

N° d'ordre : 06/2006-M/EL

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DES SCIENCES ET DE LA TECHNOLOGIE

HOUARI BOUMEDIENE

Faculté d'Électronique et d'Informatique



MEMOIRE

Présenté pour l'obtention du diplôme de MAGISTER

en : **ELECTRONIQUE**

Spécialité : Electronique des Systèmes

Par : Fatma Zohra ALLAM Née CHERGUI

Sujet

**CLASSIFICATION D'IMAGES SATELLITAIRES
PAR APPROCHE NEURONALE**

Soutenu le 23 février 2006, devant le jury composé de :

Mr. A. DJERADI

Mme. A.BELHADJ-AISSA

Mme. L.HAMAMI

Mr. S. CHITROUB

Maître de Conférences à l'USTHB

Professeur à l'USTHB

Maître de Conférences à l'ENP

Maître de Conférences à l'USTHB

Président

Directeur de Thèse

Examineur

Examineur

« Pour faire de grandes choses,
il ne faut pas être un si grand génie,
il ne faut pas être au dessus des hommes,
il faut être avec eux ».

Charles De Secondât Montesquieu (1689-1755).

DEDICACES

Je dédie honorablement ce mémoire à :

Mes très chers parents

Mon mari Rachid.

Mes enfants Lotfi Mohamed Yacine et Ilyès Youcef Islam.

Mes beaux-parents.

Mes sœurs et mes belles-sœurs.

Mes frères et mes beaux-frères.

*Mes nièces et mes neveux : Lilya, Dounia, Abderrahim, Mohamed Said,
Fatma Zohra, Ibrahim et Redouane.*

Toute ma famille, grand et petit.

Tous mes amis et mes collègues.

Fatma Zohra

REMERCIEMENTS

Le travail présenté dans le cadre de cette thèse a été développé et mis en œuvre au **Laboratoire de Traitement d'Images et Rayonnement (LTIR)** de la **Faculté d'Electronique et d'Informatique (FEI)** de l'**Université des Sciences et de la Technologie Houari Boumediene (USTHB)**.

Je voudrai remercier ma responsable de thèse Madame Belhadj-Aissa, Professeur à la **Faculté d'Electronique et d'Informatique** à l'**USTHB**, de m'avoir accueillie dans son équipe et de m'avoir proposé et dirigé ce travail de recherche. Je la remercie, également, pour sa disponibilité et pour les conseils fructueux qu'elle m'a prodigués.

Je tiens à exprimer ma reconnaissance à Monsieur Djerradi, Maître de conférences à l'**USTHB**, pour l'honneur qu'il me fait en acceptant de présider le jury de cette thèse.

J'exprime mes vifs remerciements à Madame Hammami, Maître de conférences à l'**Ecole Nationale Polytechnique (ENP)** et à Monsieur Chitroub, Maître de conférences à l'**USTHB**, pour avoir accepté de prendre part au jury de cette thèse.

Je ne saurai jamais assez remercier Mademoiselle Khedam, Attachée de recherche à l'**USTHB** et Mademoiselle Nemmour, Attachée de recherche à l'**USTHB**, pour leur aide très précieuse et pour leurs remarques et leurs discussions très constructives. Qu'elles trouvent ici l'expression de ma profonde gratitude.

Je remercie chaleureusement Madame Kemmouche, Maître de conférences à l'**USTHB**, Madame Fallek, Chargée de cours à l'**USTHB**, Monsieur Belhadj-Aissa, Maître de conférences à l'**USTHB** et Monsieur Smara, Professeur à l'**USTHB**, pour leurs encouragements et leur soutien technique et moral.

Je garderai un très bon souvenir des moments passés au laboratoire **LTIR** en compagnie de ses membres dont notamment : Afifa, Assia, Bahia, Dalila Benmehrez, Faiza, Lynda, Mounira, Nadia, Samia, Abdennour, Boularbah et Samy. Je tiens à les remercier vivement pour leur sympathie et leur gentillesse.

J'adresse mes remerciements les plus sincères à tous mes collègues de l'**Institut des Techniciens Supérieurs** avec qui j'ai passé d'excellents moments. Une pensée particulière à Akila, Habiba, Lynda, Rabea, Saida, Fadhéla, Nacera, Saliha, Amirouche, Bachir, Bali Noureddine, Bour Noureddine, Hakim Laouas, Lakhdar, Lyès, et Rachid.

Enfin, je témoigne mes remerciements à tous ceux qui n'ont pas été cités ci-dessus, mais qui ont contribué d'une manière ou d'une autre au bon déroulement de mon travail.

Résumé

Durant ces dernières années, une nouvelle technologie émergente a connu un essor considérable. Elle permet de traiter l'information plus efficacement que les systèmes conventionnels et d'en retirer des données, jusque là, indispensables afin d'aider à l'interprétation et à la prise de décision. Cette technique s'appuie sur les réseaux de neurones. Ces réseaux ont la spécificité d'apprendre par eux-mêmes à extraire des renseignements cachés dans une masse d'informations et de fournir, ainsi, des modèles puissants quant aux connaissances relatives à un problème donné. Ils représentent une classe d'algorithmes qui sont utilisés pour la classification, le regroupement des données, la prédiction, la reconnaissance de formes, etc...

Dans ce contexte, le travail que nous avons effectué consiste à mettre en oeuvre un processus de classification supervisée des images multispectrales par approche neuronale. Nous avons choisi, pour la mise en oeuvre de notre travail, deux modèles : le perceptron multicouche et le réseau de Hopfield.

Dans une première étape, nous avons étudié trois méthodes de classification d'images satellitaires, puis nous avons développé l'approche neuronale que nous avons appliquée à la classification. Dans cette partie, nous avons abordé le perceptron multicouche entraîné par l'algorithme de rétro-propagation durant la phase d'apprentissage supervisé. Cet algorithme repose sur la minimisation de l'erreur entre les sorties calculées par le réseau et les sorties désirées. Une fois la mise à jour des poids de connexions dans les couches de traitement est effectuée, vient la phase de test, durant laquelle nous avons fait varier les différents paramètres qui influent sur la convergence du réseau afin de déterminer les poids optimaux qui donnent la meilleure architecture. Nous avons développé, également, le modèle de Hopfield. La principale caractéristique de ce modèle réside dans sa fonction d'énergie, appelée fonction de Lyapunov, qui décroît au fur et à mesure que le réseau s'approche de son état stable. Nous avons assimilé cette fonction à une mesure de distance séparant le pixel à classer du centre de gravité d'une classe. Le processus est réitéré jusqu'à la convergence du réseau. A chaque itération, nous calculons le nouveau centre de gravité qui dépend étroitement du nombre de pixels affectés à la classe considérée. Les approches développées ont été testées sur des images multispectrales. Une étude comparative et une évaluation ont été effectuées.

Mots-clés : *Classification, réseau de Hopfield, Perceptron multicouche, Algorithme de Rétropropagation, Fonction d'énergie d Lyapounov.*

SOMMAIRE

INTRODUCTION GENERALE	1
CHAPITRE I : NOTIONS SUR LES RESEAUX DE NEURONES	4
Introduction	5
I. Modèle neurophysiologique	7
I.1 Structure	7
I.2 Physiologie	8
II. Modèle artificiel	8
II.1 Structure	8
II.2 Fonction de transfert	10
II.3 Structure d'interconnexion	12
II.4 Concepts d'apprentissage	14
II.5 Règles d'apprentissage	16
II.6 Propriétés des réseaux de neurones	17
III. Modèles usuels	17
III.1 Modèle de kohonen	18
III.2 Modèle de Rosenblatt	19
III.3 Modèle de Hopfield	20
Conclusion	21
CHAPITRE II : TECHNIQUES DE CLASSIFICATION D'IMAGES SATELLITAIRES	
Introduction	24
I. Principe de la classification	24
II. Méthodes de classification	26
II.1 Méthodes non supervisées	26
II.2 Méthodes supervisées	26
III. Différents types de classification	27
III.1 Classification ponctuelle	27
III.2 Classification contextuelle	28
IV. Différentes techniques d'analyse d'images	28
IV.1 Approche statistique	28
IV.2 Approche géométrique	33
IV.3 Approche génétique	34
IV.3 Approche neuronale	37
Conclusion	38

CHAPITRE III : MISE EN ŒUVRE DU PROCESSUS DE CLASSIFICATION PAR APPROCHE NEURONALE	39
Introduction	40
I. Classification par le perceptron multicouche.....	40
I.1. Description architecturale	40
I.2. Phase d'apprentissage	42
I.3. Mise en œuvre	42
I.4. Organigramme.....	48
II. Classification par le réseau de Hopfield	51
II.1. Description architecturale	51
II.2. Phase d'apprentissage	52
II.3. Mise en œuvre.....	52
II.4. Organigramme.....	55
Conclusion	57
CHAPITRE IV : TESTS ET RESULTATS	58
Introduction.....	59
I. Présentation des données images	59
II. Critères d'évaluation et interprétation	61
II.1. Appréciation visuelle	61
II.2. Appréciation statistique	61
III. Tests et résultats du perceptron multicouche.....	63
III.1. Appréciation visuelle	71
III.2. Appréciation statistique.....	71
IV. Tests et résultats de réseau de Hopfield	73
IV.1. Appréciation visuelle	78
IV.2. Appréciation statistique.....	78
V. Comparaison entre les différentes approches...	81
Conclusion	82
CONCLUSION GENERALE	83
ANNEXES.....	85
BIBLIOGRAPHIE	95

Liste des Figures

Figure I.1. Neurone biologique	7
Figure I.2. Synoptique du neurone artificiel	8
Figure I.3. Neurone artificiel	9
Figure I.4. Mise en correspondance neurone biologique / neurone artificiel	10
Figure I.5. Différents types de fonction de transfert.....	12
Figure I.6. Structures d'interconnexions	13
Figure I.7. Architecture du réseau de Kohonen.....	18
Figure I.8. Architecture du perceptron multicouche.....	20
Figure I.9. Architecture du réseau de Hopfield.....	21
Figure II.1. Etapes de classification supervisée	27
Figure II.2. Image résultat de la classification par le maximum de vraisemblance.....	30
Figure II.3. Image résultat de la classification par le minimum de distance euclidienne.....	31
Figure II.4. Image résultat de la classification par la distance de Mahalanobis.....	32
Figure II.5. Image résultat de la classification par les champs de Markov.....	33
Figure II.6. Procédure de classification par les Algorithmes Génétiques.....	36
Figure II.7. Image résultat de la classification par les Algorithmes Génétiques.....	36
Figure II.8. Comparaison des valeurs du Khat des différentes méthodes de classification.....	37
Figure II.9. Synoptique des différentes approches de classification	38
Figure III.1. Représentation schématique du perceptron multicouche	42
Figure III.2. Organigramme de la phase d'apprentissage par le perceptron.....	47
Figure III.3. Organigramme de la phase de classification par le perceptron.....	48
Figure III.4. Représentation schématique du modèle de Hopfield.....	49
Figure III.5. Organigramme de la classification par le réseau de Hopfield.....	54
Figure IV.1. Scène de Blida.....	58
Figure IV.2. Variation de l'erreur en fonction du nombre de neurones cachés.....	62
Figure IV.3. Variation du Khat en fonction du nombre de neurones cachés.....	62
Figure IV.4. Variation TMBC en fonction du nombre de neurones cachés.....	63

Figure IV.5. Variation de l'erreur en fonction du nombre d'itérations.....	64
Figure IV.6. Variation du Khat en fonction du nombre d'itérations.....	64
Figure IV.7. Variation du TMBC en fonction du nombre d'itérations.....	64
Figure IV.8. Variation de l'erreur quadratique en fonction du momentum.....	65
Figure IV.9. Variation de l'erreur quadratique en fonction du taux d'apprentissage.....	65
Figure IV.10. Variation du TMBC en fonction du momentum.....	66
Figure IV.11. Variation du TMBC en fonction du taux d'apprentissage.....	66
Figure IV.12. Variation du Khat en fonction du momentum.....	66
Figure IV.13. Variation du Khat en fonction du taux d'apprentissage.....	67
Figure IV.14. Image résultat de la classification par le perceptron.....	68
Figure IV.15. Variation du nombre de pixels par classe en fonction du nombre d'itérations...	71
Figure IV.16. Variation du nombre d'échantillons entre deux itérations successives.....	71
Figure IV.17. Variation du TMBC en fonction du nombre d'itérations.....	72
Figure IV.18. Variation du Khat en fonction du nombre d'itérations.....	72
Figure IV.19. Variation du TBC en fonction du nombre d'itérations	73
Figure IV.20. Variation du TC en fonction du nombre d'itérations.....	73
Figure IV.21. Images résultats de la classification relatives à la 1 ^{ère} et la 5 ^{ème} itération.....	74
Figure IV.22. Images résultats de la classification par le réseau de Hopfield.....	75
Figure A.1. Synoptique de la classification supervisée.....	87
Figure A.2. Représentation du perceptron à deux couches.....	92
Figure A.3. Représentation du perceptron à trois couches.....	94

Liste des Tableaux

Tableau I.1. Transition entre neurone biologique/neurone formel.....	10
Tableau II.1. Classes thématiques.....	30
Tableau II.2. Paramètres relatifs aux Algorithmes Génétiques.....	36
Tableau III.1. Topologie de la grille d Hopfield.....	51
Tableau IV.1. Classes thématiques et base d'entraînement.....	58
Tableau IV.2. Matrice de confusion normalisée relative au perceptron.....	69
Tableau IV.3. Indicateurs de surestimation et de confusion du perceptron.	69
Tableau IV.4. Matrices de confusion normalisées relatives à Hopfield.....	76
Tableau IV.5. Indicateurs de surestimation et de confusion de Hopfield.....	77

Introduction générale

« Il ne s'agit pas d'atteindre un résultat
vrai ou faux, probable ou improbable,
mais seulement profitable ou non profitable ».

Williams & Lance (1965)

Depuis des décennies, les satellites fournissent aux thématiciens de la télédétection des images de plus en plus nombreuses et de plus en plus riches en informations aussi bien spatiales que spectrales. Les images acquises par les satellites sont numériques et sont donc traitées par des ordinateurs en vue de les exploiter par des thématiciens.

L'information contenue dans ces images varie selon que les capteurs embarqués par les satellites d'observation de la Terre soient [LIB, 1989]:

- Passifs, lorsqu'ils mesurent des rayonnements électromagnétiques réfléchis ou émis par la surface observée dans plusieurs longueurs d'ondes dans le visible et le proche infra rouge du spectre électromagnétique.
- Actifs, lorsqu'ils émettent eux-mêmes les ondes dont ils mesurent le coefficient de rétrodiffusion selon une fréquence d'émission et un type de polarisation.

Les images de télédétection permettent, ainsi, de fournir des informations sur l'objet observé grâce à l'interaction matière-rayonnement déterminée par les caractéristiques du rayonnement utilisé (longueur d'onde, type de polarisation...). La caractérisation d'un objet fait intervenir des informations différentes fournies par une gamme de capteurs couvrant les régions les plus significatives du spectre électromagnétique.

L'exploitation directe de ces images est impossible en raison de la quantité d'informations acquise par les images multispectrales relative à une scène observée. A cet effet, l'analyse des informations contenues dans ces images devient plus complexe. Des traitements sont, alors, nécessaires pour exploiter efficacement ces images. De ce fait, des techniques et des méthodes plus élaborées ont été développées en relation avec les applications envisagées et la nature de l'information à extraire. Ces techniques permettent de lever les limitations de traitement et de résoudre les problèmes quant à la taille des images qui croît avec la dimension de l'espace et au temps de calcul qui augmente avec le nombre des images traitées.

Dans ce contexte, le travail que nous avons effectué consiste à mettre en œuvre un processus de classification d'images satellitaires, qui est une des techniques de traitement d'images, en utilisant l'approche neuronale. Pour cela, nous avons étudié et implémenté deux

modèles : le modèle de **Hopfield** et le **perceptron multicouche**. Nous avons effectué une comparaison des résultats obtenus.

Notre mémoire est composée de quatre chapitres organisés comme suit :

Nous avons consacré le premier chapitre à l'introduction de quelques notions de base sur les réseaux de neurones. Nous avons présenté, dans ce chapitre, les principales caractéristiques des réseaux neuronaux en se focalisant sur les différents modes d'apprentissage et les règles d'adaptation. Par la suite, nous avons exposé les différentes architectures qui établissent des topologies variées suivant l'application désirée. Nous avons exposé, également, les modèles les plus usuels, en l'occurrence, le modèle de **Kohonen**, le modèle de **Hopfield** et le **perceptron multicouche**. Ces deux derniers seront détaillés dans le troisième chapitre lors de la mise en œuvre du processus de classification.

Le deuxième chapitre est dédié à la présentation de l'une des techniques de traitement d'images la plus employée, à savoir la classification. Dans ce chapitre, nous avons, tout d'abord, présenté les différentes méthodes de classification, en l'occurrence, la méthode supervisée et la méthode non supervisée. Nous avons abordé, brièvement, le principe de la classification ponctuelle et la classification contextuelle. Par la suite, nous avons présenté les différentes approches utilisées pour la classification supervisée. Il s'agit de l'approche statistique, l'approche géométrique, l'approche génétique et l'approche neuronale, avec un intérêt particulier accordée à cette dernière qui sera détaillée dans le prochain chapitre. Finalement, nous avons soulevé les performances et les limitations de chaque approche.

Nous enchaînons notre étude par le troisième chapitre qui développera la méthodologie de classification par le réseau de **Hopfield** et le **perceptron multicouche**. Le premier modèle nécessite une phase d'apprentissage différente de celle des réseaux à couches. Néanmoins, lors de son évolution, le réseau de **Hopfield** atteint son état stable lorsqu'une fonction d'énergie, appelée fonction de **Lyapunov**, décroît vers un minimum. Dans le cadre de notre application, cette énergie correspond à la mesure de distance qui sépare le pixel à classer du centre de gravité d'une classe. Par ailleurs, le réseau de **Hopfield** converge vers un état stable lorsque cette distance est minimisée. Tandis que, pour le **perceptron multicouche**, durant la phase d'apprentissage, nous présentons au réseau les vecteurs d'entrée et de sortie. Le réseau fournit, à chaque étape, une sortie différente de celle désirée. Après

plusieurs itérations, la différence entre la sortie désirée et la sortie calculée est minimisée par modification des poids de connexions dans les couches de traitement.

Quant au quatrième chapitre, nous le consacrons à l'évaluation de chacune des méthodes développées dans le chapitre précédent. Nous effectuons plusieurs tests afin de déterminer les valeurs des paramètres les plus optimales qui donnent une meilleure architecture pour chaque modèle. Pour ce faire, nous avons agi sur les facteurs qui influent sur la convergence de chacun des deux réseaux. Nous donnons les images résultats de la classification que nous évaluons par rapport à une réalité terrain.

Après évaluations et interprétations des résultats que nous avons développées dans le quatrième chapitre, nous terminons notre travail par une conclusion générale dans laquelle nous exposons les principales difficultés rencontrées lors de la mise en œuvre de notre travail et nous proposons quelques perspectives et quelques indications bibliographiques.

Chapitre I

Notions sur les réseaux de neurones

Résumé

Dans ce chapitre, nous introduisons d'abord quelques notions rudimentaires sur les réseaux de neurones artificiels en explicitant, d'abord, le modèle artificiel inspiré du modèle biologique. Nous établissons la mise en correspondance neurone biologique/neurone artificiel au travers d'un tableau de transition. Nous décrivons, par la suite, le fonctionnement des réseaux de neurones artificiels et leurs principales propriétés d'apprentissage ainsi que les différentes structures d'interconnexions qui déterminent les différentes topologies. Il existe de nombreux modèles de réseaux de neurones artificiels, nous en présentons quelques uns qui sont, le plus couramment, utilisés, à savoir le perceptron multicouche, le modèle de Kohonen et le modèle de Hopfield.

INTRODUCTION

Bien que les ordinateurs soient, aujourd'hui, de plus en plus puissants et capables de résoudre de nombreux problèmes, il reste néanmoins des domaines où les humains ont plus de facilité de raisonner que les machines. D'où l'idée de l'Intelligence Artificielle (IA) pour essayer de se rapprocher de l'Intelligence Humaine afin de construire des machines plus puissantes et plus pratiques.

Actuellement, le traitement de l'information s'appuyant sur une approche basée sur l'Intelligence Artificielle se révèle être un moyen commode. Cependant, cette approche se limite à des applications où la modélisation d'un processus sous forme de règles est possible.

De ce fait, une autre approche basée sur le Comportement Intelligent permet d'effectuer le traitement de l'information par le cerveau. L'hypothèse principale, à la base de l'essor des réseaux de neurones artificiels, est que le Comportement Intelligent est sous-tendu par un ensemble de mécanismes mentaux basés sur des processus neurophysiologiques. Depuis une vingtaine d'années, c'est devenu un domaine de recherches très prometteur [TOU, 1992].

Le neurologue autrichien **Sigmund Freud** est l'un des premiers chercheurs et fondateurs de la psychanalyse à une époque où les notions de la physiologie étaient élémentaires et les sciences des systèmes étaient inconnues. A partir de ses réflexions sur la psychanalyse, il imaginait les prémices de la plasticité neuronale en montrant qu'il existe des neurones perméables servant à la perception et des neurones imperméables servant à la mémorisation [HER, 1994].

En **1890**, le célèbre philosophe américain **William James** introduit le concept de mémoire associative, et propose ce qui deviendra une loi de fonctionnement pour l'apprentissage sur les réseaux de neurones connue plus tard sous le nom de la loi de **Hebb** [HER, 1994].

En **1943**, les deux neurobiologistes américains **Warren McCulloch** et **Walter Pitts** de l'université de Chicago laissent leurs noms à une modélisation du neurone biologique (un neurone au comportement binaire). Ceux sont les premiers à concevoir le premier modèle du neurone formel appelé automate à seuil ou à décision. Ils montrent, par ailleurs, que les réseaux de neurones formels simples peuvent réaliser des fonctions logiques, arithmétiques et symboliques complexes (tout au moins au niveau théorique) [TOU, 1994].

En 1949, le neurophysiologiste américain **Donald Hebb** propose une formulation du mécanisme d'apprentissage sous la forme d'une règle de modification des connexions synaptiques entre les neurones [DJE, 2004].

En 1957, l'un des premiers chercheurs en connexionnisme, le psychologue **Frank Rosenblatt** s'intéressait à la possibilité d'expliquer les capacités d'association des êtres humains par des capacités équivalentes dans des modèles neuromimétiques. Ainsi, il conçoit le perceptron comme un réseau de neurone décisionnel capable d'apprendre, d'identifier des formes simples et de calculer des fonctions logiques [JOD, 1994]. Il propose le premier algorithme d'apprentissage qui permet d'ajuster les paramètres d'un neurone.

En 1960, l'automaticien **Bernard Widrow** développe le modèle d'**Adaline** (**AD**aptive **L**inear **NE**uron) (c-à-d Neurone Linéaire) qui a une structure semblable à celle du perceptron, cependant, la loi d'apprentissage est différente. Celle-ci est à l'origine de l'algorithme de rétropropagation de gradient très utilisé, aujourd'hui, avec les perceptrons multicouches [TOU, 1994] [HER, 1994].

En 1969, les deux scientifiques américains **Marvin Minsky** et **Seymour Papert** démontrent qu'un réseau de neurone de type perceptron est incapable de résoudre les problèmes non linéaires. Ils publient le livre « *Perceptrons* » dans lequel ils mettent en exergue les limitations, non seulement, du perceptron mais ils les étendent à tous les modèles de réseaux de neurones. Leur objectif, étant atteint, la plupart des chercheurs quitteront le champ de recherche sur les réseaux de neurones et il y a eu abandon financier des recherches dans ce domaine [TOU, 1994].

En 1982, le physicien américain **John Hopfield**, qui travaille dans le domaine de la physique statistique, propose des réseaux de neurones associatifs. Il démontre, à travers ses travaux, l'utilité des réseaux entièrement connectés. L'intérêt pour les réseaux de neurones renaît chez les scientifiques. Le modèle proposé par **Hopfield** se rapproche beaucoup du modèle des *verres de Spin de Ising* [TOU, 1994].

Parallèlement aux travaux de **Hopfield**, le chercheur américain **Paul John Werbos** conçoit un mécanisme d'apprentissage pour les réseaux multicouches de type perceptron : c'est l'algorithme d'apprentissage par rétropropagation d'erreurs qui fournit un moyen commode d'entraîner les neurones des couches cachées [DJE, 2004].

En 1986, les psychologues **David Rumelhart**, **Geoffrey Hinton** et **R.J. Williams** publient le livre « *Parallel Distributed Processing* ». Cet ouvrage comprend les développements mathématiques de l'algorithme de la rétropropagation de l'erreur qui permet d'optimiser les paramètres d'un réseau de neurones à plusieurs couches [DJE, 2004].

A partir de ce moment, la recherche sur les réseaux de neurones connaît un essor fulgurant et les applications de ce succès académique suivent au cours des années **90**.

I. MODELE NEUROPHYSIOLOGIQUE

I.1. Structure [TOU, 1994]

Le système nerveux comporte plus de mille milliards de neurones interconnectés qui sont répartis en quelques grandes classes selon leur forme et certaines caractéristiques.

Le neurone biologique est une cellule vivante spécialisée dans le traitement des signaux électriques. C'est l'élément fonctionnel de base du système nerveux. Il est constitué de trois régions principales :

I.1.1. Le corps cellulaire (le soma)

Il contient le noyau du neurone et les autres molécules essentielles à la vie de la cellule. Il effectue les transformations reçues par les dendrites via les synapses. Il est de forme sphérique ou pyramidale. Il mesure quelques microns de diamètre.

I.1.2. Les dendrites

Ce sont de fines extensions tubulaires qui se ramifient autour du neurone en formant une vaste arborescence, appelée chevelure dendritique (ou arborisation dendritique). Elles prolongent le neurone et reçoivent l'information au niveau des points de contact, appelés synapses, avec les autres neurones et avec l'extérieur. Elles mesurent quelques dizaines de microns de longueur.

I.1.3. L'axone (le cylindraxe)

C'est le long de l'axone que les signaux partent du neurone. L'axone est le plus long et se ramifie à son extrémité où il se connecte aux dendrites des autres neurones. Sa longueur varie entre quelques millimètres à plusieurs mètres.

I.1.4. Les synapses

Les synapses permettent la transmission des signaux entre l'axone et les dendrites. Au niveau de la jonction (c'est à dire de la synapse), il existe un espace vide à travers lequel le signal électrique ne peut pas se propager. La transmission se fait alors par l'intermédiaire de substances chimiques appelées les neuromédiateurs. Quand un signal arrive au niveau de la synapse, il provoque l'émission de neuromédiateurs qui vont se fixer sur des récepteurs de

l'autre côté de l'espace inter-synaptique. Quand suffisamment de molécules se sont fixées, un signal électrique est émis de l'autre côté et on a donc une transmission. Suivant le type de la synapse (excitatrice ou inhibitrice), l'activité d'un neurone peut renforcer ou diminuer l'activité des neurones voisins.

La figure **I.1** représente le neurone biologique :

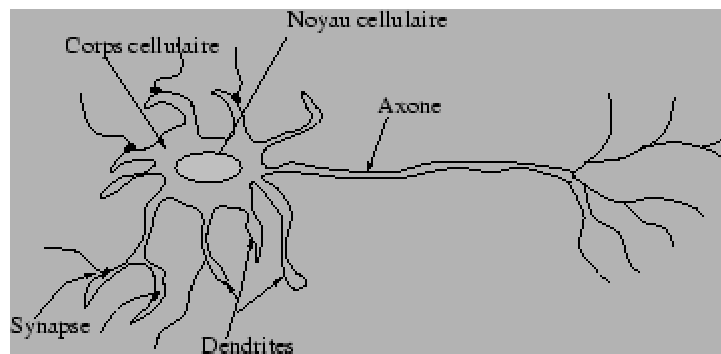


Figure I.1. Neurone biologique.

I.2. Physiologie [ARR, 1996]

La physiologie du neurone est liée aux propriétés de la membrane nerveuse et au métabolisme de la cellule. Chaque neurone est asservi au maintien d'une différence de potentiel membranaire entre le milieu extérieur et le milieu intérieur (le soma) de la cellule.

Le potentiel d'action, appelé également l'influx nerveux, est assimilé à un signal électrique se propageant dans les neurones de la manière suivante :

Les dendrites reçoivent l'influx nerveux d'autres neurones. Le neurone calcule l'ensemble de la stimulation reçue. En fonction de cette stimulation, le neurone transmet, ou non, un signal le long de son axone selon une fréquence fonction du niveau de la stimulation, il sera, ou non, excité. Cette excitation se propage le long de l'axone jusqu'aux autres neurones qui y sont connectés via les synapses.

II. MODELE ARTIFICIEL (OU FORMEL)

II.1. Structure

Le neurone artificiel est le modèle théorique qui imite le fonctionnement du neurone biologique. Il peut être décrit comme un petit automate de décision.

Le premier modèle, conçu par **Warren McCulloch et Walter Pitts** en 1943, est un modèle mathématique très simple dérivé d'une réalité biologique [DEV, 1997]. Le neurone

artificiel est modélisé par deux opérateurs ayant des fonctions distinctes [DAV, 1990] [HER, 1994] [ARR, 1996]:

- Un opérateur de sommation qui évalue l'activation reçue et élabore un potentiel égal à la somme pondérée des entrées de la cellule.
- Un opérateur qui évalue l'état interne et calcule l'état de sortie du neurone en fonction de son activation U .

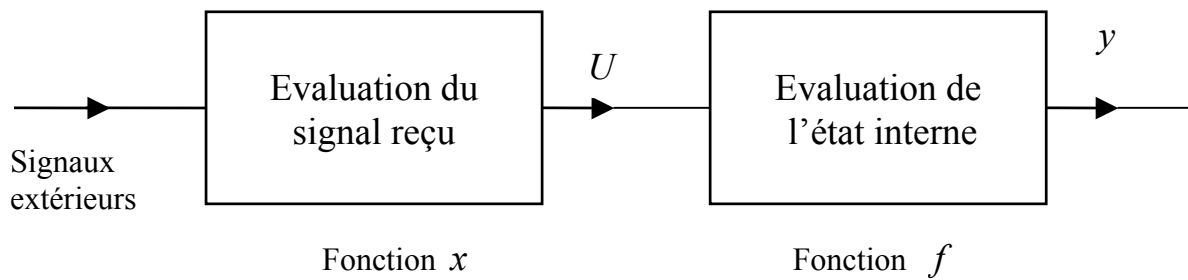


Figure I.2. Synoptique du neurone artificiel.

De manière générale, il existe quatre éléments pour caractériser un neurone:

- Sa fonction d'entrée qui définit le prétraitement effectué sur les signaux d'entrée.
- Le niveau d'activation reçue en entrée U .
- Sa fonction d'activation f .
- Son état de sortie y .

Le neurone artificiel reçoit un nombre variable d'entrées en provenance des neurones efférents (amont). A chacune d'elles, on associe un poids (weight) qui représente la force de la connexion. Chaque neurone est doté d'une sortie unique qui se ramifie ensuite pour alimenter un nombre variable de neurones afférents (aval) [TOU, 1992].

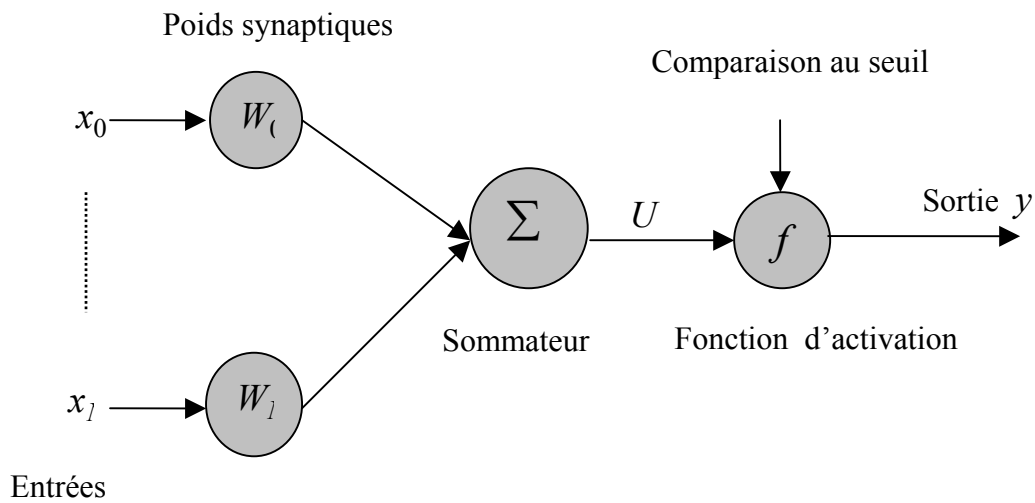


Figure I.3. Neurone artificiel.

Les entrées sont des sorties des neurones du même réseau, ou d'un autre réseau, ou encore des entrées extérieures [HER, 1994].

La figure I.4 illustre la mise en correspondance entre le neurone biologique et le neurone artificiel et le tableau I.1 donne la terminologie correspondante :

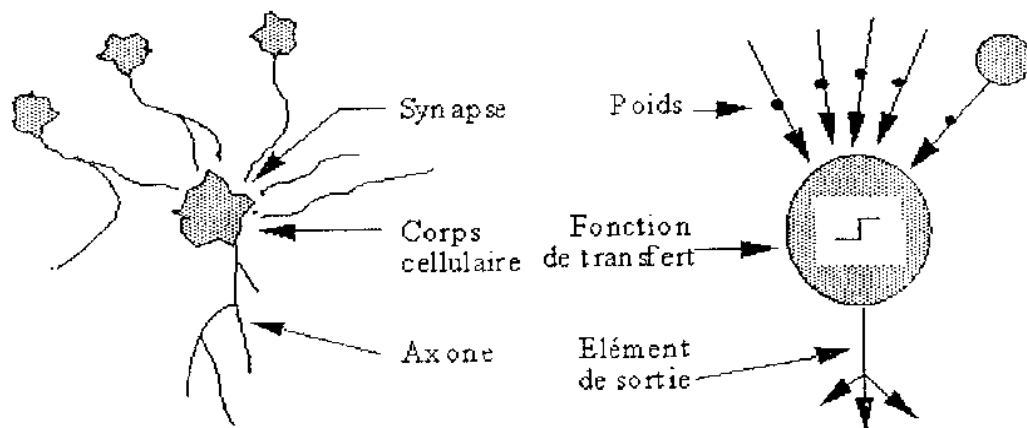


Figure I.4. Mise en correspondance neurone biologique/neurone artificiel.

II.2. Fonction de transfert

L'activité d'un neurone est calculée en fonction de l'état du neurone, de son voisinage et des poids de connexions selon la relation suivante :

(I.1)

L'état de sortie d'un neurone est une fonction de son activité. Elle est exprimée par la relation suivante :

(I.2)

Dans le cas d'un réseau constitué de plusieurs neurones, les relations (I.1) et (I.2) deviennent :

(I.3)

et :

(I.4)

Où : est la fonction d'activation.

La fonction d'activation f , appelée également fonction de transfert ou fonction de transition ou encore fonction neurone, peut prendre plusieurs formes, selon l'utilisation du réseau et selon la nature de l'état du neurone. L'ensemble des états possibles dépend de la fonction d'activation f utilisée [CAR, 1992] [TOU, 1992].

Les fonctions d'activation les plus couramment utilisées dans le cadre des réseaux de neurones sont la fonction identité, la fonction à seuil et la fonction sigmoïde [DEV, 1997]. Ces fonctions sont continues et offrent une infinité de valeurs possibles comprises dans l'intervalle $[0,1]$ ou $[-1,1]$.

II.2.1. Fonction identité :

Les neurones qui possèdent la fonction identité comme une fonction de transfert sont appelés automates linéaires. Cette fonction admet des valeurs non bornées pour les états, ceci peut entraîner des débordements lors des simulations. Ce type de neurone est, souvent, utilisé par **Kohonen** pour construire son modèle de cartes auto organisatrices (voir figure I.5 (a)) [CAR, 1992].

II.2.2. Fonction à seuil :

Les neurones qui possèdent la fonction à seuil comme une fonction de transfert sont appelés automates à seuil. Ce type de neurone est utilisé par **McCulloch** et **Pitts** dans leur modèle d'automate formel (voir figure I.5 (b)) [CAR, 1992].

II.2.3. Fonction sigmoïde :

Cette fonction est inspirée directement de l'examen du comportement des cellules nerveuses face aux signaux qui leur arrivent. Elle est, généralement, employée dans le perceptron multicouche (voir figure I.5 (c)) [CAR, 1992].

Ces fonctions sont représentées par la figure suivante :

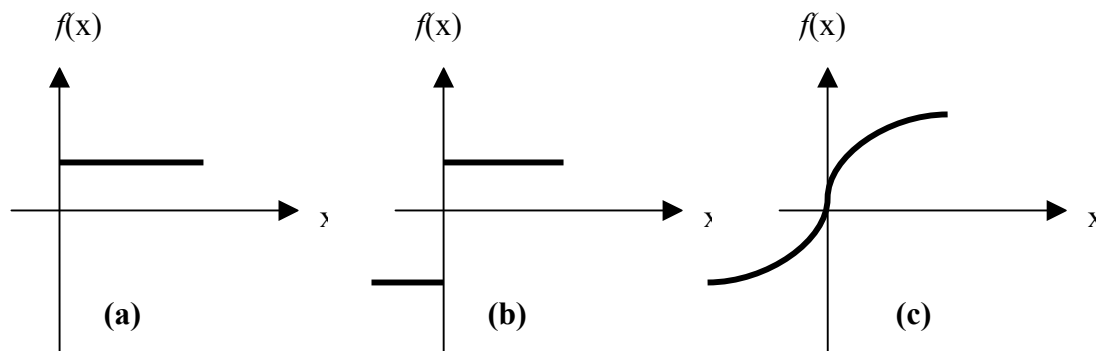


Figure I.5. Différents types de fonction de transfert pour le neurone artificiel.

II.3. Architecture du réseau

Par analogie à la biologie, la connexion représente l'axone et les différentes dendrites qui relient les neurones entre eux. Sa particularité est d'être pondérée. C'est un lien établi explicitement entre deux neurones dans un réseau ayant un sens indiquant le transfert du flux d'informations [DAV, 1990]. En effet, l'état d'un neurone est fonction des états des neurones j qui lui sont connectés. De même, les états des neurones auxquels le neurone i est connecté sont influencés par son état.

Une connexion est associée à une valeur appelée poids de connexions. On définit une matrice de poids de connexions où les lignes et les colonnes correspondent aux neurones du réseau et chaque élément de la matrice représente le poids de la connexion entre le neurone i et le neurone j dans le réseau [CAR, 1992] [TOU, 1992] [ARR, 1996].

Les connexions entre les neurones dans un réseau décrivent la topologie du modèle qui peut être très variée en modifiant seulement les règles de connexions [HER, 1994]. Habituellement, chaque neurone d'une couche est connecté à tous les autres neurones de la même couche et de la couche suivante.

On peut distinguer, selon une certaine régularité, les structures suivantes :

II.3.1. Réseau multicouche

Il n'existe pas de connexions entre les neurones d'une même couche. La connexion se fait seulement avec les neurones des couches suivantes. Ceci permet d'introduire la notion de sens de parcours de l'information et de définir les concepts de neurones d'entrée et de sortie. Cependant, l'ensemble des neurones d'entrée (respectivement de sortie) constitue la couche d'entrée (respectivement de sortie), et les couches cachées sont les couches intermédiaires n'ayant aucune relation avec l'extérieur (voir figure I.6 (a)) [TOU, 1992].

II.3.2. Réseau à connexion locale

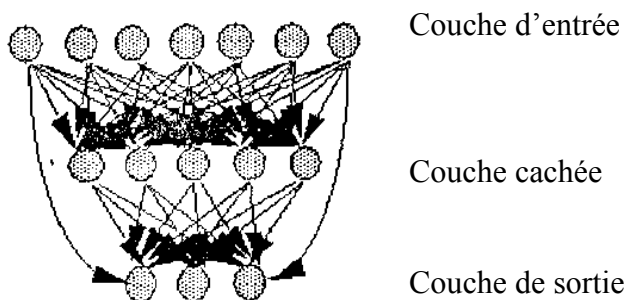
Chaque neurone entretient des relations avec un nombre réduit et localisé de neurones de la couche avale. De ce fait, les connexions sont moins nombreuses que dans le cas précédent (voir figure I.6 (b)) [TOU, 1992].

II.3.3. Réseau à connexion complète

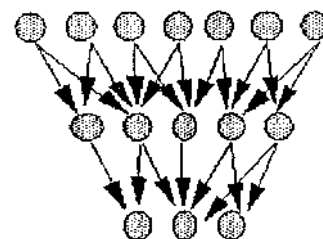
C'est la structure d'interconnexion la plus générale, car chaque neurone est connecté à tous les neurones du réseau et à lui-même (voir figure I.6 (c)) [TOU, 1992].

II.3.4. Réseau à connexions récurrentes

Les connexions récurrentes ramènent l'information en arrière par rapport au sens de propagation défini dans le réseau multicouche. Ces connexions sont le plus souvent locales (voir figure I.6 (d)) [TOU, 1992].



(a)



(b)

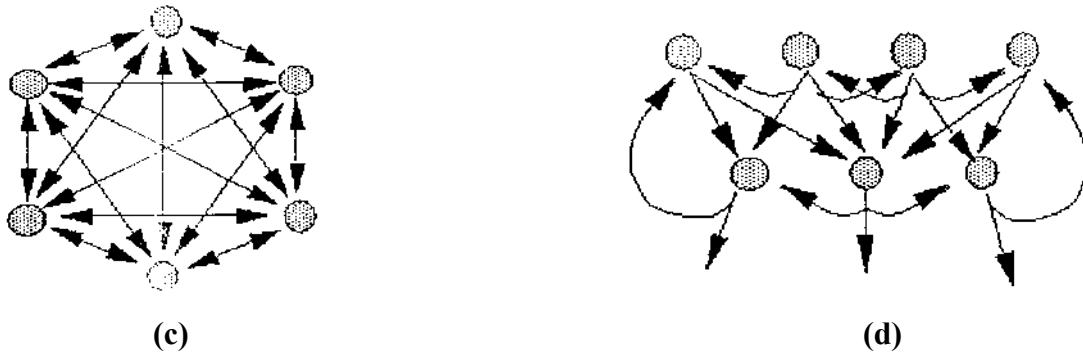


Figure I.6. Structures d'interconnexion

(a) Réseau multicouche - (b) Réseau à connexion locale - (c) Réseau à connexion complète - (d) Réseau à connexions récurrentes

II.4. Concepts d'apprentissage

L'apprentissage est une phase de développement d'un réseau de neurone durant laquelle le comportement du réseau est capable de se modifier à partir du résultat de l'expérience jusqu'à obtention du comportement désiré [TOU, 1992]. Souvent, pendant cette phase, les valeurs modifiées sont celles des poids de connexions entre les neurones dans le but de retrouver une sortie cohérente. Parfois même, il arrive que la structure du réseau soit modifiée telle que le nombre de neurones du réseau.

Les différents algorithmes d'apprentissage sont classés en trois familles selon les critères utilisés pour l'entraînement du réseau et selon l'information fournie à l'algorithme [PER, 1988] [BOW, 1992] [TOU, 1992] :

II.4.1. Apprentissage supervisé

C'est le mode d'apprentissage le plus utilisé actuellement. Son principe consiste à soumettre au réseau un nombre d'exemples pour lesquels l'entrée et la sortie associée sont connues et les poids de connexions sont modifiés afin de minimiser une fonction, appelée fonction coût, qui évalue l'erreur entre la sortie du réseau et la sortie désirée. L'apprentissage supervisé est, historiquement, le premier à avoir été utilisé pour l'entraînement des modèles neuromimétiques en **1957** avec le **perceptron** puis en **1960** avec le modèle d'**Adaline** [PER, 1988] [LAC, 1996]

II.4.2. Apprentissage non supervisé (appelé également **auto organisation**)

Contrairement au premier, seule une base d'entrée est présentée au réseau. Celui-ci doit modifier ses paramètres et déterminer lui-même ses sorties en fonction des similarités détectées entre les différentes entrées, c'est-à-dire en fonction d'une règle d'auto-organisation. Dans ce cas, on tient compte uniquement des informations locales. La règle la plus connue est celle définie par **Donald Hebb** [FRE, 1992]:

« Si deux neurones sont actifs simultanément, alors les poids des connexions les reliant doivent être augmentés. Sinon on diminue ces poids ».

La règle de **Hebb** a donné naissance à de nombreux algorithmes d'apprentissage non supervisé. Le plus simple d'entre eux est basé sur le renforcement des connexions entre les neurones d'activités corrélées [PER, 1988].

II.4.3. Apprentissage semi supervisé (appelé également **apprentissage par renforcement**)

Ce mode d'apprentissage constitue un mode intermédiaire entre les deux premiers. Dans de nombreux problèmes, on ne dispose que d'une information qualitative permettant l'évaluation de la réponse calculée sans connaissance de la réponse la plus adaptée selon certains critères. Ce mode permet d'entraîner le réseau pour qu'il puisse fournir, à chaque entrée présentée, la sortie la plus adéquate [DRE, 2002].

II.5. Règles d'apprentissage

Nous présentons dans ce qui suit quelques règles d'apprentissage :

II.5.1. Règle de Hebb

La règle d'apprentissage la plus commune est celle formulée par **Donald Hebb**. Comme nous l'avons déjà cité auparavant, la règle de **Hebb** consiste à forcer les poids des connexions entre les neurones actifs au même moment. Par contre, les poids seront diminués si les neurones sont dans des états contraires [FRE, 1992] [ARR, 1996].

L'expression mathématique proposée par **Sutton** comme une approximation de cette règle est :

$$W_i(t + 1) = W_i(t) + C_i(t) * X_i(t) * Y(t) \quad (I.5)$$

Où : w_{ij} est l'une des entrées du neurone dont la sortie est o_j .
 η est le taux d'apprentissage.

w_{ij}^{new} est le nouveau poids de connexion.

w_{ij}^{old} est l'ancien poids de connexion.

II.5.2. Règle compétitive

Le principe d'un apprentissage compétitif est, en général, de favoriser le neurone vainqueur. Cela consiste à faire une compétition entre les neurones du réseau afin de déterminer lequel sera actif à un instant donné. Parfois, il arrive qu'un neurone et son voisinage soient vainqueurs, alors ils vont bénéficier d'une adaptation de leurs poids. Les neurones vainqueurs vont agir comme des détecteurs face aux entrées à qui ils doivent leur victoire et seront pénalisés face aux autres entrées. Cependant, là où la règle de **Hebb** a la propriété de trouver automatiquement des corrélations entre les neurones, les règles d'apprentissage compétitives ont plutôt tendance à pousser à une sélection parmi les exemples. Cette capacité de découvrir automatiquement des catégories dans un ensemble de données fait de ces réseaux des techniques efficaces de classification [KOH, 1984] [ARR, 1996].

II.5.3. Règle de Widrow-Hoff (règle Delta)

Cette loi s'applique uniquement aux réseaux à une seule couche cachée. L'apprentissage se fait à partir d'un couple d'entraînement, formé d'un vecteur d'entrée \mathbf{x} et d'un vecteur de sortie \mathbf{y} . Le vecteur d'entrée \mathbf{x} (respectivement de sortie \mathbf{y}) correspond aux valeurs des connexions d'entrée (respectivement de sortie) du réseau [LAN, 1988].

L'apprentissage par la règle Delta comprend quatre étapes :

- Avant la phase d'apprentissage, il faut attribuer une valeur nulle aux poids de connexions.
- Prendre au hasard une paire d'entraînement et faire propager le vecteur d'entrée dans le réseau pour obtenir la réponse calculée par chaque neurone. Cette réponse constitue le vecteur de sortie.
- Pour chaque neurone de sortie, il faut générer la fonction d'erreur e_j par comparaison entre la réponse calculée par le réseau et la réponse désirée.
- Modifier les poids de connexions en fonction des valeurs de l'erreur e_j .

Le processus est réitéré jusqu'à ce que le réseau fournisse une réponse satisfaisante face aux différents vecteurs d'entrée. L'apprentissage se fait en modifiant chaque poids associé à une connexion qui relie un neurone de sortie, de façon à réduire la différence entre la réponse calculée et la sortie désirée. La modification des poids se fait selon l'expression suivante :

$$\partial_i = \alpha(Y_{ci} - Y_{di})X_i \quad (\text{I.6})$$

Où : est le terme d'erreur.

est le taux d'apprentissage, il correspond au temps mis par les neurones pour apprendre. Il est d'autant plus élevé que les risques d'un comportement oscillatoire augmentent [LAN, 1988].

II.5.4. Règle de Delta généralisée

Cette règle est une généralisation de la règle delta dans le sens qu'elle peut s'appliquer à un réseau multicouche. C'est la règle d'apprentissage de l'algorithme de rétropropagation [LAN, 1988] [ZAH, 1992] [MAC, 2003].

II.6. Propriétés des réseaux de neurones

Les réseaux de neurones possèdent quelques propriétés. Nous en citons les principales :

II.6.1. Capacité d'apprentissage

Comme l'être humain, un réseau de neurone est capable d'apprendre à résoudre des problèmes à partir d'exemples [TOU, 1992].

II.6.2. Capacité de généralisation

Cette capacité se résume par le fait que, à partir de quelques exemples d'apprentissage, le réseau peut apprendre à trouver les règles sous-jacentes qui permettent de résoudre le problème [DAV, 1993] [ARR, 1996].

II.6.3. Capacité d'adaptation

Cette capacité se manifeste par la capacité d'apprentissage des réseaux de neurones qui leur permet de tenir compte des nouvelles contraintes ou des nouvelles données, principalement, pour les problèmes évolutifs [DJO, 1999].

II.6.4. Fonction d'oubli

Un réseau de neurones peut apprendre N exemples sans problème et les oublier tous dès lors que l'on souhaite lui en faire apprendre un de plus.

II.6.5. Mémoire associative

La mémoire associative est une des premières applications issues des réseaux de neurones. Il s'agit de réaliser un dispositif capable de mémoriser des informations que l'on peut ensuite retrouver, non pas par leur adresse comme dans le cas d'une mémoire classique, mais en fournissant des données même incomplètes ou bruitées relatives aux informations stockées. Les premiers modèles de mémoires associatives sont ceux de **Widrow** et de **Rosenblatt**, suivis dans les années **70** par les modèles de **Kohonen** et **d'Anderson** notamment [ARR, 1996].

III. MODELES USUELS

Il existe plusieurs modèles utilisant le neurone artificiel comme organe de décision. Parmi les modèles les plus courants, on trouve :

- Le modèle de Kohonen.
- Le perceptron de Rosenblatt.
- Le modèle de Hopfield.

III.1. Modèle de Kohonen :

Les modèles d'auto organisation, proposés dès les années **70**, notamment par **Von der Malsburg**, puis par **T. Kohonen** au début des années **80**, sont inspirés par l'organisation corticale des vertébrés. Ces modèles, actuellement très utilisés, ont pour objectif de représenter des données complexes, souvent bruitées, appartenant généralement à un espace de grande dimension, dans un espace discret dont la topologie est limitée à une ou deux dimensions. Ils ont la propriété de s'adapter à la structure des données qui leur est présentée [HER, 1994].

Ces réseaux peuvent être considérés comme dynamiques dans le sens où ils ont tendance à créer ou à détruire des neurones. De plus, ces réseaux fonctionnent dans leur très grande majorité dans le mode non supervisé [ARR, 1996]. Ils sont constitués de deux couches : une couche d'entrée et une couche compétitive dont les éléments sont disposés sur une grille à deux dimensions [KOH, 1990] [BOW, 1992] [TOU, 1992].

La figure suivante donne l'architecture du réseau de **Kohonen** :

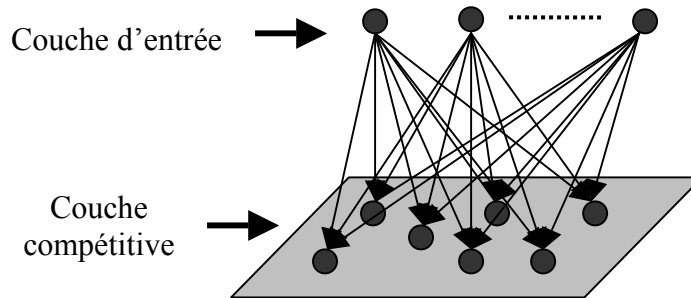


Figure I.7. Architecture du réseau de Kohonen

Les neurones dans la couche compétitive réagissent différemment aux entrées afin d'élire un neurone vainqueur. Cette condition est souvent remplie en choisissant une distribution de poids aléatoire. Nous pouvons soit prendre le premier neurone le plus activé et modifier ses connexions en conséquence, soit imaginer une relation de désactivation d'un neurone vers les autres et d'auto activation sur lui-même [LEM, 1994] [ARR, 1996].

On utilise le réseau en présentant une entrée qui va provoquer des réponses des neurones de la couche suivante. C'est à ce niveau que l'on peut lancer une compétition entre les neurones. Cette compétition aura pour but d'élire le neurone vainqueur.

- Dans un cadre supervisé, on peut considérer le neurone vainqueur le neurone le plus activé mais aussi le neurone qui minimisera l'erreur avec la sortie souhaitée. Ce neurone va subir, par conséquent, une modification de ses connexions [ARR, 1996].
- Dans un cadre non supervisé, on peut imaginer que le neurone le plus activé l'emporte. Il arrive assez souvent que l'on considère le neurone se rapprochant le plus du vecteur d'entrée comme neurone gagnant suivant certaines règles telles que la minimisation de distance [ARR, 1996].

La structure topologique provient d'une mesure de distance définie sur les neurones du réseau. On définit, par suite, le voisinage du neurone. Dans cette structure, les neurones sont liés sur une surface élastique. Il en résulte que le comportement des neurones varie graduellement d'une région à une autre. Selon **Kohonen**, le voisinage du neurone doit être choisi assez large au début de l'apprentissage pour le rétrécir par la suite [PER, 1988] [BOW, 1992] [DAV, 1993].

L'intérêt majeur de ces réseaux est de présenter sur un petit nombre de dimensions la structure présente dans des données de haute dimensionnalité. C'est une méthode d'analyse et de traitement de données performante [ARR, 1996].

III.2. Modèle de Rosenblatt :

Le perceptron de Rosenblatt se compose d'un seul neurone artificiel possédant des capacités décisionnelles très limitées, d'où l'idée d'associer plusieurs perceptrons en réseau connu sous le nom de **perceptron multicouche** (en anglais **Multi Layer Perceptron**) [FRE, 1992].

Ce réseau est constitué d'une couche d'entrée qui transmet les données au réseau, d'une couche de sortie qui fournit le résultat du traitement et d'une ou plusieurs couches cachées qui sont spécifiques au traitement des données. Les neurones sont organisés en couches tels que les neurones d'une même couche sont entièrement connectés avec les neurones de la couche précédente et de la couche suivante [FRE, 1992] [TOU, 1992] [JOD, 1995] [BOR, 2002].

La figure (I.8) donne l'architecture du **perceptron multicouche** :

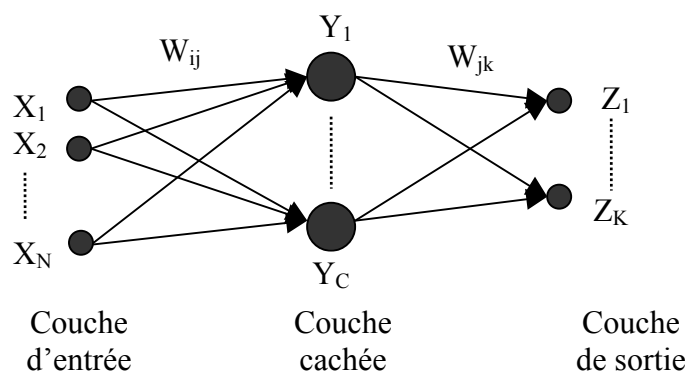


Figure I.8. Architecture du Perceptron multicouche

L'algorithme d'apprentissage de ce réseau, dit algorithme de rétropropagation de gradient, repose sur la minimisation d'une fonction dépendante d'une erreur quadratique entre la sortie fournie par le réseau et la sortie désirée. Cette minimisation est réalisée par une configuration des poids adéquate. Durant la phase d'apprentissage supervisé, un couple (entrée, sortie désirée) est présenté au réseau. Cette phase se déroule comme suit :

L'échantillon à apprendre est présenté aux neurones de la couche d'entrée, il sera propagé vers l'avant à travers le réseau. La réponse des neurones de sortie sera comparée à la réponse désirée. Compte tenu des valeurs des poids de connexions, la sortie obtenue est différente de celle désirée. Cette différence est associée à une erreur quadratique. La minimisation de cette erreur permet la convergence du réseau. Plusieurs itérations sont nécessaires pour atteindre la stabilité du réseau [CAR, 1992] [TOU, 1992] [ARR, 1996] [GRA, 2001].

III.3. Modèle de Hopfield :

Le modèle de **Hopfield** fut présenté en **1982**. Ce modèle est basé d'une part sur le modèle des *verres de spin de Ising* et d'autre part sur le principe des mémoires associatives. D'après les travaux effectués par **Hopfield**, le système nerveux tend à rechercher des états stables, attracteurs dans son espace d'états, et les états voisins, quant à eux, cherchent à se rapprocher d'un état stable, ce qui permet la correction des erreurs et la capacité à compléter des informations manquantes [BRU, 1998]. Pour ces raisons, ce type de modèle est dit « associatif ». Ce modèle a contribué à relancer les recherches sur les réseaux de neurones [ARR, 1996]. La figure (I.9) donne l'architecture du réseau de **Hopfield**:

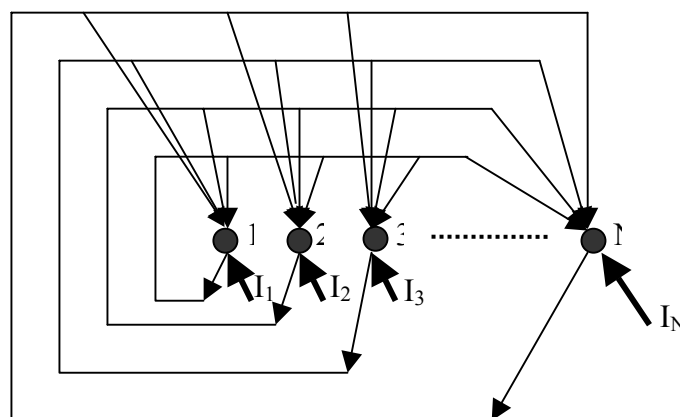


Figure I.9. Architecture du réseau de Hopfield.

Les réseaux de Hopfield sont des réseaux à rétroaction et à connexions symétriques. Les sorties sont fonctions des entrées et du dernier état pris par le réseau. Au cours de la phase d'apprentissage, les valeurs des poids de connexions sont modifiées et ne seront plus modifiées. Par suite, vient la phase de généralisation, durant laquelle on présente le vecteur d'entrée au réseau. Celui-ci calcule la sortie correspondante et la réinjecte à l'entrée. Le processus est réitéré jusqu'à ce que l'entrée à l'étape n et la sortie correspondant à l'entrée à l'étape $n-1$ soient identiques.

Outre les caractéristiques de ce modèle, **Hopfield** a démontré que, lors de l'évolution de son réseau, une fonction de **Lyapunov**, équivalente à la fonction d'énergie de **Ising**, décroît vers un minimum jusqu'à ce que le réseau arrive à un point de convergence. A ce moment là, le réseau devient stable [TSA, 1992].

CONCLUSION

Les réseaux de neurones sont très utilisés de nos jours en traitement de l'image et de l'information. Ceci est dû, principalement, aux capacités multiples que possèdent les réseaux de neurones, entre autres, leur capacité d'apprendre des exemples et leur capacité de s'adapter aux nouvelles contraintes.

Dans ce chapitre, nous avons présenté quelques notions rudimentaires sur les réseaux neuronaux. En premier lieu, nous avons défini l'aspect structural du neurone formel et ses fonctions de base. Par la suite, nous avons présenté les différentes structures topologiques et les différents modes d'apprentissage. Nous avons, également, cité les principales propriétés des réseaux de neurones. A la fin de ce chapitre, nous avons exposé quelques modèles.

Cette étude constitue un outil de base essentiel pour la mise en œuvre de notre travail qui consiste à classifier des images satellitaires en utilisant les réseaux de neurones. Notre choix s'est porté sur le **perceptron multicouche** et le modèle de **Hopfield**. Nous allons développer la mise en œuvre de notre travail dans le troisième chapitre.

Dans le prochain chapitre, nous présenterons quelques méthodes de classification d'images satellitaires en utilisant les approches statistiques, géométriques et génétiques.

Chapitre II

Techniques de classification d'images satellitaires

Résumé

Ce chapitre est dédié particulièrement à la classification des images multispectrales par différentes approches. En premier lieu, nous abordons l'approche statistique dans laquelle nous exposons les méthodes de classification les plus utilisées en traitement d'images, en l'occurrence, celles basées sur le maximum de vraisemblance, le minimum de distance euclidienne, la distance de Mahalanobis et les champs de Markov. Nous enchaînons par l'approche de classification géométrique où nous spécifions les avantages ainsi que les limites d'utilisation de cette approche. Par la suite, nous rappelons brièvement le principe méthodologique de l'approche génétique. Nous exposons, également, l'approche neuronale. Celle-ci sera détaillée et développée dans le prochain chapitre. Nous citons les performances et les limites d'utilisation de chaque méthode. En collaboration avec les travaux effectués dans le laboratoire LTIR, nous donnons les images résultats de la classification par chaque approche.

INTRODUCTION

La télédétection spatiale connaît une évolution très rapide et très puissante sur le plan technologique. Grâce à la haute résolution et la large gamme spectrale des différents capteurs embarqués par les satellites qui enregistrent à différentes échelles les signaux électromagnétiques réfléchis ou émis par les surfaces observées, les données images fournies sont de plus en plus riches en informations et peuvent être utilisées à des fins d'observation et de surveillance de la Terre, de l'atmosphère et de l'espace.

Pour une meilleure exploitation de ces données, il est intéressant de développer des techniques plus élaborées issues du domaine de traitement d'image afin d'investir les avantages des caractéristiques multisources et multitemporelles. Parmi les techniques de traitement d'image s'inscrit la classification. C'est une méthode d'analyse quantitative qui consiste à regrouper dans une même entité ou classe tous les pixels ayant les mêmes propriétés spectrales, spatiales et texturales, tout en respectant leur localisation spatiale, pour en extraire des informations thématiques qui servent à l'interprétation de la nature des surfaces observées.

La classification permet, ainsi, d'attribuer à chaque pixel de l'image une étiquette ou classe parmi un ensemble d'étiquettes qui correspond à l'ensemble des classes présentes dans l'image. Par ailleurs, chaque pixel portera l'information d'appartenance à une classe. L'image obtenue sera formée de nuages de pixels homogènes ayant la même signature spectrale, représentant des classes distinctes correspondant aux différents thèmes au sol. Cette image servira d'appui pour l'élaboration de cartes géographiques et thématiques.

Cependant, le problème posé par la classification des images satellitaires est de trouver la manière la plus efficace qui permet de traduire l'information contenue dans les images multispectrales en un nombre de classes distinctes et de faire correspondre ces classes aux différents thèmes des surfaces observées.

I. PRINCIPE DE LA CLASSIFICATION

La classification d'images est envisagée comme une procédure de décomposition de l'espace image en régions étiquetées sur la base d'une ressemblance de propriétés qui peuvent être [BEL, 1998] :

- Les luminances ou les réflectances mesurées par les capteurs satellitaires (images brutes).
- Les mesures texturales dérivées des traitements et d'analyse de texture des images brutes.

- Les données dérivées des transformations linéaires telles que l'analyse en composante principale, l'analyse canonique, etc...
- Les valeurs des combinaisons linéaires des images telles que les indices de végétation, les indices de couleur, etc....

Une des propriétés qui décrit un pixel de l'image est de le représenter par un vecteur dans un espace image à n dimensions dont les composantes sont les valeurs radiométriques du pixel dans les images multispectrales à classifier. L'ensemble constitue l'image multispectrale, où n représente la dimension spatiale de l'image.

Formellement, les classes sont obtenues par la partition de l'ensemble (image) en sous-ensembles non vides (classes) disjointes et dont la réunion forme l'ensemble [MAN, 1994]. Autrement dit :

Pour tout

(II.1)

L'élaboration d'un classifieur nécessite la connaissance de certaines informations sur l'ensemble S . En revanche, cette information est incomplète et ne peut être disponible que sur un sous-ensemble de sites tel que [BEL, 1998] :

Où : n représente le nombre de classes de l'image.

Le site s constitue l'ensemble des échantillons d'entraînement. Il est utilisé pour établir les règles de décision de classification et généraliser la connaissance sur tout l'ensemble S . La localisation, l'identification et la définition du site s constituent la première étape dans le processus de classification supervisée [BEL, 1998].

II. METHODES DE CLASSIFICATION

Les méthodes de classification des images multispectrales peuvent être regroupées en deux grandes familles. La première, appelée classification supervisée, repose sur une information *a priori* sur les classes importantes (leur nombre, leur signification, leurs caractéristiques statistiques...) issue d'une base de données, ou acquise lors d'une étape d'apprentissage. Par opposition, la seconde, appelée classification non supervisée, ne nécessite pas d'informations *a priori* sur les classes présentes dans l'image.

II.1. Méthodes non supervisées (non dirigées)

La classification non supervisée, appelé également classification automatique, est une technique relativement primitive car elle ne fait appel à aucune connaissance extérieure du problème telle que le nombre de classes, la structure des classes, etc. Les pixels sont classés seulement à partir d'un critère de regroupement et de distance défini dans l'espace spectral. La classification non supervisée fournit une classification contenant un nombre prédéterminé de classes, qui peut être identifié *a posteriori*. Ce type de classification ne fournit pas à l'utilisateur une carte dont les étiquettes sont thématiquement connues à l'avance.

On distingue deux orientations possibles quant aux procédures de classification non supervisée : les méthodes hiérarchiques et les méthodes non hiérarchiques. Ce sont ces dernières qui sont, le plus souvent, appliquées en traitement d'images satellitaires. En ce sens, les données sont regroupées en classes en se fondant sur la ressemblance entre ces données suivant un critère de similarité. Les méthodes hiérarchiques et non hiérarchiques sont développées en annexe pour plus d'informations.

II.2. Méthodes supervisées (dirigées)

Dans la classification supervisée, on dispose d'une connaissance *a priori* des caractéristiques des objets fournie par l'utilisateur. L'intégration de cette connaissance est appelée apprentissage, et consiste à sélectionner dans les images quelques échantillons représentatifs de chaque classe recherchée. Les pixels à classer sont, alors, comparés aux échantillons connus pour fournir une classification contenant les classes présélectionnées lors de l'apprentissage. Généralement, une classe de rejet est utilisée pour regrouper les pixels dont la ressemblance avec les échantillons d'apprentissage est trop faible.

Les étapes de la classification supervisée sont schématisées par la figure suivante:

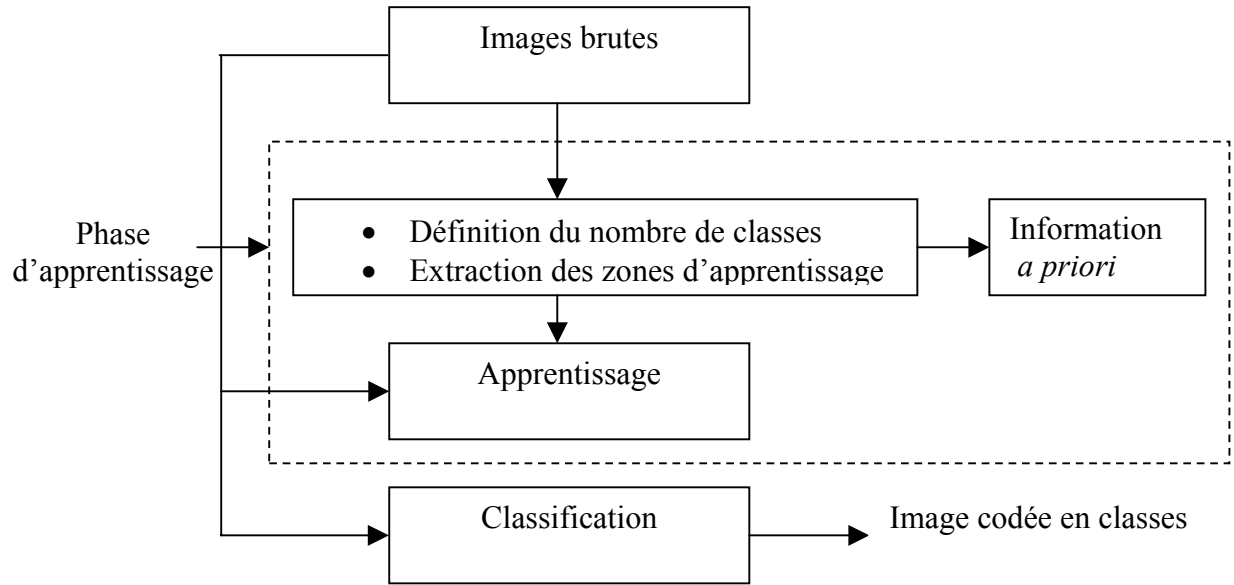


Figure II.1 Etapes de classification supervisée

Les processus de classification supervisée ou non supervisée sont basés sur des modèles mathématiques qui permettent de déterminer la ressemblance entre le pixel à classifier et les signatures des classes.

III. DIFFERENTS TYPES DE CLASSIFICATION

Chacune des méthodes de classification supervisée et non supervisée utilise deux types de classification [CHI, 1994] [BEL, 1998] [KHE, 2001] [CHI, 2003]:

III.1. Classification ponctuelle

Le principe de la classification ponctuelle consiste à décider d'attribuer un pixel de l'image à une classe donnée à laquelle il ressemble le plus suivant un modèle décisionnel. Ce dernier est établi à partir de l'information estimée des classes définies *a priori* en utilisant la valeur radiométrique du pixel, indépendamment des valeurs des pixels de son voisinage.

L'inconvénient majeur de la classification ponctuelle réside dans le fait que l'image résultante est souvent bruitée et que certains pixels sont isolés dans des régions homogènes. L'image résultat présente l'apparence d'un phénomène communément appelé « sel et poivre ». Ce phénomène nuit considérablement à la lisibilité et à l'interprétation de l'image.

III.2. Classification contextuelle

Les méthodes de classification contextuelle se proposent d'être plus robustes dans le sens où elles consistent à optimiser les résultats de la classification ponctuelle afin d'élaborer

une carte thématique plus explicite et plus précise sur l'état de surface. Ces méthodes reposent sur la notion de l'information de texture d'une région ainsi que sur l'information spatiale (contextuelle) apportée à un pixel par un voisinage qui lui est approprié.

Les méthodes de classification contextuelle consistent à modifier l'étiquette de chaque pixel en tenant compte de son voisinage immédiat.

IV. DIFFERENTES TECHNIQUES D'ANALYSE D'IMAGES

Afin de réaliser une cartographie de l'état de surface, plusieurs approches de classification ont été développées dans la littérature. Parmi elles, on distingue :

- ◆ Approche statistique.
- ◆ Approche géométrique.
- ◆ Approche génétique.
- ◆ Approche neuronale.

IV.1. Approche statistique

Le processus de classification est basé sur les approches soit paramétriques, dont la fonction de densité de probabilité décrivant la classe d'objets à estimer a une forme connue, soit non paramétriques, dans lesquelles cette fonction n'est pas connue. Dans ces deux cas, il utilise les caractéristiques statistiques des régions d'intérêt. Dans cette approche, on distingue principalement :

- **Classification par le Maximum de Vraisemblance (MV).**
- **Classification par le Minimum de Distance Euclidienne (MDE).**
- **Classification par la Distance de Mahalanobis (DM).**
- **Classification par les champs de Markov.**

Les trois premières méthodes sont utilisées dans le cadre d'une classification ponctuelle, tandis que la dernière est utilisée dans le cadre d'une classification contextuelle. Nous avons repris les travaux concernant la classification par approche statistique développée au sein de notre laboratoire **LTIR**.

IV.1.1. Classification par le Maximum de Vraisemblance

IV.1.1.1. Principe

Cette méthode est la plus utilisée en traitement d'images. Des recherches considérables ont été focalisées sur cette méthode [FOO, 1992] et sur les facteurs qui influent sur les performances de cette méthode, à savoir le nombre, la taille et la localisation des sites

d'entraînement, la nature des variables discriminantes et l'évaluation de la précision de la classification [CAM, 1981] [ARO, 1982] [ROS, 1986] [FOO, 1988].

La classification d'un pixel de l'image peut être réalisée par la recherche d'une mesure de ressemblance entre ce pixel et les différents paramètres caractéristiques calculés durant la phase d'apprentissage pour chaque classe. Cette mesure est traduite par le calcul de la probabilité d'appartenance du pixel à chaque classe et la décision de lui attribuer la classe qui lui offre la probabilité maximale [BEL, 1998] [KHE, 2001].

Le lien statistique entre la donnée observée et la classe recherchée est donné par le calcul de la probabilité conditionnelle qui s'écrit sous la forme suivante [KHE, 2001] :

(II.2)

Où : μ_c est le vecteur moyenne.

Σ_c est la matrice de covariance de la classe c .

n est la dimension de l'espace des variables.

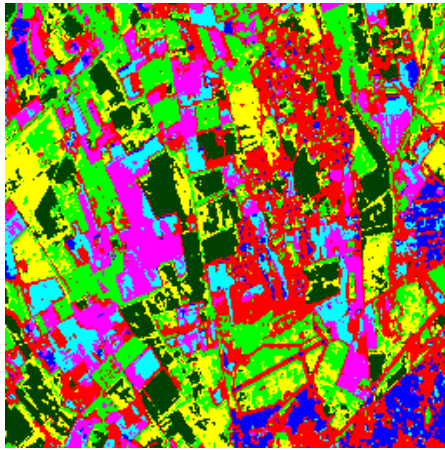
IV.1.1.2. Contrainte

L'efficacité de l'algorithme dépend de la représentativité des paramètres statistiques estimés lors de la phase d'apprentissage pour chaque classe. Dans le cas où le nombre et l'exactitude des zones d'apprentissage sont limités, on a une mauvaise estimation des vecteurs moyennes et des matrices de covariance des classes.

Le tableau II.1 représente les différentes classes thématiques relatives à la scène de Blida et la figure II.2 représente l'image résultat de la classification par le Maximum de Vraisemblance de la scène de Blida [KHE, 2001] :

Classes	Thèmes
■ 1 (rouge)	Zone Urbaine Moins Dense (UMD)
■ 2 (jaune)	Végétation naturelle Moins Dense (VMD)
■ 3 (vert brillant)	Sol Nu, aérodrome de Blida (SN)
■ 4 (bleu turquoise)	Champs Non Cultivés (CNC)
■ 5 (rose)	Zone Urbaine Dense (UD)
■ 6 (bleu)	Champs Cultivés (CC)
■ 7 (vert)	Végétation naturelle Dense (VD)

Tableau II.1. Classes thématiques.



Classes	Thèmes
■ 1	UMD
■ 2	VMD
■ 3	SN
■ 4	NC
■ 5	UD
■ 6	CC
■ 7	VD

Figure II.2. Image résultat de la classification par le Maximum de Vraisemblance.

IV.1.2. Classification par le Minimum de Distance Euclidienne

IV.1.2.1. Principe

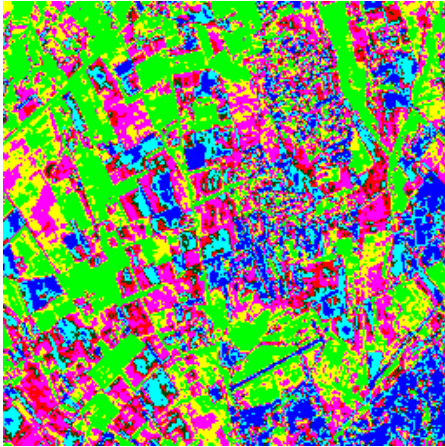
Cette méthode consiste à affecter un point de l'image à la classe la plus proche, suivant les distances par rapport au centre de gravité de chaque classe. Les données d'apprentissage sont utilisées pour estimer le vecteur moyenne de chaque classe. L'estimation est relativement stable par rapport à celle de la matrice de covariance [SWA, 1978]. Ce vecteur est calculé lors de la phase d'apprentissage. La distance euclidienne entre un pixel et une classe est définie dans l'espace vectoriel comme suit [BEL, 1998] [KHE, 2001] :

(II.3)

IV.1.2.2. Contrainte

L'algorithme de classification par le Minimum de Distance Euclidienne est moins sensible quant à la précision des données d'apprentissage. Cependant, si on ne tient pas compte de l'information apportée par les matrices de covariance des classes, cela diminue, considérablement, la discrimination entre les différentes classes.

La figure II.3 représente l'image résultat de la classification par le Minimum de Distance Euclidienne relative à la scène de Blida [KHE, 2001] :



Classes	Thèmes
■ 1	UMD
■ 2	VMD
■ 3	SN
■ 4	NC
■ 5	UD
■ 6	CC
■ 7	VD

Figure II.3. Image résultat de la classification par le Minimum de Distance.

IV.1.3. Classification par la distance de Mahalanobis

IV.1.3.1. Principe

Dans le cas où les classes définies ont des probabilités *a priori* égales, la fonction discriminante au sens du maximum de vraisemblance, à un signe moins près, est une mesure de distance au carré, définie par la relation suivante [BEL, 1998] [KHE, 2001] :

(II.4)

Les fondements de la classification par le maximum de vraisemblance sont similaires à ceux de la classification par le minimum de distance dont la distance est sensible à la dispersion des classes via la matrice de covariance. Sous l'hypothèse que toutes les classes ont la même matrice de covariance, le deuxième terme de l'expression précédente sera éliminé lors de la discrimination. Par ailleurs, la mesure de distance se réduit à :

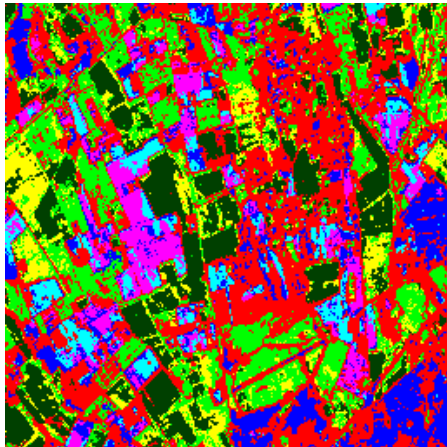
(II.5)

Cette distance est appelée Distance de **Mahalanobis**.

IV.1.3.2. Contrainte

L'algorithme de classification par la Distance de **Mahalanobis** présente l'avantage d'être plus rapide, néanmoins, il est sensible à la dispersion des données des classes via la matrice de covariance.

La figure **II.4** représente l'image résultat de la classification par la Distance de **Mahalanobis** relative à la scène de Blida [KHE, 2001] :



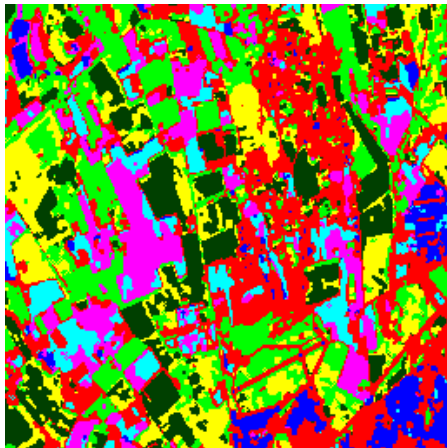
Classes	Thèmes
1	UMD
2	VMD
3	SN
4	NC
5	UD
6	CC
7	VD

Figure II.4. Image résultat de la classification par la Distance de Mahalanobis.

IV.1.4. Classification par les champs aléatoires de Markov

Les fondements de la théorie de classification contextuelle par les champs de Markov ont fait objet de plusieurs recherches en traitement d'images. Notre appui s'est porté sur les travaux effectués au sein de notre laboratoire **LTIR**. A cet effet, nous donnons, seulement, l'image résultat de la classification. Nous orientons les lecteurs aux références bibliographiques [BEL, 1998] [KHE, 2001] [CHI, 2003].

La figure **II.5** représente l'image résultat de la classification par les champs aléatoires de Markov relative à la scène de Blida [KHE, 2001] :



Classes	Thèmes
■ 1	UMD
■ 2	VMD
■ 3	SN
■ 4	NC
■ 5	UD
■ 6	CC
■ 7	VD

Figure II.5. Image résultat de la classification par les champs de Markov.

IV.2. Approche géométrique

A partir des travaux développés dans la littérature concernant la classification par approche statistique, il s'avère qu'une bonne précision des résultats des thèmes, à savoir la correspondance classes-vérité terrain, dépend essentiellement de la localisation et de la sélection des données d'apprentissage. Un problème similaire est rencontré lors de la sélection du nombre, de la dispersion et de la localisation des noyaux de départ dans la classification non supervisée [SWA, 1978] [CHU, 1988] [BEL, 1998] [SMA, 1998].

Plusieurs chercheurs ont proposé des solutions à ces problèmes en développant une nouvelle technique de visualisation, d'analyse et de classification d'images [COG, 1973] [ANU, 1977] [JEN, 1979] [EYT, 1983] [MOR, 1989] [HOD, 1989]. Il s'agit de définir, d'abord, les fonctions de distribution multidimensionnelles, ensuite, de représenter graphiquement les fréquences de ces fonctions. Ceci permet, ainsi, de faciliter la localisation des classes dominantes. Par ailleurs, le processus de classification par approche géométrique est produit par la combinaison de l'approche statistique et de l'information apportée par la représentation graphique des fréquences de chaque point.

Le principe de cette approche est basé sur des mesures de distances séparant deux observations dans l'espace des variables telles que la détermination des distances radiométriques entre un point de l'image et deux points de référence dans un modèle de données à n dimensions. Par la suite, une fonction de distribution est définie pour toute l'image multispectrale. La particularité de cette procédure est qu'on ne dispose d'aucun moyen de test démontrant que les données sont classées avec un taux d'erreur minimale, contrairement aux procédures statistiques [BEL, 1998].

En outre, cette approche traite les deux modes de classification, supervisé et non supervisé de l'espace des données en utilisant les fonctions de densité de probabilité à n dimensions [BEL, 1998].

Remarque

Pour cette approche, nous ne disposons pas d'image résultat relative à l'image test. Cette approche a été testée en utilisant des images simulées et a montré de bonnes performances dans le cas d'images de faibles tailles [BEL, 1998].

IV.3. Approche génétique

Les principes fondamentaux des algorithmes génétiques ont été développés par **John Holland** [HOL, 1975]. Ces algorithmes s'inspirent du fonctionnement de l'évolution naturelle, notamment la sélection de Darwin, et la procréation selon les règles de Mendel [CHI, 2003].

Les algorithmes génétiques fournissent des solutions à un problème n'ayant pas de solutions analytiques ou algorithmiques. Selon ce principe, plusieurs solutions, appelées génotypes, plus ou moins bonnes, sont créées au hasard, puis sont soumises à une imitation de l'évolution des espèces, afin que les plus adaptées survivent plus que les autres. De ce fait, la population évolue par générations successives. Un mécanisme d'encouragement des éléments les plus aptes assure que les générations successives sont de plus en plus adaptées à la résolution d'un problème [CHI, 2003].

En effet, la sélection naturelle que Darwin appelle l'élément propulseur de l'évolution, favorise les individus d'une population qui sont les plus adaptés à l'environnement. La sélection est suivie de la procréation, réalisée à l'aide de croisements et de mutations au niveau du patrimoine génétique des individus, constitué d'un ensemble de gènes. Ainsi, deux individus parents, qui se croisent, transmettent une partie de leur gène à leurs descendants. Le génotype de l'enfant fait que celui-ci est plus ou moins bien adapté à l'environnement. S'il est bien adapté, il a une plus grande chance de procréer dans la génération future. Au fur et à mesure des générations, on sélectionne les individus les plus adaptés, et l'augmentation du nombre d'individus bien adaptés fait évoluer la population entière [CHI, 2003].

Le processus de classification par les algorithmes génétiques est constitué de deux phases principales :

- Phase d'apprentissage.
- Phase de classification.

1. Phase d'apprentissage

Durant cette phase, on extrait, d'abord de chaque classe, un chromosome qui la caractérise le mieux. Ce dernier est un vecteur de taille variable qui pointe vers un tableau contenant les positions des pixels dans l'image support. Par la suite, on procède à la localisation et la sélection des sites d'entraînement. Cette étape nécessite l'utilisation des données de référence. On fait correspondre les points de l'image à traiter avec les données de référence. Les sites d'entraînement doivent être représentatifs de toute la variabilité du thème observé sur l'image. Si deux régions distinctes spatialement représentent la même classe, il faudrait extraire une ou plusieurs parcelles de chaque région, ce qui traduit le mieux la variabilité de la classe. L'étape suivante est l'extraction des échantillons. C'est une étape très importante car la variabilité des données d'échantillons influe sur la précision de la classification. Cependant, l'intégration des opérations de prétraitement peut être nécessaire pour l'amélioration de l'information de base durant cette phase.

Pour chaque classe, il est nécessaire de rassembler et caractériser l'ensemble des pixels situés dans les zones d'apprentissage sélectionnées. Ceci permet de sélectionner les bons échantillons représentatifs de la classe, ainsi une base de données sera générée.

Après codage et génération de la population initiale d'une classe, on utilise les opérateurs génétiques, à savoir, le croisement et la mutation, pour faire évoluer la population initiale et d'en extraire le meilleur chromosome de chaque classe.

2. Phase de classification

Une fois que les bons chromosomes caractérisant les classes sont obtenus, vient la phase de classification, pendant laquelle on fait appel, une seconde fois, aux opérateurs génétiques pour affecter chaque pixel à la classe appropriée.

La procédure de classification est schématisée par les étapes suivantes :

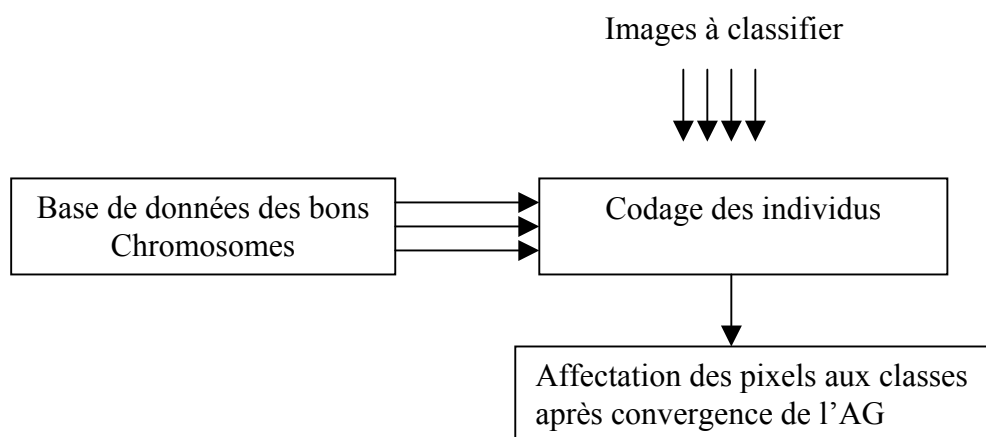


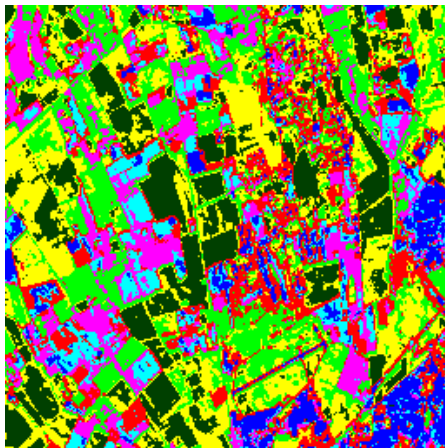
Figure II.6 Procédure de la classification.

Les paramètres choisis pour la classification par les Algorithmes Génétiques sont résumés par le tableau suivant :

Paramètres	Phase d'apprentissage	Phase de classification
Taille de la population	50	5
Probabilité de croisement	0,5	0,03
Probabilité de mutation	0,8	0,5
Nombre de génération	50	10
Taux d'exploration	100%	100%

Tableau II.2. Paramètres relatifs aux Algorithmes Génétiques.

La figure II.7 représente l'image résultat de la classification par les algorithmes génétiques relative à la scène de Blida [KHE, 2003] :



Classes	Thèmes
■ 1	UMD
■ 2	VMD
■ 3	SN
■ 4	NC
■ 5	UD
■ 6	CC
■ 7	VD

Figure II.7 Image résultat de la classification par les Algorithmes Génétiques.

IV.4. Approche neuronale

La classification d'images par approche neuronale fait objet de notre travail. Cette approche sera détaillée dans le prochain chapitre. Notons, cependant, que la procédure de classification par les réseaux de neurones se fait en deux étapes :

La première étape est la phase d'apprentissage durant laquelle on effectue le calcul des poids de connexions entre les neurones du réseau en minimisant l'erreur entre la sortie désirée et la sortie calculée par le réseau.

La deuxième étape est la phase de reconnaissance du réseau durant laquelle le réseau est prêt à évaluer la classification de la scène traitée.

La figure II.9 illustre le synoptique de la classification par les différentes approches développées auparavant :

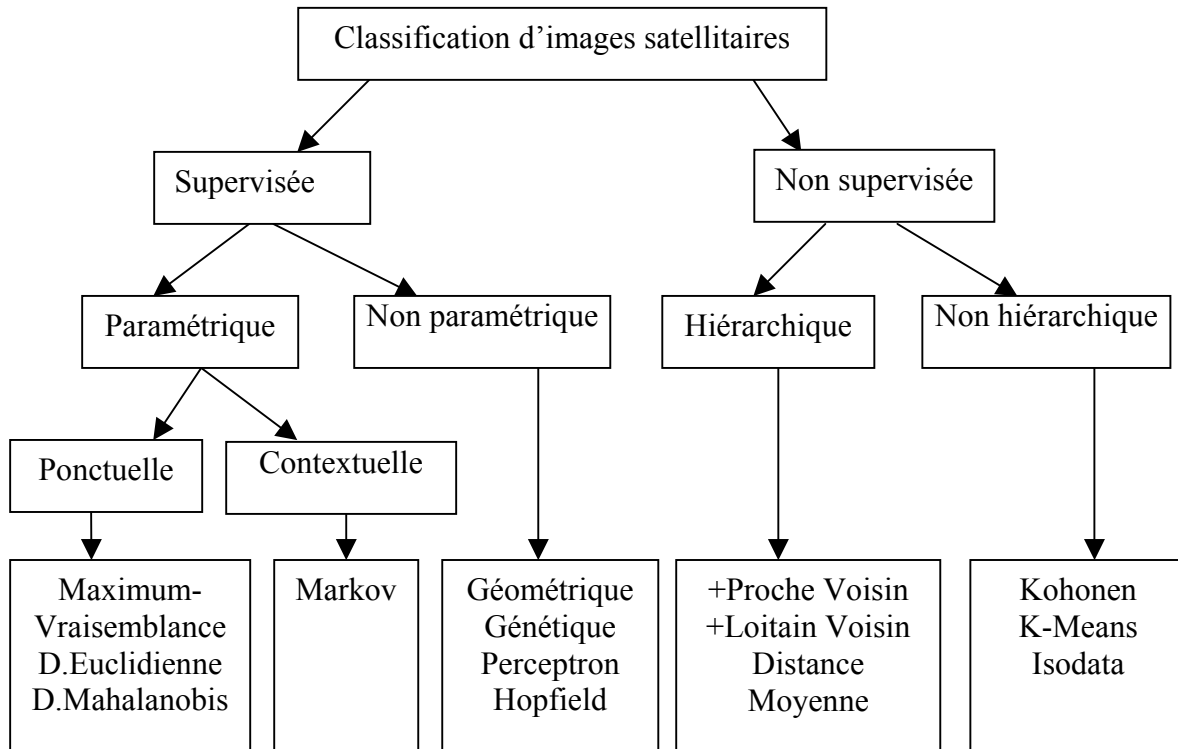


Figure II.9. Différentes approches de classification

CONCLUSION

Les techniques de classification, utilisant les approches statistiques, nécessitent des hypothèses puissantes sur les données et font appel à des modélisations particulières telles que la distribution gaussienne sur l'ensemble des classes. En revanche, ces approches sont gourmandes en calculs et en données, néanmoins, elles aboutissent à des résultats performants. De même, les techniques de classification par approche génétique sont coûteuses en temps de calcul, et sont difficiles à mettre en œuvre du fait que la détermination des paramètres, en l'occurrence, la taille de la population et la fonction d'évaluation, permettant la convergence du processus vers la solution optimale, est une tâche ardue. Hormis l'inconvénient de temps de calculs longs et pénibles, les techniques de classification par approche géométrique ne sont appliquées qu'aux images de faibles dimensions. Par ailleurs, ces techniques, dont la performance est très limitée, ne sont pas très utilisées en traitement d'images surtout pour des images de dimensions élevées. En revanche, la technique de classification par approche neuronale qui est utilisée, actuellement, avec succès, possède une

grande capacité de généralisation et d'adaptation à l'application sans pour autant nécessiter une modélisation du problème.

Nous allons développer deux modèles de réseaux de neurones, dans le prochain chapitre, le premier étant le **perceptron multicouche** et le second étant le réseau de **Hopfield**. Nous déterminerons les facteurs essentiels qui influent sur les performances de chaque modèle.

Chapitre III

Mise en œuvre du processus de classification par approche neuronale

Résumé

*Nous présentons dans ce chapitre la méthodologie de développement des réseaux de neurones pour élaborer le processus de classification. En premier lieu, nous avons utilisé le modèle de **perceptron multicouche** entraîné par l'algorithme de rétropropagation lors de la phase d'apprentissage. En deuxième lieu, nous proposons la mise en œuvre du réseau de **Hopfield**. Celui-ci nécessite une phase d'apprentissage différente de la première. En effet, pendant la phase d'apprentissage, le réseau évolue en se rapprochant de son état stable quand une fonction d'énergie est minimisée. Pour notre cas, cette énergie correspond à la distance minimale entre le pixel à classer et le centre de gravité de chaque classe. Le processus est réitéré jusqu'à la convergence du réseau. Ces deux modèles nécessitent une connaissance a priori sur les données de l'image à classer.*

INTRODUCTION

La technique de classification d'images multispectrales par approche neuronale se révèle être une méthode récente, très utilisée en traitement d'images. Actuellement, de nombreuses recherches ont été effectuées dans ce domaine en se focalisant sur la manière d'adapter un modèle à l'application. De ce fait, plusieurs facteurs permettent d'agir sur le modèle pour le rendre adéquat, tels que l'architecture du modèle, les poids de connexions, les valeurs des vecteurs d'entrée, etc. La plupart des réseaux de neurones font appel à des règles d'apprentissage sur les données afin d'ajuster les poids de connexions. Généralement, ils sont élaborés à partir d'un ensemble d'exemples, ils présentent, ensuite, une capacité de généralisation pour des données non présentes dans la base d'apprentissage.

La classification des images multispectrales en utilisant les réseaux de neurones se fait en deux étapes :

- La première, étant la phase d'apprentissage, durant laquelle on effectue le calcul des poids de connexions en minimisant l'erreur quadratique entre la sortie calculée par le réseau et la sortie désirée.
- La deuxième, étant la phase de reconnaissance et de généralisation, durant laquelle on évalue le processus de classification.

I. CLASSIFICATION PAR LE PERCEPTRON MULTICOUCHE

Nous allons développer, dans cette section, la méthodologie de classification par le **perceptron multicouche**. Ce réseau, souvent utilisé en traitement d'images, a montré ces performances quant à la classification d'images satellitaires.

I.1. Description architecturale

Sur le plan architectural, le **perceptron multicouche** est un réseau neuronal dont l'élément de base est le neurone de traitement qui correspond au neurone artificiel. Les neurones sont organisés de manière à ce que les neurones d'une même couche soient indépendants et que tout neurone d'une couche donnée soit connecté à tous les neurones de la couche précédente et de la couche suivante.

Dans le cadre de la classification, les couches du **perceptron multicouche** sont organisées comme suit [HER, 1994] :

- La couche d'entrée qui reçoit le vecteur d'entrée de dimension n composé des valeurs radiométriques de l'échantillon correspondant à chaque bande spectrale. Le nombre de neurones, dans cette couche, correspond au nombre de canaux spectraux.

- Les couches cachées qui sont spécifiques au traitement. Le nombre de neurones de cette couche peut varier.
- La couche de sortie qui fournit le résultat du traitement. Cette couche contient autant de neurones que le nombre de classes présentes dans l'image.

Notation

Pour mettre en œuvre la méthodologie de classification par le **perceptron multicouche**, nous adopterons la notation suivante :

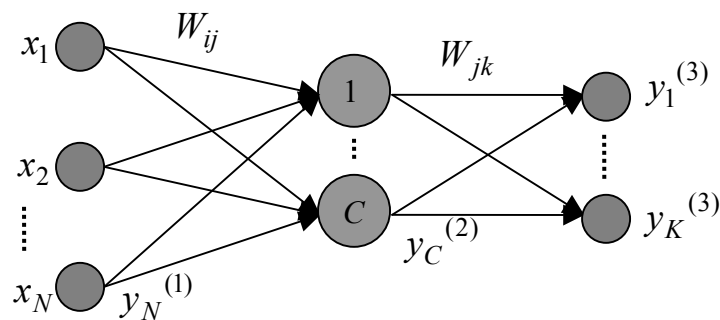


Figure III.1. Représentation schématique du perceptron.

Pour :

étant le nombre de canaux spectraux.

étant le nombre de neurones cachées.

étant le nombre de classes présentes dans l'image.

correspond aux valeurs des composantes du vecteur d'entrée .

correspond aux valeurs des composantes du vecteur de sortie de la couche .

correspond aux valeurs des poids de la couche d'entrée vers la couche cachée.

correspond aux valeurs des poids de la couche cachée vers la couche de sortie.

Les sorties du neurone de la couche cachée, et du neurone de la couche de sortie, sont données, respectivement, par les relations suivantes :

(III.1)

et :

(III.2)

Où : est la fonction de transfert.

I.2. Phase d'apprentissage :

L'algorithme de rétropropagation est la méthode d'apprentissage la mieux connue et l'une des plus efficaces pour les réseaux de neurones multicouches [HER, 1994]. C'est une méthode qui nécessite un mode supervisé car on fournit au réseau, à la fois, le couple « entrée-sortie désirée ». Cette méthode permet d'ajuster les poids de connexions en commençant par les dernières couches jusqu'aux premières couches. Elle se déroule comme suit :

A chaque itération, un ensemble des couples « entrée-sortie » est présenté au réseau. L'algorithme apprend cet ensemble en utilisant un cycle de deux phases : propagation et adaptation [WEI, 1999].

L'échantillon à apprendre active les neurones de la couche d'entrée. Il sera propagé vers l'avant à travers chaque couche du réseau. La réponse des neurones de sortie est, ainsi, générée et sera comparée aux réponses désirées. Compte tenu des valeurs des poids de connexions, la sortie calculée par le réseau est différente de celle désirée. Cette différence est associée à l'erreur quadratique, notée . Elle sera calculée pour chaque neurone de la couche de sortie. Cette erreur sera, donc, transmise en sens inverse de la couche de sortie vers chaque neurone de la couche intermédiaire.

Cette procédure est répétée couche par couche jusqu'à ce que chaque neurone dans le réseau ait reçu un signal d'erreur qui décrive sa contribution relative à l'erreur générée. Les poids de connexion sont, donc, mis à jour pour faire converger le réseau vers un état stable.

I.3. Mise en œuvre :

Pour la mise en œuvre du **perceptron multicouche**, la démarche à suivre est composée, principalement, de quatre étapes :

Première étape : Fixer le nombre de couches cachées.

En l'absence de couches cachées, le réseau n'offre que de faibles possibilités de convergence et donc d'adaptation. Par contre, en présence de couches cachées, le réseau devient capable, avec un nombre suffisant de neurones, d'approximer toute fonction continue.

➤ **Deuxième étape : Déterminer le nombre de neurones dans la couche cachée.**

Chaque neurone supplémentaire permet de prendre en compte des profils spécifiques des neurones d'entrée. Néanmoins, il existe un compromis quant au nombre important des neurones de la couche intermédiaire qui, d'une part, permet de mieux s'adapter aux données présentées au réseau, et d'autre part, amoindrit la capacité de généralisation du réseau.

➤ **Troisième étape : Choisir la fonction d'activation.**

Pour le passage d'une couche à une couche, la fonction sigmoïde bipolaire est souvent choisie. Le choix de cette fonction est dicté par le fait que la fonction d'activation de l'algorithme de rétropropagation doit être continue et croissante car elle conduit à une convergence mieux que celle fournie par les autres fonctions.

➤ **Quatrième étape : Choisir le critère d'arrêt.**

Lors de la phase d'apprentissage, les paramètres d'ajustement des poids de connexions sont déterminés à chaque itération. Un critère d'arrêt doit être défini pour que la convergence du réseau ne passe pas par des minimums locaux.

L'algorithme de rétropropagation repose sur la minimisation de l'erreur quadratique entre les sorties calculées par le réseau et les sorties désirées [WEI, 1999]. Le terme rétropropagation du gradient provient du fait que l'erreur calculée en sortie est transmise en sens inverse vers l'entrée.

L'erreur quadratique commise sur le neurone de sortie est donnée par la relation suivante :

(III.3)

Par conséquent, l'erreur globale à minimiser est exprimée par :

(III.4)

Pour minimiser l'erreur E , on calcule le gradient de cette erreur par rapport à chaque poids W_{ik} , puis on modifie les poids dans le sens inverse du gradient.

➤ **Mise à jour des poids de connexions**

Chaque vecteur d'entrée est associé à un vecteur de sortie désirée Y . Puisque les poids ont des valeurs aléatoires, le vecteur de sortie calculée \hat{Y} est *a priori* différent du vecteur Y . La différence est associée à l'erreur quadratique moyenne donnée par l'expression (III.6). Le gradient de cette erreur par rapport au poids W_{ik} est donné par:

(III.5)

La dérivée partielle du membre de droite est nulle pour tout W_{ik} , car seule la sortie est fonction du poids W_{ik} . De plus, la sortie désirée ne dépend d'aucun poids. Nous obtenons ainsi :

(III.6)

En posant : $\delta_k = -\frac{\partial E}{\partial \hat{Y}_k}$, le gradient de l'erreur quadratique moyenne E s'écrit comme suit :

(III.7)

La mise à jour des poids, selon le principe du gradient, s'écrit :

(III.8)

Où : η : étant un gain d'adaptation (ou pas d'apprentissage).

Ainsi, la règle générale de modification des poids de la couche de sortie à la $n^{\text{ième}}$ itération est donnée par :

(III.9)

Calculons le gradient de cette erreur par rapport au poids w_{hk} associé à la connexion venant du neurone h de la couche 1 vers le neurone k de la couche 2, nous aurons :

(III.10)

Le premier terme du membre de droite peut être exprimé par l'expression **(III.11)** :

(III.11)

et le second terme est donné par l'expression suivante :

(III.12)

Nous remarquons que le seul terme qui dépend de w_{hk} est la sortie z_k obtenue pour $j = k$. Ainsi, l'équation **(III.12)** devient :

(III.13)

Le seul terme de la somme sur m dont la dérivée non nulle par rapport à w_{hk} est le terme obtenu pour $m = h$. On écrit alors :

(III.14)

Le gradient de l'erreur E s'écrit donc :

(III.15)

On pose :

(III.16)

L'équation **(III.15)** devient :

(III.17)

On pose aussi :

(III.18)

L'équation **(III.18)** montre que l'erreur à la sortie du neurone k de la couche (2) est égale à la somme des erreurs des neurones de la couche suivante (3). On déduit, enfin, que :

(III.19)

D'une manière générale, dans n'importe quelle couche cachée du réseau, l'incrément d'un poids $w_{ij}^{(j)}$ d'une couche j est le produit de l'entrée $x_i^{(j-1)}$ provenant de la couche précédente j-1 par l'erreur attribuée au neurone k. L'erreur $\delta_k^{(j)}$ du neurone k est la somme des erreurs $\delta_i^{(j+1)}$ associées aux neurones i de la couche suivante j+1 pondérées par les poids synaptiques $w_{ki}^{(j)}$ qui connectent les neurones i avec le neurone k. On a alors :

(III. 20)

Où l'erreur vaut :

pour les neurones de la dernière couche.

pour les neurones d'une couche interne.

La règle d'adaptation des poids s'écrit finalement à l'itération $t + 1$:

(III.21)

La forme d'adaptation des poids donnée dans l'équation **(III.21)** est rarement utilisée telle en pratique. Le fait que l'algorithme soit fondé sur la minimisation d'une erreur instantanée et non pas d'une erreur calculée sur toute la base d'apprentissage, rend la convergence assez lente car la modification des poids à chaque mise à jour n'est pas importante. De plus, lors du choix du pas d'apprentissage, il existe un compromis entre la vitesse de convergence et les instabilités (ou les minimums locaux). C'est pourquoi il est préférable d'utiliser une valeur assez petite du pas d'apprentissage pour que les oscillations de l'erreur soient négligeables. En revanche, pour augmenter la vitesse de convergence du réseau, un facteur de filtrage sur les incréments d'adaptation connu sous le nom momentum est introduit. Ce facteur apporte une pondération exponentielle sur l'erreur comme l'indique l'équation suivante :

(III.22)

La règle d'adaptation finale est donnée par :

(III.23)

Où: μ est le momentum.

α est le pas d'apprentissage.

➤ **Lissage de la règle d'adaptation**

L'algorithme de rétropropagation de gradient est fondé sur la minimisation d'une erreur locale et non pas sur une erreur calculée sur toute la base d'apprentissage. Par ailleurs, l'introduction du pas d'apprentissage α , appelé également pas d'adaptation, est nécessaire pour éviter l'instabilité. La valeur de ce paramètre a un effet très significatif quant aux performances du réseau. Généralement, elle est comprise entre **0** et **1** [HER, 1994]. Si cette valeur est choisie trop petite, l'algorithme convergera lentement. En revanche, si elle est choisie trop grande, l'algorithme générera des oscillations. Par ailleurs, le choix de la valeur du taux est d'une importance cruciale pour assurer la convergence de l'algorithme vers une solution optimale.

Un autre paramètre est rajouté dans les expressions précédentes. Il s'agit du momentum μ qui permet d'accélérer l'apprentissage, et donc, la convergence de l'algorithme. Il permet, en outre, de sortir des minimums locaux. En pratique, la valeur du momentum est inférieure et voisine de **1** [HER, 1994].

I.4. Organigramme

Le processus d'apprentissage par le **perceptron multicouche** est illustré par l'organigramme de la figure **III.2**, tandis que la phase de classification, elle est développée par l'organigramme de la figure **III.3** :

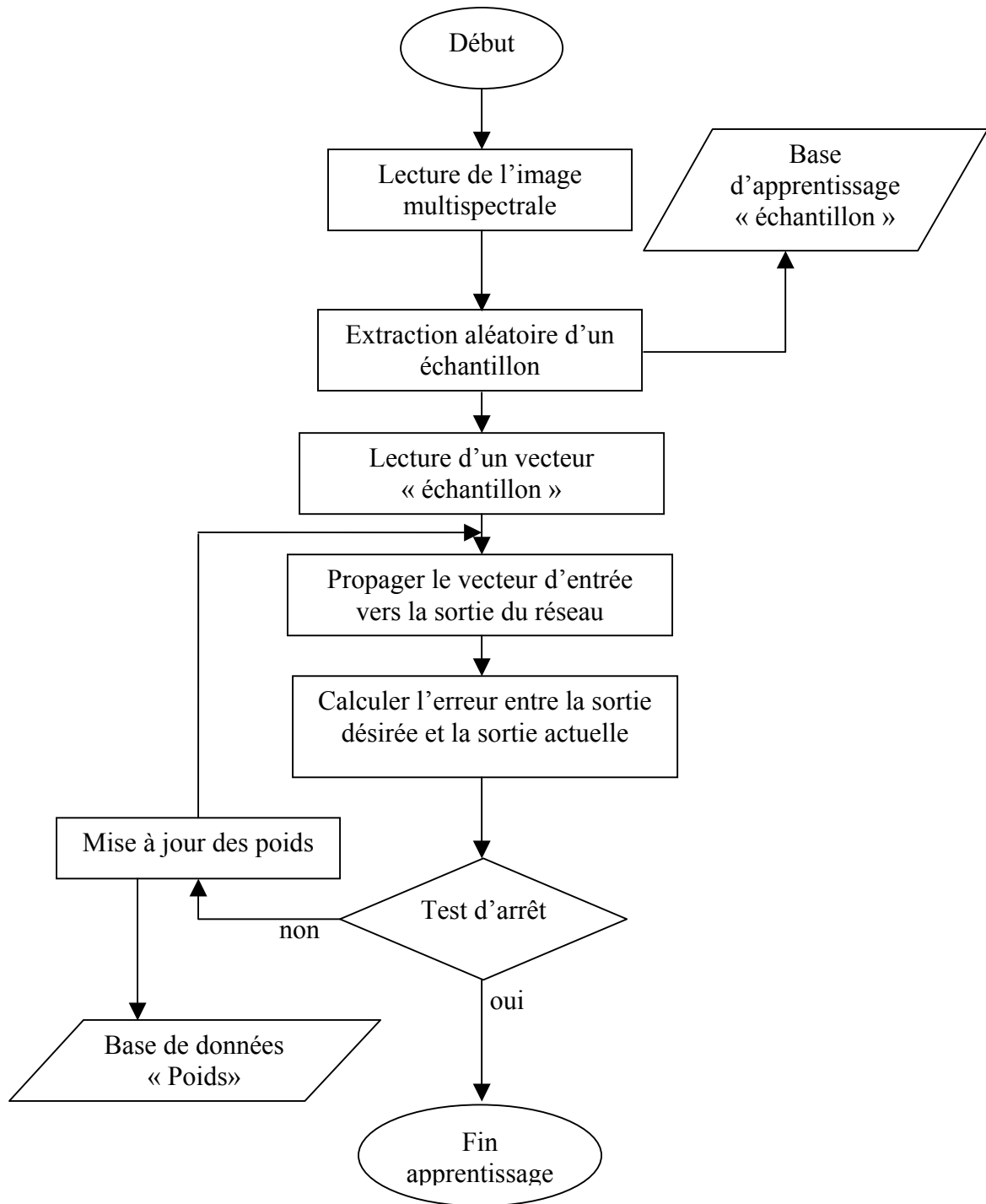


Figure III.2. Organigramme de la phase d'apprentissage par le perceptron.

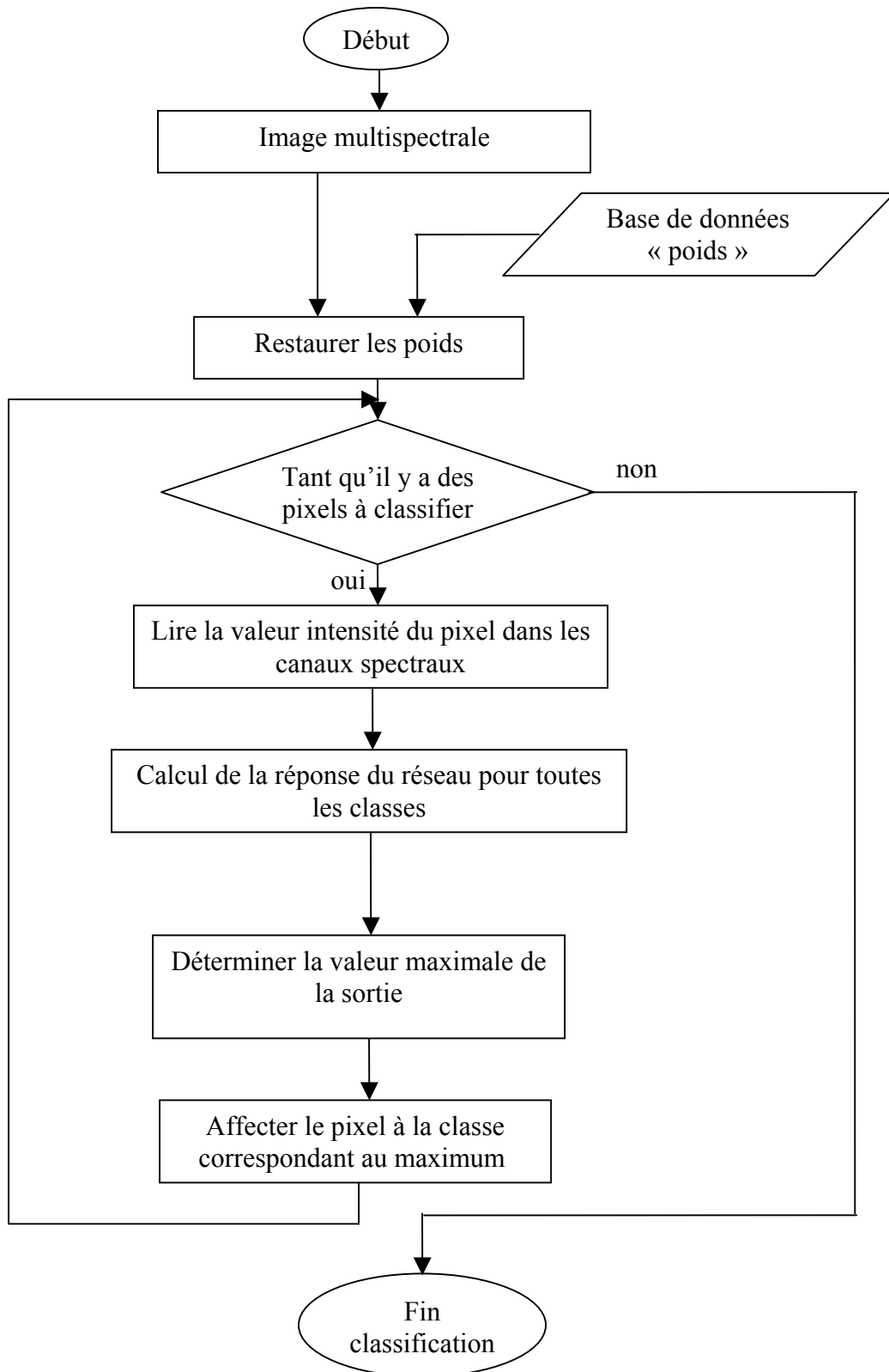


Figure III.3. Organigramme de la phase de classification par le perceptron.

II. CLASSIFICATION PAR LE RESEAU DE HOPFIELD

II.1. Description architecturale

Le réseau de **Hopfield** est relativement différent des modèles à couches du fait que les neurones d'entrée et de sortie sont vus de la même manière. En plus, tous les neurones du réseau sont entièrement connectés et les poids de connexions entre ces neurones forment une matrice de poids W . La matrice des poids W d'un tel réseau est symétrique et les éléments de la diagonale sont nuls [HER, 1994]. Autrement dit :

et pour tout :

Les données sont appliquées simultanément à tous les neurones du réseau. Ces neurones possèdent un fonctionnement dynamique [MAH, 2000] c'est-à-dire qu'après avoir appliqué une nouvelle entrée, la sortie calculée par le réseau est réintroduite à l'entrée. Celle-ci, étant modifiée, permet de recalculer une nouvelle sortie. Le réseau réitère, alors, par une succession d'états qui réduisent les variations de la sortie jusqu'à ce qu'il converge vers un état stable. Cet état sera atteint lorsque, au cours du temps, il n'y a aucun changement d'aucun de ces neurones, et donc, la sortie calculée est constante.

La figure III.4 donne une représentation schématique du réseau de **Hopfield** :

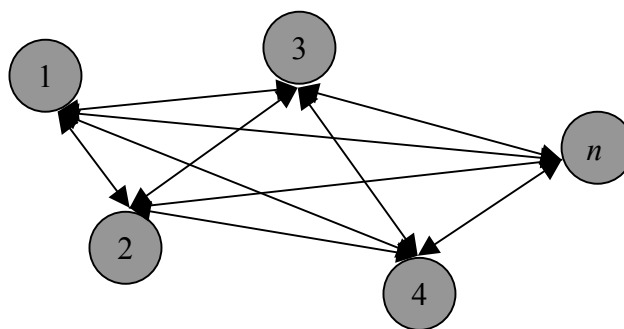


Figure III.4. Représentation schématique du réseau de Hopfield.

L'entrée totale du neurone provient de deux sources : des entrées externes et des entrées des autres neurones. Elle est donnée par :

(III.9)

Où : est le poids de connexions entre les neurones et .

est l'état du neurone .

est une entrée externe.

II.2. Phase d'apprentissage

Dans les réseaux de **Hopfield**, l'apprentissage est souvent qualifié d'apprentissage par cœur [MUH, 2004] par opposition aux algorithmes d'apprentissage par essai ou par erreur. Les neurones sont interrogés de manière asynchrone.

La mise à jour des sorties des neurones est donnée par les règles suivantes :

(III.10)

et :

(III.11)

Où : est le nouvel état du neurone du neurone.

est la fonction d'activation du $i^{ième}$ neurone.

est une entrée externe au $i^{ième}$ neurone.

est la valeur du seuil, considérée comme étant nulle [TSA, 1992].

II.3. Mise en oeuvre

Dans notre application, un réseau de **Hopfield** à N dimensions est utilisé pour affecter chaque pixel de l'image multispectrale à la classe qui lui est appropriée, où représente la dimension de l'espace image.

II.3.1. Topologie de la grille :

La topologie du réseau de **Hopfield** est une grille de neurones tous interconnectés. La grille possède N_p lignes et N_{cl} colonnes, où N_p correspond au nombre total de pixels de l'image à classifier et N_{cl} correspond au nombre de classes présentes dans l'image.

Le tableau suivant représente la topologie du réseau :

		Nombre de classes			
		classe 1	classe 2	...	classe k
Nombre total de pixels à classer	n_1	0	0	...	1
	n_2	1	0	...	0

	N_p	1	0	...	0

Tableau III.1. Topologie de la grille.

La valeur **1** dans la grille indique que le neurone est actif, tandis que la valeur **0** indique que le neurone est inactif. Le neurone est affecté à la classe où il est actif.

Les poids de connexions entre les neurones sont déterminés par une fonction d'énergie qui atteint un minimum quand les conditions suivantes sont satisfaites :

- Chaque ligne doit avoir un seul neurone actif.
- La somme de chaque ligne doit être égale à la valeur **1**. Cela signifie que chaque pixel doit être affecté à **100%** à une seule classe.
- La distance entre le pixel affecté à une classe et le centre de gravité de la classe doit être minimale.

II.3.2. Expression de la fonction d'énergie

Un autre apport essentiel du réseau de **Hopfield** au domaine des réseaux de neurones est l'analogie avec la physique statique. Il montre que durant l'évolution du réseau, qui tend à se rapprocher de son état stable, une fonction d'énergie décroît vers un minimum [KOS, 1999]. Cette fonction est la fonction de **Lyapounov** [HOP, 1982].

Pour un réseau de **Hopfield** à N dimension, elle est donnée par l'expression suivante :

$$(III.13)$$

Où : w_{ij} est le poids de connexion entre les neurones i et j .
 s_i et s_j sont les états de sortie des neurones i et j .
 x_i est une entrée externe au neurone i .
 N est le nombre total des pixels de l'image à classifier.

II.3.3. Simplification de la fonction d'énergie :

Pour des raisons de commodité de calcul, nous avons considéré un réseau de **Hopfield** sans entrée externe et sans seuil [TSA, 1992]. Par ailleurs, le second terme de l'équation (III.13) s'annule. La fonction d'énergie aura pour expression :

$$(III.14)$$

Une variation δs_i de l'état d'un neurone V_{ik} entraîne forcément une variation de l'énergie globale δE [NAS, 1992]. Par ailleurs, nous calculons le gradient de cette énergie par rapport à V_{ik} , ainsi nous déterminons la contribution de l'énergie pour chaque neurone pris isolément. Nous trouvons :

$$(III.15)$$

Cette variation décroît vers un minimum si nous arrivons à trouver la valeur des poids de connexions w_{ij} qui représente la distance minimale entre le pixel à classer et le centre de gravité de chaque classe. Nous calculons, par ailleurs, cette distance pour chaque itération

lorsque le système évolue pour atteindre son état stable. A ce moment, aucun neurone ne change d'état.

Les équations **(III.9)**, **(III.10)** et **(III.11)** deviennent :

(III.16)

(III.17)

(III.18)

A partir des équations **(III.17)** et **(III.18)**, nous effectuons la mise à jour des sorties des neurones. L'équation **(III.18)** permet de déterminer les neurones actifs et les neurones inactifs.

II.4. Organigramme

La figure **III.5** donne l'organigramme de la classification par le réseau de **Hopfield** :

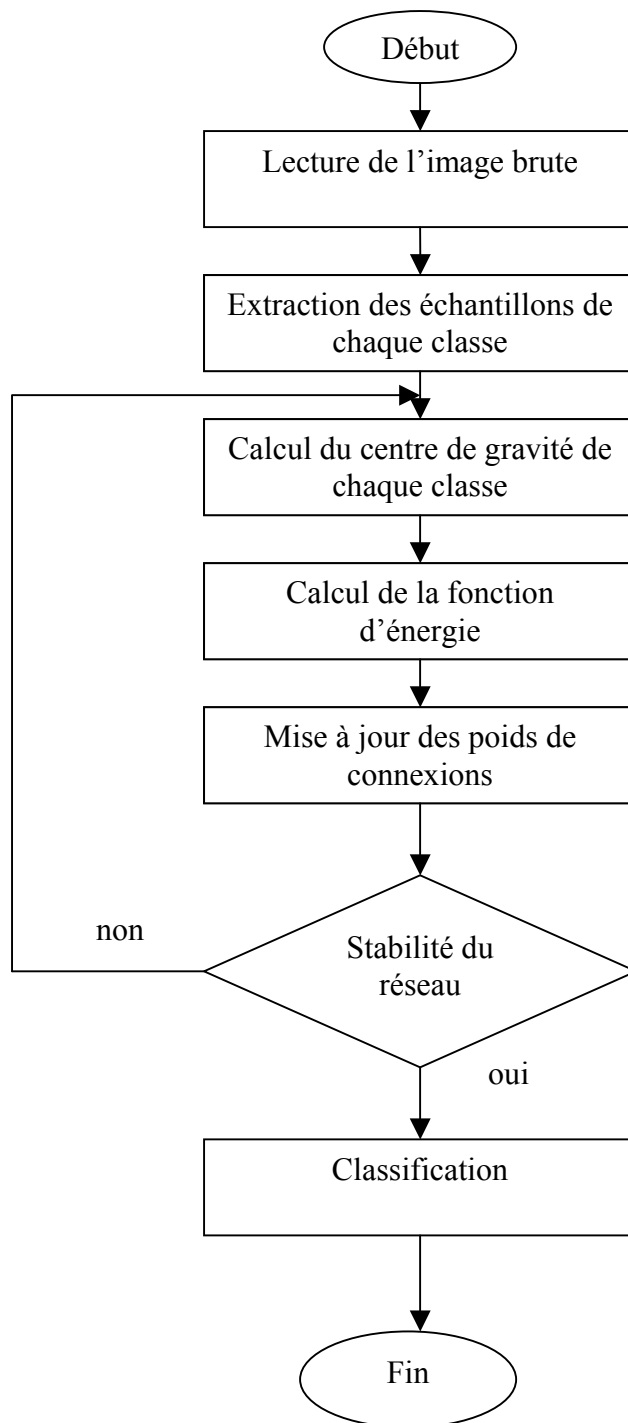


Figure III.5. Organigramme de la classification par le réseau de Hopfield.

CONCLUSION

La méthode de classification par les réseaux de neurones constitue, sans doute, une très bonne approche pour notre travail grâce à la capacité d'apprentissage et la capacité d'adaptation de ces réseaux.

Nous avons développé, dans ce chapitre, la méthodologie de la mise en œuvre du processus de classification par deux modèles de réseaux de neurones : le **perceptron multicouche** et le réseau de **Hopfield**. Le bon choix des paramètres requis nous permet de trouver la meilleure architecture de chaque modèle.

A l'issue de l'étude expérimentale que nous allons développer dans le chapitre suivant, nous pouvons déterminer les capacités et les performances de chaque modèle.

Chapitre IV

Tests et résultats

Résumé

*Ce chapitre est consacré à l'évaluation des résultats obtenus. La qualité de la classification des images résultats est évaluée pour chacune des deux méthodes développées. Pour le **perceptron multicouche**, les principaux paramètres qui influent sur la convergence de l'algorithme de rétropropagation du gradient sont le momentum, le taux d'apprentissage, le nombre d'itérations ainsi que le nombre de neurones dans la couche cachée. Le choix de ces paramètres est d'une importance cruciale. Tandis que pour le réseau de **Hopfield**, l'intérêt majeur s'est porté, d'abord, sur la façon de trouver une architecture adéquate à l'application. Par suite, nous avons exploité la principale caractéristique du modèle, à savoir, la fonction d'énergie qu'il faut minimiser afin que le réseau se rapproche de son état stable. Cette fonction est assimilée à la mesure de distance qui sépare le pixel à classer des centres de gravité de chaque classe. Ce qui revient à déterminer la distance minimale pour pouvoir décider d'affecter le pixel à la classe qui lui est la plus proche. Le processus est réitéré jusqu'à la convergence.*

INTRODUCTION

Le domaine de traitement des données par les réseaux de neurones se base principalement sur les essais. Pour cela, nous avons effectué plusieurs tests afin de présenter les résultats obtenus dans ce chapitre.

Lors de la phase d'apprentissage de chacune des deux méthodes, plusieurs paramètres sont pris en considération afin de trouver la meilleure architecture. Il n'existe pas de règles spécifiques quant au bon choix des valeurs des paramètres, cela se fait de manière empirique.

Lors de la phase de classification, les pixels à classer sont présentés à chacun des deux réseaux avec les paramètres choisis durant la phase d'apprentissage.

I. PRESENTATION DES DONNEES IMAGES

Pour les tests que nous avons effectués, les images utilisées sont des images fournies par le satellite **SPOT**. Une fenêtre est extraite couvrant la région de Blida sur une surface de 5Km*5Km environ et d taille 256*256 pixels.

Le satellite **SPOT** fournit trois images en mode multispectral de résolution 20m*20m et une image en mode panchromatique de résolution 10m*10m. Nous avons effectué les tests sur les trois images multispectrales, soient : **XS1** (vert), **XS2** (rouge), **XS3** (proche infrarouge).

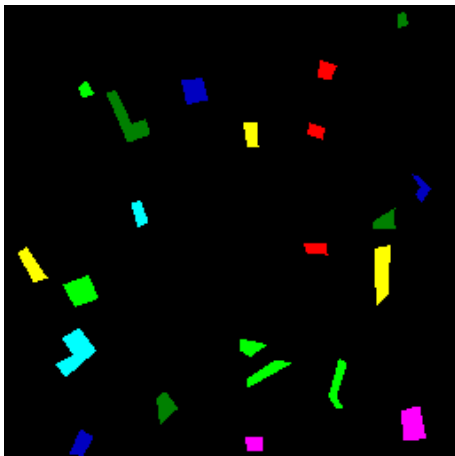
L'image de Blida est de taille 256 pixels*256 pixels. C'est une région agricole, constituée de grands champs à végétation dense et moins dense ainsi que de parcelles cultivées et non cultivées. Elle est constituée, également, d'une région urbaine moins dense regroupant quelques villes et villages isolés à proximité de l'agriculture. On définit, ainsi, sept classes reportées dans le tableau **IV.1** en indiquant le nombre de pixels appartenant à chaque classe. L'image de la figure **IV.1-(a)** représente la scène de Blida en composition colorée des trois canaux. Tandis que, l'image de la figure **IV.1-(b)** représente les échantillons d'apprentissage et de contrôle.

Classes	Thèmes	Nombre de pixels (Apprentissage)	Nombre de pixels (Contrôle)
1 (rouge)	Zone Urbaine Moins Dense (UMD)	220	357
2 (jaune)	Végétation Moins Dense (VMD)	483	1052
3 (vert brillant)	Sol Nu, aérodrome de Blida (SN)	620	1348
4 (bleu turquoise)	Champs Non Cultivés (CNC)	378	773
5 (rose)	Zone Urbaine Dense (UD)	290	643
6 (bleu)	Champs Cultivés (CC)	349	984
7 (vert)	Végétation naturelle Dense (VD)	519	1100

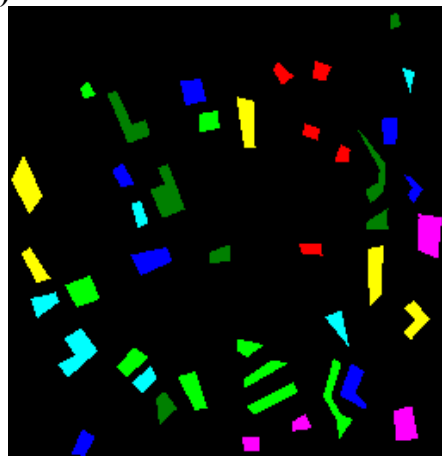
Tableau IV.1. Classes thématiques et base d'entraînement.



(a)



(b)



(c)

Figure IV.1. Image SPOT.

(a) Composition colorée des trois canaux - (b) Echantillons d'apprentissage –

(c) Echantillons de contrôle.

II. CRITERES D'EVALUATION ET INTERPRETATION

Plusieurs travaux et recherches ont été effectués afin d'évaluer la qualité de la classification en déterminant, entre autres, le nombre de pixels bien classés et, éventuellement, les erreurs commises dans la méthode de classification [CON, 1983] [GIR, 1999].

Généralement, la qualité d'une image classifiée est évaluée par une appréciation visuelle suivie d'une appréciation statistique basée sur l'estimation de la matrice de confusion.

II.1. Appréciation visuelle

Elle permet un premier jugement sur la qualité d'une image par rapport à la vérité terrain ou par rapport à une autre image classifiée. Cependant, cette appréciation est subjective et d'autant plus compliquée à réaliser que le nombre de thèmes présents dans l'image est élevé. En outre, cette appréciation diffère d'une personne à une autre. Pour cela, nous avons besoin de méthodes plus rigoureuses afin de quantifier la précision du processus de classification.

II.2. Appréciation statistique

Elle est basée sur l'estimation d'une matrice de confusion établie entre l'image résultat de la classification et la réalité terrain. Elle permet d'estimer, d'une part, le taux de bonne classification correspondant au nombre de pixels échantillons bien classés, et d'autre part, les erreurs d'inclusion (ou de commission) et d'exclusion (ou d'omission) correspondant au nombre de pixels échantillons mal classés. Il s'agit d'établir un tableau indiquant le nombre de pixels affectés à chaque classe relativement à leur appartenance aux échantillons de vérité terrain.

La matrice de confusion est une matrice carrée dont la taille est égale au nombre de classes et dont les éléments représentent le nombre de pixels affectés à chaque classe relativement à leur appartenance aux classes de la vérité de terrain. Nous disposons en lignes les classes de référence, et en colonnes les classes résultantes de la classification.

A partir d'une matrice de confusion, nous pouvons calculer des indicateurs de précision locaux (calculés pour chaque classe) ou globaux (calculés sur l'ensemble des classes). Nous adoptons la notation suivante pour le calcul de ces indicateurs :

: sont les éléments de la matrice de confusion.

: est la somme totale des éléments en ligne.

: est la somme totale des éléments en colonne.

N : est le nombre total des pixels de la matrice de confusion.

M : est le nombre des classes présentes dans l'image.

Avec :

II.2.1. Précision pour le producteur (Producer's Accuracy)

Cette mesure de précision locale indique le pourcentage de pixels d'une classe de référence affectés à la même classe par la classification. Aussi, on trouve la mesure (100% - PA), appelée erreur d'omission qui indique le pourcentage de pixels d'une classe de référence affectés à d'autres classes par la classification. Elle est donnée par :

(IV.1)

II.2.2. Précision pour l'utilisateur (User's Accuracy)

Cette mesure de précision locale indique le pourcentage de pixels d'une classe issue de la classification correspondante à la même classe dans les données de référence. Aussi, on trouve la mesure (100% - UA), appelée erreur de commission qui indique le pourcentage de pixels d'une classe issue de la classification qui appartiennent à d'autres classes dans les données de référence. Elle est donnée par :

(IV.2)

II.2.3. Taux de surestimation ou Taux d'erreur

Cette mesure est calculée, pour chaque classe, en prenant la différence absolue entre la précision pour le producteur et la précision pour l'utilisateur [BAR, 2002]. Soit :

(IV.3)

II.2.4. Taux de confusion :

Cette mesure exprime le rapport du nombre de pixels appartenant à une classe donnée et qui ont été affectés à d'autres classes. Elle est donnée par :

(IV.4)

II.2.5. Précision moyenne ou Taux Moyen de Bonne Classification (Overall Accuracy)

Cette mesure de précision globale représente le rapport entre le nombre total des pixels correctement classés et le nombre total des pixels de la matrice de confusion. C'est, donc, la somme des termes diagonaux de la matrice par le nombre total des pixels de cette matrice. Elle est donnée par :

(IV.5)

II.2.6. Khat :

Le Khat est un paramètre souvent utilisé pour l'évaluation et l'interprétation de la classification. Il prend en compte la précision pour le producteur et la précision pour l'utilisateur. Il est calculé à partir de la matrice de confusion établie entre l'image résultat et la réalité terrain de toute la scène. Il est donné par l'expression suivante :

$$Khat = \frac{N \sum_{i=1}^M X_{ii} - \sum_{i=1}^M (X_{i+} * X_{j+})}{N^2 - \sum_{i=1}^M (X_{i+} * X_{j+})} \quad (IV.6)$$

Ces grandeurs sont calculées pour chacune des deux méthodes développées afin d'évaluer la qualité de la classification mise en œuvre dans notre travail. Nous entamons les tests, d'abord, pour le **perceptron multicouche**, puis, pour le modèle de **Hopfield**.

III. TESTS ET RESULTATS DU PERCEPTRON MULTICOUCHE

En choisissant les bons paramètres requis tels que le nombre d'itérations et le nombre de neurones dans la couche cachée, on peut accélérer la convergence de l'algorithme de rétropropagation du gradient, car la convergence de cet algorithme dépend étroitement du bon choix des ces paramètres entre autres.

Nous avons effectué différents tests pour essayer de trouver le réseau le plus optimal ainsi que ses meilleurs paramètres. Parmi les principaux paramètres, nous avons fixé, au préalable, le momentum, le taux d'apprentissage et le nombre d'itérations, et avons fait varier le nombre de neurones dans la couche cachée afin de déterminer l'influence de ce nombre sur la convergence de l'algorithme.

- Le nombre d'itérations :

pour le réseau **1** noté **R1**

pour le réseau **2** noté **R2**

pour le réseau **3** noté **R3**

- Les paramètres propres au gradient :

Le momentum :

Le pas d'apprentissage :

Nous avons représenté les résultats de nos tests par les courbes illustrées par les figures **IV.2**, **IV.3** et **IV.4**. Nous avons constaté une variation de l'*erreur quadratique*, du et du lorsque le nombre de neurones cachés varie. Ce qui montre l'influence de ce dernier sur la convergence du réseau.

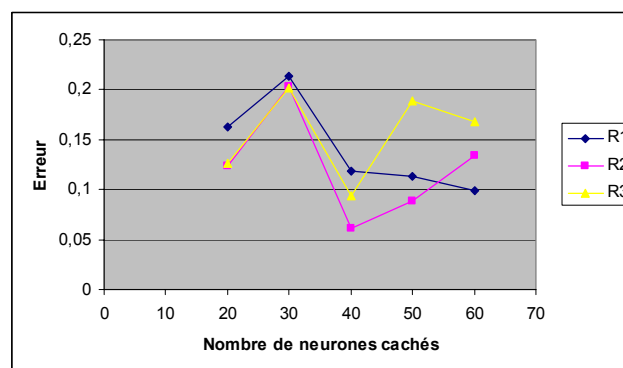


Figure IV.2. Variation de l'erreur en fonction du nombre de neurones cachés.

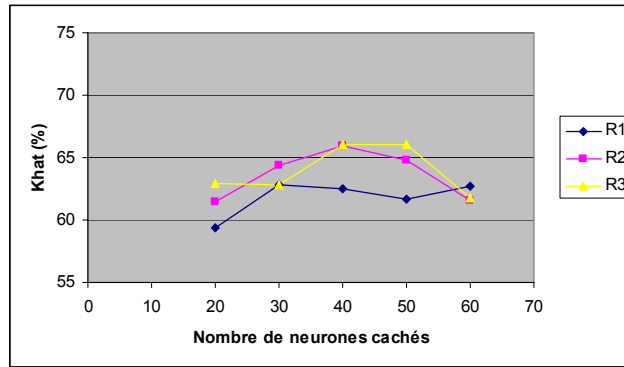


Figure IV.3. Variation du *Khat* en fonction du nombre de neurones cachés.

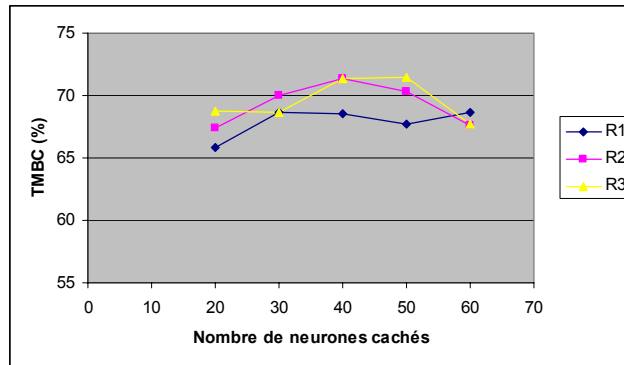


Figure IV.4. Variation *TMBC* en fonction du nombre de neurones cachés.

D'après ces courbes, nous avons constaté que *l'erreur quadratique* est relativement plus faible, soit : 0.0608 et 0.0879 (à la 150^{ième} itération) lorsque le nombre de neurones dans la couche cachée est de 40 et 50 neurones respectivement. En revanche, le *TMBC* et le sont plus élevés lorsque le nombre de neurones cachés est de 50 neurones cachés, soient : 71.43% et 66.09% (à la 200^{ième} itération). Par conséquent, l'architecture la plus adaptée comporte 50 neurones cachés.

Par suite, nous avons fixé le momentum, le taux d'apprentissage et le nombre de neurones dans la couche cachée, et nous avons fait varier le nombre d'itérations afin de déterminer l'influence de ce nombre sur la convergence de l'algorithme.

- Le nombre de neurones dans la couche cachée :

- Les paramètres propres au gradient :
 - Le momentum :
 - Le pas d'apprentissage :

Les courbes données par la figure IV.5 (respectivement IV.6 et IV.7) illustrent la variation de l'erreur quadratique (respectivement du *Khat* et du *TMBC*) en fonction du nombre d'itérations :

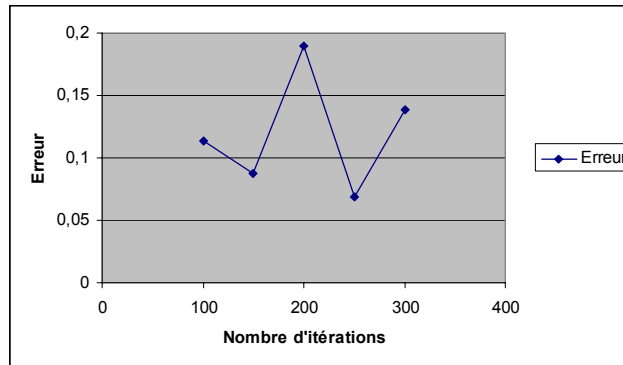


Figure IV.5. Variation de l'erreur en fonction du nombre d'itérations.

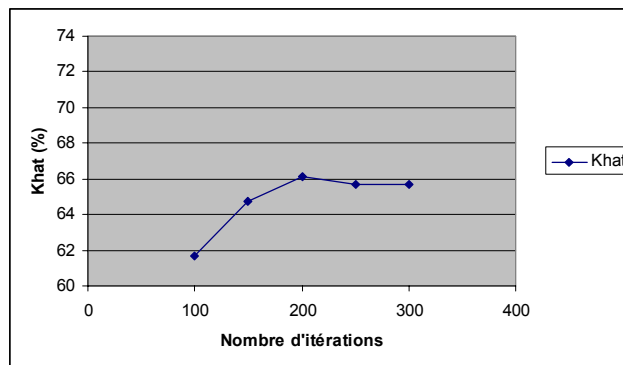


Figure IV.6. Variation du *Khat* en fonction du nombre d'itérations.

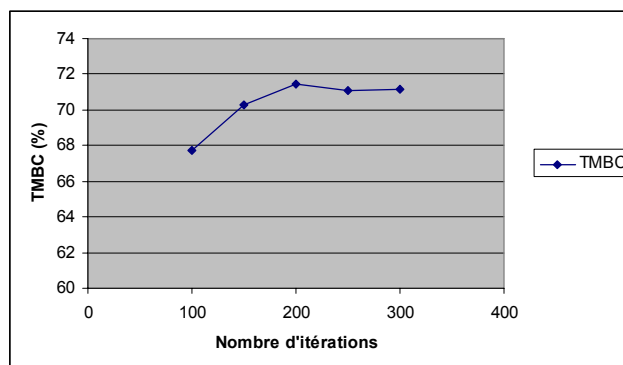


Figure IV.7. Variation du *TMBC* en fonction du nombre d'itérations.

Nous avons remarqué que l'augmentation des itérations permet d'améliorer le *TMBC* et le *Khat*, mais l'*erreur quadratique* du réseau est relativement plus faible, soit : 0.0691 à la 250^{ième} itération.

En optant pour le choix des valeurs du nombre d'itérations à 250 itérations et du nombre de neurones cachés à 50 neurones, nous avons effectué d'autres tests en faisant varier, d'abord, le momentum μ puis le taux d'apprentissage α pour trouver les paramètres optimaux qui aboutissent à une meilleure architecture.

Les résultats obtenus sont représentés graphiquement comme suit :

- La variation de l'*erreur quadratique* en fonction du momentum μ (respectivement du taux d'apprentissage α) est illustrée par la courbe de la figure IV.8 (respectivement la figure IV.9).
- Les figures IV.10 et IV.11 donnent la variation du *TMBC* en fonction du momentum et du taux d'apprentissage respectivement.
- Tandis que, les figures IV.12 et IV.13 donnent la variation du *Khat* en fonction du momentum et du taux d'apprentissage:

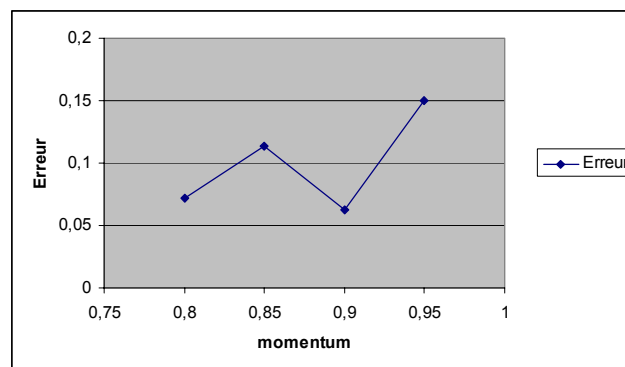


Figure IV.8. Variation de l'*erreur quadratique* en fonction du momentum.

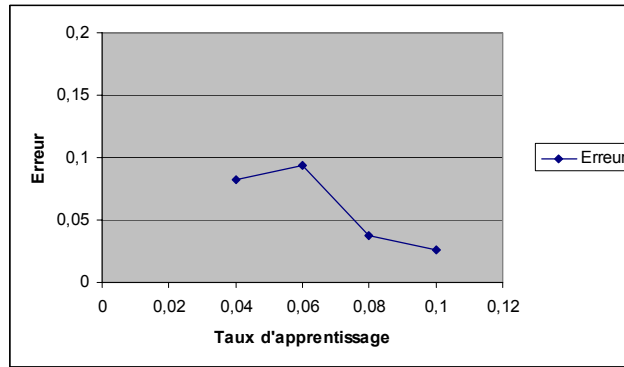


Figure IV.9. Variation de l'erreur quadratique en fonction du taux d'apprentissage.

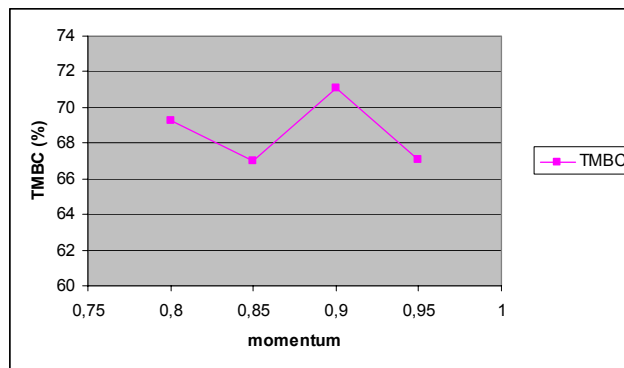


Figure IV.10. Variation du *TMBC* en fonction du momentum.

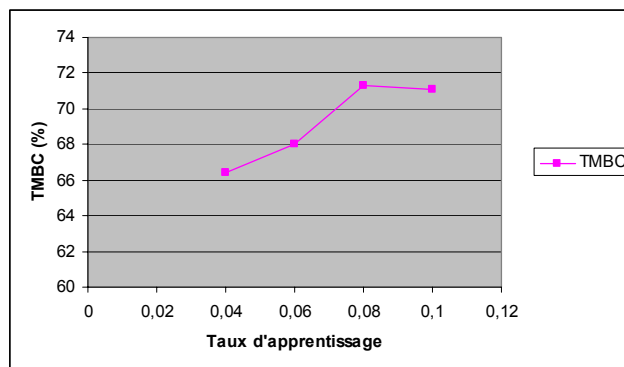


Figure IV.11. Variation du *TMBC* en fonction du taux d'apprentissage.

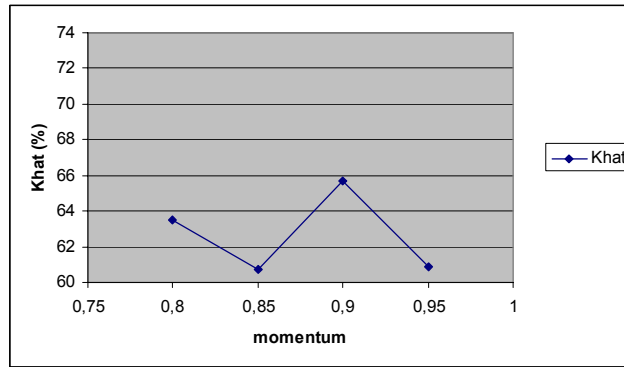


Figure IV.12. Variation du *Khat* en fonction du momentum.

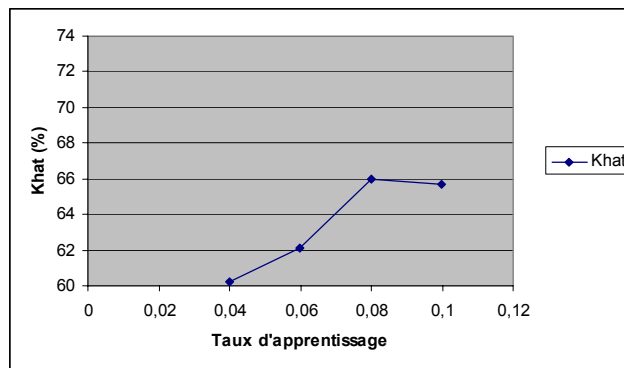


Figure IV.13. Variation du *Khat* en fonction du taux d'apprentissage.

Les paramètres que nous avons utilisés, pendant la phase d'apprentissage, sont choisis de manière empirique et aléatoire afin de chercher à retrouver les meilleures valeurs qui aboutissent à une bonne architecture, car, en pratique, il n'existe pas de critères pour trouver et fixer ces valeurs, d'où la nécessité d'effectuer plusieurs tests afin de tenter de s'approcher le plus de la meilleure optimisation possible.

Nous avons constaté que, en augmentant le nombre des neurones cachés, la convergence de l'algorithme de rétropropagation est plus appréciable mais en un temps de calcul plus important. Nous avons trouvé que les indicateurs *Khat* et *TMBC* sont plus importants, soient *Khat* = 66.02% (à la 150^{ième} itération) et 66.09% (à la 200^{ième} itération) et *TMBC* = 71.39% (à la 150^{ième} itération) et 71.43% (à la 200^{ième} itération), quand le nombre des neurones cachés varie de 40 à 50 neurones. En augmentant, également, le nombre d'itérations, soit de 100 à 300 itérations, nous avons constaté l'amélioration de la convergence de l'algorithme. Nous avons constaté une légère variation entre la 200^{ième} itération et la 250^{ième} itération quant aux valeurs du *Khat* (66.09% et 65.58%) et du *TMBC* (71.43% et 71.13%). En revanche, l'erreur quadratique varie d'une valeur maximale de

0.1894 à une valeur minimale de 0.0691. Raisons pour lesquelles nous avons opté pour les valeurs des paramètres comme suit :

Pour ces deux valeurs, nous avons une erreur quadratique faible, tandis que le *Khat* et le *TMBC* sont élevés, soient :

En optant pour le choix requis de ces paramètres, l'image résultat de la classification est donnée par la figure IV.14 :

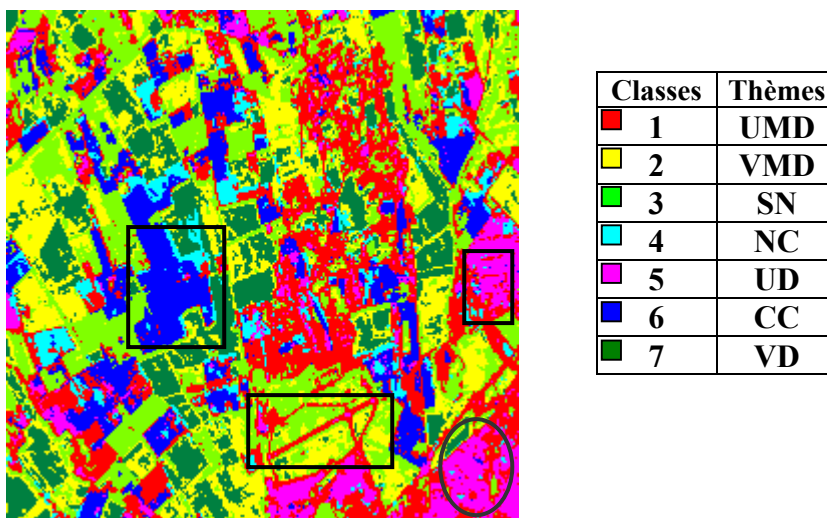


Figure IV.14. Image résultat de la classification par le perceptron.

A présent, nous allons effectuer une évaluation visuelle puis statistique de l'image résultat obtenue de la classification par le **perceptron multicouche**.

III.1. Appréciation visuelle

A partir de l'image classifiée obtenue, nous constatons que certaines parcelles sont bien définies telles que les zones à champs cultivés (classe 6) et les zones urbaines denses (classe 5). Nous remarquons, également, que l'aérodrome de Boufarik est bien visible, ses contours sont bien déterminés (zones encadrées sur la figure IV.14). A proximité de cet aérodrome, nous constatons une très légère confusion entre le sol nu (classe 3) et les zones à végétation moins dense (classe 2) (zone encerclée sur la figure IV.14). Mêmes constatations sont faites pour les zones urbaines denses (classe 5) et moins denses (classe 2). Ceci est dû, probablement, aux ressemblances des thèmes qui existent entre ces zones.

III.2. Appréciation statistique

Nous établissons une matrice de confusion normalisée entre le résultat de la classification et la base de contrôle que nous illustrons par le tableau IV.2. A partir de cette matrice, nous calculons, par la suite, les taux de surestimation et de confusion, soient : le et le donnés au tableau IV.3 :

	1	2	3	4	5	6	7
1 (Urbain moins dense)	60.22	0.00	0.29	16.42	11.97	3.86	0.00
2 (Végétation moins dense)	1.68	77.66	27.37	0.00	0.00	0.00	11.81
3 (Sol nu)	7.00	9.88	64.16	0.00	0.00	0.30	0.00
4 (Champs non cultivés)	4.20	0.20	0.22	28.58	1.86	1.11	7.72
5 (Urbain dense)	19.88	0.00	0.00	9.57	84.29	1.52	0.00
6 (Champs cultivés)	7.00	0.00	0.00	45.40	1.86	93.19	0.00
7 (Végétation dense)	0.00	12.26	7.93	0.00	0.00	0.00	80.36

Tableau IV.2. Matrice de confusion normalisée relative à la classification par le perceptron multicouche.

classe	TS (%)	TC (%)
1	13.69	39.77
2	15.85	12.45
3	22.50	35.83
4	34.92	36.43
5	07.10	15.70
6	22.95	06.80
7	01.43	09.63

<i>Khat</i>	65.96%
<i>TMBC</i>	71.29%

Tableau IV.3. Indicateurs de surestimation et de confusion relatifs aux résultats de la classification par le perceptron multicouche.

Nous constatons que les valeurs du taux de surestimation sont relativement plus faibles par rapport aux valeurs du taux de confusion, de même que le *Khat* et le *TMBC*. Nous constatons, également, que pour la classe 7 (végétation dense), nous avons la plus faible valeur du taux de surestimation, soit : 01.43%. Ce qui nous incite à conclure que la classe 7 est bien discriminée. Contrairement à la classe 4 (champs cultivés), le taux de surestimation de cette classe est le plus élevé, soit : 34.92%. Ce qui signifie que cette classe présente de fortes ambiguïtés avec les autres classes. Cependant, nous signalons que, pour certaines classes, il existe une grande ressemblance de représentativité des différents thèmes entre elles, telles que les classes 4 et 6 (champs cultivés et champs non cultivés), les classes 1 et 5 (urbain dense et urbain moins dense) et les classes 2 et 7 (végétation moins dense et végétation dense). Ce qui se traduit par les valeurs élevés du taux de confusion. Les valeurs, auxquelles nous avons abouties, sont en corrélation avec les variations des thèmes et leur interaction mutuelle. D'un point de vue global, nous pouvons dire que le résultat de la classification est satisfaisant dans l'ensemble des classes présentes dans l'image.

Finalement, nous pouvons conclure que notre choix pour les paramètres, qui influent sur la convergence de l'algorithme et donc sur la stabilité du réseau, est dicté par la comparaison de l'erreur quadratique, du *Khat* et du *TMBC* en faisant varier le nombre d'itérations, le nombre de neurones cachés, le momentum et le taux d'apprentissage afin d'avoir une erreur plus faible et un *TMBC* ainsi qu'un *Khat* plus élevés.

IV. TESTS ET RESULTATS DU RESEAU DE HOPFIELD

Pour le réseau de **Hopfield**, la procédure adoptée pour effectuer les tests est totalement différente de celle du **perceptron multicouche**. La procédure se déroule comme suit :

Les exemples correspondant aux pixels échantillons à classer sont présentés au réseau. Les neurones sont entièrement ou partiellement interconnectés. Les poids de connexions sont déterminés en calculant les distances qui séparent les pixels à classer des centres de gravité des classes présentes dans l'image. Parmi toutes les valeurs des distances que nous avons calculées, nous cherchons le minimum pour décider d'attribuer le pixel à la classe qui lui est la plus proche. A chaque itération, nous déterminons le nouveau nombre de pixels affectés à chaque classe. Nous présentons, à la figure **IV.15**, l'histogramme correspondant. Par conséquent, nous calculons les nouveaux centres de gravité des classes et les nouvelles valeurs des poids de connexions. Nous considérons, à chaque fois, le poids dont la valeur est la plus faible pour affecter le pixel à la nouvelle classe. Plus cette valeur est minimisée, plus le poids de connexion renforce la liaison, et ainsi, le pixel est attribué à la classe qui lui est la plus proche. Le processus est réitéré jusqu'à ce qu'on obtienne un taux de changement de pixels faible ou, idéalement, nul entre deux itérations successives. La figure **IV.16** donne la variation moyenne du nombre d'échantillons qui ont changé de classe entre deux itérations successives.

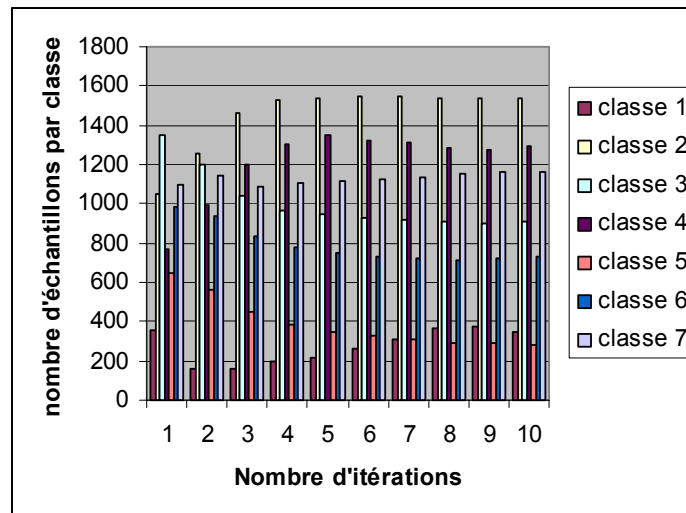


Figure IV.15. Variation du nombre de pixels par classe en fonction du nombre d'itérations.

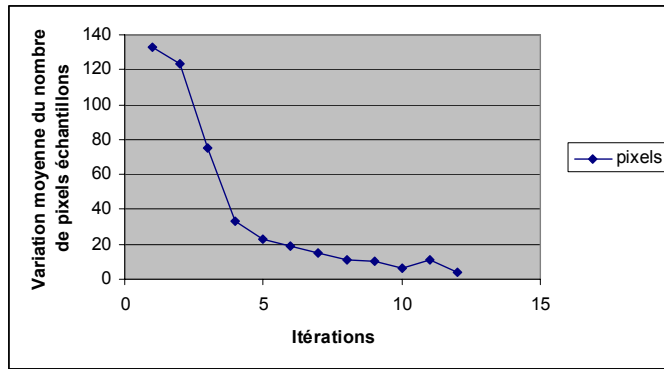


Figure IV.16. Variation moyenne du nombre de pixels entre deux itérations successives.

De même que pour le **perceptron multicouche**, nous calculons les taux de précision et de confusion pour l'évaluation de la classification par le réseau de **Hopfield**.

D'après les tests que nous avons effectués, nous avons obtenu les courbes des figures **IV.17** et **IV.18** :

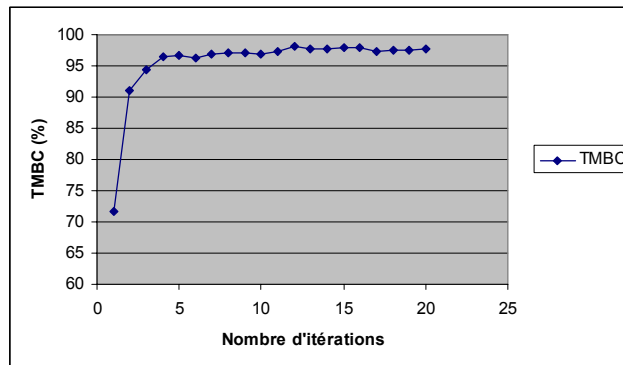


Figure IV.17. Variation du TMBC en fonction du nombre d'itérations.

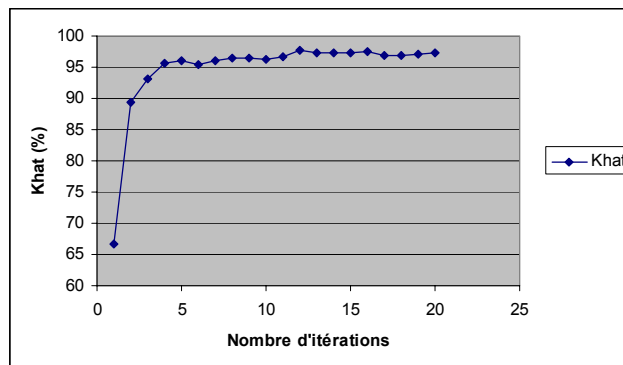


Figure IV.18. Variation du Khat en fonction du nombre d'itérations.

Nous avons constaté que le réseau a commencé à atteindre sa convergence au bout de la 5^{ème} itération pour laquelle les valeurs du *Khat* et du *TMBC* sont :

Au-delà de cette itération, nous avons remarqué des oscillations au niveau de ces deux valeurs (voir figures **IV.17** et **IV.18**) En outre, nous avons constaté que, seulement 33 pixels échantillons qui ont changé de classe entre la 4^{ème} et la 5^{ème} itération comme le montre la figure **IV.16**. Un autre constat que nous signalons concernant la 12^{ème} itération pour laquelle les valeurs du *Khat* et du *TMBC* sont au maximum, soient : 97.63% et 98.05%, et 11 pixels échantillons ont changé de classe. Néanmoins, il y a eu plusieurs oscillations avant d'atteindre ces valeurs, c'est la raison pour laquelle nous avons estimé que la convergence du réseau est obtenue au bout de la 5^{ème} itération et que le résultat de la classification est suffisamment appréciable.

Nous présentons, par les courbes des figures **IV.19** et **IV.20**, les valeurs des taux de précision et de confusion locaux relatives à chaque classe. Nous constatons une amélioration importante entre la 1^{ère} et la 5^{ème} itération. Ceci confirme qu'au fil des itérations, le processus a évolué en convergeant vers une solution optimale. Nous avons trouvé que le *TBC* a varié de 24.36% à 81.20%, par exemple, pour la classe 1 (urbain moins dense) et que le *TBC* a varié de 75.63% à 18.80% pour la même classe.

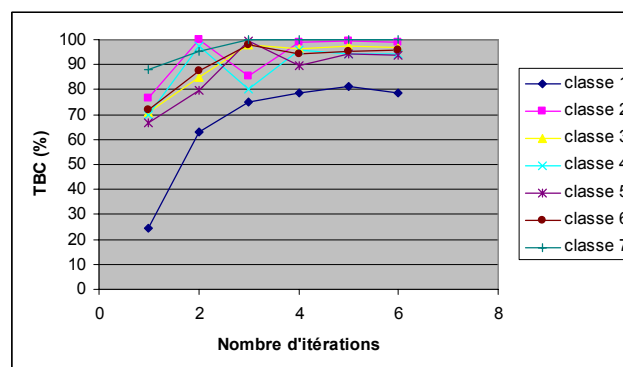


Figure IV.19. Variation du *TBC* en fonction du nombre d'itérations.

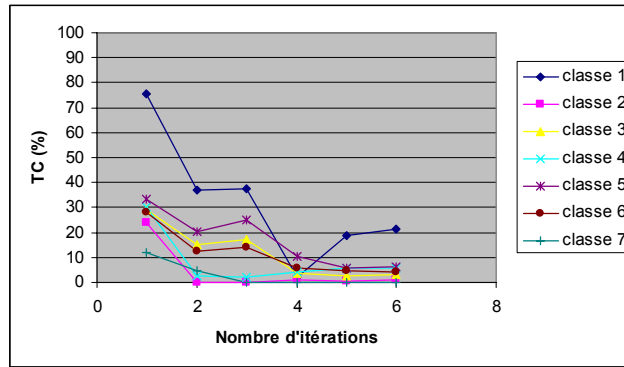
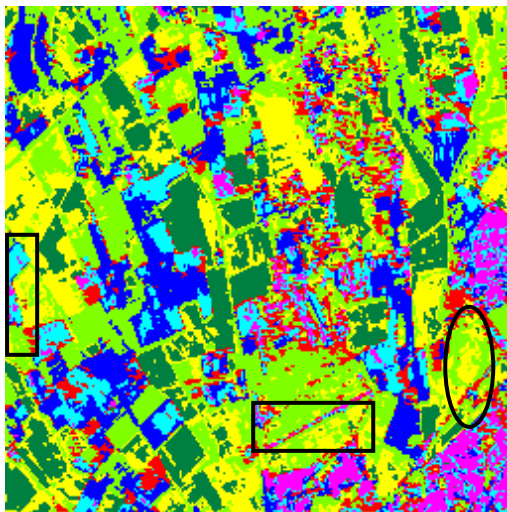
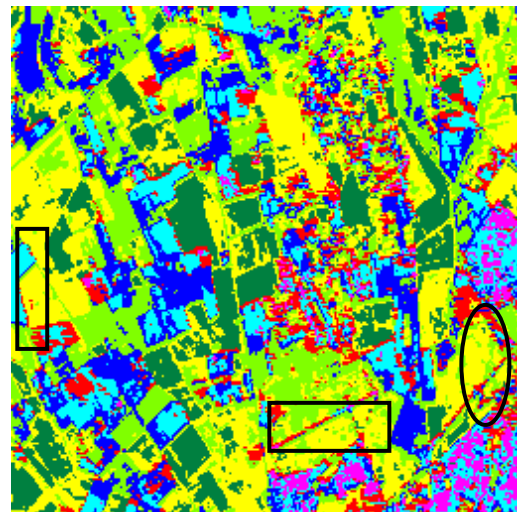


Figure IV.20. Variation du *TC* en fonction du nombre d'itérations.

Nous donnons à la figure IV.21 les images résultantes de la classification relatives à la 1^{ère} et la 5^{ème} itération afin de mieux distinguer la différence et constater une amélioration visible au fil des itérations, et à la figure IV.22, nous donnons quelques images résultats relatives aux itérations indiquées en dessous de chaque image.

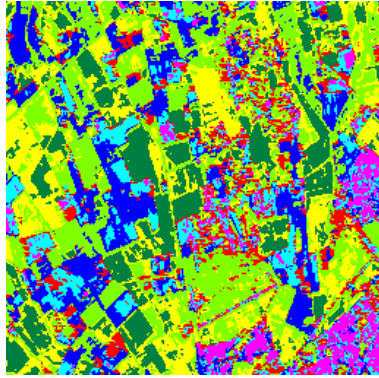


1^{ère} itération.



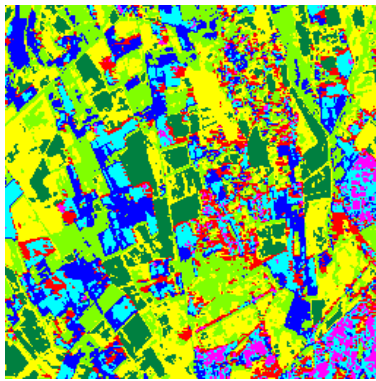
5^{ème} itération.

Figure IV.21. Images résultats de la classification relatives à la 1^{ère} et la 5^{ème} itération.

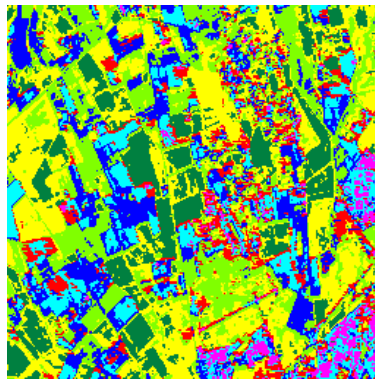


1^{ère} itération.

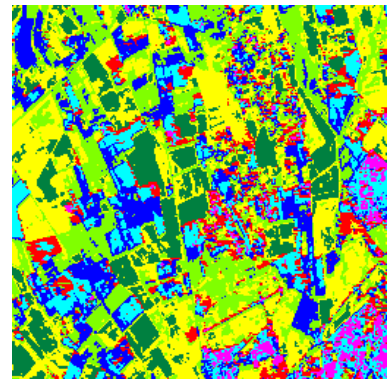
Classes	Thèmes
■ 1	UMD
■ 2	VMD
■ 3	SN
■ 4	NC
■ 5	UD
■ 6	CC
■ 7	VD



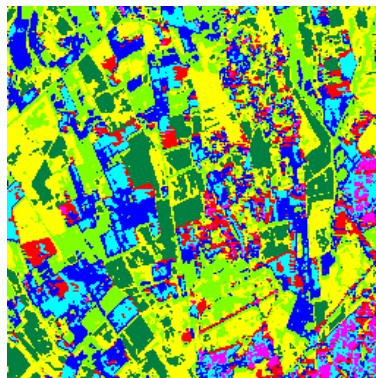
4^{ième} itération.



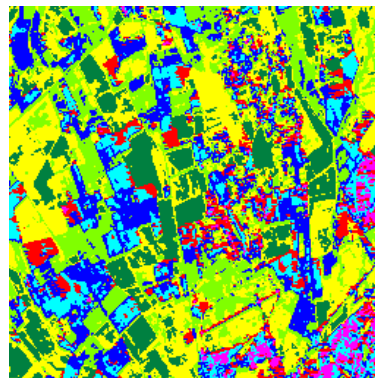
5^{ième} itération.



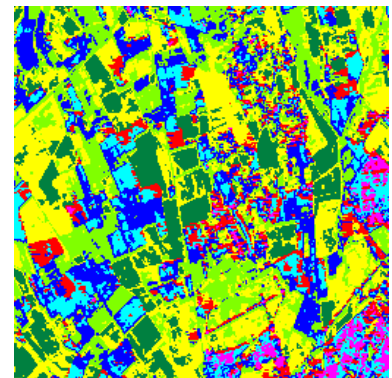
6^{ième} itération.



11^{ième} itération.



12^{ième} itération.



13^{ième} itération.

Figure IV.22. Images résultats de la classification par le réseau de Hopfield.

Nous allons effectuer une évaluation visuelle puis statistique de l'image résultat obtenue de la classification par le réseau de **hopfield**.

IV.1. Appréciation visuelle

A l'issue du résultat obtenu de la première itération, nous constatons une confusion visible entre les champs cultivés (classe 4) et les zones urbaines denses (classe 5) et les contours de l'aérodrome de Boufarik (classe 3) ne sont pas bien définis (zones encadrées sur la figure IV.21). Nous remarquons, également, que, dans certaines parcelles, il existe une confusion entre la végétation moins dense (classe 2) et le sol nu (classe 3) (zone encerclée sur la figure IV.21). A partir de l'image issue de la cinquième itération, nous pouvons constater, clairement, qu'il y a eu une amélioration visible quant aux confusions obtenues à la première itération. En effet, les contours de l'aérodrome sont bien déterminés et les zones à végétation dense (classe 7) ainsi que les champs cultivés (classe 6) sont mieux représentés (zones encadrées sur la figure IV.21). Dans l'ensemble de l'image obtenue à la 5^{ième} itération, nous remarquons une amélioration de la classification comparativement à celle obtenue à la 1^{ère} itération.

IV.2. Appréciation statistique

Nous avons présenté les matrices de confusion normalisées relatives à la 1^{ère}, la 5^{ième} et la 12^{ième} itération par les tableaux IV.4 ((a), (b) et (c)), ainsi que les valeurs des taux de surestimation et de confusion, soient : le et le par les tableaux IV.5 ((a), (b) et (c)) respectivement.

Classes	1	2	3	4	5	6	7
1 (Urbain moins dense)	24.37	0	0	0	0	0	12.09
2 (Végétation moins dense)	75.63	76.14	0	0	0	0	0
3 (Sol nu)	0	23.89	71.01	0	0	0	0
4 (Champs non cultivés)	0	0	29.01	69.85	0	0	0
5 (Urbain dense)	0	0	0	30.14	66.72	0	0
6 (Champs cultivés)	0	0	0	0	33.28	72.05	0
7 (Végétation dense)	0	0	0	0	0	27.94	87.91

(a)

Classes	1	2	3	4	5	6	7
1 (Urbain moins dense)	81.19	0	0	0	0	0	0.11
2 (Végétation moins dense)	18.81	99.48	0	0	0	0	0
3 (Sol nu)	0	0.52	97.14	0	0	0	0
4 (Champs non cultivés)	0	0	2.85	94.21	0	0	0
5 (Urbain dense)	0	0	0	5.81	94.46	0	0
6 (Champs cultivés)	0	0	0	0	5.54	95.39	0
7 (Végétation dense)	0	0	0	0	0	5.53	99.89

(b)

Classes	1	2	3	4	5	6	7
1 (Urbain moins dense)	85.82	0	0	0	0	0	0
2 (Végétation moins dense)	14.17	99.73	0	0	0	0	0
3 (Sol nu)	0	0.26	95.33	0	0	0	0
4 (Champs non cultivés)	0	0	4.66	97.22	0	0	0
5 (Urbain dense)	0	0	0	2.77	96.51	0	0
6 (Champs cultivés)	0	0	0	0	3.49	98.11	0
7 (Végétation dense)	0	0	0	0	0	1.88	100

(c)

Tableau IV.4. Matrices de confusion normalisées relatives à :
(a) la 1^{ère} itération. (b) la 5^{ème} itération. (c) la 12^{ème} itération.

Classes	TS(%)	TC (%)
1	15.17	60.45
2	01.36	23.21
3	08.21	20.77
4	11.84	30.14
5	01.91	35.19
6	11.02	04.31
7	10.06	22.14

<i>Khat</i>	66.60%
<i>TMBC</i>	71.75%

(a)

Classes	TS(%)	TC (%)
1	18.22	11.40
2	02.10	00.76
3	01.99	02.01
4	03.70	05.81
5	13.87	02.95
6	02.02	03.45
7	02.85	00.08

<i>Khat</i>	94.95%
<i>TMBC</i>	95.67%

(b)

Tableau IV.5. Indicateurs de surestimation et de confusion relatifs aux résultats de la classification par le réseau de Hopfield.

Nous avons remarqué que les valeurs des taux de surestimation et de confusion ont largement varié de la 1^{ère} itération à la 5^{ième} itération. Par exemple, pour la classe 7 (végétation dense), le taux de surestimation a varié de 10.06% à 2.85%, et le taux de confusion a varié de 22.14% à 0.08% pour la même classe. Ce qui s'explique par l'amélioration de la discrimination de la classe 7. Même constat est fait pour les classes 3 (sol nu), 4 (champs non cultivés) et 6 (champs cultivés). En revanche, la classe 5 (urbain dense) a subi des dégradations au niveau de la discrimination, le taux de surestimation de cette classe a varié de 1.91% à 13.87%. Néanmoins, sur l'ensemble de l'image classifiée, nous pouvons déduire que certaines classes sont mieux représentées et mieux discriminées que d'autres. Les valeurs des taux de surestimation et de confusion ont diminuées de la 1^{ère} itération à la 5^{ième} itération pour la plupart des classes. Ceci est dû, principalement, au fait que nous avons utilisé l'image résultat de la 1^{ère} itération comme nouvelle base de données dans le but d'améliorer et d'affiner la classification. Par conséquent, chaque image résultat issue d'une itération a servi comme base d'entraînement pour générer une nouvelle image classifiée. Les résultats auxquels nous sommes arrivés montrent qu'il y a une nette amélioration entre la 1^{ère} itération et la 5^{ième} itération.

Avant de conclure, nous pouvons dire que la mise en œuvre du processus de classification d'images satellitaires par le **perceptron multicouche** a donné de bons résultats.

Nous arrivons à distinguer, à travers l'image classifiée, les différentes structures. Les contours et les frontières sont bien définis et visibles. La principale difficulté, lors de la mise en œuvre de cette méthode, réside dans le choix de la configuration adéquate telle que le nombre de neurones cachés, aussi, dans le choix du nombre d'itérations afin de minimiser au maximum la différence entre la sortie calculée par le réseau et la sortie désirée. Le choix des paramètres relatifs au gradient tels que le momentum et le pas d'apprentissage est d'une importance cruciale. Ces paramètres influent étroitement sur la convergence de l'algorithme de rétropropagation et donc sur la stabilité du réseau. Par conséquent, nous avons été appelés à effectuer plusieurs essais et tests avant d'aboutir à un choix judicieux.

Quant à la classification par le réseau de **Hopfield**, nous avons amélioré l'image résultat en la réinjectant au processus itératif comme nouvelle base de données afin d'affiner au mieux la classification. Effectivement, nous constatons une amélioration entre la 1^{ère} et la 5^{ème} itération dans certaines parcelles telles que l'aérodrome de Blida (classe **3**), quelques parcelles à végétation moins dense (classe **2**). Par contre, dans d'autres parcelles, telles que les champs cultivés (classe **6**) et non cultivés (classe **4**), il existe, visiblement, une importante confusion qui est due, principalement, à la similitude entre les deux thèmes. Nous remarquons, aussi, que certaines zones urbaines (classe **1**) sont confondues avec les zones considérées comme étant sol nu (classe **3**). En revanche, à travers l'image résultat obtenue, nous arrivons à discerner certaines classes discriminantes telles que les zones à végétation dense (classe **7**). Toutefois, cette image nécessite d'autres améliorations afin de représenter, au mieux, les différentes classes.

V. COMPARAISON ENTRE LES DIFFERENTES APPROCHES

A l'issue de cette étude, nous présentons les valeurs du Khat calculées à partir de la matrice de confusion établie entre l'image résultat de la classification de chaque méthode, étudiée auparavant, et la vérité terrain, relative à la scène de Blida. Nous résumons les valeurs obtenues par l'histogramme de la figure **VI.8** :

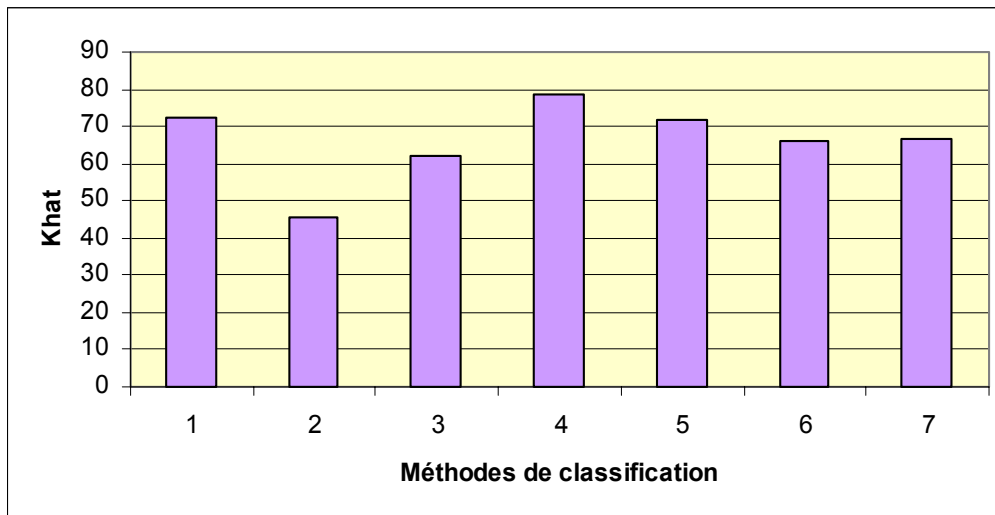


Figure VI.8. Comparaison des valeurs du Khat des différentes méthodes de classification.

- (1) Maximum de Vraisemblance. (2) Minimum de Distance Euclidienne. (3) Distance de Mahalanobis. (4) Markov(ICM). (5) Algorithmes Génétiques. (6) Perceptron Multicouche. (7) Réseau de Hopfield.**

CONCLUSION

En développant les méthodes de classification par le **perceptron multicouche** et par le réseau de **Hopfield**, nous avons constaté que l'image classifiée par le **perceptron multicouche** était plus nette dans certaines zones que celle classifiée par le réseau de **Hopfield**, et inversement. Visiblement, c'est le résultat auquel nous sommes arrivé.

Néanmoins, la mise en œuvre de la première méthode était ardue, ceci est dû, principalement, au choix des bons paramètres requis qui conduisent à une meilleure architecture. En outre, cette méthode est gourmande en temps de calcul, notamment, durant la phase d'apprentissage. Pendant cette phase, nous avons effectué plusieurs essais en agissant sur la variation de certains paramètres tels que la taille de la couche cachée, le momentum et le taux d'apprentissage.

Quant à la deuxième méthode, nous n'avons pas eu à modifier plusieurs paramètres. Il fallait, d'abord, trouver l'architecture adéquate pour l'adapter à notre application. Puis, en présentant les données au réseau, une fonction d'énergie a été minimisée pour que le réseau atteigne son état stable. A cet effet, nous avons remarqué que, au bout de quelques itérations, peu de pixels ont changé de classe. Ce qui nous a permis de dire que le réseau a convergé vers son état stable.

Conclusion générale

« d'une image d'un objet, le cerveau humain retire une multitude d'informations,
est-il possible de doter une machine d'une telle capacité d'analyse ? »

Jean Serra.

Vu la grande variété et la richesse d'information contenue dans les images de télédétection satellitaire, les thématiciens font appel, de plus en plus, aux méthodes issues de champs différents : traitement de signal, traitement d'images, reconnaissance de formes, intelligence artificielle, etc...

Durant cette dernière décennie, des recherches considérables se sont focalisées sur le traitement numérique de l'information, en particulier, le traitement d'images par différentes approches. Au fait, ceci est dû, principalement, à l'intérêt d'exploiter efficacement les images classifiées pour la surveillance de la surface de la Terre ainsi que l'évolution des changements terrestres sur des surfaces étendues.

La classification d'images est une des méthodes de traitement et d'analyse d'images qui consiste à organiser l'image multibande en classes ou groupes homogènes pour en extraire une image servant à l'interprétation thématique des zones observées.

L'introduction des réseaux de neurones dans le domaine de traitement numérique des données a montré ses performances quant à la résolution des problèmes non prévisibles par des modèles mathématiques. Cela découle de la capacité d'adaptation et de généralisation des réseaux neuronaux.

L'objectif de notre mémoire était la mise en œuvre d'un processus de classification supervisée des images de télédétection satellitaire en utilisant les réseaux de neurones qui font partie des méthodes de classification non paramétriques. Notre choix s'est porté sur l'utilisation de deux modèles : le **perceptron multicouche** et le réseau de **Hopfield**.

Au préalable, nous nous sommes intéressés au concept général de la classification par différentes méthodes, en l'occurrence, statistiques, géométriques, génétiques et neuronales. Nous avons mis en évidence la classification par les réseaux neuronaux.

Pour la mise en œuvre de notre travail, nous avons développé, dans un premier temps, le **perceptron multicouche** entraîné par l'algorithme de rétropropagation durant la phase d'apprentissage supervisé. Cet algorithme, dans son principe, nous a permis de minimiser l'erreur entre les sorties calculées et les sorties désirées par modification des poids de connexions entre les neurones. Nous avons fait varier les paramètres qui influent sur la convergence de ce réseau afin de trouver les paramètres optimaux qui nous donnent une meilleure architecture. Dans un second temps, nous avons développé le modèle de **Hopfield**. Lors de l'évolution de ce réseau, nous avons utilisé une des propriétés de l'énergie entre les neurones du réseau qui décroît au fur et à mesure jusqu'à atteindre un minimum. Nous avons assimilé cette énergie au calcul de distance séparant le pixel du centre de gravité de chaque classe. A chaque itération, nous avons déterminé la distance la plus minimale pour assigner

les pixels à leurs classes les plus proches. A ce moment là, le réseau a convergé et nous a fourni une sortie stable.

Par la suite, nous avons évalué la qualité de la classification de chaque méthode en calculant les indicateurs de précision globaux sur l'image classifiée, à savoir, le Precision et le Recall , et les indicateurs de précision locaux de l'image sur chaque classe, soient le taux de surestimation Overestimation et le taux de confusion Confusion . Ces taux permettent d'estimer la robustesse de la méthode de classification.

Le **perceptron multicouche** demeure délicat à paramétrer et le réseau de **Hopfield** délicat à adapter sa fonction d'énergie aux contraintes de l'application envisagée. En effet, le premier utilise la règle de rétropropagation qui présente une grande sensibilité quant au choix des bons paramètres qui influent sur sa convergence. Tandis que le deuxième, trouver la fonction d'énergie adéquate et satisfaire les contraintes imposées pour minimiser cette fonction d'énergie est une tâche ardue.

Les tests effectués, nous ont permis de constater que le **perceptron multicouche** et le réseau de **Hopfield** ont donné des résultats intéressants et satisfaisants. Toutefois, il aurait pu être profitable de tester d'autres types de modèles connexionnistes tels que les modèles compétitifs de **Kohonen** qui semblent s'adapter à un grand nombre de problèmes, entre autres, la classification.

Les perspectives éventuelles à ce travail sont nombreuses. Nous en citons quelques uns. Nous suggérons, par exemple, inclure un terme qui modélise le contexte spatial du pixel à classifier lors de la phase d'apprentissage afin d'affiner au mieux la classification. En effet, la contribution du pixel et de son voisinage améliore la qualité de la classification. Nous suggérons, également, l'utilisation conjointe des réseaux de neurones et des algorithmes génétiques pour la mise en œuvre du processus de classification ou encore l'emploi des algorithmes génétiques pour le développement des réseaux de neurones.

Enfin, il est indispensable de dire qu'à travers ce travail, nous constatons que les réseaux de neurones constituent une famille d'outils très diversifiés. Chaque modèle possède ses propres caractéristiques et ses propres paramètres, ainsi, il peut être destiné à des opérations variées. L'utilisation des réseaux de neurones dans le traitement des données ouvre tant de portes à explorer...

Annexes

« ne crains pas d'avancer lentement, crains seulement de t'arrêter »

Sagesse chinoise

Annexe A

Classification supervisée et non supervisée

Nous introduisons, dans cette annexe, les principes méthodologiques de la classification supervisée et la classification non supervisée.

Les méthodes de classification des images multispectrales peuvent être regroupées en deux grandes familles. La première, appelée classification supervisée, repose sur une information *a priori* sur les classes importantes (leur nombre, leur signification, leurs caractéristiques statistiques...) issue d'une base de données, ou acquise lors d'une étape d'apprentissage. Par opposition, la seconde, appelée classification non supervisée, ne nécessite pas d'informations *a priori* sur les classes présentes dans l'image.

A.1. Classification automatique ou non supervisée

La classification automatique ou non supervisée est une technique relativement primitive du fait qu'elle s'effectue sans aucune connaissance *a priori* ni sur le nombre de classes ni sur la structure de ces classes. Le groupement des individus, réalisé sur la base de similarité, est conditionné par le choix du nombre de classes et d'une configuration initiale liée au tirage des noyaux (pixels représentant les classes initiales). On distingue deux orientations possibles quant aux procédures de classification non supervisée regroupant les méthodes hiérarchiques et les méthodes non hiérarchiques. Se sont les méthodes non hiérarchiques qui sont appliquées en imagerie satellitaire.

A.1.1. Principe général

La classification regroupe des objets en classes, en se fondant sur la ressemblance entre ces objets. Tout algorithme de classification comprend, donc, les critères essentiels suivants :

- Un critère de ressemblance.
- Un critère de décision.

Une grande partie des problèmes de classification automatique non hiérarchique peut s'énoncer en termes d'optimisation d'un critère bien défini. La **Méthode des Nuées Dynamiques (MND)**, aussi appelée **Isodata**, fournit un cadre général permettant d'obtenir des algorithmes qui donnent des solutions appropriées.

La caractéristique principale de la **MND** est l'optimisation d'un critère qui exprime l'adéquation entre une classification d'objets et un mode de représentation des classes. Le problème se pose, alors, en termes de recherche simultanée de la classification et de la représentation des classes parmi un ensemble de partitions et de représentations possibles qui optimisent le critère. La première étape dans ce processus est la détermination de la configuration initiale qui dépend du nombre de classes et de la position des premiers noyaux qui définissent les centres de gravité des classes dans la première itération.

.1.2. Configurations initiales

Les méthodes non hiérarchiques commencent avec une partition initiale des données dans des groupes ou avec un ensemble de points autour desquels les groupes vont se former. Un ensemble de k points peut être utilisé comme un groupe de noyaux autour duquel l'ensemble des m données peuvent être groupées suivant un critère de similarité. Plusieurs procédés sont utilisés pour le tirage des premiers noyaux, appelés aussi centroïdes ou centres de classes, la valeur de k définit, également, le nombre de classe. En effet, chaque méthode de classification utilise un des types d'initialisation suivante :

1. Choisir les k premières données de l'ensemble des pixels.
2. Etiqueter les données de 1 à m et choisir celles étiquetées par $m/k, 2m/k, \dots, (k-1)m/k$.
3. Choisir subjectivement k données de l'ensemble des données.
4. Etiqueter les données de 1 à m et choisir de manière arbitraire k données de 1 à m .
5. Prendre une partition de k groupes mutuellement exclusifs et calculer les centres de gravité des groupes comme des données de classement.

Remarques

- Les méthodes partagent la propriété essentielle que chaque noyau est lui-même un élément de l'ensemble. Toute classe aura, donc, au moins un élément.
- Les méthodes ont des éléments aléatoires soit à travers un ordre aléatoire implicite des noyaux dans l'ensemble des données, soit à travers une sélection aléatoire explicite.

- Les méthodes les plus adaptées à la classification d'images multispectrales sont basées sur une procédure qui consiste à alterner le calcul des noyaux successifs et l'affectation des données à des classes jusqu'à la convergence vers une configuration stable.

A.1.3. Critères de similarité ou ressemblance

La similarité est, par définition, toute application à des valeurs numériques qui permet de mesurer le lien entre les individus d'un même ensemble. Ce lien est d'autant plus fort que sa valeur est grande. Dans le cas de la télédétection, la classification implique la partition de pixels dans un espace multispectral. La mesure de similarité entre deux éléments est quantifiée par la distance radiométrique entre eux.

Plusieurs distances (ou indices de similarité) métriques existent, chacune d'elles a son importance en analyse de données. Les distances les plus couramment utilisées sont :

- **Distance euclidienne simple**

La distance euclidienne est la distance la plus répandue dans les procédures de groupement. Son expression analytique est donnée par la relation suivante :

$$d^2(x, k_j) = \sum_{i=1}^n (x^i - k_j^i)^2 \quad (\text{A.1})$$

Où : k_j est le noyau de la classe j .

x est le pixel à classer, il a n composantes qui correspondent au nombre d'images :

(A.2)

- **Distance de Mahalanobis**

La distance de Mahalanobis est utilisée fréquemment en analyse de données comme une fonction discriminante. Son expression analytique est la suivante :

(A.1)

Où : S^{-1} est la matrice de variance-covariance inverse des noyaux des classes.

A.1.4. Etapes d'un algorithme de classification automatique

- On détermine k centres provisoires de chaque classe. Ces centres induisent une partition de C en k classes $\{C_1^0, C_2^0, \dots, C_q^0, \dots, C_k^0\}$. Ainsi le pixel P appartient à la classe C_i^0 s'il est plus proche du centre C_i^0 que tous les autres centres.
- On détermine k nouveaux centres de classes C_i^j en prenant les centres de gravité des classes $\{C_1^0, C_2^0, \dots, C_q^0, \dots, C_k^0\}$. Ces nouveaux centres induisent une nouvelle partition construite selon la même règle de décision et forme de nouvelles classes notées C_i^j . Le nouveau noyau, pour une classe donnée à l'itération j , est donné par la relation :

$$(A.3)$$

Où : P_i^j sont les pixels de la classe C_i^j à l'itération j .

N_i^j est le nombre total des pixels affectés à la classe C_i^j à l'itération j .

- Le processus se poursuit sur le même principe en alternant les étapes de calcul des nouveaux centres et d'affectation des points P_i^j dans les nouvelles répartitions des classes C_i^j .
- L'algorithme s'arrête lorsque deux itérations successives conduisent à la même partition au critère de précision près. Ce critère peut être défini et fixé de deux manières :

1^{er} cas : on fixe *a priori* le nombre d'itération sans regarder les changements de la partition de l'itération $j - 1$ à l'itération j .

2^{ième} cas : on se base sur les mouvements des pixels entre les classes de l'itération $j - 1$ à l'itération j et on fixe un seuil correspondant au nombre total des pixels ayant changé de classes entre deux itérations successives. Plus ce nombre est petit, plus le nombre d'itérations est grand.

A.2. Classification supervisée

Pour ce type de classification d'images, on dispose d'une connaissance *a priori* des caractéristiques des objets qui doivent être discriminés sur la zone d'étude. A cet effet, la première étape dans un processus de classification supervisée est la localisation et l'identification des échantillons représentatifs des thèmes sur la base de cette connaissance. Les méthodes d'extraction de ces échantillons ou de la base d'entraînement dite « base de connaissance » sont de deux types : l'extraction automatique et l'extraction interactive.

Dans la première méthode, difficile à mettre en œuvre, la base d'entraînement est générée par élargissement de régions en utilisant des critères statistiques sur le voisinage de chaque pixel analysé, le résultat de cette phase est sous forme d'image « masque », c'est à dire chaque thème est représenté par des petites régions prises comme échantillons de la classe.

La deuxième méthode, la plus répandue, est une méthode interactive. Elle consiste à localiser visuellement sur l'image les classes d'intérêt en utilisant des données extérieures à l'image telles que les photographies aériennes, les cartes thématiques ou des connaissances terrain.

Pour le choix des zones d'entraînement, la méthode interactive consiste à localiser les thèmes d'intérêt et délimiter des parties de ces derniers en respectant certains critères qui sont : la répartition sur l'image, le nombre de pixels par partie d'une classe et la représentativité en termes d'occupation du sol. Pour ce dernier point, il faut vérifier soit sur le terrain, soit sur des documents existants de la correspondance entre le thème et les échantillons associés.

Il est à noter que cette phase est très importante dans le processus de classification supervisée. La précision des résultats de la classification est conditionnée, principalement, par cette étape. Généralement, la base d'entraînement extraite est divisée en deux parties. Une partie des échantillons est utilisée pour établir le modèle de décision (règle de décision) et la deuxième partie dite base de contrôle est utilisée pour contrôler et valider les résultats de la classification « étape de post-classification ».

Le processus de classification supervisée est schématisé par le synoptique de la figure (A.1) :

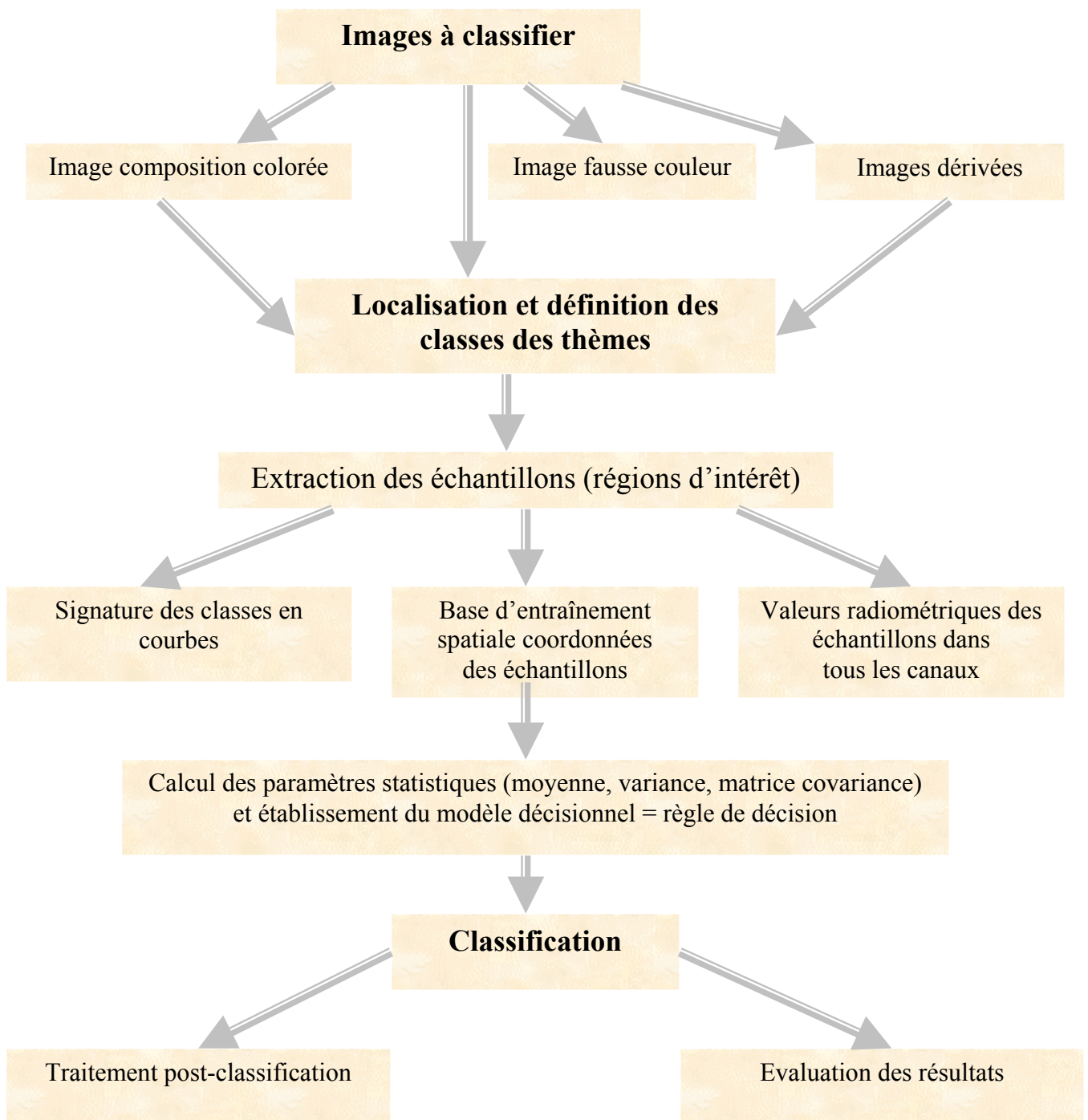


Figure A.1. Synoptique de la classification supervisée.

A.2.1. Base d'entraînement

La méthode interactive d'extraction des échantillons dépend du type de logiciel de traitement d'images utilisé. Dans certains logiciels, la délimitation des échantillons se fait par des formes géométriques telles que le rectangle ou le carré, d'autres utilisent le tracé par des polygones, des polygones ou des points. Par ailleurs, d'autres logiciels tels que l'ENVI offre à

l'utilisateur un outil d'extraction des échantillons par la délimitation des nuages des points d'une classe sur le scattérogramme et le report de l'information radiométrique de ce nuage sur les canaux à classifier en générant ainsi les régions réparties spatialement sur les images, les signatures des classes ainsi que des données radiométriques de chaque classe en ASCII.

A.2.2. Etablissement du modèle de décision

Les procédures de mise en œuvre de méthodes et de définition des règles de décision de classification sont à la base de nombreuses techniques d'analyse d'images. Ces techniques sont utilisées pour segmenter l'espace spectral afin de permettre de définir à quelle classe un pixel doit être affecté. Elles sont dites modèles décisionnels ou règles de décisions ou encore fonctions discriminantes. Les modèles de décision sont divisés en deux catégories essentielles : les modèles ou règles de décision paramétriques et règles de décision non paramétriques.

La première catégorie s'appuie sur une connaissance des paramètres statistiques des classes. Quant à la deuxième catégorie, elle n'utilise aucune connaissance *a priori* des statistiques associées aux classes considérées.

A.2.2.1. Techniques non paramétriques

Parmi ces méthodes, on trouve la méthode parallélépipédique (hypercubes) et la méthode des minima de distances :

1. Méthode de parallélépipèdes

Le principe de cette méthode, dite aussi arithmétique, est la localisation sur l'histogramme de chaque canal les valeurs radiométriques minimales et maximales de chaque classe. Ces valeurs servent comme bornes pour délimiter le parallélépipède de la classe. Tous les pixels dont leurs coordonnées radiométriques sont à l'intérieur du parallélépipède sont affectés à la même classe. Alors que les pixels qui se trouvent à l'extérieur des hypercubes constituent une classe de rejet et enfin les pixels qui se trouvent dans les zones de chevauchement des hypercubes sont souvent mal classés (erreur d'affectation). Il est à noter que cette méthode n'obéit pas à une règle de décision mathématique et plus les zones de confusion sont importantes, plus le risque d'erreur est grand.

2. Méthodes des minima de distances

Ces méthodes sont basées sur la mesure d'une distance entre le pixel à classer et le paramètre caractéristique d'une classe. Le principe reste le même dans le cas de la classification non supervisée, à la différence que dans un cadre supervisé, les moyennes des classes (centres de gravité) sont connues. Dans ce cas, nous avons la règle de décision du minimum de distance euclidienne dont l'expression est la suivante :



(A.4)

(A.5)

est le vecteur moyenne de la classe i .

n est le nombre de canaux.

3. Règle de décision de classification

Un point X d l'image appartient à la classe C_i si le carré de la distance par rapport à cette classe est la plus petite que toute autre distance. Autrement dit :

si pour tout (A.6)

L'expression de cette distance est dite fonction discriminante.

Remarques

- On suppose pour ces modèles de décision que les données des classes sont distribuées selon une loi normale. Dans le cas du minimum de distance, les variances sont considérés égaux dans toutes les composantes.

- Les classes sont représentées par des sphères dans l'espace nD dont le rayon dépend de la dispersion du nuage de points de la classe. A cet effet, on peut introduire un seuil sur la distance pour laquelle toute distance supérieure implique un pixel de rejet (l'ensemble forme une classe de rejet).

A.2.2.2. Technique paramétriques

La classification est un processus de prise de décision judicieux sur les données présentant une grande variabilité. Pour mieux réaliser ce processus, il faut associer à cette décision une certaine probabilité de commettre une erreur d'affectation. En effet, la démarche de prise de décision d'affectation de chaque pixel doit être contrôlée par un critère de minimisation d'erreur d'affectation sous ces hypothèses. Les fonctions paramétriques vont donc s'appuyer sur les lois probabilistes et la règle de décision sur un pixel est liée à une théorie de probabilité du pixel lui-même et des classes considérées *a priori*. La définition mathématique des lois de probabilité est soumise à une hypothèse essentielle qui est la forme de distribution des données traitées. Dans le cas général, les données d'une classe suivent une distribution gaussienne.

Hypothèse gaussienne

Chaque classe est caractérisée par un certain nombre de pixels choisis *a priori*, si ce nombre est suffisamment grand (condition statique), on peut établir l'histogramme de la classe qui aura la forme d'une fonction de densité de probabilité $P(x / C_i)$ donnée par :

(A.7)

Pour chaque pixel, on peut calculer la fonction de densité associée à la classe C_i , caractérisé par sa moyenne et sa variance. Ce pixel est affecté à cette classe si la probabilité $P(x / C_i)$ est supérieure à toutes probabilités du pixel x associées aux autres classes à savoir :

$$P(x / C_i) > \dots \text{ pour : } \dots \quad (\text{A.8})$$

Seuil de probabilité minimale

Comme le processus de classification est initié par un certain nombre de classes, il y a donc toujours une des probabilités du pixel associée aux classes qui est supérieure. De ce fait, certains pixels qui ne correspondent à aucune classe seront malgré tout affectés à une des classes. Pour palier à ce problème, on introduit un seuil de rejet. La valeur la plus forte de $P(x / C_i)$ trouvée doit être supérieure à une valeur minimale pour que le pixel soit affecté à une classe. Dans le cas où cette condition n'est pas vérifiée, le pixel est affecté à une classe de rejet.

A.2.3. Evaluation de la classification et post-traitement

Cette partie portera sur la validation des résultats de la classification pour la caractérisation et l'identification des thèmes par rapport à la réalité terrain. L'évaluation est soit statistique (quantitative) ou qualitative par comparaison aux données existantes.

Les résultats de la classification sont analysés ou traités pour leur exploitation dans la réalisation des couches thématiques. A cet effet, des étapes de vectorisation automatiques ou manuelles seront effectuées ainsi que la génération des masques propres à chaque thème sur une scène satellitaire.

Bibliographie

« car il est parfois facile une fois qu'on acquiert une certaine
connaissance des questions, d'en imaginer ensuite la démonstration,
que si l'on cherchait sans aucune notion au préalable»

Archimède (-287 (avant l'ère chrétienne)-212)

- [ANU, 1977] **Anuta P.E.** « Computer-Assisted Analysis Techniques for Remote Sensing Data Interpretation ». *Geophysics*, 42 (3), pp: 468-481. (1977).
- [ARO, 1982] **Aronoff S.** « The Mapp Accuracy Report ». a users view *Photogrammetric Engineering and Remote Sensing*. 48, pp: 1309-1312. (1982).
- [ARR, 1996] **Arrouy William** « Les Réseaux de Neurones ». Paris, (1996). site <http://www.tls.mms.fr>.
- [BAR, 2002] **Bardossy A. & Samaniego L.** «Fuzzy Rule-Based Classification of Remotely Sensed Imagery ». *IEEE Transactions on Geoscience and Remote Sensing*.vol.40, Number 2. pp 362-374. February 2002.
- [BEL, 1998] **Belhadj-Aissa A.** « Contribution au logiciel d'Analyse et de Traitement d'Images Satellitaires (LATIS). Analyse spatiale-Spectrale d'Images Satellitaires Appliquées à la cartographie Thématique ». Thèse de doctorat d'Etat en Electronique, spécialité : Traitement d'Images et Télédétection. USTHB. p.268. Avril 1998
- [BOW, 1992] **Bow Sing-Tze** « Pattern Recognition and Image Processing ». Edition Marcel Decker. (1992).
- [CAM, 1981] **Campbell N.A.** « Spatial Correlation Effects Upon Accuracy of Supervised Classification of Land Cover ». *Photogrammetric Engineering and Remote Sensing*. 47, pp: 355-363. (1981).
- [CAR, 1992] **Carlos E. Mejia** « Architectures Neuronales pour l'Approximation des Fonctions de Transfert : Application à la Télédétection ». Thèse de Doctorat. Université de Paris Sud Centre d'Orsay. Octobre 1992.
- [CHI, 1994] **Chitroub S.** « Implémentation d'un Processus d'Analyse et de Classification d'Images Satellitaires. Application à la Caractérisation des Etats de Surfaces ». Thèse de Magister en Electronique des Systèmes. Institut d'Electronique. USTHB. Mai 1994.
- [CHI, 2003] **Chitroub S.** « Modélisation, Analyse, Classification et Fusion d'Images de Télédétection Multisources Optiques et Radars ». Thèse de doctorat d'Etat en Electronique, spécialité : Electronique des systèmes. USTHB. p.237. Février 2003.
- [CHU, 1988] **Chuvieco E. & Congalton R.G.** « Using Cluster Analysis to Improve the selection of Training Statistics in Classifying Remotely Sensed Data ». *Photogrammetric Engineering and Remote Sensing*. 54 (9), pp: 1275-1281. (1988).

- [COG, 1973] **Coggeshall M.E. & Hoffer R.M.** « Basic forest and Cover Mapping Using Digitized Remote Sensor Data and ADP Techniques ». LARS Information Note 030573. Purdue University. (1973).
- [DAV, 1990] **Davalo E. et Patrick N.** « Des Réseaux de Neurones ». Edition Eyrolles. Paris 1990.
- [DEV, 1997] **Devaux F.** « Filtrage d'images par réseaux de neurones ». Maîtrise d'Informatique. Option Micro-informatique Micro-électronique. Département MIME. Université de Paris. Septembre 1997.
- [DJE, 2004] **Djerriri K.** « Approche Neuro-Génétique pour la Classification des Images de Télédétection Intégrant l'Information de Texture ». Mémoire de Magister en Techniques spatiales et Appliquées, option : Télédétection. Centre National des Techniques Spatiales (CNTS). p.116. Octobre 2004.
- [DRE, 2002] **Dreyfus G., Martinez J.M., Samuelides M., Gordon M.B., Badran F., Thiria S. & Hérault L.** « Réseaux de Neurones – Méthodologies et Applications ». Edition Eyrolles. Paris 2002.
- [EYT, 1983] **Eyton J.R.** « A Hybrid Image Classification Instructional Package ». Photogrammetric Engineering and Remote Sensing. 49 (8), pp: 1175-1181. (1983).
- [FOO, 1988] **Foody G.M.** « The Effects of Viewing Geometry on Image Classification ». International Journal of Remote Sensing. 9, pp: 1909-1915. (1988).
- [FOO, 1992] **Foody G.M., Campbell N.A., Trodd N.M. & Wood T.F.** « Derivation and Applications of Probabilistic Measures of Class Membership from the Maximum-Likelihood Classification ». Photogrammetric Engineering and Remote Sensing. 8 (9), pp: 1335-1341. (1992).
- [FRE, 1992] **Freeman J.A. & Skapura D.M.** « Neuronal Networks: Algorithms application and programming techniques ». Addison-Wesley Publishing Company (1992).
- [HER, 1994] **Hérault J. et Christian J.** « Réseaux neuronaux et traitement du signal ». Edition HERMES. Paris 1994.
- [HOD, 1989] **Hodgson M.E. & Plews R.W.** « N-Dimensional Display of Cluster Means in Feature space ». Photogrammetric Engineering and Remote Sensing. 55 (5), pp: 613-619. (1989).
- [HOL, 1975] **Holland J.H.** « Adaptation in Natural and Artificial Systems ». University of Michigan Press, Ann Arbor, MI, Internal report. (1975).

- [HOP, 1982] **Hopfield J.J.** « Neural Networks and Physical Systems with Emergent Collective Computational Abilities ». *Proc. Nat. Sci.* vol.79. Avril 1982.
- [JEN, 1979] **Jensen J.R.** « Computer Graphics Feature Analysis and Selection ». *Photogrammetric Engineering and Remote Sensing.* 45 (11), pp: 1507-1512. (1979).
- [JOD, 1995] **Jodouin J-F.** « Les Réseaux de Neurones : Principes et Définitions ». Edition Hermes. Paris 1995.
- [KHE, 2001] **Khedam R.** « Etude et développement de méthodes de classification supervisée contextuelle d'images satellitaires ». Thèse de Magister en Electronique, spécialité : Electronique des systèmes. Laboratoire de traitement d'images et rayonnement. USTHB. p.154. Novembre 2001.
- [KHE, 2003] **Khedam R., Belhadj-Aissa A.** « Contextual Fusion by Genetic Approach Applied to the Classification of Satellite Images ». Earsel Symposium Ghent, Belgium. June 2003.
- [KOH, 1984] **Kohonen T.** « Self Organization and Associative Memory ». Second Edition, Springer Series in Information Sciences. Vol. 8, Springer Verlag, Berlin, ISBN 3-540-18314-0. (1987).
- [KOS, 1999] **Koss J.E., Newman F.D., Johnson T.K. & Kirch D.L.** « Abdominal organ segmentation using texture transforms and a Hopfield Neuronal Network ». *IEEE Transactions on Medical Imaging.* vol.18. Number 7. pp.640-648 (1999).
- [LAC, 1996] **Lacaille J.** « Réseaux de neurones ». Rapport de recherche, DIAM – CMLA ENS de Cachan. Septembre 1996.
- [LAN, 1988] **Langlois J.** « Les Réseaux de Neurones ». Rapport de Recherche. Université de Sherbrooke. (1988).
- [LIB, 1989] **LibbeyEurotext J.** « Le traitement des images de télédétection : aperçus et perspectives ». Edition AUPELF-UREF. p.97-110. Paris 1989.
- [MAH, 2000] **Mahiddine L., Zenati N. et Achour K.** « Algorithme de Mise en Correspondance Stéréo Par Approche Neuronale ». Centre de Développement des Techniques Avancées (CDTA). Laboratoire de Robotique et d'Intelligence Artificielle. Article interne. (2000).
- [MAN, 1994] **Mangolini M.** « Apport de la Fusion d'Images Satellitaires Multicapteurs au Niveau Pixel en Télédétection et Photo-Interprétation ». Thèse de Doctorat en Sciences. Spécialité Sciences de l'Ingénieur. Université de Nice – Sophia Antipolis.. p.174. Novembre 1994

- [MOR, 1989] **Mori M. & Goth K.** « Tree-Dimensional Analysis of SPOT HRV Data ». Proceeding of the 12th Canadian Symposium on Remote Sensing, 9th Annual International Geoscience & Remote Sensing Symposium. I'GARSS 89, 2: 467-470. (1989).
- [MUH, 2004] **Muhlenbach F.** « Réseaux de neurones artificiels - Apprentissage automatique ». Master 1. Informatique. (2004).
- [NAS, 1992] **Nasrabadi N. & Choo C.Y.** « Hopfield Network for Steréovision correspondance ». IEEE. Transactions on Neuronal Networks. vol.3. Number 1 (1992).
- [PER, 1988] **Personnaz L., Drefys G. et Guyon I.** « Les Machines Neuronales » La Recherche – Spécial les Nouveaux Ordinateurs. Numéro 204. Novembre 1988.
- [ROS, 1986] **Rosenfield G.M. & Fitzpatrick-Lins K.** « A Coefficient of Agreement as a Measure of Thematic Classification Accuracy ». Photogrammetric Engineering and Remote Sensing. 52, pp: 223-227. (1986).
- [SMA, 1998] **Smara Y.** « Contribution au Système LATIS. Evaluation de Correspondance Données Terrain et Classification d'Images Satellitaires ». Thèse de doctorat d'Etat en Electronique, spécialité : Traitement d'Images et Télédétection. Institut d'Electronique. USTHB 1998.
- [SWA, 1978] **Swain P.H. & Davis S.H.** « The Quantitative Approach ». Remote Sensing. McGraw-Hill. (1978).
- [TOU, 1992] **Touzet C.** « Les Réseaux de Neurones Artificiels. Introduction au Connexionnisme. Cours, Exercices et Travaux Pratiques ». Juillet 1992.
- [TSA, 1992] **Tsai W. & Yu S.** « Relaxation by the Hopfield Neuronal Network ». Pattern Recognition. Vol.25. Number 2. pp.197-200 (1992).
- [WEI, 1999] **Weizang Z.** « Verification of the Nonparametric Characteristics of Backpropagation Neuronal Networks for Image Classification ». IEEE, Transaction Geoscience and Remote Sensing, vol.37. Num 2 (1999).