigne la procédure d'ajustement des poids synaptiques.

Le Perceptron multicouches est un réseau à propagation comprenant une ou plusieurs couches cachées. Ce type de réseau consiste en une couche d'entrée, au moins une couche de neurones de traitement (couche cachée), et une couche de sortie. Les signaux d'entrée sont propagés vers l'avant d'une couche à l'autre jusqu'à la couche de sortie. Chaque couche dans un PMC a une fonction spécifique. La couche d'entrée reçoit des signaux de l'extérieur et les distribue à tous les neurones de la couche cachée. Avec un PMC à une seule couche cachée, on peut représenter toute fonction continue des signaux d'entrée (Negnevitsky 2005). Avec un réseau à deux couches cachées même les fonctions discontinues peuvent être représentées. La couche de sortie reçoit les patrons de stimuli provenant de la couche cachée et établit la sortie du réseau (Negnevitsky 2005).

Les neurones de la couche cachée déterminent diverses caractéristiques qui sont dissimulées dans les patrons d'entrée et codées dans leurs poids synaptiques. Ces caractéristiques sont alors utilisées par la couche de sortie pour déterminer chaque patron de sortie du réseau.

Règle d'apprentissage

Plusieurs algorithmes existent pour l'apprentissage d'un réseau PMC, mais la méthode la plus populaire est la retropropagation d'erreur avec descente de gradient (appelée souvent simplement retropropagation d'erreur). Dans cette méthode, le réseau calcule le patron de sortie, et s'il y a une erreur, les poids synaptiques des différentes couches sont ajustés pour la réduire, en partant de la couche de sortie. L'erreur représente la différence entre le pattern de sortie actuel et le pattern de sortie désirée, et est propagée vers la/les couche cachée, d'où le nom de l'algorithme (Negnevitsky 2005).

Pour illustrer la règle de minimisation de l'erreur, on suit un exemple tiré de (Haykin 1997). Il s'agit d'un simple neurone k comprenant un seul nœud de traitement dans la couche de sortie d'un réseau à rétropropagation (voir la figure 25). Le neurone k est excité par le vecteur de signaux $x(n)$ produit par une ou plusieurs couches de neurones cachés, qui sont excités à leur tour par un vecteur d'entrée (stimulus) affecté à la couche d'entrée du réseau. L'argument n indique le temps discret qui représente le pas d'un processus itératif d'ajustement des poids synaptiques du neurone k . Le signal de sortie du neurone k est dénoté par $y_k(n)$. Ce signal de sortie est comparé à la réponse désirée (*Target*), dénotée par $d_k(n)$. Un signal d'erreur dénoté par $e_k(n)$ est ensuite produit. On pourrait représenter ce processus à travers l'équation suivante :

$$e_k(n) = d_k(n) - y_k(n)$$

Le signal d'erreur $e_k(n)$ entraîne une séquence d'ajustements aux poids synaptiques du neurone k . Ces ajustements ont pour but de rendre le signal de sortie $y_k(n)$ plus proche de la réponse désirée $d_k(n)$ d'étape en étape.

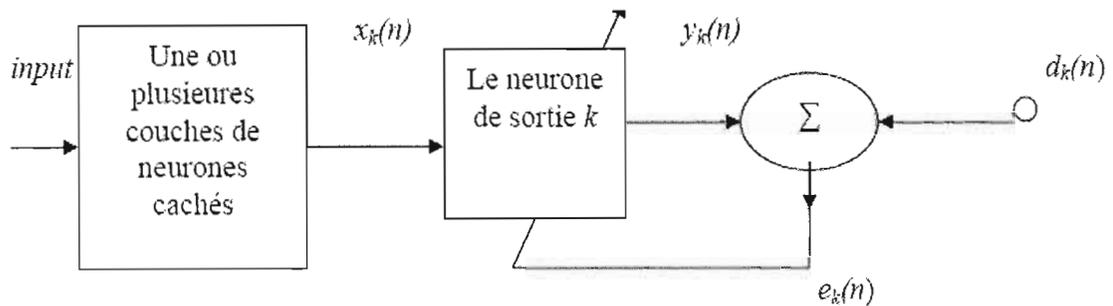


Figure 25. Illustration du l'apprentissage à minimisation de l'erreur (Haykin 1997)

Cet objectif est réalisé en minimisant la fonction de coût comme suit :

$$E(n) = 1/2 e_k^2(n)$$

$E(n)$ représente la valeur instantanée de l'erreur de l'énergie. Les ajustements des poids synaptiques du neurone k se renouvellent jusqu'à ce que le réseau atteigne un état stable, le processus d'apprentissage est de ce fait terminé.

Le processus d'apprentissage que l'on vient de décrire représente une fonction de correction de l'erreur. En théorie, la minimisation de la fonction de coût $E(n)$ mène à la méthode d'apprentissage appelée la règle Delta (règle de Widrow-Hoff). À partir de la figure 25 on peut observer que l'apprentissage avec correction de l'erreur se fait à un niveau local du neurone, ce qui veut dire que les ajustements synaptiques faits par la règle Delta sont localisés autour du neurone k .

On note $w_{kj}(n)$ la valeur du poids synaptique w_{kj} du neurone k excité par l'élément $x_j(n)$ du vecteur de signaux $x(n)$ pour un pas à l'instant n . d'après la règle Delta, l'ajustement $\Delta w_{kj}(n)$ appliquée au poids synaptique w_{kj} à l'instant n est défini par :

$$\Delta w_{kj}(n) = \eta e_k(n) x_j(n)$$

où η est une constante positive qui détermine le paramètre d'apprentissage. On peut noter à partir de la figure 25 que l'apprentissage avec correction de l'erreur est un exemple d'un système de contre-réaction à boucle fermée. Théoriquement, on sait que la stabilité d'un tel système est déterminée par les paramètres qui constituent les boucles de contre-réaction du système. Dans ce cas-ci, on a une simple boucle dont le paramètre η doit être sélectionné d'une manière appropriée pour pouvoir favoriser la convergence du réseau. En pratique, le paramètre du taux d'apprentissage η joue un rôle clef dans la performance de l'apprentissage à minimisation de l'erreur.

Dans le cadre de ce travail, on entraîne le PMC mis en œuvre en utilisant une version améliorée de la règle d'apprentissage à rétropropagation de l'erreur conventionnelle, soit, la propagation résiliente (Riedmiller et Braun, 1993). Cette dernière procède à un calcul d'une adaptation locale des poids synaptiques selon l'information locale de la fonction d'énergie $E(n)$. À cet effet, on introduit un nouveau terme pour chaque poids synaptique qui représente sa propre valeur d'ajustement. Cette valeur adaptative change tout au long du processus d'apprentissage en se basant sur la valeur que prend la fonction d'énergie à un niveau local, selon la règle suivante :

$$\Delta w_{ij}^{(t)} = \begin{cases} -\Delta_{ij}^{(t)} & , \text{ if } \frac{\partial E}{\partial w_{ij}}^{(t)} > 0 \\ +\Delta_{ij}^{(t)} & , \text{ if } \frac{\partial E}{\partial w_{ij}}^{(t)} < 0 \\ 0 & , \text{ else} \end{cases}$$

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} + \Delta w_{ij}^{(t)}$$

À chaque instant t , si la dérivée partielle de la fonction d'énergie change de signe, cela signifie que la dernière valeur d'ajustement était trop grande et que l'algorithme a dépassé un minimum de la fonction. La valeur d'ajustement est donc diminuée. Si la dérivée garde le même signe, la valeur d'ajustement est légèrement augmentée dans le but d'accélérer la convergence dans les régions de cavité. Une fois que les valeurs d'ajustements sont adaptées pour chaque poids synaptique, si la dérivée est positive, le poids est diminué par sa valeur d'ajustement, et si la dérivée est négative, la valeur d'ajustement lui est ajoutée. Les valeurs d'ajustements des poids et les poids en tant que tels changent à chaque époque du cycle d'apprentissage.

3.5 Conclusion

On a présenté dans ce chapitre les notions fondamentales sur lesquelles est basée la conception des trois modèles de réseaux de neurones que l'on propose dans le chapitre 4. Trois modèles de réseaux neuronaux sont mis en œuvre et présentés dans le chapitre expérimental : un BAM, un BAM à fonction de sortie chaotique, et un PMC à rétropropagation d'erreur. À cet effet, les notions citées dans ce chapitre incluent des concepts de base sur les mémoires associatives bidirectionnelles, des notions sur le chaos et les systèmes dynamiques dans les mémoires associatives, ainsi qu'une introduction sommaire du Perceptron multicouche.

Par ailleurs, mise à part les notions sur les modèles classiques, on a secondairement présenté les caractéristiques que constituent les trois modèles implémentés, notamment, leur règle d'apprentissage et leur fonction de sortie. On détaille les expérimentations sur les trois différents modèles dans le chapitre qui suit.

CHAPITRE IV

EXPÉRIMENTATIONS ET RÉSULTATS

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 étiopathologiques. 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.

Il est donc intéressant de pouvoir palier à ces inconvénients. Pour ce faire, on se propose de mettre au point des architectures de complexité moindre d'une part, et permettant de dépasser largement la limite des 2/3 classes, en l'occurrence 175 classes pour chacun des trois modèles, le tout avec des performances meilleures que les approches présentées au chapitre 2.

On a pour but dans ce travail de tester l'efficacité d'un réseau de neurones chaotique par rapport à deux autres modèles de réseau, et ce, pour faire la classification de cas de tumeur dans le cadre du cancer du sein. Dans ce contexte, on a exploité une base de données en premier lieu sur une mémoire associative bidirectionnelle. Puis sur une mémoire associative bidirectionnelle chaotique en reprenant l'architecture et la règle d'apprentissage du premier modèle, et en le testant avec les différentes fonctions de sorties chaotiques évoquées dans le chapitre précédent. Pour le troisième modèle, il s'agit d'un réseau multicouche à rétropropagation de l'erreur, soit, un PMC utilisé comme référence de base dans l'étude comparative.

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

4.2 Description et précisions sur les données utilisées

Les données cliniques ont été recueillies par Matjaz Zwitter & Milan Soklic (physicians) à l'institut d'Oncologie en Yougoslavie (*University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia*). Ces données contiennent 286 cas de tumeur, chacun codé par les 10 paramètres suivants : récurrence de la tumeur, âge de la patiente, type de la ménopause, taille de la tumeur, node-inv, node-caps, deg-malig, position du sein, quadrant du sein, et l'irradiation.

A1. Événement

Événement Récurrent	Oui	Non
Code	1	2

A2. Âge:

Âge	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99
Code	1	2	3	4	5	6	7	8	9

A3. Type de la ménopause:

Type	lt40	ge40	premeno
Code	1	2	3

A4. Taille de la tumeur :

Taille	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59
Code	1	2	3	4	5	6	7	8	9	10	11	12

A5. Inv-nodes:

Inv-nodes	0-2	3-5	6-8	9-11	12-14	15-17	18-20	21-23	24-26	27-29	30-32	33-35	36-39
Code	1	2	3	4	5	6	7	8	9	10	11	12	13

A6. Node-caps:

Node-caps	yes	no
Code	1	2

A7. Deg-malig:

Deg-malig	1	2	3
Code	1	2	3

A8. Sein (breast) :

Sein	gauche	droit
Code	1	2

A9. Sein - Quadrant :

Sein - Quadrant	gauche- haut	gauche-bas	droit-haut	droit-bas	centre
Code	1	2	3	4	5

A10. Irradiation:

Irradiation	yes	no
Code	1	2

Une fois les données codées, la base de données résultante contient des nombres compris dans l'intervalle [1,13]. Pour les besoins des expérimentations, on a normalisé les données en effectuant une division par 10. Le chiffre 0 n'a pas été utilisé pour ce codage afin d'éviter la génération de points fixes instables. Les données utilisées durant les expérimentations qui suivent sont représentées par 286 vecteurs à valeurs réelles de dimension 10.

Les échantillons de la base de données n'étant pas catégorisés, il était nécessaire d'accomplir une phase de catégorisation pour pouvoir déterminer les différentes classes à travers lesquelles la classification devait être effectuée. Pour ce faire, on a repris les résultats d'une carte auto-organisatrice de Kohonen

obtenu dans (Tabari 2005). Dans ce travail, les catégories ont été établies par ce réseau tel qu'illustré dans la figure 12.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	16	28	94		159	172	147	47	148	56	9	100	78	97		91			89	115
2	19			104		98	93		107			27	80	1		149	87	156		90
3	12		163					20	22		69	133				143	92	162		166
4	7	38		18	74		55	129	23		135	110			175	112				105
5		70	68		102		17			76	32		36			101	154	127		145
6	41		165			146		15	117											103
7	174		26	108		122	88		111		21	150	120				167	86		153
8								24	161	168	72		79		71	109				126
9	54				119								134	14		53		171		164
10		57	4		66			77									62			158
11	106			65		29		10	114	3	81		2	67						169
12	142	152	125	113			116	40			43	42		128		45	95			140
13	46				49	31	130				75		51	59			50			141
14						64			25					11						137
15							84	37	60		35	96		5	33	39	85	34		
16							13						48					82		131
17			52					8	121	61	30	63							58	
18								44											83	
19							6		139		73	173	123		160	99				155
20								132	136	138	170		144	118	157	124		151		

Figure 26. Distribution des neurones sur la carte SOM [20 20] (Tabari 2005)

Le réseau *SOM* (*Self-Organized Map*) a été exploité sur les 286 cas de tumeur de la base de données « cancer du sein », la figure 26 illustre les résultats de catégorisation obtenus après la convergence de l'algorithme. La carte SOM étant composée de 400 neurones (20*20 neurones) tel qu'indiqué à travers le nombre de lignes et de le nombre de colonnes. Chaque case de la grille représente une unité de traitement et le numéro inscrit dans chacune de ces

cases fait référence au numéro de la classe déterminée par l'algorithme. En observant la carte obtenue dans la figure 12, on remarque que le réseau a distingué 175 catégories. Chacune de ces catégories est représentée par un neurone qui s'est spécialisé selon les spécifications médicales. Dès lors, on utilise ces 175 classes dans ce qui suit, en prenant en considération les interprétations établies dans (Tabari 2005) pour effectuer la tâche de classification.

Pour les besoins des expérimentations qui suivent, on a consacré deux tiers des échantillons de la base de données « cancer du sein » à la phase d'apprentissage, et le tiers restant à la phase de test. En effet, chaque modèle mis en œuvre a été évalué en utilisant des données qu'il n'a pas apprises et qui ne lui ont donc pas été présentées durant la phase d'apprentissage. Il s'agit d'une validation à trois tiers où un test est effectué sur chacun des tiers de la base de données. Une fois les tests complétés, une moyenne des résultats obtenus avec les trois tiers est calculée, cette moyenne représente en l'occurrence la performance de classification du réseau. On dispose de 286 échantillons dans la base de données, 191 patrons sont consacrés à l'apprentissage, et les 95 patrons restants à la phase de test.

4.3 Expérimentation du réseau BAM

Le premier modèle de réseau de neurones qu'on a employé pour procéder à la classification de tumeurs, est une mémoire associative bidirectionnelle. Selon (Chartier et Boukadoum, 2006), un BAM efficace doit avoir certaines propriétés telles que :

- ✚ Le réseau doit pouvoir faire un apprentissage en-ligne puisque dans la plupart des problèmes réels, on dispose rarement de tous les stimuli en même temps.
- ✚ Aussi, en associant le feed-back de la fonction de sortie, le BAM doit itérativement générer des connexions synaptiques qui convergent vers une solution locale stable et optimale.
- ✚ L'autre propriété que devrait avoir ce réseau, est que son processus d'apprentissage soit entièrement basé sur la corrélation sans aucun besoin d'avoir une couche cachée, une couche virtuelle, ou une quelconque information a priori sur les stimuli ou sur le changement effectué sur les paires de patrons.
- ✚ Ensuite, la règle d'apprentissage employée sur un BAM doit rester aussi simple que possible et elle doit être apte à faire l'association des patrons de différentes dimensions aussi bien que de dimensions identiques.
- ✚ De plus, la fonction de sortie du réseau et sa règle d'apprentissage, doivent être choisies de manière à ce que la matrice de poids synaptiques génère des attracteurs à points fixes stables aussi bien pour des patrons bipolaires que pour des patrons à valeurs réelles.

4.4.1 Architecture du BAM

Dans le but de pouvoir bénéficier des propriétés du modèle BAM citées ci-haut, on a opté pour reprendre les mêmes paramètres du réseau employé dans

(Chartier et Boukadoum, 2006). La topologie du réseau est indiquée à la figure 27.

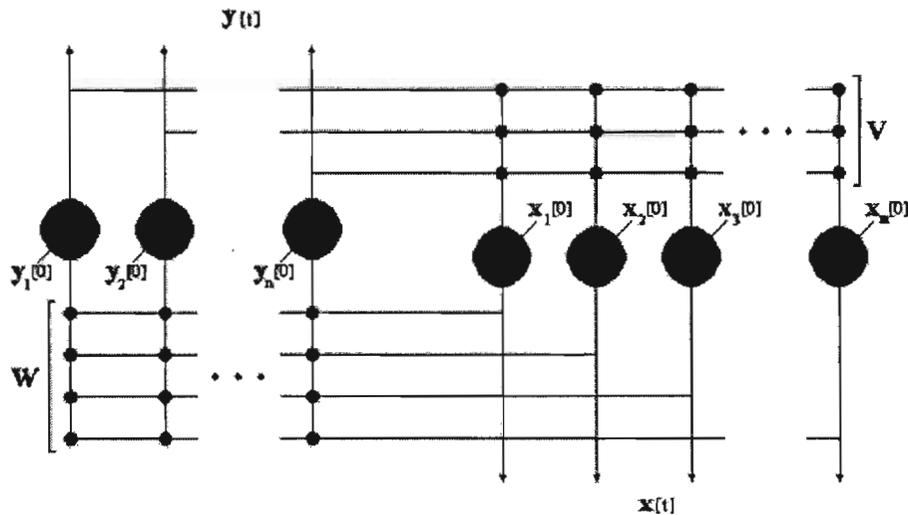


Figure 27. Architecture du BAM (Chartier et Boukadoum, 2006)

Il s'agit de deux réseaux de type Hopfield interconnecté tête-bêche, de manière à créer un flux récurrent de l'information entre eux, tels que $X[0]$ et $Y[0]$ représentent les vecteurs d'états initiaux. W et V sont les matrices de poids synaptiques qui définissent les connexions entre les couches X et Y dans les deux directions, et t est le numéro de l'itération courante.

4.4.2 Fonction de sortie et règle d'apprentissage

La fonction de sortie proposée par Chartier et Boukadoum est une carte logistique cubique, qui a été utilisée dans le but d'avoir deux attracteurs situés à 1 et -1. Cette fonction est définie par l'équation dynamique : $\frac{dz}{dt} = Gz(1-z^2)$ où

G est un paramètre général. En faisant une approximation et un changement de variables on abouti à la fonction de sortie (Chartier et Boukadoum, 2006):

$$\forall i, \dots, N, y_i(t+1) = \begin{cases} 1 & \text{if } a_i > 1 \\ -1 & \text{if } a_i < -1 \\ (\delta+1)a_i - \delta a_i^3 & \text{else} \end{cases}$$

$$\forall i, \dots, N, x_i(t+1) = \begin{cases} 1 & \text{if } b_i > 1 \\ -1 & \text{if } b_i < -1 \\ (\delta+1)b_i - \delta b_i^3 & \text{else} \end{cases}$$

dans ces équations $y_i(t+1)$ et $x_i(t+1)$ représentent la sortie d'un neurone i de la couche X ou Y à l'instant $t+1$; et $a_i[t]$ et $b_i[t]$ sont les fonctions d'activation correspondantes à l'instant t ($a_i[t]=[W.X[t]]_i$; $b_i[t]=[V.Y[t]]_i$) et δ est un paramètre de sortie général ; le réseau peut converger vers des attracteurs stables, cycliques ou chaotiques si la valeur de δ est trop grande, et il exhibe une approche monotone à des états d'équilibres si la valeur de δ est entre 0 et 0.5.

La règle d'apprentissage utilisée est dérivée de l'approche Hébbienne/anti-Hébbienne (Chartier et Boukadoum, 2006) :

$$W_{[k+1]} = W_{[k]} + \eta (y_{[0]}x_{[0]}^T + y_{[0]}x_{[1]}^T - y_{[1]}x_{[0]}^T - y_{[1]}x_{[1]}^T), \quad (1)$$

$$V_{[k+1]} = V_{[k]} + \eta (x_{[0]}y_{[0]}^T + x_{[0]}y_{[1]}^T - x_{[1]}y_{[0]}^T - x_{[1]}y_{[1]}^T) \quad (2)$$

On pose $t=1$ de manière à avoir :

$$W(k+1) = W(k) + \eta [y(0) - y(1)] [x(0) + x(1)]^T$$

$$V(k+1) = V(k) + \eta [x(0) - x(1)] [y(0) + y(1)]^T$$

où :

$$X(1) = f(V[k] * Y_0)$$

$$Y(1) = f(W[k] * X_0)$$

Dans les équations ci-dessus, W et V représentent les matrices de poids dans les deux directions ; X_0 et Y_0 sont les entrées initiales à associer, η est le paramètre d'apprentissage et k est numéro du pattern d'apprentissage. Comme cette règle d'apprentissage inclut une boucle de contre-réaction à partir des sorties non linéaires via $X(1)$ et $Y(1)$, ceci permet au réseau d'effectuer un apprentissage en ligne et de contribuer dans la convergence des poids de connections. Pour plus de détails, voir (Chartier et Boukadoum, 2006).

Le paramètre δ de la fonction de sortie a été mis à la valeur 0.1 comme en (Chartier et Boukadoum, 2006) dans le but d'avoir des attracteurs fixes. Le paramètre η d'apprentissage a été mis à 0.01. Aussi, pour limiter le temps de chaque simulation, on mettait le nombre d'itérations à 1 avant chaque ajustement de la matrice de poids. Noter qu'avec la valeur utilisée de δ , le réseau opère en mode non chaotique.

La phase d'apprentissage s'effectue selon les étapes suivantes :

- 1) Sélection aléatoire d'une paire de patrons à partir de la série d'apprentissage de la base de données « cancer du sein ».

- 2) Calcul de la fonction de sortie au niveau des unités du réseau.
- 3) Calcul de la nouvelle matrice des poids synaptiques selon (1) et (2).
- 4) Répéter le processus à travers les étapes 1), 2) et 3) tout au long de la phase d'apprentissage jusqu'à ce que la matrice des poids converge, cela prenait entre 900 et 1700 époques.

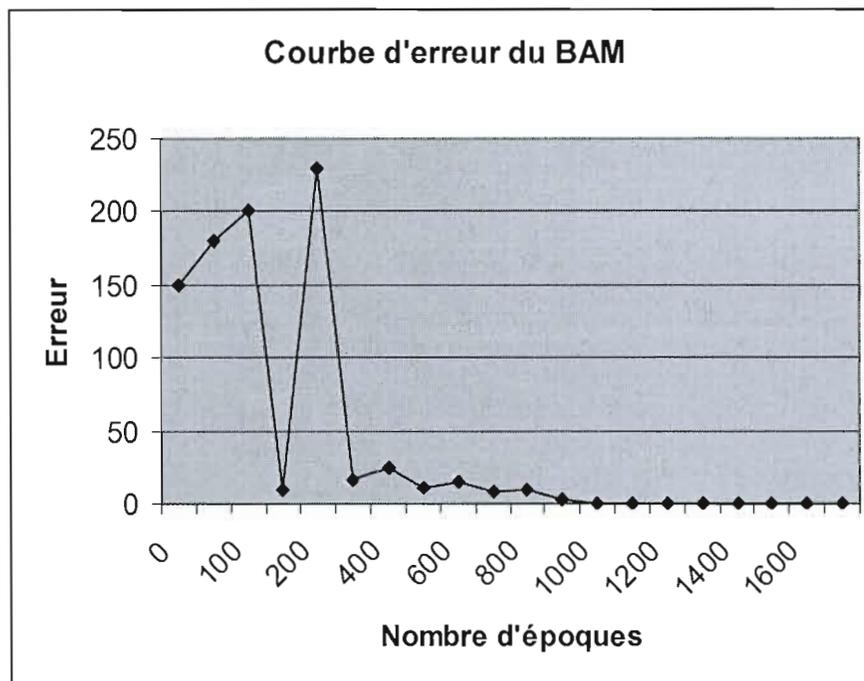


Figure 28. Courbe d'erreur au rappel en fonction du nombre de cycles d'apprentissage.

La figure 28 montre que la matrice des poids a pu atteindre un état stable après un nombre raisonnable de cycles durant la phase d'apprentissage. Après 1700 époques, la moyenne de l'erreur valait moins de 0.0005.

Les patrons d'apprentissage comportent les caractéristiques des classes qui leurs correspondent. Après la réalisation de l'étape d'apprentissage, le réseau BAM a été testé sur les patrons de test, c'est-à-dire, des échantillons qui ne lui ont pas été présentés durant la phase d'apprentissage. Il s'agit d'une série de 95 cas de tumeurs de la base de données « cancer du sein » qui sont consacrés au test. Le but est d'évaluer l'aptitude du réseau à faire le rappel du stimulus appris à partir d'un stimulus nouveau.

En effectuant le test de performance sur ce réseau, 85 cas de tumeurs ont été correctement classés, ce qui donne un pourcentage de reconnaissance exacte de 90.17%. Par conséquent, la matrice des poids synaptiques a pu converger pour les échantillons de test bien qu'ils soient tout à fait inconnus pour le réseau. Cette performance face aux nouveaux échantillons prouve que le réseau a une bonne capacité de généralisation. De plus, le fait que l'architecture du réseau soit aussi simple est un avantage certain pour la réalisation du réseau.

4.4 Expérimentations du réseau CNN-BAM

Le deuxième modèle de réseaux de neurones qu'on a employé pour effectuer la classification dans le cadre du cancer du sein est un BAM au même titre que le modèle précédent, à la différence que la fonction de sortie du réseau est une carte chaotique. Ce BAM chaotique emploie les mêmes paramètres que le BAM précédent, avec des valeurs qui le font opérer dans le mode voulu. On rappelle que le but de ce travail est d'évaluer la performance d'un réseau de neurones chaotique par rapport à celle de deux autres modèles, dont le BAM précédemment présenté. L'usage de paramètres communs est un atout pour l'étude comparative que l'on veut d'établir.

4.4.1 Architecture du CNN-BAM

L'architecture employée dans le travail de Chartier et Boukadoum (Chartier et Boukadoum, 2006) a été reprise pour le BAM chaotique (voir la figure 27) mais avec une fonction de sortie chaotique.

4.4.2 Fonctions de sortie testées et règle d'apprentissage

Initialement on a réutilisé la carte logistique cubique employée dans le modèle BAM précédent, mais en augmentant la valeur du paramètre δ à 1.4, elle opère en l'occurrence en mode chaotique. Ensuite, on a utilisé chacune des cartes chaotiques définies dans le chapitre précédent comme fonction de sortie du BAM chaotique. Les paramètres de chacune de ces fonctions étaient tels que spécifiés au chapitre précédent. Le paramètre d'apprentissage η a été fixé à 0.01 avec toutes les cartes chaotiques.

À la phase de rappel, on a évalué la performance du réseau, en utilisant la série d'échantillons de la base de données « cancer du sein » qui ont été prévus à cet effet, soit les 95 patrons de test. Dans tous les cas, le réseau a convergé après 3 à 10 époques d'apprentissage. Les résultats obtenus durant les simulations avec les différentes cartes sont illustrés dans la table 12.

Carte chaotique	Taux de reconnaissance	Identifiant de la fonction
Hénon	100%	1
Bernoulli	100%	2
Logistic3	98.59%	3
Mira1	98.59%	4
Spikin Map3	96.48%	5
Spikin Map	95.43%	6
Logistic2	94.38%	7
Tent	91.92%	8
Logistic	89.47%	9
Spikin Map2	84.90%	10
PWAM2	77.89%	11
TailedTent1	75.78%	12
Logistic1	72.62%	13
PWAM4	70.17%	14
PWAM3	65.96%	15
Tent1	62.45%	16
PWAM1	53.32%	17
Logistic-Cubic	44.21%	18
Mira2	22.45%	19
Spikin Map1	18.94%	20
Ikeda	18.94%	21
Mira-Gumolski	12.27%	22
Tent2	4.56%	23

Table 12. Classification des cartes chaotiques selon le taux de succès à la reconnaissance.

Les résultats obtenus montrent que sur les 23 fonctions chaotiques testées sur le réseau BAM chaotique, la carte Hénon et la carte Bernoulli ont engendré un taux de reconnaissance parfait, soit 100% des échantillons de test ont été identifiés bien qu'ils ne fassent pas partie de l'ensemble d'apprentissage.

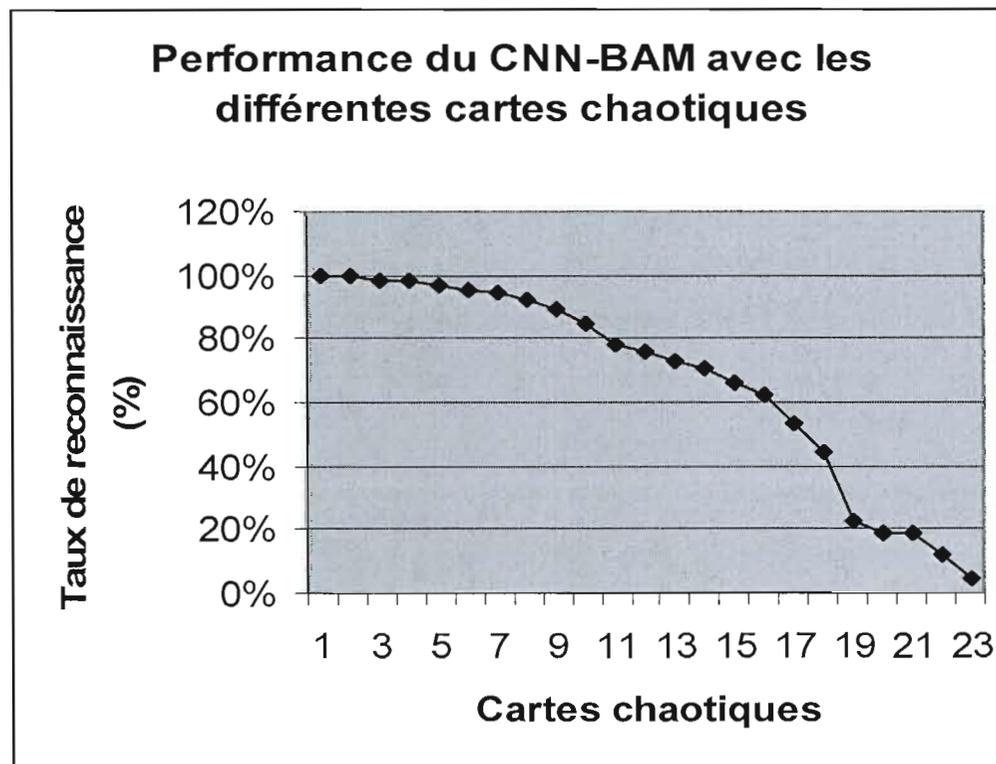


Figure 29. Performance du BAM chaotique avec les différentes cartes telles que classées dans la table 12.

Par ailleurs, d'autres fonctions ont donné des résultats moins parfaits mais qui restent quand même très satisfaisants, c'est notamment le cas des cartes Logistic3, Mira1 et Spikin_Map3. Par ailleurs, d'autres fonctions ont donné des résultats médiocres voir quasi-nuls ; c'est le cas des cartes Mira-Gumolski et

Tent1. La performance du BAM chaotique avec les différentes cartes est illustrée à la figure 29.

4.5 Expérimentation du PMC

Le dernier modèle de réseaux de neurones mis en œuvre pour effectuer la tâche de classification dans le cadre du cancer du sein, est différent des modèles à mémoire précédemment vus, il s'agit d'un Perceptron multicouche à rétropropagation de l'erreur. Pour les caractéristiques du modèle, on s'est principalement inspiré du PMC qui a été employé dans (Naoum et al., 2008), et ce, vu les résultats intéressants qui y ont été atteints. On amène plus de détails tout au long de cette section.

Pour des besoins de simulations, en plus de la normalisation que l'on a effectué sur les données auparavant avec les deux autres modèles (voir section 1 du présent chapitre), on a dû apporter un changement dans les patrons de la base de données « cancer du sein ». En effet, étant donné que l'apprentissage que l'on exerce sur un réseau à rétropropagation est de type supervisé, les échantillons que l'on présente à ce type de réseaux doivent comporter des données qui représentent le but (résultat) recherché par les simulations, soient les valeurs désirées (*targets*).

La conception du réseau ainsi que les différentes simulations ont été faites en utilisant le simulateur de réseaux de neurones *Stuttgart Neural Network Simulator (SNNS)*. Pour les besoins expérimentaux, on a dû introduire la valeur désirée (Target) pour chaque pattern de la base de données « cancer du sein ». À cet effet, suivant les résultats de catégorisation obtenus dans (Tabari 2005) avec les cartes auto-organisatrices SOM, on a attribué à chaque cas de tumeur l'identifiant de la classe qui lui est appropriée, les classes étant numérotées de 1 à 175. Par conséquent, on a ajouté un paramètre à chaque échantillon, qui

représente l'identifiant de la classe à laquelle il appartient, soit un 11^{ème} paramètre.

4.5.1 Architecture du Perceptron multicouche

Pour ce qui est de l'architecture du modèle PMC qu'on a employé, certains de ses paramètres sont étroitement liés aux propriétés des données dont on dispose. En effet, un échantillon comprenant 11 paramètres, le vecteur d'entrée au réseau va être de taille 10. Le 11^{ème} paramètre, soit celui qui représente l'identifiant de la classe à laquelle appartient l'échantillon en question, définit la longueur du vecteur de sortie du PMC. D'où, la couche d'entrée du réseau contient 10 unités d'entrées, et la couche de sortie du réseau en contient une seule.

Une des caractéristiques du PMC sur le plan architectural est qu'il doit avoir un intermédiaire entre la couche d'entrée et la couche de sortie, soit au moins une couche cachée. Cette couche contient des unités de traitement (neurones) et leur nombre doit être fixé d'une manière empirique. Dans (Naoum et al., 2008), le réseau contient deux couches cachées ayant chacune le même nombre de neurones que la couche d'entrée. Selon Fausett (Fausett 1994), pour ce genre de réseaux, on devrait avoir un maximum de deux couches cachées, avec une précision arbitraire pouvant être obtenue en consacrant suffisamment de neurones par couche. Pour les deux raisons évoquées, on a opté pour la conception de deux différentes architectures, la première contenant une seule couche cachée, et la deuxième deux. Dans chaque cas, une couche comporte autant d'unité que la couche d'entrée, soit 10 unités de traitement chacune.

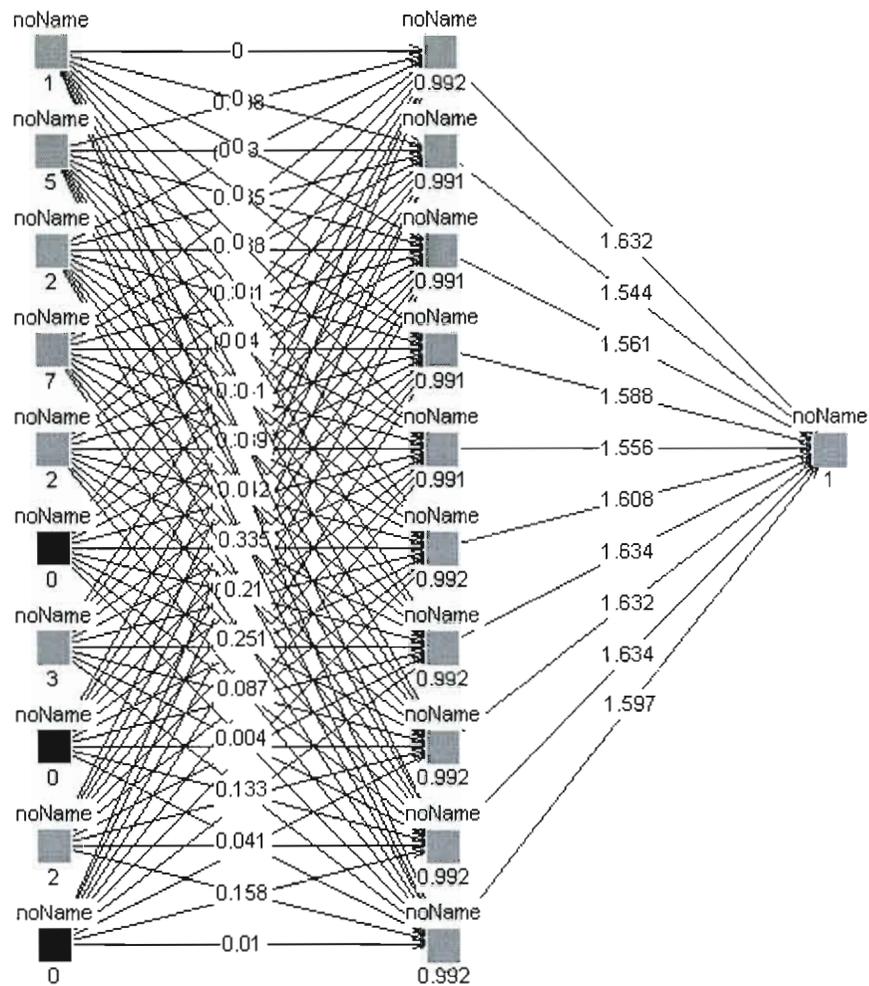


Figure 30. Architecture du Perceptron multicouche (10-10-1)

La figure 30 illustre la première architecture du PMC, celui-ci étant en phase d'apprentissage, on remarque que les unités des différentes couches sont reliées via des connexions qui comportent les valeurs des poids synaptiques. L'architecture du réseau avec deux couches cachées est illustrée à la figure 31.

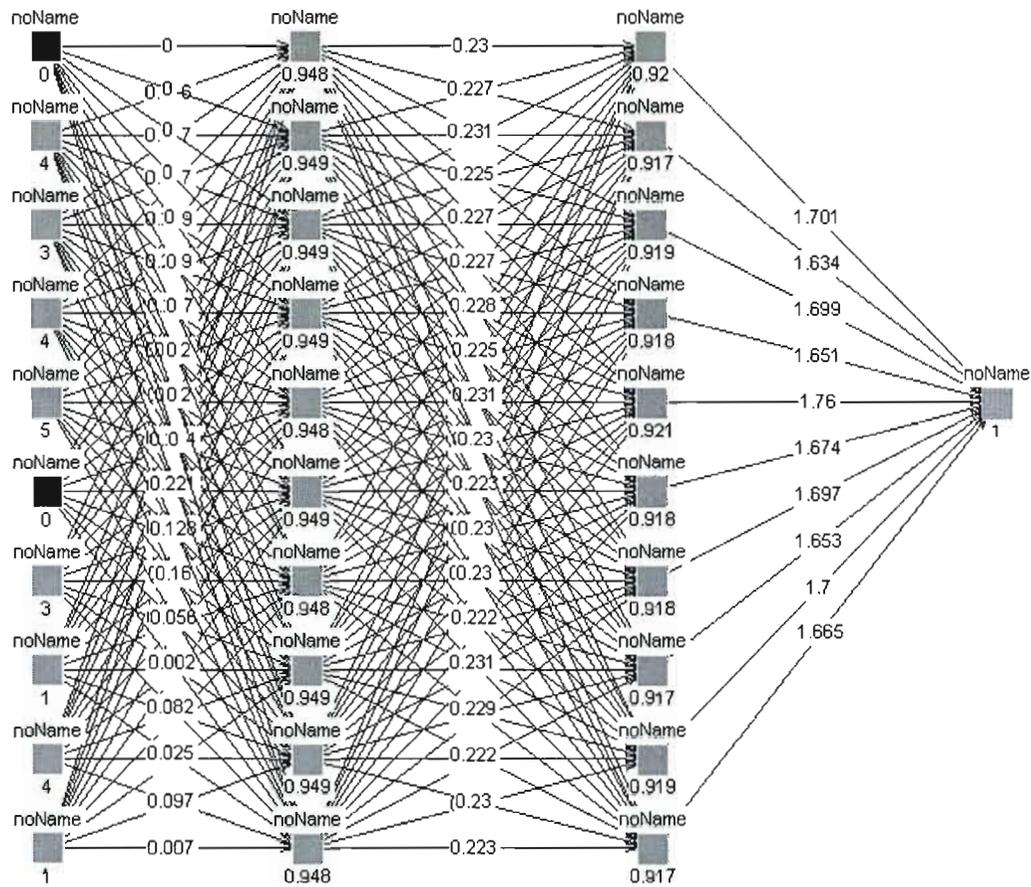


Figure 31. Architecture du Perceptron multicouche (10-10-10-1)

Dans chacune des deux architectures, une couche cachée contient autant de neurones que la couche d'entrée, soit 10.

4.5.2 Fonction de sortie et règle d'apprentissage

La fonction d'activation utilisée pour les neurones de traitement, soit les unités dans la couche cachée, est la fonction Logistique, appelée aussi Sigmoide qui est définie par

$$f(x) = 1 / (1 + \exp(-x)).$$

Pour les couches d'entrée et de sortie, la fonction identité a été utilisée.

La règle d'apprentissage utilisée est dérivée de l'algorithme de rétropropagation de l'erreur détaillé dans le chapitre précédent, soit l'algorithme de la rétropropagation résiliente (*RProp*). C'est un algorithme qui revient dans la littérature car il engendre un apprentissage plus rapide et une meilleure performance globale du réseau (pour plus de détails voir (Riedmiller et Braun, 1993)).

Pour ce qui est des paramètres du *RProp*, δ_0 représente la valeur initiale de mise à jour, δ_{\max} est la limite supérieure pour la valeur de mise à jour, et α est la déclinaison du poids, ce paramètre détermine la relation entre l'erreur de sortie et la réduction dans la valeur des poids (*The weight decay*). L'apprentissage du réseau a été accompli avec ces paramètres ayant comme valeurs : 0.01, 50 et 4 respectivement. Le nombre de cycles d'apprentissage variait en époque de 10, 100 puis 1000, avec des pas aléatoires entre les différentes époques.

On dispose de deux fichiers de données, l'un contenant les échantillons d'apprentissage, et l'autre contenant les échantillons de test (validation). Les données ont été divisées aléatoirement en deux parties, une partie contenant 2 tiers des échantillons pour l'apprentissage et l'autre contenant le tiers restant pour la validation. À la fin de chaque simulation, *SNNS* génère deux fichiers de résultats, l'un représentant les résultats obtenus à la phase d'apprentissage et l'autre les résultats obtenus à la phase de la validation. À travers ces deux fichiers de résultats, on calcule la performance du réseau dans les deux différentes phases. La performance du PMC a été évaluée en associant

l'activation de la sortie à la valeur catégorique la plus proche et en comparant le résultat à la valeur désirée.

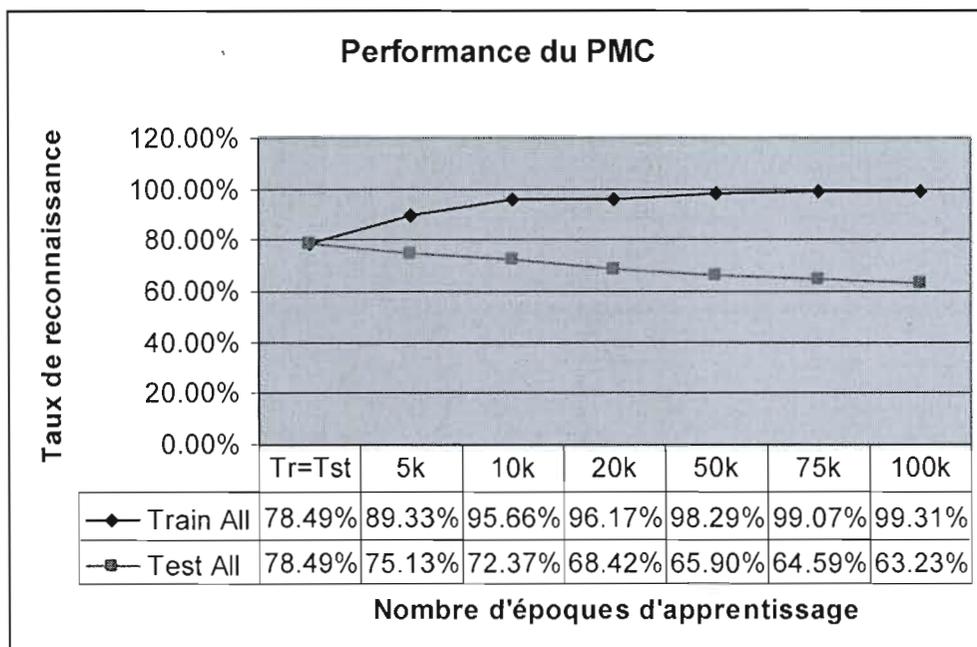


Figure 32. Taux de reconnaissance du PMC avec une couche cachée

Les résultats obtenus avec le PMC à une seule couche cachée sont illustrés à la figure 32, et ceux du PMC à deux couches cachées à la figure 33. Les courbes bleues représentent le taux de convergence du réseau à la phase d'apprentissage, quant aux courbes roses, elles représentent le taux d'exactitude à la classification à la phase de rappel.

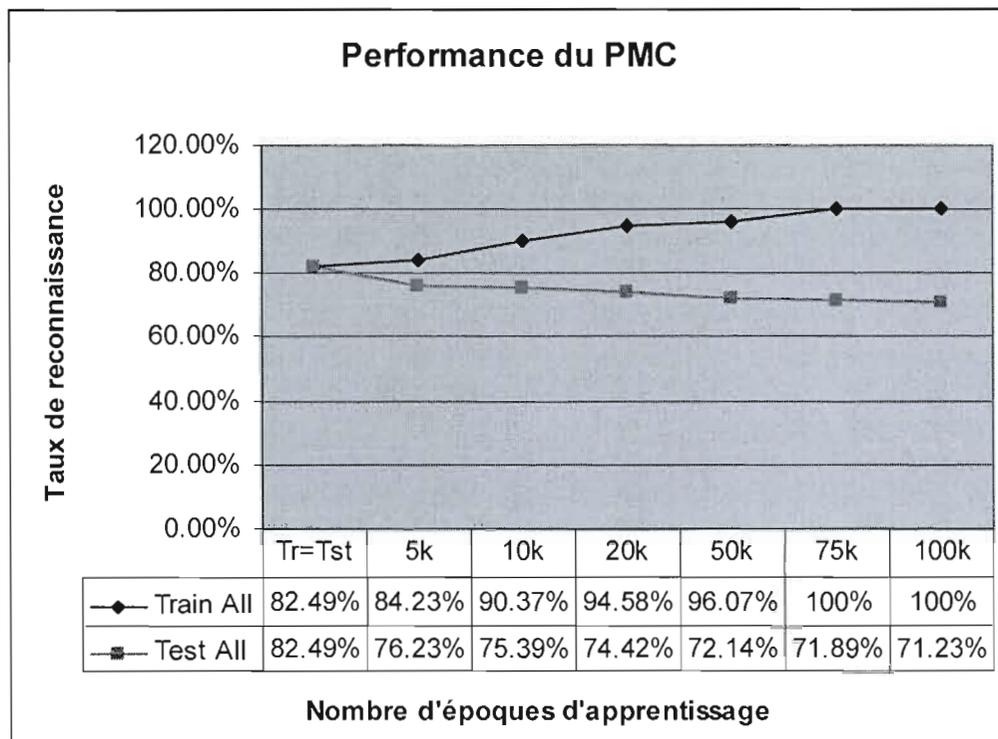


Figure 33. Taux de reconnaissance du PMC avec deux couches cachées

On peut remarquer, à travers les courbes et les taux de classification respectifs à chaque architecture, qu'avec deux couches cachées le PMC a mieux convergé. En effet, au moment où les deux courbes (rose et bleue) se croisent, 82.49% de reconnaissance a été atteinte à la phase de rappel pour la deuxième architecture, puis ce taux a baissé au fil des époques jusqu'à atteindre le pourcentage de 71.23% après 100k époques d'apprentissage. Un maximum de 78.49% à l'identification a pu être atteint avec la première architecture, ce taux de rappel a ensuite diminué jusqu'à 63.23% après 100k d'époques d'apprentissage.

4.6 Étude comparative

Dans cette partie du chapitre expérimental, on établit une étude comparative sur les trois modèles de classification qu'on a conçu, soient la mémoire associative bidirectionnelle (BAM), la mémoire associative bidirectionnelle à sortie chaotique (CNN-BAM), et le Perceptron multicouche (PMC). On en vient aussi à discuter quelques points évoqués dans le chapitre 2 dans le but de mettre en avant l'apport de ce travail.

4.6.1 Architectures

Il est important de rappeler que les deux modèles à mémoire associative sont similaires sur certains paramètres, notamment leurs architectures. On pourrait remarquer que l'architecture employée pour le BAM et pour le BAM chaotique est très simple. Le nombre de couches et de neurones est fixé d'une manière implicite. En effet, c'est la taille du vecteur d'entrée qui permet de décider du nombre d'unités de traitement dans chaque couche, le nombre de couches étant de 2 par défaut. Ceci est une propriété intéressante vu que dans d'autres types de réseaux de neurones, on se retrouve souvent face à la difficulté de déterminer la topologie optimale pour le réseau d'une manière empirique. Ceci est d'ailleurs le cas pour le Perceptron multicouche dont il est particulièrement délicat de fixer le nombre de couches et d'unités cachées. C'est pour cette raison que l'on s'est inspiré des expérimentations réalisées en (Naoum et al., 2008) pour déterminer l'architecture du PMC proposée.

4.6.2 Règle d'apprentissage

La règle d'apprentissage utilisée dans les deux modèles de mémoire associative est inspirée de l'approche hebbienne/anti-hebbienne proposée initialement par Bégin et Proulx (Bégin et Proulx, 1996). Cette règle est intéressante par le fait qu'elle permet d'assurer une bonne stabilisation du système. En effet, l'utilisation d'une règle d'apprentissage strictement hebbienne ne permet pas d'établir un contrôle autonome de l'apprentissage. Il a été montré dans (Bégin et Proulx, 1996) que la règle hebbienne définit une fonction strictement positive croissante alors que la modification inverse, ou l'anti-hebbienne, engendre une fonction strictement négative décroissante. Cependant, sous certaines conditions des paramètres impliqués, la combinaison additive de ces deux types d'ajustement permet de formuler une nouvelle règle d'apprentissage qui définit une fonction partiellement décroissante et qui admet un point d'équilibre. Ces conditions impliquent que le changement hebbien soit de plus grande ampleur et se produise avant l'application de la modification anti-hebbienne. Cette règle d'apprentissage a procuré un temps d'apprentissage relativement raisonnable.

Par ailleurs, avec le PMC, malgré que la règle d'apprentissage à rétropropagation résiliente soit connue comme étant plus rapide que la rétropropagation classique (Riedmiller et Braun, 1993), l'apprentissage du PMC a été tout de même lent, et ce, davantage avec la topologie comprenant deux couches d'unités cachées.

4.6.3 Fonction de sortie

Dans les trois réseaux de neurones conçus, des fonctions de sortie non linéaires ont été employées. Ceci favorise ces modèles en termes de plausibilité

biologique. Pour le cas du PMC, c'est une fonction logistique (Sigmoide) qui a été utilisée vu la performance intéressante qu'elle procure selon ce qui revient dans la littérature. Pour le réseau BAM, la carte cubique a été employée dans le but de reprendre les paramètres utilisés dans (Chartier et Boukadoum, 2006), cette carte a été employée en faisant en sorte qu'elle opère en mode non chaotique dans un premier temps.

L'implémentation du réseau chaotique se caractérise de par sa simplicité architecturale, mais aussi sa fonction de sortie. En plus de la carte cubique opérant en mode chaotique, plusieurs cartes chaotiques ont été employées en vue de les évaluer en termes de performance de classification. Celle-ci a été parfaite pour la carte Hénon et Bernoulli, très bonne pour d'autres cartes notamment Logistic3 et Mira1, et moins bonne avec d'autres voir quasi-nulle. Il s'agit là d'une caractéristique qui le distingue des modèles présentés dans la littérature, qui utilisent des algorithmes d'apprentissage et de rappel complexes et difficiles à mettre en œuvre (voir chapitre 2).

À la lumière des résultats obtenus avec le modèle BAM, on pourrait remarquer que l'apport des fonctions chaotiques dans le CNN-BAM soit dans l'atteinte de taux de classification parfaits avec deux des 23 différentes cartes chaotiques utilisées, mais aussi des taux quasi-parfaits dépassant les 98% avec d'autres, ce qui n'a point pu être réalisable avec le premier BAM.

Remarquons aussi que c'est avec la carte Hénon et la fonction Bernoulli que l'on a obtenu le taux de succès le plus haut, une parfaite reconnaissance des 95 cas de tumeur de la série de test. Ceci peut être expliqué par le fait que la carte Hénon soit une carte caractérisée par une orbite chaotique attractrice contrairement par exemple à la logistique3 qui a un point fixe attracteur. Cette

propriété peut être traduite par le fait que le bassin d'attraction de la carte Hénon soit large. Aussi, avec la fonction Bernoulli, on a pu classer chaque cas de tumeur bien que les échantillons de test n'aient pas été appris à la phase d'apprentissage. De plus, cette fonction étant définie dans l'intervalle $] 0,1[$, les valeurs entre -1 et 0 ont été en l'occurrence ignorées, cela n'a néanmoins pas affecté sa performance à générer un taux de reconnaissance parfait. Ceci prouve la robustesse de la fonction.

Quant à la performance du réseau PMC, elle a été meilleure avec l'architecture à deux couches cachées avec un taux de 82.49% contre 78.49% avec une seule couche. Cette performance est tout à fait remarquable : petite taille de la série d'apprentissage (191 patrons) relativement au nombre de classes qu'elle contient (175). En fait, la base de données « cancer du sein » contient 286 cas de tumeurs qui appartiennent à 175 classes différentes, ceci implique que l'on dispose d'à peine 1.63 échantillons pour chaque classe. Il s'agit d'une insuffisance considérable, et pourtant, les modèles de PMC proposés ont fait preuve d'une bonne capacité de généralisation. L'importance de la taille de la série d'apprentissage est illustrée en (Naoum et al., 2008). Dans ce travail il était question d'effectuer une classification de données à travers 19 classes différentes, on disposait d'une série de 2103 patrons consacrée à la phase d'apprentissage, soit chaque classe était représentée par 110 échantillons. Les résultats de classification atteignaient les 94% et ce taux a été estimé perfectible avec une série de données plus grande.

Par ailleurs, pour la performance du PMC obtenue au fil des époques d'apprentissage, on pourrait noter que le meilleur taux à la phase de rappel est obtenu au moment où les deux courbes d'erreur (erreur à l'apprentissage et

erreur au test) se croisent. Une valeur de 82.49% de précision a été atteinte à la phase de rappel pour l'architecture à deux couches, puis ce taux a baissé au fil des époques jusqu'à atteindre le pourcentage de 71.23% après 100k époques. Un maximum de 78.49% à l'identification a pu être atteint avec l'architecture à une couche cachée. Ce taux de rappel a ensuite diminué jusqu'à 63.23% après 100k d'époques d'apprentissage. Ces chiffres montrent la tendance du réseau au surapprentissage.

Enfin, on pourrait remarquer à travers les travaux présentés au chapitre 2, notamment pour le cas des approches utilisant les réseaux de neurones artificiels, que les taux de reconnaissance que l'on a obtenu avec les trois modèles proposés, sont compétitifs voir meilleurs que ceux obtenus dans plusieurs travaux de recherche. 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. Ceci dit, il est tout aussi important de souligner le fait qu'aucun des travaux présentés précédemment dans la revue de littérature n'a effectué une classification comprenant plus de 3 catégories. Or, dans ce travail, on a mis en œuvre trois classifieurs pour accomplir une tâche de classification comprenant 175 catégories. Ceci est remarquable en termes de pluralité de classes par rapport à ce qui revient dans la recherche dans le cadre du dépistage du cancer du sein.

CONCLUSION

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 de neurones artificiels ont largement fait leurs preuves au sein de la communauté scientifique et industrielle.

Chaque modèle de réseau de neurones 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. On s'est intéressé dans ce travail à étudier la performance d'un modèle de réseau de neurones chaotique sur une tâche de classification reliée à la détection du cancer du sein, soit, une mémoire associative bidirectionnelle à fonction de sortie chaotique. À cet effet, on a 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. On a constaté en conséquence que les classifieurs proposés dans le domaine font preuve d'une certaine complexité à la réalisation de par les hybridations que comportent la plupart sans pour autant que leurs performances soit irréprochables. Le réseau chaotique

que l'on a proposé se caractérise de par sa simplicité mais aussi de par sa performance de classification. Celle-ci a été parfaite avec la fonction de sortie Hénon et la fonction Bernoulli, atteignant un taux 100% de précision. Une bonne performance a été atteinte par ailleurs avec d'autres cartes chaotiques, notamment Logistic3 et Mira1.

Notons aussi que dans les travaux présentés dans la revue de littérature, il était question de classification binaire ou triple au plus. Or, dans ce travail, on a 286 cas de tumeurs à identifier à travers 175 classes différentes, ce qui est remarquable en termes de pluralité de classes par rapport à ce qui revient dans la recherche dans le cadre du dépistage du cancer du sein.

On a outre cela mis en œuvre deux autres modèles de réseaux de neurones classiques, à savoir, une mémoire associative bidirectionnelle et un réseau multicouche à rétropropagation d'erreur pour effectuer la même tâche de classification. On a procédé secondairement à une étude comparative à travers laquelle on met en avant l'apport fait par le réseau chaotique relativement aux deux autres classificateurs. La performance de la mémoire associative chaotique a souvent été irréprochable comparativement aux deux autres modèles. Notamment avec les deux fonctions de sortie Hénon et Bernoulli pour lesquelles le rappel du réseau a été parfait à l'égard de tous les échantillons de test. Ceci met en avant la bonne capacité de généralisation et de rappel du réseau en utilisant ces deux cartes.

On considère qu'à travers les différentes réalisations citées, il a été possible de palier à certaines limites que connaissent les classificateurs de cancer du sein déjà existants dans la littérature. Ces palliatifs favorisent en particulier l'applicabilité des réseaux de neurones chaotiques pour l'aide au diagnostic du cancer du sein. Toutefois, on estime pouvoir élargir l'exploitation de cette applicabilité, et il serait dans cette optique intéressant de mettre plus d'investigations pour expérimenter d'autres

modèles de réseaux chaotiques et d'autres fonctionnalités du chaos au service de l'aide au diagnostic du cancer du sein.

On note par ailleurs que l'on disposait dans le cadre de ce travail d'un nombre relativement petit de données, à savoir, 286 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 de neurones chaotique face à un plus grand nombre de cas de tumeurs et un plus grand nombre de patrons pour chaque cas.

De plus que ces dernières considérations, on souligne le fait que les résultats que l'on a obtenu à travers les expérimentations du CNN-BAM sont prometteurs, on retiendra particulièrement ceux obtenus avec la cartes Hénon et la carte Bernoulli. Ces résultats sont similaires à ceux obtenus avec un réseau BAM ayant une fonction de sortie chaotique employé dans (Tabari et Boukadoum, 2006), où il était question de faire la reconnaissance d'expressions faciales. De ce fait, il serait tout à fait avantageux d'exploiter les RNAs chaotiques pour le dépistage du cancer du sein avec d'autres natures de données telles que, des images. Dans cet esprit, la reconnaissance de microcalcifications est envisageable.

BIBLIOGRAPHIE

- McCulloch, W. S., et Pitts W., 1943. « A logical calculus of the ideas immanent in nervous activity ». *Bull. Math. Biophysics*, vol. 5, pp. 115-133.
- Guevara M.R., Glass L., Mackey M.C., et Shrier A., 1983. « Chaos in Neurobiology ». *IEEE Transactions on Systems, Man, and Cybernetics*, SMC vol 13, pp. 790-798.
- Babloyantz A., Nicolis C., Salazar J.M., 1985. « Evidence of chaotic dynamics of brain activity during the sleep cycle ». *Physic Letters*, 111 A, pp. 152-156.
- Freeman W.J., 1987. « Simulation of Chaotic EEG Pattern with a dynamic model of the olfactory system ». *Biol. Cyber*, 56, pp. 139-150.
- Freeman W.J., Yao Y., Burke B., 1988. « Central pattern generating and recognizing in olfactory bulb : a correlation learning rule ». *Neural Networks*, 1, pp. 277-288.
- Adachi, M., et Aihara, K., 1997. « Associative dynamics in a chaotic neural network ». *Neural Networks*, 10, pp. 83-98.
- Aihara, K., Takabe, T., et Toyoda, M., 1990. « Chaotic neural networks ». *Physical Letters A*, 144, pp. 333-340.
- Tokuda, I., Nagashima, T., et Aihara, K., 1997. « Global bifurcation structure of chaotic neural networks and its application to traveling salesman problem ». *Neural Networks*, 10, pp. 1673-1690.

- Cheng S.N.C., Chan H.P., Helvie M.A., Goodsitt M.M., Adler D.D., St. Clair D.C., 1994. « Classification of mass and non-mass regions on mammograms using artificial neural networks ». *J. Imaging Sci, Technol*, 38.
- Baeg S., Kehtarnavaz N., 2000. « Texture based classification of mass abnormalities in mammograms ». *Proceedings of the IEEE Symposium on Computer-based Medical Systems*, pp. 163–168, 2000.
- Floyd C.E., Lo J.Y., Yun A.J., Sullivan D.C., Kornguth P.J., 1994. « Prediction of breast cancer malignancy using an artificial neural network ». *Cancer* 74, pp. 2944–2948.
- Bovis K., Singh S., Fieldsend J., Pinder C., 2000. « Identification of masses in digital mammograms with MLP and RBF nets ». *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks Com*, pp. 342–347.
- Christoyianni I., Dermatas E., Kokkinakis G., 2000. « Fast detection of masses in computer-aided mammography ». *IEEE Signal Process*, pp. 54–64.
- Asoke Hong G., Nandi K., 2006. « Breast cancer diagnosis using genetic programming generated feature ». *Elsevier Pattern recognition*, vol 39, pp. 980-987.
- Cheng H. D., Cui Muiyi, 2004. « Mass lesion detection with a fuzzy neural network ». *Elsevier Pattern recognition*, vol 37, pp. 1189-1200.
- Wojciech M., Grohman, Atam P., Dhawan, 2001. « Fuzzy convex set-based pattern classification for analysis of mammographic microcalcifications ». *Elsevier Pattern recognition*, vol 34, pp. 1469-1482.

- Abbass Hussein A., 2002. « An evolutionary artificial neural networks approach for breast cancer diagnosis ». *Artificial intelligence in medicine*, vol 25, pp. 265-281.
- Cheng H.D., Shi X.J, Min R., Hu l.m., X.P., Cai, H.N., 2006. « Du. Approaches for automated detection and classification of masses in mammograms ». *Pattern recognition*, vol 39, pp. 646-668.
- Bovis K., Singh S., Fieldsend J., Pinder C., 2000. « Identification of masses in digital mammograms with MLP and RBF nets ». *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks Com*, pp. 342-347.
- Moghaddam Baback, et Shakhnarovich Gregory, 2002. « Boosted dyadic kernel discriminants. Mitsubishi electric research laboratories ». *Advances in Neural information processing systems (NIPS 15)*. Vancouver B.C, Canada.
- Fisher Igor, et Poland Jan, 2005. « Amplifying the block matrix structure for spectral clustering ». *Technical Report, IDSIA*, pp. 03-05.
- Subhash C., Bagui, Sikha Bagui, Kuhu Pal, et Nikhil R., Pal, 2003. « Breast cancer detection using rank nearest neighbor classification rules ». *Elsevier Pattern recognition*, vol 36, pp. 25-34.
- Fei Sha, Lawrence K., Saul, Daniel D., Lee, 2003. « Multiplicative updates for nonnegative quadratic programming in support vector machines ». *Advances in Neural Information Processing Systems 15, Sebastian and K. Obermayer, Eds. Cambridge, MA: MIT Press*.
- Huang Kaizhu, Yang Haiqin, King Irwin, Lyu Michael R., Chan Laiwan, 2004. « Biased minimax probability machine for medical diagnosis ». *The 8th International Symposium on Artificial Intelligence and Mathematics*, pp. 4-6.

- Madden Michael G., 2002. « Evaluation of the Performance of the Markov Blanket Bayesian Classifier Algorithm » . *CoRR*, cs . LG/0211003.
- Esmeir Saher, Markovitch Shaul, 2004. « Lookahead-based Algorithms for anytime induction of decision trees ». *21th International conference on machine learning*, vol 69, pp. 33.
- De Santo Massimo, Molinara Mario, Tortorella Francesco, Vento Mario, 2003. « Automatic classification of clustered microcalcifications by a multiple expert system ». *Elsevier Pattern Recognition*, 36, pp. 1467-1477.
- Karnan M., Thangavel K., Ezhilarasu P., 2008. « Ant colony optimization and a new particle swarm optimization algorithm for classification of microcalcifications in mammograms ». *The 6th International Conference on Advanced Computing and Communication* .
- Liu Bo, Abbass Hussein A., McKay Bob, 2004. « Classification rule discovery with ant colony optimization ». *IEEE Computational intelligence bulletin*, Vol.3 No.1.
- Parepinelli R. S., Lopes H. S., Freitas A, 2002. « An Ant Colony Algorithm for Classification Rule Discovery ». *Data Mining: Heuristic Approach: Idea Group Publishing*, H. A. a. R. S. a. C. Newton Edition.
- Jaganathan P., Thangavel K., Pethalakshmi A., Karnan M., 2007. « Classification rule discovery with ant colony optimization and improved quick reduct algorithm ». *IAENG International journal of computer science*, 33-1, IJCS_33_1_9.
- Negnevitsky Michael. Artificial Intelligence, 2005. « A Guide to Intelligent Systems ». Addison Wesley, Second Edition.

- Chartier S., Boukadoum M., 2006. « A Bidirectional Heteroassociative Memory for Binary and Grey-Level Patterns ». *IEEE Transactions on Neural Networks*, 18-2.
- Alligood, Kathleen T., Tim D. Sauer, James A. Yorke, 2006. « Chaos: An Introduction to Dynamical Systems », *Springer*, 1996.
- Gleick James, 1999. « La Théorie du Chaos ». *Flammarion*.
- Adachi M., Aihara K., 1997. « Associative dynamics in a chaotic neural network ». *Neural Networks*, 10, 83-98.
- Kobori.E.R, Ikoda. K et Nakayama. K, 1996. Model of dynamic associative memory ». *IEEE International Conference on Neural Networks Conference Proceedings*, 2, 804-809.
- Guoguang He, Cao Zhitong, Chen Hongping, Zhu Ping, 2003. « Controlling Chaos in a Neural Network Based on The phase Space Constraint ». *International Journal of Modern Physics B*, 17, 22, 23 & 24, 4209-4214.
- Guoguang He a,b,c, Manish Dev Shrimali, Kazuyuki Aihara, 2007. « Partial state feedback control of chaotic neural network and its application ». *Elsevier*, 0375-9601.
- Kushibe Masanori, Liu Yun, et Ohtsubo Junji, 1996. « Associative memory with spatiotemporal chaos control ». *Physical Rreview E* , 53, 5.
- Watanabe Masataka, Takahashi Yohei, Aihara Kazuyuki, 2000. « Associative memory model with temporal spike coding and active dendrite ». *Elsevier BioSystems* 58, 259–263.

- Guoguang He, Zhitong Cao, Ping Zhu, Hisakazu Ogura, 2003. « Controlling chaos in a chaotic neural network », *Neural Networks* 16, 1195–1200.
- Guoguang He, Jousuke Kuroiwa, Hisakazu Ogura, Ping Zhu, Zhitong Cao, Hongping Chen, 2004. « A type Delay Feedback Control of Chaotic Dynamics in a Chaotic Neural Network ». *IEICE Trans, Fundamentals*, E87-A, 7.
- Guoguang He, Luonan Chen and Kazuyuki Aihara, 2007. « Associative Memory with Controlled Chaotic Neural Network ». *Neurocomputing*, PII:S0925-2312(07)00304-9, 10884.
- Shin Mizutani, Takuya Sano, Tadasu Uchiyama, Noboru Sonehara, 1998. « Controlling Chaos in Chaotic Neural Networks ». *Electronics and communications in Japan, Part III: Fundamental Electronic Science*, pp. 73-82.
- Ott, E., Grebogi, C. & Yorke, J. A., 1990. « Controlling chaos ». *Physical Review Letters* 64, 1196–1199.
- W-E Wikipedia, the free encyclopedia,
http://en.wikipedia.org/wiki/List_of_chaotic_maps.
- England J. P., Krauskopf B., Osinga H. M., 2004. « Computing One-Dimensional Stable Manifolds and Stable Sets of Planar Maps without the Inverse ». *SIAM J. Applied Dynamical Systems, Society for Industrial and Applied Mathematics*, 3, 2, 161–190.
- England J.P., Krauskopf B., Osinga H.M., 2005. « Bifurcations of stable sets in noninvertible planar maps ». *Bifurcation and Chaos* 15-3 891-904.

- Hauptmann C., Touchette H., Mackey M. C., 2003. « Information capacity and pattern formation in a tent map network featuring statistical periodicity ». *Physical Review Letters*, 67, 026217.
- Yoshida T., Mori H., et Shigematsu H., 1983. « Analytic Study of Chaos of the Tent Map: Band Structures, Power Spectra, and Critical Behaviors ». *Journal of Statistical Physics*, 31-2.
- Hasler M., Maistrenko Yu. L., 1997. « An Introduction to the Synchronization of Chaotic Systems: Coupled Skew Tent Maps ». *IEEE Transactions, CAS 1*, 44, 856-866.
- Shilnikov A.L., Rulkov N.F., 2003. « Origin of Chaos in a Two-Dimensional Map Modeling Spiking-Bursting Neural Activity ». *International Journal of Bifurcation and Chaos*, 13, 3325.
- Haykin Simon, 1997. « Neural Networks: A Comprehensive Foundation ». *Prentice-Hall, Inc.*
- Riedmiller M., Braun H., 1993. « A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm ». *IEEE Proceedings, International Conference on Neural Networks*, San Francisco, CA.
- Tabari K., 2005. « Rapport de projet pour DIC9310 ».
- Naoum H., Boukadoum M., 2008. « Intelligent Classifier Module for Fluorescence-based Measurements ». *Fifth International Workshop on Signal Processing and its Applications*, 18-20.

- Fausett Laurene, 1994. « Fundamentals of Neural Networks: Architectures, Algorithms, And Applications ». *Prentice-Hall, Inc.*
- Bégin J., Proulx R., 1996. « Categorization in Unsupervised Neural Networks: The Eidos Model ». *IEEE Transactions on Neural Networks*, 7, 147-154.
- Gumowski I., Mira C., 1980. « Recurrences and discrete dynamic systems ». *Springer-Verlag, New York.*
- Baek Seung-Jong et Ott Edward, 2004. « Onset of synchronization in systems of globally coupled chaotic maps ». *Physical Review Letters* 69, 066210.
- Chen Chi-Chung, Yao Kung, Umeno Ken, et Biglieri Ezio, 2001. « Design of Spread-Spectrum Sequences Using Chaotic Dynamical Systems and Ergodic ». *IEEE Transactions on circuits and systems, Fundamental Theory and applications*, Vol 48, No 9.
- Tomoki Inoue, 1994. « Ergodic theorems for piecewise affine Markov maps with indifferent fixed points ». *Hiroshima Math. J. Volume 24, Number 3*, 447-471.
- Giovaiznrdi A., Mazziizi G., 2001. « Frequency domain chaotic watermarking ». *Circuits and systems, ISCAS, The IEEE International Symposium*, 521-524 vol. 2.
- Rovatti Riccardo, Setti Gianluca, et Mazzini Gianluca, 1998. « Chaotic Complex Spreading Sequences for Asynchronous DS-CDMA—Part II: Some Theoretical Performance Bounds ». *IEEE Transactions on circuits and systems—I: Fundamental theory and applications*, vol 45, No 4.
- Tabari K., Boukadoum M., S. Chartier, Lounis H., 2006. « Reconnaissance d'expressions faciales à l'aide d'une mémoire associative bidirectionnelle à fonction de sortie chaotique ». *Proceedings of Maghrebian Conference on Software Engineering and Artificial Intelligence*, Agadir, Morocco, 422-426.