

Commande Contribution à la classification automatique des données multidimensionnelles en utilisant les réseaux de neurones compétitifs Hebbians : Application à la segmentation des images couleurs

L'augmentation du volume de données, de plus en plus de champs scientifiques ont besoin de leurs données pour des objectifs décisionnels ou descriptifs, d'autant plus de sommes demandés à traiter des données plus complexes (images, vidéos ou textes ...), l'analyse d'étude très important. La reconnaissance de forme est une large discipline issue de l'analyse de données, qui englobe toutes les techniques de classification automatique, dans le but de donner à des machines la capacité de perception aussi performante que possible à celle des humains (ou animal), voir même plus. La classification automatique est relativement difficile, surtout quand les données sont placés dans un contexte non supervisé. Par conséquent, de nombreuses techniques ont été développées dans ce domaine, mais la classification utilisant les réseaux de neurones compétitifs est en termes de qualité et de temps. Dans notre étude, nous nous sommes concentrés sur la classification de données en termes de qualité des résultats de la classification des données en utilisant des réseaux de neurones compétitifs, qui se caractérise par un phénomène de compétition et une conscience de la forme. L'étude que nous mènerons consiste à développer une approche locale non supervisée de classification, basée sur un réseau de neurones à apprentissage compétitif qui permettra de dépasser l'approche Neuromimétique. A savoir, la faiblesse de la classification des données est due à la concurrence des neurones et le seuillage, ce qui influence grandement l'automatisme et l'utilité de la classification. La nouvelle approche est constituée de quatre étapes : (i) l'estimation de la fonction de densité de probabilité (fdp), (ii) l'extraction des maxima locaux par un réseau de neurones compétitif avec la méthode de Mahalanobis comme une fonction d'activation, (iii) la détection des modes en utilisant une approche basée sur la méthode d'apprentissage compétitif Hebbien (CHL), (iv) l'affectation du reste des données aux classes désignées par les modes détectés. L'approche proposée est appliquée au début, sur un ensemble d'échantillons de données multidimensionnelles. Pour être appliquée à la suite sur la classification des images couleur et en particulier, les images TEP (Tomographie par Emission de Positons) et les images médicales. En comparaison avec plusieurs méthodes de classification, notre approche a fait preuve de sa bonne performance, en vertu d'un ensemble d'exemples synthétiques et de données réelles, sans aucun seuillage et ne nécessiter aucune information préalable sur le nombre des classes ou la structure de leurs distributions dans l'ensemble de données.

Classification automatique, fonction de densité de probabilité (fdp); Réseaux de neurones compétitifs Hebbians, Classification de Mahalanobis, Competitive Hebbian Learning (CHL), seuillage, Image couleur,

Année : 2017



École Nationale Supérieure d'Informatique et d'Analyse des systèmes
Centre d'Études Doctorales en Sciences des Technologies de l'Information et de l'Ingénieur

THÈSE DE DOCTORAT

CONTRIBUTION A LA CLASSIFICATION AUTOMATIQUE DES DONNEES MULTIDIMENSIONNELLES EN UTILISANT LES RESEAUX DE NEURONES COMPETITIFS HEBBIANS : APPLICATION A LA SEGMENTATION DES IMAGES COULEURS

Présentée par

Meriem TIMOUYAS

Le 29/12/2017

Formation doctorale : Informatique

Structure de recherche : Laboratoire de Génie Electrique ENSET de Rabat

JURY

Professeur Mohamed OUADOU

Faculté des sciences, Université Mohammed V de Rabat

Président

Professeur Ahmed HAMMOUCH

CNRST, Rabat

Directeur de thèse

Professeur Souad EDDAROUICH

Centre d'Education Régional, Rabat

Co-Encadrant de thèse

Professeur Ouadoudi ZYTOUNE

ENSA, Université Ibn Tofail, Kénitra

Rapporteur

Professeur Youssef FAKHRI

Faculté des sciences, Université Ibn Tofail, Kénitra

Rapporteur

Professeur Abdennaser BOUROUHOU

ENSET, Université Mohammed V de Rabat

Examineur

Meriem TIMOUYAS

Contribution à la classification automatique des données multidimensionnelles en utilisant les réseaux de neurones compétitifs Hebbians : Application à la segmentation des images couleurs.

Année : 2018 N° thèse : 102/ST2I

A mes parents

A mes sœurs et frères

A ma grande famille et mes très chers amis

REMERCIEMENTS

Ce travail a été effectué au Laboratoire de recherche de génie électrique LRGE de l'école ENSET sous la direction du Professeur Ahmed HAMMOUCH.

Je remercie tout d'abord mon Dieu pour tous ces dons myriades, Merci mon Dieu.

Je tiens à exprimer ma gratitude à mon directeur de thèse Pr. Ahmed HAMMOUCH pour son soutien scientifique, pour ses qualités d'encadrements et pour ses conseils permanents tout au long de mon travail.

J'adresse également mes remerciements au Pr. Souad EDDROUICH pour m'avoir co-encadré durant cette thèse pour ses conseils précieux, sa disponibilité, son soutien ainsi que pour ses conseils très utiles qui m'ont aidé à enrichir et à mener à bien ce travail.

Par ailleurs j'adresse mes gratitudes à Pr. Mohamed ESSAIDI directeur de l'école ENSIAS pour sa serviabilité et sa gentillesse.

Je tiens aussi à remercier les membres du jury, pour avoir accepté de me faire profiter de leurs compétences pour l'évaluation de ce travail.

J'adresse vivement mes remerciements à ma famille, et plus particulièrement mes petites sœur Sakina et Yasmine, qui ont été toujours une source de bonheur.

Je tiens aussi à adresser mes gratitudes à tous mes professeurs durant toutes les années de mes études, en particulier mon cher professeur des Maths Mr Mohamed AJAAMOUD pour sa méthodologie et sa qualité d'enseignement.

Et finalement j'adresse mes remerciements à tous ceux qui, de loin ou de près, m'ont aidé, que ce soit par leurs compétences, leurs savoir-faire ou leurs amitiés.

RESUME

Devant l'augmentation du volume de données, de plus en plus de champs scientifiques ont besoin de catégoriser leurs données pour des objectifs décisionnels ou descriptifs, d'autant plus qu'aujourd'hui, nous sommes demandés à traiter des données plus complexes (images, vidéos ou textes ...), ce qui incite un domaine d'étude très important.

La reconnaissance de forme est une large discipline issue du champ d'analyse de données, qui englobe toutes les techniques de classification automatique, dans l'objectif de donner à des machines la capacité de perception aussi performante que possible à celle des êtres vivants (humain ou animal), voir même plus.

La classification automatique est relativement difficile, surtout lorsque nous sommes placés dans un contexte non supervisé. Par conséquent, de nombreuses méthodes ont été développées dans ce domaine, mais la classification utilisant les réseaux de neurones reste la plus efficace en termes de qualité et de temps.

Dans notre étude, nous nous sommes concentrés sur l'amélioration de la qualité des résultats de la classification des données en utilisant des réseaux de neurones à apprentissage compétitif, qui se caractérise par un phénomène de compétition et une conscience entre les neurones.

L'étude que nous mènerons consiste à développer une approche locale non supervisée pour la classification, basée sur un réseau de neurones à apprentissage compétitif qui permettra de surmonter les défis posés par l'approche Neuromimétique. A savoir, la faiblesse de la classification des distributions non gaussiennes et le seuillage, ce qui influence grandement l'automatisme et l'utilité de la classification.

Notre nouvelle approche est constituée de quatre étapes : (i) l'estimation de la fonction de densité de probabilité (fdp), (ii) l'extraction des maxima locaux par un réseau de neurones compétitif avec la distance de Mahalanobis comme une fonction d'activation, (iii) la détection des modes en utilisant une méthode développée de la méthode d'apprentissage compétitif Hebien (CHL), (iv) l'affectation du reste des données à leurs classes désignées par les modes détectés.

L'approche proposée est appliquée au début, sur un certain nombre d'échantillons de données multidimensionnelles. Pour être appliquée à la suite sur la segmentation d'images couleur et en particulier, les images TEP (*Tomographie par Emission de Positons*) de la maladie Alzheimer.

En comparaison avec plusieurs méthodes de classification, notre approche a fait preuve de son efficacité et de sa bonne performance, en vertu d'un ensemble d'exemples synthétiques et réels, sans passer par aucun seuillage et ne nécessiter aucune information préalable sur le nombre des classes, ni sur la structure de leurs distributions dans l'ensemble de données.

Mots Clés : Classification automatique, fonction de densité de probabilité (fdp); Réseaux de neurones compétitifs, Distance de Mahalanobis, *Competitive Hebbian Learning* (CHL), seuillage, Image couleur, Image TEP.

ABSTRACT

Given the increasing volume of data, more and more scientific fields need to categorize their data for decision-making or descriptive purposes, especially as today we are asked to process more complex data (images , Videos or texts,...), which incites a very important field of study.

Form recognition is a wide discipline of data analysis, encompassing all techniques of clustering, with the aim of giving machines the perception ability as efficiently as possible to that of living beings (human or animal), or even more

Cluster analysis is relatively difficult, especially when we are placed in an unsupervised context. Therefore, many methods have been developed in this area, but the clustering using neural networks remains the most efficient in terms of quality and time.

In our study, we focused on improving the quality of data classification results using competitive learning neural networks, which is characterized by a competition phenomenon and conscience between neurons.

The study that will lead consist to the development of an unsupervised local approach for clustering, based on a competitive learning neural network, that will overcome the challenges posed by the neuromimetic approach. Namely, the weakness of non-Gaussian distributions clustering and thresholding, which greatly influence the automaticity and utility of the cluster analysis.

Our new approach consists of four phases: (i) estimation of the probability density function (pdf), (ii) extraction of local maxima by a competitive neural network with the Mahalanobis distance as an activation function, (iii) mode detection using a developed method of the competitive Hebian learning (CHL) method, (iv) assignment of the remainder of the data to their classes designated by the detected modes.

The proposed approach is applied first to a number of multidimensional data samples. To be applied further on the segmentation of color images, in particular, PET (*Positron Emission Tomography*) images of Alzheimer's disease.

Compared with several clustering methods, our approach has demonstrated its efficiency and good performance, using a set of synthetic and real examples, without going through any

thresholding and requiring no prior information either on the number of classes, nor on the structure of their distributions in the dataset.

Keywords: Cluster Analysis, Probability density function (pdf); Competitive neural networks, Mahalanobis Distance, Competitive Hebbian Learning (CHL), Thresholding, Color image, PET image

TABLE DES MATIERES

INTRODUCTION GENERALE.....	1
CHAPITRE I : Analyse des données : de la classification automatique à la segmentation d'image	5
I.1. Introduction	5
I.2. Classification automatique.....	6
I.2.1. Préparation des données	7
I.2.1.1 Représentation des données.....	7
I.2.1.2 Sélection des caractéristiques pertinentes.....	8
I.2.1.3 Critère de ressemblance.....	10
I.2.2. Méthodes de classification	12
I.2.2.1 Méthodes de classification classiques	12
I.2.2.1.1 Procédures globales	13
I.2.2.1.2 Procédures Locales.....	18
I.2.2.2 Méthodes de classification floue	24
I.2.3. La validation de la classification	25
I.3. Image et segmentation.....	26
I.3.1 Préambule.....	26
I.3.2 Image numérique.....	26
I.3.2.1 Définition.....	26
I.3.2.2 Image niveaux de gris.....	27
I.3.2.3 Image couleur	27
I.3.2.3.1 Définition.....	27
I.3.2.3.2 Espaces de représentation de couleur	28
I.3.2.4 Les images de TEP scan.....	31
I.3.3 Segmentation d'image numérique.....	32
I.3.3.1 Définition.....	32
I.3.3.2 Méthodes de segmentation	32
I.3.3.2.1 Approche Région.....	33
I.3.3.2.2 Approche Contour	35
I.3.3.3 Applications de Segmentation.....	35
I.3.3.3.1 Segmentation des images médicales : Image TEP	35
I.3.3.3.2 Segmentation de l'iris.....	38

I.4.	Conclusion	40
CHAPITRE II : Application des réseaux de neurones à la classification automatique des données multidimensionnelles		41
II.1.	Introduction	41
II.2.	Historique	42
II.3.	Du neurone biologique au neurone formel	43
II.3.1.	Le neurone biologique.....	43
II.3.1.1.	Définition.....	43
II.3.1.2.	Fonctionnement neurobiologique.....	44
II.3.2.	Le neurone formel	45
II.3.2.1.	Définition.....	45
II.3.2.2.	Fonctionnement d'un neurone formel	46
II.4.	Réseaux de Neurones Artificiels (RNA).....	49
II.4.1.	Architecture du réseau de neurones.....	49
II.4.1.1.	Les réseaux récurrents (Feedback)	50
II.4.1.2.	Les réseaux non bouclés (Feed-forward).....	50
II.4.2.	Apprentissage des réseaux de neurones.....	51
II.4.2.1.	Types d'apprentissage	51
II.4.2.2.	Règles d'apprentissage	53
II.5.	Apprentissage Compétitif (AC).....	53
II.5.1.	Apprentissage compétitif dur	54
II.5.1.1.	Principe d'apprentissage compétitif du réseau multicouche	55
II.5.1.2.	Algorithme d'apprentissage compétitif dur.....	57
II.5.1.3.	Fonction d'apprentissage.....	58
II.5.1.4.	Les extensions des modèles à apprentissage compétitif dur.....	60
II.5.2.	Apprentissage Compétitif Doux	62
II.5.2.1.	AC doux Avec Fixation de Dimensionnalité de Réseau	62
II.5.2.2.	AC doux Sans Fixation de Dimensionnalité de Réseau	64
II.6.	Conclusion	71
CHAPITRE III : nouvelle approche neuronale statistique et non paramétrique pour la classification des données multi dimensionnelles		72
III.1.	Introduction	72

III.2.	Estimation de la fonction de densité de probabilité.....	73
III.2.1.	Principe de l'estimation de densité.....	73
III.2.2.	Estimation par la méthode du noyau	75
III.2.3.	Algorithme rapide d'estimation de la fdp.....	77
III.2.3.1	Normalisation	78
III.2.3.2	Discrétisation.....	79
III.2.3.3	Estimation de la densité locale	80
III.2.3.4	Exemple modèle	81
III.3.	Extraction des maxima locaux de la fdp par apprentissage compétitif.....	83
III.3.1.	Architecture du Réseau de Neurones.....	84
III.3.2.	Apprentissage du Réseau.....	85
III.3.3.	Discussion	86
III.3.3.1	Choix de la distance.....	86
III.3.3.2	Collecte des échantillons	87
III.3.3.3	Comparaison entre les résultats de la distance euclidienne et la distance de Mahalanobis	89
III.3.4.	Algorithme d'apprentissage du réseau de neurones à apprentissage compétitif (RNAC).	90
III.4.	Détection des modes significatifs de la fdp	92
III.4.1.	La détection des modes par l'apprentissage compétitif Hebbien (CHL).....	92
III.4.1.1	Diagramme de Voronoi	93
III.4.1.2	Triangulation de Delaunay	94
III.4.1.3	Triangulation de Delaunay induite	95
III.4.1.4	Pseudo algorithme d'apprentissage	96
III.4.2.	Discussion	96
III.5.	Classification des données	103
III.5.1.	Assignation simple des observations au noyau le plus proche (ASNP)	103
III.5.2.	Pseudo algorithme	104
III.5.3.	Discussion	105
III.6.	Conclusion	106
CHAPITRE IV: Performance et Résultats expérimentaux de l'approche		108
IV.1.	Introduction	108
IV.2.	Application pour la classification des données.....	109

IV.2.1.	Exemple 1.....	109
IV.2.1.1.	Estimation des maxima locaux.....	110
IV.2.1.2.	Détection des modes : Comparaison	110
IV.2.1.3.	Classement des données	111
IV.2.2.	Exemple 2.....	112
IV.2.2.1.	Détection des modes.....	114
IV.2.2.2.	Classement des données	116
IV.2.3.	Exemple 3.....	116
IV.2.3.1.	Classification de la distribution.....	117
IV.2.3.2.	Comparaison entre l'approche proposée et l'approche Neuromorphologique	118
IV.3.	Application pour la segmentation des images.....	120
IV.3.1.	Ajustement de la procédure de détection des modes.....	120
IV.3.2.	Image couleur	121
IV.3.2.1.	Images synthèses	122
IV.3.2.1.1.	Exemple modèle 1	122
IV.3.2.1.2.	Exemple modèle 2	124
IV.3.2.2.	Images réelles.....	126
IV.3.2.2.1.	Évaluation qualitative.....	127
IV.3.2.2.2.	Évaluation quantitative.....	128
IV.3.3.	Images TEP scan de la maladie d'Alzheimer.....	131
IV.3.3.1.	Image TEP synthétique.....	131
IV.3.3.1.1.	Évaluation qualitative.....	133
IV.3.3.1.2.	Évaluation quantitative.....	134
IV.3.3.2.	Images TEP réelles.....	135
IV.4.	Discussion	136
IV.5.	Conclusion	138
	CONCLUSION GENERALE	140
	LISTE DES PUBLICATIONS.....	142
	BIBLIOGRAPHIE	143

TABLE DES FIGURES

Figure I.1: les différentes étapes de la classification automatique des données.....	6
Figure I. 2: les deux types de matrice représentant les données.....	7
Figure I.3: Schéma des procédures de classification automatique non supervisées.....	13
Figure I.4: Représentation graphique d'une distribution Gaussienne et son partitionnement.....	14
Figure I.5: les différentes représentations graphiques de la classification hiérarchique	15
Figure I. 6: Représentation graphique des données et de leur fonction de densité de probabilité.....	19
Figure I.7: Représentation d'une image niveaux de gris.....	27
Figure I.8: (a) synthèse additive, (b) synthèse soustractive.....	28
Figure I.9: Présentation spatiale du modèle RVB	29
Figure I.10: équipement et résultat d'une machine TEP-scanner ; (a) Appareil physique de TEP-scan, (b) image TEP d'une coupe transversale du cerveau	31
Figure I.11: présentation de quelques indices de quantification d'une image TEP [BUV-11]	37
Figure I.12: Vue frontale d'un œil humain avec ses différentes caractéristiques	38
Figure I.13: exemple de deux images d'iris segmentées par les deux approches (a) méthode de modélisation par cercle [HAN-07], (b) méthode de modélisation par contours elliptiques [GRAG-15]	39
Figure II.1: Schéma d'une cellule neuronale.....	44
Figure II.2: Le modèle du neurone formel	45
Figure II.3: neurone distance.....	47
Figure II.4 : Neurone formel avec deux entrées	48
Figure II.5: Interprétation géométrique de la réponse d'un neurone formel	48
Figure II.6: topologie des réseaux récurrents	50
Figure II.7: topologie des réseaux non bouclés	51
Figure II.8: Mode d'apprentissage supervisé	52
Figure II.9: Mode d'apprentissage non supervisé	52
Figure II.10 : connexion entre deux neurones.....	53
Figure II.11: Apprentissage de réseau à deux couches.....	55
Figure II.12 : Principe de l'apprentissage compétitif standard.....	57
Figure II.13 : Carte auto organisatrice.....	62
Figure II.14: apprentissage des cartes SOM de Kohonen	63
Figure II.15 : Diagramme de Voronoi représenté sur R^2	65
Figure II.16: la triangulation de Delaunay d'un ensemble de six points.....	66
Figure II.17: topologie des réseaux non bouclés	67
Figure II.18: la préservation de topologie des données par l'apprentissage du réseau de Gaz neuronal68	

Figure II.19: (a) Le DV et la TD de 14 points de données 2D. (b) TD induite des 14 points de données	70
Figure III.1: Espace de représentation des observations bidimensionnelles	78
Figure III.2: Plage de variation des observations normalisée	79
Figure III.3: Réseau de points d'échantillonnage (L=10)	79
Figure III.4: Estimation de la fdp dans l'espace discret	80
Figure III.5: Présentation tridimensionnelles des valeurs de la fdp sous-jacente à la distribution	81
Figure III.6: Ensemble des observations brutes	82
Figure III.7: Présentation de la fdp sous-jacente au quatre classes (L=25)	82
Figure III.8: Réseau compétitif à deux couches	84
Figure III.9: Maxima locaux extraits par réseau de neurones pour k=17	85
Figure III.10: analyse discriminante entre la distance DE et DM	86
Figure III.11: diagramme de collecte de données pour le calcul de la DM	88
Figure III.12: Présentation des 4 classes non gaussiennes et leur estimation de densité	89
Figure III.13: Comparaison entre les résultats de l'apprentissage utilisant la DE et celui utilisant la DM	90
Figure III.14: Diagramme de Voronoi des 17 maxima locaux de l'exemple modèle	94
Figure III.15: la triangulation de Delaunay des 17 maxima locaux de l'exemple modèle	95
Figure III.16: la triangulation de Delaunay induite de l'exemple modèle	95
Figure III.17: présentations des triangulations de Delaunay induite de l'exemple modèle pour différents nombres K de neurones de sortie	98
Figure III.18: présentation de la relation entre une distribution gaussienne et son écart-type	99
Figure III.19 : présentation de la relation entre deux une distribution gaussienne	100
Figure III.20: présentations des triangulations de Delaunay induite de l'exemple modèle pour différents nombres K de neurones de sortie par la nouvelle méthode	102
Figure III.21: Représentation de la distribution classée de l'exemple modèle	106
Figure IV.1: représentation des données brute de l'exemple 1	109
Figure IV.2 : La détection des maxima après l'estimation de la fdp sous-jacent au modèle 1	110
Figure IV.3: Les résultats des deux techniques de détection des modes de l'approche neuromimétique et l'approche proposée	110
Figure IV.4: Comparaison entre les résultats de l'approche neuromimétique et ceux de la nouvelle procédure	111
Figure IV.5: (a) représentation des données brute de l'exemple 2, (b) Estimation de la fdp sous-jacent à la distribution	113
Figure IV.6: La détection des 10 maxima après l'estimation de la fdp sous-jacent du modèle 2	114
Figure IV.7: distribution classifiée par l'approche proposée	116

Figure IV. 8: Classification par l'approche proposée: (a) Données brut, (b) Estimation de la fdp sous jacente, (c) 12 Maxima locaux détectés après 4 cycles; (d) Les composante connexes des maxima représentant les 3 classes, (e) distribution classifiée	118
Figure IV.9: (a) Observations classifiées par l'approche Neuromorphologique, (b) Observations classifiées par la méthode proposée	118
Figure IV.10: Effet du pas de discrétisation R sur le nombre de modes détectés	121
Figure IV. 11 : (a) image couleur originale; (b) présentation des pixels dans un espace RVB; (c) Estimation de la fdp sous-jacente; (d) Prototypes des classes dans l'espace RVB; (e) Prototypes connectés par des arrêtes ; (f) image couleur segmentée.....	123
Figure IV. 12: Les quatre classes de l'image modèle.....	124
Figure IV.13 : segmentation de l'exemple 2 par l'approche proposée.....	125
Figure IV.14: illustration des six régions de l'image segmentée séparées	125
Figure IV. 15: (a) image Originale; (b) Hypercubes de Pixels dans l'espace colorimétrique RVB discrétisé; (c) Modes détectés; (d) image segmentée	127
Figure IV.16: segmentation de deux images par l'approche proposée.....	128
Figure IV.17: Comparaison entre les paramètres statistiques des trois images.....	130
Figure IV. 18: Les différentes phases de la segmentation de l'image PET par la nouvelle approche : (a) Image MA originale ; (b) Pixels dans l'espace couleur RVB ; (c) Estimation de la fdp sous-jacente ; (d) 30 Maxima du pdf détecté après 7 cycles ; (e) Les modes détectés ; (f) Image MA segmentée avec des étiquettes de groupe colorées.....	132
Figure IV.19: La visualisation des différentes régions de l'image MA séparées (a) arrière-plan de l'image MA en niveaux de gris; De (b) à (e) la délimitation des quatre classes de premier plan.....	133
Figure IV.20: Comparaison des paramètres statistiques	134
Figure IV.21: Segmentation des images PET qui représentent 4 coupes axial du cerveau affecté d'MA par la nouvelle approche : de (a) à(d) Images TEP originales; de (a') à (d') Images TEP segmentées	135
Figure IV.22: La visualisation des différentes régions de l'image IV.16.(c) séparées : (a) arrière-plan de l'image MA en niveaux de gris; De (b) à (f) la délimitation des cinq classes de premier plan	136

LISTE DES TABLEAUX

Tableau II.1 : les fonctions d'activation courantes selon l'état de neurone	46
Tableau III.1: Paramètres statistiques de l'exemple modèle.....	82
Tableau III.2: Paramètres statistiques de l'exemple de comparaison	89
Tableau IV.1: paramètres statistiques de la classe gaussienne.....	109
Tableau IV.2: paramètres statistiques des 3 classes non gaussiennes.....	109

Tableau IV.3: Comparaison entre les matrices de confusion et les taux d'erreur des deux approches appliquées sur l'exemple1	112
Tableau IV. 4: Paramètres statistiques des deux classes gaussiennes.....	112
Tableau IV. 5: paramètres statistiques des 2 classes non gaussiennes.....	113
Tableau IV. 6: résultats des deux phases d'apprentissage pour différents nombres de neurones de sortie	115
Tableau IV. 7: Les paramètres statistiques des trois distributions gaussiennes de l'exemple 3	117
Tableau IV.8: Les résultats statistiques des deux approches appliquées sur des trois distributions gaussiennes de l'exemple 3.....	119
Tableau IV. 9: Paramètres statistiques des trois images segmentées par les deux méthodes	129
Tableau IV.10: Paramètres Statistiques.....	134

LISTE DES ABREVIATIONS

Abréviatiion	signification
DE	Distance euclidienne
DM	Distance de Mahalanobis
FCM	Fuzzy c-means
SOM	les cartes auto-organisatrices de Kohonen
GNG	Growing Neural Gaz
fdp	Fonction densité de probabilité
TEP	tomographie par émission de positons
SVM	Machine à vecteur de support
SUV	standardized uptake value
RNA	Réseau de neurones artificiel
AC	Apprentissage compétitif
DV	diagramme de Voronoi
TD	triangulation de Delaunay
CHL	Competitive Hebbian Learning
sd	Standard deviation
ASNP	Assignation Simple des Observations au Noyau le plus Proche
MA	maladie d'Alzheimer
ROI	Region Of Interest

INTRODUCTION GENERALE

Il est certes que l'apparition de l'informatique a fait une révolution dans le monde de l'analyse des données, qui était avant limitée à la statistique. Aujourd'hui, l'évolution de la technologie et l'existence des bases de données robustes, ont rendu l'analyse des données, notamment la classification automatique, un domaine de recherche très fécond appliqué presque dans toutes les disciplines.

La première étude théorique sur la classification a été consacrée au domaine de la biologie pour nous permettre de préciser les classes des espèces animales à partir de descriptions simples de ces animaux. Après cela, les biologistes, les historiens, les sociologues et les médecins ont exprimé leurs besoins pour classer les maladies ou leurs reliques.

Aujourd'hui, la classification automatique est intégrée dans les processus de traitement des données dans plusieurs domaines, utilisant parfois la classification sous différents noms dans différents contextes, tels que l'apprentissage compétitif (intelligence artificielle), la taxonomie numérique (biologie, écologie), la typologie (sciences sociales) et la partition (théorie des graphes) [THE-99].

Les éléments de notre environnement sont en effet classés par comparaison avec leurs semblables dans des modes de représentations relationnelles hiérarchiques. Cette démarche est relativement difficile à formaliser, surtout quand on se place dans un contexte non supervisé. C'est à dire, lorsqu'il s'agit d'identifier les classes en présence dans un échantillon à partir de la seule information qui peut être extraite des objets à classer.

Au cours de dernières décennies, de nombreuses méthodes de classification automatique ont été développées, ayant le même but, celui de classer les objets semblables dans la même classe et les objets dissemblables dans des classes différentes. En général, les objets sont caractérisés par des attributs qu'il est commode de représenter par des points dans un espace multidimensionnel.

Pour atteindre leur objectif, ces méthodes de classification adoptent diverses stratégies, certaines optimisent des critères qui reflètent les distances interclasses et intra-classes entre les objets [DEV-82]. Tandis que d'autres se réfèrent à des notions statistiques où l'on fait explicitement référence à la fonction de densité de probabilité (fdp) sous-jacente à la distribution des données à classer [POS-89].

En effet, sous l'hypothèse non paramétrique, cette dernière approche suppose que chaque groupement d'observations correspond à un mode de la fonction de densité de probabilité sous-jacente et le problème de la classification se trouve posé en termes de détection des modes, ce qui encourage la recherche dans ce sens.

Généralement, les méthodes classiques de classification sont inadaptées aux données enchevêtrées et non-sphériques, coûteuses en temps de calcul et leurs performances dépendent de plusieurs facteurs et nécessitent le seuillage ou le réglage de certains paramètres difficiles à fixer. Ce sont ces raisons qui ont motivées de plusieurs recherches, en essayant des nouvelles alternatives pour contourner les limitations de ces méthodes, une de ses alternatives les plus prometteuses est l'approche connexionniste. Cela est dû au fait qu'elle permet souvent d'allier la rigueur mathématique des méthodes classiques à la souplesse et la simplicité d'utilisation des réseaux de neurones artificiels [BIE-99].

En effet, les réseaux de neurones artificiels, étant inspirés des réseaux de neurones biologiques, ont pour objectif la simulation des comportements associés aux tâches sensorielles, afin de former un complémentaire qui s'occupe des tâches répétitives et fastidieuses et parfois impossible pour l'humain. Ils impliquent d'avantages des unités de processeurs nommées neurones formels connectés entre eux, dans des architectures statiques ou dynamiques, pour effectuer des opérations de traitement d'information en parallèle. Ce qui permet d'atteindre des vitesses de calcul très importantes.

Par ailleurs, la propriété la plus marquante du réseau de neurones est sa capacité d'apprentissage fortement autonome. Ce concept a permis d'importants progrès en intelligence artificielle. En effet, les algorithmes d'apprentissage modélisent un processus dynamique de discrimination complexe à partir d'exemples en entrée, permettant de modifier les paramètres d'un réseau en réaction avec les stimuli qu'il reçoit des exemples, cela fait des réseaux de neurones des méthodes d'emploi très général, robustes et tolérantes aux fautes.

La classification par les réseaux de neurones s'est avérée bien adaptée à la reconnaissance de formes, notamment la segmentation d'images. En considérant l'analogie entre la classification et la segmentation, les méthodes de segmentation par classification de pixels construisent les classes en privilégiant les propriétés colorimétriques des pixels.

La segmentation d'images couvre un domaine très vaste, y compris l'imagerie médicale, qui ne cesse de se développer technologiquement, précisément les récentes modalités d'imagerie, telles que l'Imagerie par Résonance Magnétique fonctionnelle (IRMf) et la Tomographie par

Emission de Positons (TEP), qui fait le thème de l'application de notre travail, et la plus récente l'IRM-TEP.

C'est dans ce contexte qu'a été conduit notre travail de recherche. Ainsi, nous proposons, dans cette thèse, une nouvelle approche neuronale de classification automatique non supervisée, fondée sur la détection des modes de la fdp et l'apprentissage compétitif, qui serait ensuite appliqué sur la segmentation des images TEP de la maladie Alzheimer. Ce qui distingue notre approche est qu'elle ne passe par aucun seuillage et ne nécessite aucune information à priori sur le nombre de classes ni sur la structure de leurs distributions dans l'ensemble de données.

Organisation de la thèse :

Cette thèse est composée de quatre chapitres organisés sur deux parties ; partie bibliographique et partie expérimentale :

La première partie débute par un chapitre divisé à son tour en deux sections ; la première section introduit l'analyse de données et expose une typologie des méthodes de classification automatique adoptée dans notre travail. La deuxième section est dédiée à la présentation de l'image numérique, des différentes méthodes de segmentation et certaines applications de la segmentation d'images.

Dans le deuxième chapitre, nous abordons tout d'abord les notions fondamentales nécessaires à la définition d'un réseau de neurones et à l'étude de sa dynamique. Ensuite nous développons plus en détail les principaux réseaux de neurones que l'on trouve dans la littérature, en particulier, les réseaux de neurones à apprentissage compétitif doux et dur utilisés en classification

La deuxième partie expérimentale est composée aussi de deux chapitres :

Dans le premier chapitre nous détaillons notre nouvelle approche de classification fondée sur le concept compétitif, divisée en quatre étapes de traitement. Elle commence par l'estimation de la fonction de densité de probabilité, suivie d'un réseau de neurones à apprentissage compétitif par la distance de Mahalanobis comme fonction d'activation. Cette étape permet de détecter les maxima locaux de la fdp. Ensuite, nous utilisons une méthode d'apprentissage compétitif Hebbien (CHL) que nous avons développée pour analyser la connectivité entre les maxima détectés du fdp par le biais de la distance de Mahalanobis aussi. Les modes ainsi détectés sont alors utilisés pour le processus de classement du reste des données non classées.

Le dernier chapitre est consacré à la validation de notre procédure de classification à travers la présentation des résultats expérimentaux obtenus des échantillons multidimensionnels générés artificiellement dans la première section du chapitre. La deuxième section a été dédiée à l'application de l'approche ; elle a été réalisée, dans un premier temps, sur des images couleurs synthétiques et réelles. Ensuite, sur deux exemples d'images TEP de la maladie Alzheimer. Les résultats obtenus ont été analysés et comparés aux résultats de certaines méthodes de classification similaires telles que l'approche neuromorphologique [MUT-12a], K-moyennes et c-moyennes flou (*Fuzzy c-means* (FCM)).

Enfin, nous clôturons notre thèse par une conclusion générale qui résume les travaux réalisés et donne quelques perspectives.

CHAPITRE I: ANALYSE DES DONNEES : DE LA CLASSIFICATION AUTOMATIQUE A LA SEGMENTATION D'IMAGE

I.1. Introduction

Selon J-P. Fénelon "l'analyse des données est un ensemble de techniques pour découvrir la structure, éventuellement compliquée, d'un tableau de nombres à plusieurs dimensions et de le traduire par une structure plus simple et qui la résume au mieux. Cette structure peut le plus souvent, être représentée graphiquement" [FEN-81]. C'est ainsi que nous pouvons considérer ce domaine comme un enchaînement de procédures de traitement des données, dont la classification automatique est une procédure parmi plusieurs.

En terme de définition générale, la classification automatique est le regroupement d'une population d'objets semblables dans des classes dissemblables entre eux, dans un contexte supervisé ou non supervisé. Jusqu'à présent, il n'existe pas une seule méthode qui est capable de résoudre tous les problèmes de la classification et donc qui sert à classer tous les types des données. Ainsi, on regroupe schématiquement les procédures de classification automatique en deux catégories, selon que toutes les observations disponibles sont prises en compte simultanément pour découvrir l'existence de classes ou, qu'au contraire, on s'attache à ne considérer que les relations entre les observations et leurs voisines, pour découvrir la structure de leur distribution.

Une démarche essentielle pour le développement et l'amélioration des techniques de classification, est l'évaluation de ces méthodes à travers la validation de ses résultats.

La vision intervient dans un grand nombre d'activités humaines, ce qui donne l'intérêt à la reconnaissance de formes, précisément la phase de segmentation. Cette dernière occupe une grande partie des applications de la classification, ses méthodes trouvent des applications nombreuses dans divers domaines de la médecine, de la sécurité, de la vision robotique, de la reconnaissance de la parole, ...etc.

La recherche en imagerie médicale est un domaine très actif, ce qui exige des techniques de segmentation adaptables, robustes et fiables.

I.2. Classification automatique

La classification automatique (en anglais "*clustering*" ou "*cluster analysis*" et non pas "*classification*" qui signifie classement) est l'opération d'arranger, selon certains critères, une distribution des données composée d'un ensemble d'observations hétérogènes dans des groupes homogènes, appelés classes. Pour donner une définition claire, il faut d'abord poser la question, qu'est-ce qu'une classe ?

En réalité, cette question a été posée plusieurs fois sans avoir la même réponse [IZE-08]. Cela revient au fait que les méthodes de classification ne traitent pas le même type de données, et donc n'auront pas les mêmes résultats. Mais on se permet de la définir superficiellement et dire qu'une classe est un groupe constitué des éléments possédants presque les mêmes propriétés, et ceux qui ont des propriétés différentes sont dans d'autres groupes.

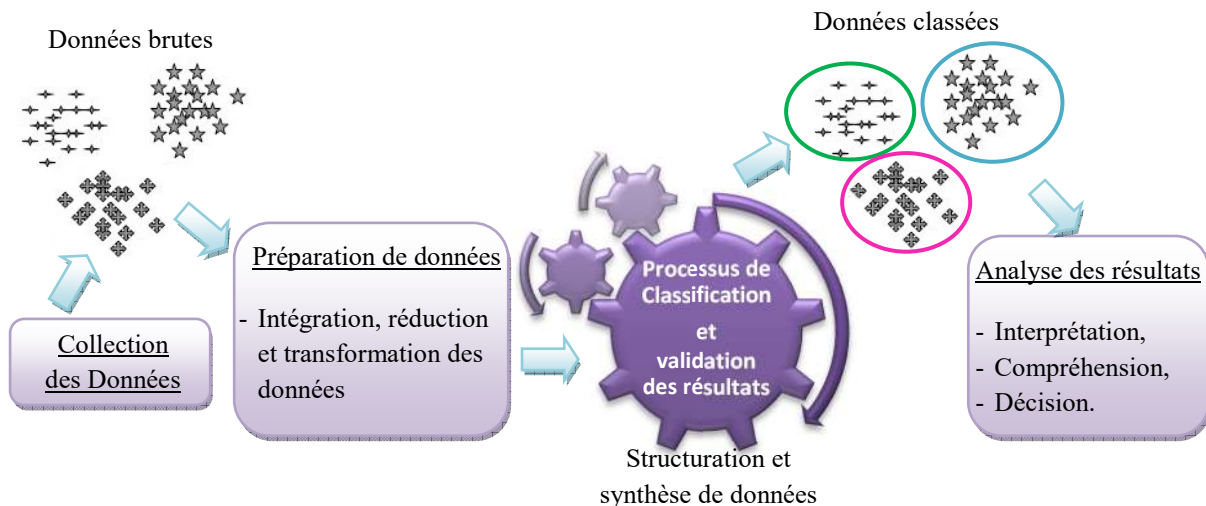


Figure I.1: les différentes étapes de la classification automatique des données

Nous pouvons donc inférer les étapes principales de la classification automatique à travers le schéma ci-dessus :

- La préparation des données,
- La structuration et synthèse de données,
- L'interprétation et l'exploitation des classes.

Dans ce chapitre nous allons détailler les trois étapes principales de la classification des données multidimensionnelles.

I.2.1. Préparation des données

Le but de la classification est de grouper les objets les plus similaires dans la même classe et les plus dissimilaires dans des classes différentes. Pour un humain, cette opération est triviale et donc il apparait de prime à bord, que l'automatisation de la classification le serait aussi.

Cependant, la classification automatique non supervisé, contrairement au classement, est basée seulement sur les informations préliminaires extraites des données qui décrivent les objets et les relations entre eux. En effet, après la collection des données qui prennent différentes formes (image, parole, signal, statistiques,..., etc), ces dernières peuvent être inadaptées, non représentatives, bruitées, aberrantes, manquantes ou de dimensionnalité très élevée ; ces données, qui sont de plus en plus volumineuses dans divers domaines, sont donc souvent imparfaites.

On peut décomposer l'imperfection dans les données en trois catégories (non exclusives) : l'incertitude (probabiliste, possibiliste, crédibilise), l'inconsistance (conflit) et l'imprécision (ensemble, intervalle, flou, incomplétude) [MAR-05]. Ces données brutes sont après présentées sous différents types de représentation [AND-73], permettant ainsi de les traiter par différentes méthodes de classification.

I.2.1.1 Représentation des données

Dans cette étape d'analyse, les données sont organisées soit dans une matrice d'observation ou dans une matrice de proximité ; dans une *matrice d'observation* chaque ligne définit une observation (unité de données) et chaque colonne présente un paramètre (attribut ou caractéristique) qui caractérise l'ensemble des observations, tandis que la *matrice de proximité* regroupe les indices de proximité ou de similarité par paire d'observation [ESS-97].

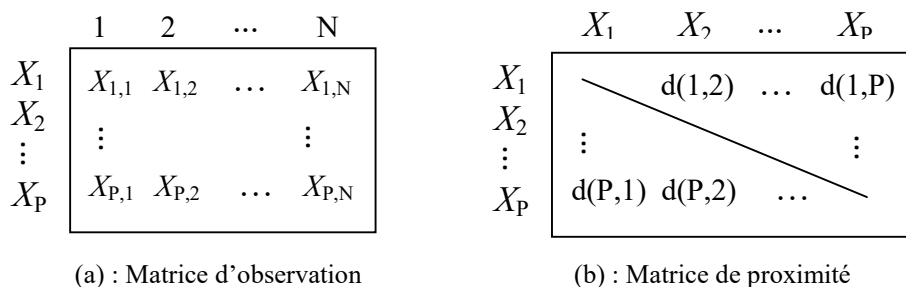


Figure I. 2: les deux types de matrice représentant les données

Le nombre des paramètres dans une matrice d'observation représente la dimension des données. Un paramètre peut être continu (Ex: intervalle), discret (Ex: niveau de gris) ou binaire dans une échelle multidimensionnelle.

En général, on représente un ensemble de Q observations par Q points dans un espace multidimensionnel \mathbb{R}^N , en associant à chacune de ces observations un vecteur X_q , $q=1,2,\dots,Q$, tel que $X_q = [x_{q,1}, x_{q,2}, \dots, x_{q,n}, \dots, x_{q,N}]^T$, où $x_{q,1}, x_{q,2}, \dots, x_{q,n}, \dots, x_{q,N}$ sont des attributs de l'observation X_q .

On outre, ces paramètres peuvent prendre une valeur unique comme le cas des *données classiques* ou plusieurs valeurs comme le cas des *données symboliques* qui peuvent être quantitatives ou qualitatives, ces dernières demandent des méthodes d'extraction de caractéristiques plus complexes [HAV-13] [SHA-15].

1.2.1.2 Sélection des caractéristiques pertinentes

Au cours des dernières années, le nombre des caractéristiques des données multidimensionnelles a pris un essor considérable. Dans plusieurs domaines (reconnaissance de formes, Data Mining, chimie,...etc.) le nombre des caractéristiques peut arriver à des centaines voire plus, ce qui rend le traitement de ces données presque impossible à cause de leur dimension très importante. Mais ces caractéristiques sont souvent redondantes et non pertinentes. Par conséquent, en vue d'une exploitation efficace, plusieurs techniques de réduction de la dimensionnalité ont été mises en place pour extraire des caractéristiques potentiellement plus fondamentales, plus compactes et plus explicatives, même cela peut conduire à une légère perte d'information.

Les techniques de réduction de la dimensionnalité sont classées en deux catégories principales ; l'*extraction des caractéristiques* et la *sélection des caractéristiques* [FAR-13] [YUD-14] [LIU-06] :

- Les méthodes de *l'extraction des caractéristiques*, sont souvent basées sur la transformation des caractéristiques de l'espace de départ pour pouvoir représenter les données dans un espace de faible dimension. Les méthodes classiques de cette catégorie, qui sont des méthodes linéaires citant l'analyse en composante principale (ACP), l'analyse Linéaire Discriminante (ALD) et la méthode de positionnement multidimensionnel (*Multi Dimensional Scaling* (MDS)), ont été la base de plusieurs méthodes non linéaires plus récentes citant Isomap [TEN-00] et Plongement localement linéaire (LLE) [ROW-00].

- L'objectif des méthodes de *la sélection des caractéristiques* est de choisir parmi l'ensemble des caractéristiques originales celles les plus pertinentes et les plus adaptées à la résolution du problème. En général, ces méthodes commencent d'abord par l'initialisation du point de départ de la recherche, puis une procédure de génération des sous-ensembles de caractéristiques, suivie par une méthode d'évaluation, ces deux dernières étapes se répètent jusqu'à l'obtention d'un critère d'arrêt. Les algorithmes de sélection sont catégorisés selon leur méthode d'évaluation en quatre groupes :

1- Le modèle "*filter*" étant le premier utilisé, n'introduit aucun algorithme d'apprentissage à l'évaluation de pertinence d'une caractéristique mais il exploite ces propriétés, il est plutôt considéré comme une phase prétraitement,

2- Le modèle "*wrapper*", par contre, se base sur des algorithmes d'apprentissage et arrive effectivement à éliminer les caractéristiques redondantes et non pertinentes, sauf qu'il est limité par le temps de calcul important et la dépendance des caractéristiques pertinentes,

3- Le modèle intégré (*embedded*) est plus avantageux que le modèle Wrapper par son pouvoir d'utiliser tous les exemples d'apprentissage pour établir le système et par sa rapidité,

4- Le modèle *hybride* profite de l'efficacité du modèle *Filter* et de la performance du modèle *Wrapper*. Toutefois, ce modèle peut souffrir en termes de précision, puisque les deux modèles *Filter* et *Wrapper* sont considérés comme deux étapes séparées [PAR-15].

Il convient de mentionner que les méthodes de réduction des caractéristiques sont aussi classifiées comme :

- Supervisées, qui guident les processus de recherche vers les caractéristiques pertinentes en se basant sur plusieurs algorithmes tel que l'algorithme génétique [CHO-11],
- Non supervisées qui ont connues récemment un large épanouissement sur différentes méthodes telles que les réseaux de neurones et la classification des graphes [PAR-15].

Toutes ces techniques de réduction des caractéristiques, et même celle de classification, ne peuvent exister sans la notion de ressemblance, qui impose aux classes un ensemble de critères à réaliser pour faire la distinction entre eux.

1.2.1.3 Critère de ressemblance

Parmi les facteurs les plus marquants qui définissent l'homogénéité dans un groupe de données, on trouve la similarité, la proximité et la connectivité ; vu que ces facteurs sont mesurés à travers les variables descriptives des données, pour dire qu'un ensemble d'objets est homogène, ils doivent partager quelques propriétés de ressemblance. Afin de mesurer ces dernières, deux démarches sont anticipées, *monothétique* et *polythétique*:

- Dans la démarche *monothétique*, il faut et il suffit de partager une seule propriété ou un ensemble restreint de propriétés pour qu'un objet appartienne à une classe.
- Quant à la démarche *polythétique*, la mesure de ressemblance est basée sur une matrice constituée d'une combinaison de plusieurs propriétés, comme le cas de la mesure de proximité réalisée par le calcul de distance. Cette démarche est la plus répandue.

Quel que soit la mesure de ressemblance utilisée, il doit vérifier les indices de similarité et de dissimilarité suivants :

✓ Critères de similarité :

Soit un échantillon d'observation $Q = \{x_1, x_2, \dots, x_n\}$ tel que $x_2 \neq x_1$:

- $\text{sim}(x_1, x_2) = \text{sim}(x_2, x_1) \geq 0$,
- $\text{sim}(x_1, x_1) \geq \text{sim}(x_1, x_2)$.

Ces indices permettent de regrouper les objets les plus similaires au sein d'une même classe.

✓ Critères de dissimilarité:

- $\text{diss}(x_1, x_2) = \text{diss}(x_2, x_1) \geq 0$,
- $\text{diss}(x_1, x_2) = 0 \Rightarrow x_1 = x_2$.

Ils permettent de créer des classes les plus dissimilaires possible.

Généralement quatre groupes de mesures de ressemblance peuvent être distingués [BOU-06] :

1. ***Coefficients de corrélation*** : le plus courant est celui de Bravais-Pearson, ce coefficient représente un indice statistique qui exprime l'intensité et le sens (positif ou négatif) de la relation linéaire entre deux variables quantitatives [ZAR-12], il a montré son efficacité en tant que outil de mesure dans plusieurs algorithmes d'apprentissage

[MOK-11][MOR-15]. Néanmoins, l'utilisation de ce coefficient peut aboutir à des résultats incorrects sur l'absence ou la présence d'une relation, pour les distributions non gaussiennes ou comportant des valeurs exceptionnelles.

2. **Mesures de distance** : la mesure de distance introduit la notion de voisinage entre les données, chaque distance respecte certaines règles géométriques ou statistiques qui définissent la relation entre les données. La chose qui explique l'existence d'une grande variété de distance, parmi lesquelles nous citons les plus célèbres :

- ✓ La *distance Euclidienne* (DE) calculée par la relation mathématique suivante :

$$DE(X, Y) = \sqrt{\sum_{i=1}^N (X_i - Y_i)^2} \quad (I.1)$$

Ce n'est pas au hasard que cette distance est la plus populaire, elle n'est pas seulement caractérisée par son efficacité et son temps de calcul réduit, mais aussi par sa capacité de surmonter le phénomène connu par "la malédiction de la dimension" [KRI-08] [XIA-15]. Cependant, la (DE) n'est pas toujours fiable lorsqu'il s'agit des distributions non gaussiennes, vue qu'elle néglige la variance entre les observations.

- ✓ La *distance de Manhattan* entre deux observations, calcule la somme de la différence entre les dimensions:

$$DH(X_i, Y_i) = \sum_{k=1}^K |X_{i,k} - Y_{i,k}| \quad (I.2)$$

Graphiquement, Manhattan calcule le chemin parcouru en forme de grille entre deux points de données, d'où son deuxième nom '*city block*'. Cette métrique a prouvé une meilleure performance par rapport à la DE et la similarité de cosinus pour la reconnaissance de signature [ISM-11], elle a été aussi adaptée à la mesure de distance dans certains protocoles de routage [ELG-15].

- ✓ La *distance de Mahalanobis* défini mathématiquement comme suivant :

$$DM(X, Y) = \sqrt{(X - Y)^t \Sigma^{-1} (X - Y)} \quad (I.3)$$

Avec Σ est la matrice de covariance.

Cette distance, introduite par Mahalanobis [MAH-36], est basée sur la corrélation entre les variables. En conséquent, malgré qu'elle demande un temps d'exécution plus élevé par rapport à la (DE), elle donne des résultats plus précis concernant les distributions non gaussiennes [Mao-94] [TOR-12] [RIC-15], compte tenu qu'elle prend en considération la dispersion entre une série des données. Parmi ses variantes, la distance de Mahalanobis localement centrée [TOD-13] qui est centrée sur chaque échantillon, et non pas sur le centre de gravité de données, pour la détection des données aberrantes.

3. *Mesures dérivées* soit des mesures de distance soit des coefficients de corrélation, telle que la similarité de cosinus qui est parmi les mesures les plus populaires dans le domaine de text-mining [MAL-09].
4. D'autres mesures développées pour des raisons spéciales, principalement pour les variables binaires.

Le choix entre toutes les mesures existantes dépend de la nature des données et de l'objectif de la classification, ces deux facteurs sont aussi parmi les facteurs principaux qui engendrent toutes une variété des méthodes de classification.

1.2.2. Méthodes de classification

Bien que les méthodes de classification automatique ont toutes le même but, étant de regrouper les objets homogènes dans une même classe qui est hétérogène avec les autres classes, la diversité des données à traiter et la variété de sa nature, le problème de la prise de décision automatisée et la façon qu'un problème de classification est abordé, obligent l'analyste à choisir une stratégie parmi un ensemble de méthodes très variées et rarement comparables selon certains critères. L'ambiguïté de la notion de classe est aussi un facteur [EST-00]. Tout ce qui précède explique l'existence dans la littérature de différentes catégorisations (statistique/métrique, hiérarchique/partitionnement, basées sur les prototypes /basées sur les graphes/basées sur la densité, ...etc.).

Ainsi, nous pouvons distinguer entre deux types généraux des méthodes de classification ; les méthodes de classification classiques et les méthodes de classification floues.

1.2.2.1 Méthodes de classification classiques

D'après notre travail, les méthodes classiques sont divisées en deux catégories des procédures de classification automatique : procédures globales et procédures locales. Le

diagramme de la figure I.3 illustre l'organisation de ces deux catégories sur lesquelles on se base pour poursuivre notre étude :

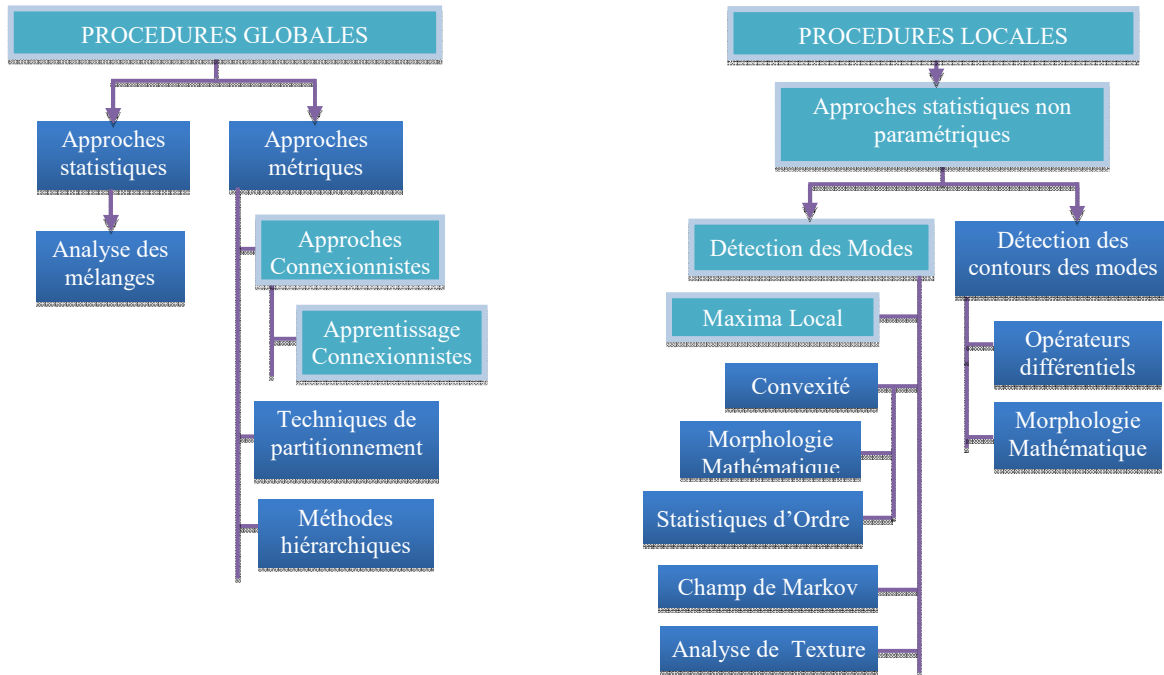


Figure I.3: Schéma des procédures de classification automatique non supervisées

I.2.2.1.1 Procédures globales

La procédure globale consiste à analyser une distribution des observations en leur totalité et vise une analyse profonde et exhaustive des données, cela implique souvent, l'exigence d'établir des hypothèses restrictives au préalable. Ce type de méthodes fait appel à deux approches : approches métriques, approches statistiques.

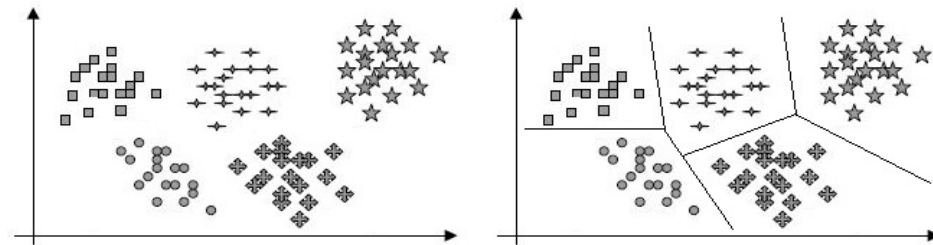
I.2.2.1.1.1 Approches métriques

Le principe de la technique métrique est basé sur la définition d'une métrique sur l'espace de représentation des données, compte tenu de certaines propriétés de ressemblance en commun entre les observations de la même classe. On distingue trois principaux types d'approches métriques globales permettant de classer les observations. La première consiste à effectuer un partitionnement de l'espace de représentation des observations, la deuxième établit une hiérarchie de classes et la dernière analyse la connectivité entre les observations.

I.2.2.1.1.1.1 Techniques de partitionnement

La technique de partitionnement est basée sur la définition d'une fonction d'homogénéité ou d'un critère de qualité, qui est souvent un critère de proximité, ce qui rend

le problème de classification un problème d'optimisation. Le choix de cette fonction reflète la performance de la classification par partition, la raison pour laquelle de nombreux critères de proximité ont été proposés [FUKU-70] [ALI-07] [FIL-08]. La recherche des extrema de ces critères conduit en général à des classifications qui maximisent la dispersion interclasse tout en minimisant la dispersion intra-classe [EDD-07b].



(a) : Représentation des données brutes

(b) : Représentation des données partitionnées

Figure I.4: Représentation graphique d'une distribution Gaussienne et son partitionnement

Ainsi donc plusieurs méthodes cherchent à satisfaire autant d'homogénéité à l'intérieure des classes aussi bien que de séparabilité entre les classes [RAM-09]. Il s'agit de la méthode *d'agrégation autour de centres mobiles*, cette dernière, développée par Forgy [FOR-65], vise à partitionner n observations à K classes par l'affectation de chaque observation au centre le plus proche. Il existe de nombreuses méthodes qui adoptent le même principe que celui du centre mobile :

- La méthode ***K-moyennes*** [MAC-67], étant plus rapide, est parmi les algorithmes les plus largement utilisés pour la classification. Plusieurs extensions de cette méthode ont été proposées pour surmonter certaines limites, telles que ***Fuzzy c-moyens (FCM)*** [DUN-73], ***K-médoïdes*** [KAU- 90] et ***k-modes*** [HUA-98] qui permet de supporter différents types des données, ou ***global k-moyennes*** [LIK-03] qui permet de réduire l'influence de l'initialisation sur les résultats. On outre, les chercheurs ne cessent de développer ces méthodes pour leur donner plus de flexibilité [CAR-12] [RAZ-13] [MEI-14] [MOE-14].
- La méthode des ***nuées dynamiques*** [DID-71] est une généralisation de k-moyennes. En fait, cette méthode est basée sur la notion de noyaux (centre de gravité) qui peuvent être plus complexes que des moyennes ou des centres. Le choix arbitraire des noyaux de la population à traiter, a sollicité l'introduction de la notion de formes fortes pour améliorer les performances de ces méthodes [DID-79]. Cependant, l'étude des formes fortes a permis de s'affranchir d'un autre inconvénient de la méthode qui est de nécessiter le choix à priori du nombre de classes. En effet le nombre de formes fortes peut être très variable et ne dépend pas directement du nombre de classes choisi [ROU-06].

I.2.2.1.1.2 Classification hiérarchique

Si la classification par partitionnement est basée sur la recherche directe des partitions, la classification hiérarchique fournit une suite de partitions emboîtées les unes dans les autres, présentées soit par une structure arborescente appelée *Dendrogramme* ou par un *graphe orienté acyclique (GAD)*.

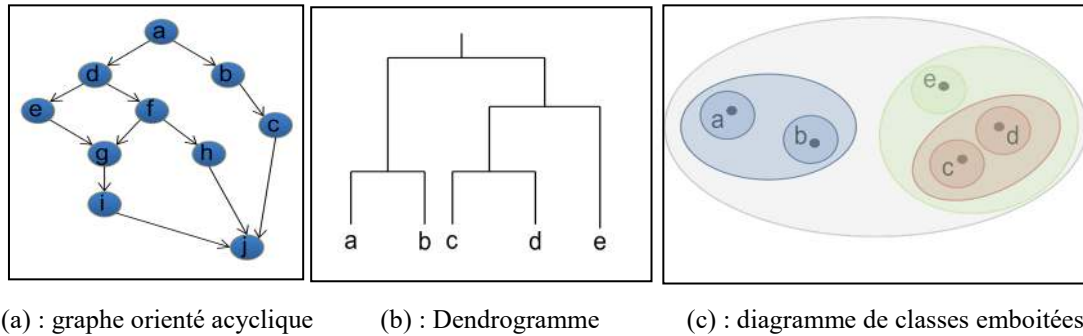


Figure I.5: les différentes représentations graphiques de la classification hiérarchique

Dans ce type de classification, l'ensemble des classes peut être représenté par une structure hiérarchique ascendante, descendante ou de recouvrement [JAMB-78] :

- La démarche **ascendante (agglomerative)**, étant la plus utilisée vue sa simplicité par rapport aux autres, consiste à considérer initialement chaque objet comme une classe, puis à regrouper, à chaque étape, les observations deux à deux suivant une matrice symétrique de distance comme critère de similarité. De nombreux développements ont été réalisés au cours des dernières années concernant entre autres la réduction de taux d'exécution par l'introduction de plusieurs techniques, comme la création d'une hiérarchie basée sur des centroïdes au lieu des points disjoints ou l'utilisation des K plus proches voisins dynamique [BOUG-15] [LAI-11], y compris l'utilisation de nouvelles mesures de similarité [LU-16].
- La démarche **descendante (divisive)** consiste à grouper toutes les observations dans une seule classe, puis à diviser itérativement les classes suivant un critère de dissimilarité. Le nombre de classes est alors déterminé par un seuil prédéfini [BAYN-80] [LANE-67] [LUKA-79].
- La démarche **de recouvrement (pyramidale)**, proposée par Diday [DID-86], est considérée comme une généralisation des démarches hiérarchiques, ce modèle a été conçu par Brito [BRI-91] par la suite, pour classifier les données symboliques, grâce à la spécificité de

donner aux classes, appelées aussi *paliers*, la possibilité de s'intersecter sans être incluses l'une dans l'autre.

Une autre méthode consiste à réduire la hiérarchie des parties, ou l'arbre des classifications, sous forme d'un ensemble de nœuds significatifs, avant d'effectuer la partition de l'ensemble des objets à classer [GOR-87] [LER-91].

Une limitation importante de ce type de méthodes est le problème d'incohérence: Une erreur de classification, à tout niveau de la hiérarchie, sera propagée à tous ses descendants [HER-14]. L'utilisation de classificateurs binaires est la façon la plus commune pour faire face au problème de classification hiérarchique [FAG-08] [VAL-09], mais les classificateurs multi-classes sont plus adaptables à des problèmes avec les taxinomies de taille [HER-14].

I.2.2.1.1.3 Approches connexionnistes

Les approches connexionnistes font appel à la classification par la théorie des graphes. Elles représentent les données sous forme de graphe constitué des nœuds, qui simulent les observations, connectés entre eux par des liens qui représentent les relations entre ces observations, où chaque composante connexe représente une classe. Cette structure leur permet de préserver la topologie des données.

Parmi les modèles connexionnistes les plus simples, *Delta Rule* proposé par Widrow and Hoff (1960) est une loi de correction des erreurs qui a donné naissance à plusieurs modèles d'apprentissage de concept [LAM-15]. Un autre modèle, très célèbre dans la théorie des graphes, le *MST (Minimal Spanning Tree)* de Zahn (1971) qui fait la base des méthodes *Single-link* de la classification hiérarchique. L'algorithme *MST* est aussi utilisé comme un outil très fort pour la performance de la classification automatique [WEI-14].

Ces approches font partie des méthodes de classification incrémentale, c'est-à-dire qu'elles sont capables de prendre en compte les formes qui ne sont pas disponibles au moment de la constitution du modèle initial par les données initiales [BEN-07], contrairement aux approches basées sur les prototypes (Ex: k-moyennes) qui font un seul niveau de partitionnement des données.

L'étude de ces approches conduit directement vers le domaine des réseaux de neurones, plus précisément les réseaux connexionnistes, à titre d'exemple *les cartes auto-organisatrices SOM* de Kohonen (1995) et *GNG (Growing Neural Gaz)* de Fritzke (1994). Ce qui donne à ces approches leur force, en plus de tous les avantages des réseaux de neurones, c'est qu'elles

permettent de mettre à jour les modèles de classification d'une façon dynamique. Cette partie sera plus détaillée dans le chapitre II.

1.2.2.1.1.2 Approches Statistiques

Étant donné que les observations sont des distributions de points dans l'espace de représentation des données, distinguées par leurs fonctions de densité de probabilité, ce type d'approches, en tant que des approches paramétriques, consiste à donner une hypothèse concernant les modèles de distribution des observations et à estimer les paramètres inconnus de cette dernière (moyenne, variance, ..., etc) à partir des données disponibles. Le problème d'analyse de données peut alors être ramené à celui de la détermination des paramètres d'un mélange de fonctions de densité représentant les distributions des observations provenant de chacune des classes en présence dans l'échantillon analysé.

1.2.2.1.1.4 Analyse de mélange

Une large variété des phénomènes aléatoires est modélisée par le biais des mélanges finis de distributions de probabilité. L'estimation des paramètres d'un mélange de fonctions de densité peut être obtenue par des techniques d'apprentissage Bayésien ou par des procédures d'estimation par maximum de vraisemblance [DUD-73] à travers une procédure itérative connue comme l'*espérance-maximisation (EM)* [DEM-77], Cet algorithme a connu des variantes sur plusieurs niveaux pour plus de performance des distributions plus complexes comme *CEM*, *k-MLE* et *Fuzzy EM* [CEL-91][NIE-12][QUO-15].

Cependant, outre la possibilité d'utiliser un modèle paramétrique pour décrire la fonction de densité sous-jacente, ces techniques statistiques nécessitent des hypothèses restrictives. Ainsi, la connaissance du nombre de classes en présence est souvent obligatoire [SCH-76], ce nombre pouvant même être limité à deux classes dans certains cas [MAK-77][MIZ-75]. D'autres hypothèses restrictives, telles que l'égalité des matrices de covariance ou la connaissance des probabilités a priori des différentes classes sont parfois exigées.

Dans cette optique, J. -G Postaire et C. Vasseur ont proposé d'analyser la convexité de la fonction de densité de probabilité sous-jacente pour approcher tous les paramètres nécessaires à la description d'un mélange totalement inconnu. Bien que limitée aux distributions normales, cette approche a le mérite de ne nécessiter aucune hypothèse restrictive [POS-81].

Toutes ces procédures globales sont particulièrement adaptées à des ensembles d'objets qui se regroupent en classes de formes globulaires. Dans le cas de classes de formes et tailles variées, les procédures locales semblent mieux adaptées.

I.2.2.1.2 Procédures Locales

Dans ce type d'approche, il s'agit d'analyser de manière plus ponctuelle la répartition des observations, avec pour objectif de mettre en évidence soit des concentrations locales d'observations qui correspondent à des classes, soit des zones circonscrites de l'espace relativement vides d'observations qui correspondent aux vallées séparant ces classes. Plus précisément, ces procédures reposent sur l'analyse de la fonction de densité de probabilité (fdp) sous-jacente à la distribution des observations disponibles, afin d'extraire l'information nécessaire à la classification de ces dernières [EDD-07b]. Cela explique la capacité de ces approches à détecter les classes dans des distributions de forme arbitraire.

I.2.2.2.1.1 Détection des modes

I.2.2.2.1.1.1 Estimation de la fonction de densité de probabilité

Ces procédures font appel à des techniques non paramétriques pour l'estimation de la fdp tel que l'estimateur des *K plus proches voisins* [COV-67] ou l'estimateur du *noyau de Parzen* [PAR-62] ou même par la *transformation d'histogramme* [LPR-14] considérée comme une approximation de la fdp. Une autre variation récente utilise les *polygones de fréquence* pour développer un estimateur de la fdp continu pour des données multivariées permettant ainsi une complexité réduite et une haute précision d'estimation [LPR-15].

Dans ce chapitre nous traitons plus en détail la méthode la plus utilisée, celle de l'estimateur du noyau de Parzen.

- *Estimation de Parzen*

La méthode de noyau de Parzen ou fenêtre de Parzen est un outil très puissant pour l'estimation de la fonction de densité de probabilité, en effet, cet estimateur continu a été la meilleure solution pour estimer la densité des échantillons continus multivariés au lieu d'un estimateur constant par morceaux comme les histogrammes.

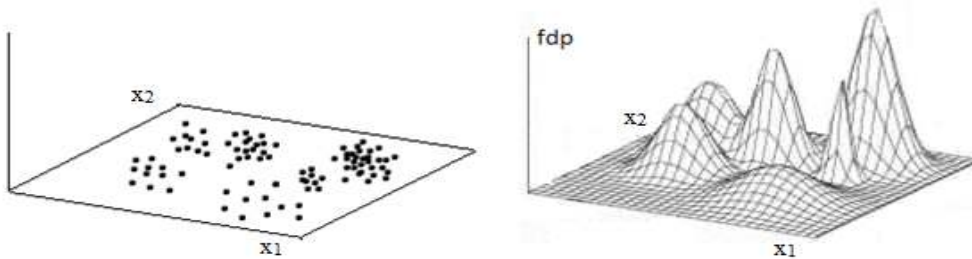
La densité estimée d'un variable aléatoire $X(x_1, x_2, \dots, x_N)$ s'écrit sous la forme :

$$f_N(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - x_i}{h}\right) \quad (I.4)$$

Avec K est le noyau gaussien et $h > 0$ est la fenêtre qui régit le degré de lissage de l'estimation.

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (I.5)$$

Le calcul de l'estimateur de la fdp d'un échantillon bidimensionnel est représenté comme suivant :



(a) : distribution de données multi-classes en 2D

(b) : représentation de la fdp sous-jacente à la distribution

Figure I. 6: Représentation graphique des données et de leur fonction de densité de probabilité

L'inconvénient majeur de cette méthode est sa faiblesse devant la malédiction de la dimensionnalité, pour contourner cette difficulté, un algorithme rapide a été développé par Postaire et Vasseur [POS-82]. Récemment, une autre approche a été élaboré par la modification du noyau de l'estimateur afin d'optimisé le temps de calcul pour les données de dimensionnalité élevé [HAMS-15].

Les procédures locales posent ainsi le problème de la classification en termes de détection des modes ou de régions modales, en se basant sur l'hypothèse selon laquelle chaque mode de la fdp sous-jacente correspond à une classe dans l'échantillon. Dans ce contexte, de nombreuses approches ont été proposées :

1.2.2.2.1.1.2 Détection des modes par recherche des maxima locaux

Dans un espace toujours métrique, les modes sont assimilés à des maxima locaux de la fdp qui peuvent être détectés en remontant les pentes de la fonction densité de probabilité, en utilisant le calcul de gradient local de cette fonction à partir des observations [FUKU-75], ou encore en déplaçant progressivement les observations jusqu'à ce que chacune d'elles se stabilise en des endroits indiquant les modes de cette fonction [BOCK-79].

Cependant, et comme ces méthodes sont basées sur l'utilisation d'opérateurs différentiels, elles sont sensibles aux irrégularités locales des distributions des observations, et tendent à générer de nombreux modes parasites. Pour diminuer l'effet de ces irrégularités, Touzani [TOU-88] a développé une technique d'étiquetage probabiliste, ou relaxation, directement applicable à la fonction de densité estimée.

1.2.2.2.1.1.3 Analyse de la convexité

L'analyse de la convexité a conduit C. Vasseur et J. G. Postaire [VAS-80] à assimiler les modes à des régions localement concaves de la fdp séparées par des régions convexes, au lieu de les considérer comme des maxima locaux.

Dans cette approche, l'analyse de la convexité de la fonction de densité est effectuée sur des domaines d'observations de taille variable [POS-82a]. Cette analyse améliore considérablement la robustesse de la méthode par rapport aux techniques faisant appel aux notions de gradient, mais elle reste sensible aux irrégularités de la distribution des données.

La procédure de base a été améliorée, soit grâce à une technique d'étiquetage probabiliste itératif [POS-94], soit grâce à des processus adaptatifs basés sur des opérateurs morphologiques, soit encore en faisant appel aux statistiques d'ordre [SBI-05].

- Morphologie mathématique

La plupart des méthodes locales présentées jusqu'alors sont fondées sur des considérations purement statistiques. Mais il existe des travaux qui tendent plutôt à lier le problème de détection des régions modales à la théorie ensembliste en transformant tout ensemble fini d'observations multivariées en un ensemble mathématique dans l'espace Euclidien multidimensionnel de représentation des données.

Les techniques basées sur la morphologie mathématique binaire définissent les régions modales comme des régions de cet espace Euclidien qui présentent des fortes concentrations locales d'observations et sont détectées par des techniques d'analyse structurale basées sur la théorie de la morphologie mathématique binaire [BOT-93] [BEN-96]. La robustesse de ces techniques est cependant mise à l'épreuve lorsque les classes en présence sont enchevêtrées.

Quant aux techniques basées sur la morphologie mathématique numérique, elles assimilent les régions modales aux maxima régionaux de la fdp et procèdent à leur détection par le biais d'opérateurs homotopiques moyennant une distance géodésique [SBI-94] [SBI-95]. Toutefois,

ces techniques nécessitent un nombre d'observations d'autant plus important que la dimension des données est élevée.

La morphologie floue a été aussi intégrée pour améliorer ces procédures de classification, La morphologie floue postule qu'une observation appartient à toutes les classes avec un degré d'appartenance à chacune d'elles. Son application dans la recherche des modes de la fdp [TUR-00] utilise la combinaison d'opérateurs d'érosion et de dilatation floue pour extraire les différentes régions modales, et plus particulièrement les observations modales les constituant.

En 2012, *F. Ghanimi et al* ont proposé une approche de détection des modes basée sur la théorie de re-topologie [GHA-12].

- Statistiques d'ordre

Les filtres de statistiques d'ordre (fso) définis sur des ensembles ont été conçus au début pour le traitement d'images et le traitement de la parole. En utilisant ces fso, une approche statistique de classification automatique non supervisée, a été proposée par *Sbai* [SBAI-99], elle est basée sur la détection des modes de la fdp sous-jacente à une distribution de données multidimensionnelles.

Cette technique permet la séparation des modes lorsque le degré de chevauchement est important mais nécessite un grand nombre d'observations lorsque les données sont de dimension élevée [EDD-07b].

Toutefois, malgré toutes ces améliorations, cette approche atteint rapidement ses limites dans le cas d'échantillons de petite taille et de dimension élevée. De plus, il arrive que la convexité locale de la fdp ne soit pas parfaitement définie, mettant en défaut les hypothèses de travail [SBI-05] [ALR-12].

1.2.2.2.1.1.4 Détection des modes par les champs de Markov

La théorie de modélisation Markovienne a été adaptée au problème de la détection des modes par *A. Moussa* et *A. Sbihi*. Cette technique décrit un ensemble d'observations multidimensionnelles par un sous-ensemble discret d'éléments à valeurs binaires qui modélise le champ observable. De ce sous-ensemble, les composantes connexes qui représentent les noyaux ou prototypes, sont extraites modélisant le champ des classes [SBI-00] [MOU-01].

En se basant sur la même approche, *A. Moussa* a considéré un champ discret dérivé de l'ensemble de données brutes, comme un domaine de mesures. Un champ caché, calculé à

partir d'une nouvelle fonction de potentiel, est utilisé pour détecter les modes correspondant à des domaines de concentrations locales élevées d'observations [MOU-08].

L'approche Markovienne consiste, dans un premier temps, à détecter les régions modales de cet histogramme pour assigner ensuite, suivant une règle de décision adaptée, les observations aux différentes classes associées à ces régions. En effet, cette approche induit des effets de régularisation qui permettent de s'appuyer sur un simple histogramme multidimensionnel qui peut être obtenu grâce à l'estimateur de *Parzen-Rosenblatt* avec le noyau cubique [POS-82]. Cette procédure a l'avantage d'accélérer substantiellement les calculs [SBI-05].

Récemment, une nouvelle approche pour la détection des régions modales d'histogrammes représentant les distributions de données multidimensionnelles, basée sur la théorie du champ de Markov, a été proposée. Cette approche considère les histogrammes comme un champ Y de mesures. Le champ caché X est utilisé pour modéliser les régions modales et les vallées de l'histogramme. Grâce à l'algorithme *des modes conditionnelle itératif (ICM)*, combiné avec l'algorithme *d'estimation-maximisation (EM)*, la détection des maxima locaux de l'histogramme est devenue possible et permet donc de déterminer le champ caché identifié X de Y , qui est composé des régions modales et des vallées de l'histogramme [ALR-12].

Une approche morphologique utilise l'épaississement numérique pour détecter les régions modales de la fdp [MUT-12a].

1.2.2.2.1.1.5 Détection des modes par l'analyse de texture

Le principe de cette méthode est de considérer le noyau de la classe comme des domaines caractérisés par ses spécifiques textures dans l'espace des données [HAM-06]. La distribution des données est d'abord présentée par un histogramme multidimensionnel défini dans un ensemble multidimensionnel régulier d'échantillon.

Ensuite, afin d'évaluer localement une texture multidimensionnelle, une matrice de cooccurrence est introduite, ce qui caractérise la distribution locale des points de données dans l'espace de données multidimensionnelles. Plusieurs caractéristiques de texture locales peuvent être calculées à partir de cette matrice de cooccurrence, qui accumule des informations spatiales et statistiques sur la distribution des données dans les voisins de points d'échantillon.

Les caractéristiques de texture sont après choisies en fonction de leur capacité à discriminer les différentes distributions de points de données. Les points de l'échantillon, où la texture locale sous-jacente, est évaluée sont répartis en différentes classes de texture. Les points attribués à ces classes tendent alors à former des composantes connexes dans l'espace de données, qui sont considérés comme les noyaux des classes [HAM-09].

1.2.2.2.1.1.6 Détection des modes par les opérateurs récurrents hyperboliques et de DERICHE

Ces deux techniques de détection des modes de la fdp, proposées par A. Hamid, R. Allaoui et A. Sbihi [ALL-02] [HAM-05] [HAM-09], montrent la possibilité d'exploiter les propriétés de deux filtres séparables récurrents à réponse impulsionnelle infinie à savoir : le filtre différentiel à sinus hyperbolique et le filtre différentiel de Deriche, usuellement utilisés pour la segmentation d'images en contours, pour les généraliser à une fonction multivariable. Les dérivées directionnelles première et seconde sont ensuite exploitées différemment et selon deux usages différents de la notion de convexité pour aboutir à deux approches différentes pour la détection des modes de la fdp qui sont considérés ici comme des régions où la fdp sous-jacente présente des valeurs élevées, contournés par des régions où les valeurs de la fdp sont faibles (vallées) [MUT-12b].

1.2.2.2.1.2 Détection des contours des modes

La détection des contours des modes est une approche alternative qui considère les régions modales comme des régions délimitées par leurs contours et aborde le problème de détection des modes en terme de localisation de ces contours. En effet, les régions modales sont considérées comme des régions de l'espace de données où la fonction de densité de probabilité sous-jacente prend des valeurs élevées délimitées par des régions où cette fonction présente des valeurs plus faibles.

Quelques approches ont été développées dans ce contexte, soit par l'utilisation des opérateurs différentiels après lissage de la fdp par un filtre médian multidimensionnel [TOU-87] développés par l'introduction de la technique d'étiquetages probabilistes itératifs [POS-89], sans pouvoir résoudre le problème de leur sensibilité aux irrégularités de la distribution des observations. Soit par l'application de transformations morphologiques binaires [ZHA-91a], qui ne s'adapte qu'à des problèmes de classification de faibles degrés de

chevauchement. Ou même par l'introduction du gradient morphologique de la fdp et la ligne des partages des eaux [TOU-99], qui a montré sa robustesse et son efficacité pour l'analyse des données de dimension peu élevée.

Toutes ces méthodes, qu'elles soient globales ou locales, sont de base des méthodes exclusives, c'est-à-dire que quel que soit la technique de classification, un objet ne peut appartenir qu'à une seule classe, ce qui n'est pas toujours le cas en réalité.

1.2.2.2 Méthodes de classification floue

Parmi les techniques très utiles pour l'optimisation des méthodes de classification, la logique floue (Fuzzy logic), formalisée en 1965 par Lotfi Zadeh [ZAD-65], est devenue intégrée dans diverses méthodes de classification. Cette technique est très utile pour les données qui connaissent un chevauchement très élevé en comparaison aux méthodes exclusives [TAN-06].

Dans une classification floue, chaque objet appartient à chaque classe avec une probabilité d'existence entre 0 et 1 et en tant que probabilité leur somme doit être égale à 1. Le but de la classification floue n'est pas de traiter des données multi-classe, mais de rendre la classification plus flexible en laissant la possibilité à un objet d'appartenir aux autres classes au lieu de l'affecter à une seule, alors qu'il peut être plus proche à une autre au cours du processus de la classification.

Jusqu'à aujourd'hui, la grande majorité des approches de classification floue est fondée sur la généralisation initiale de k-moyennes à *Fuzzy c-moyens* (FCM) comme il a été proposé par Dunn [DUN-73] et Bezdek [BEZ-73][BER-81], telle que k-harmonique moyennes [FIH-15], fuzzy k-modes avec l'information d'entre-classe [BAI-13] et la méthode hybride basée sur la FCM et l'optimisation par essais particuliers (*particle swarm optimization (PSO)*) [WU-15].

Une étude faite dans cette optique, a démontré que le principal avantage de la classification floue est d'éviter les minima locaux qui correspondent aux résultats de classification indésirables, ce qui permet de définir des algorithmes dédiés à des structures de classification plus complexe. Mais en pratique, cela peut très souvent échouer à cause du grand nombre de minima locaux qu'un système complexe de classification introduit dans la fonction objective. Par conséquent, même si l'on est intéressé par une méthode de classification classique (k-

moyennes, ..), cet algorithme plus sophistiqué doit toujours être exécuté le premier avec des degrés d'appartenance et une défuzzification ne doit être effectuée qu'à la fin [KLA-15].

Tous les essais d'amélioration des méthodes de classification plus ou moins réussis, n'auraient pas avancées sans que ces dernières subissent une évaluation de leurs résultats.

I.2.3. La validation de la classification

Le développement d'un modèle de classification, quel que soit-il, exige la mise en place d'un système d'évaluation de ce modèle, ce qui assure le progrès de la procédure de classification et la qualité de ses résultats.

En effet, les résultats de la classification doivent accomplir certains critères pour définir une "bonne" classification. Plus précisément, certains paramètres, comme le nombre de classes, la densité et la séparabilité, sont pris en compte pour déduire le meilleur partitionnement d'un ensemble de données [HAL-01]. Puisqu'un algorithme extrait les classes qui ne sont pas connus à priori, les partitions finales d'un ensemble de données exigent une certaine forme d'évaluation dans la plupart des applications [REZ-98].

Généralement, la validation de classification impose l'accomplissement de cinq indices [THE-08][CHR-14]:

- **Validité relative** : ce critère demande l'évaluation de la structure de classification en faisant varier les différentes valeurs de paramètres pour le même algorithme (Ex : changement du nombre de classes k). Il est généralement utilisé pour déterminer le nombre optimal de classes,
- **Validité interne** : c'est la vérification des résultats au niveau de la structure interne, il s'agit de l'homogénéité interclasse, l'hétérogénéité intra-classe et l'adaptation aux données.
- **Validité externe** : une méthode de classification doit avoir les meilleurs résultats en comparaison avec d'autres méthodes externes. Ce critère permet de savoir quelle méthode à adopter pour quel type de données.
- **Interprétabilité** : les résultats de la classification doivent être interprétables et répondent aux objectifs de la classification.
- **Validité de stabilité** : une méthode est stable si un changement léger ne cause pas le changement des résultats.

Cependant, selon la méthode de classification utilisée, sa validation ne nécessite pas forcément l'établissement de tous ces indices de validité, sauf celui de l'interprétabilité qui relève du principe de la classification.

Parmi les applications les plus courantes de la classification automatique est la reconnaissance de forme et plus particulièrement la segmentation d'image, dont nous exposons dans le reste du chapitre.

I.3. Image et segmentation

I.3.1 Préambule

Motivées par la performance du système visuel en terme de rapidité et d'efficacité d'analyse d'une image, de nombreuses recherches dans ce domaine ont été menées à imiter ces compétences biologiques, pour parvenir à l'extraction automatique de l'information à partir d'une image numérique et son interprétation après. Cela fait appel à un vaste sujet d'étude qui fait partie des grands thèmes du traitement d'image; la segmentation d'image.

A ce titre, plusieurs méthodes et techniques ont été consacrées à la segmentation en tenant compte les domaines d'applications. La nature de l'information qui est contenue dans chaque élément de l'image peut être: Colorimétrique, Densitométrique, Métrique, Multispectrale, Photométrique, etc... Dans le cadre de notre travail, nous nous sommes intéressés à l'application de la segmentation sur les TEP scan (*Tomographie par Émission de Positrons*).

Afin de mieux cerner notre sujet, nous exposons tout d'abord un rappel sur les images numériques, puis nous abordons les différentes méthodes de segmentation d'image, pour terminer par une présentation sur la segmentation des images de TEP scan qui peuvent faire la matière de l'application réelle de notre approche, et nous présentons finalement un aperçu sur la segmentation des images d'iris comme une proposition d'un sujet d'application.

I.3.2 Image numérique

I.3.2.1 Définition

Une image numérique est une fonction à support discret et limité en bidimensionnel ou tridimensionnel, et à valeurs discrètes d'unité appelée pixel, la nature de ces valeurs distingue entre trois types d'image : image binaire, image à niveau de gris (valeurs scalaires) et image couleur (valeurs vectorielles), une image peut se distinguer aussi par sa dimension qui peut être 3D (images volumétriques), 2D (images matricielles) ou 1D (images linéaires).

La création d'une image se fait par l'absorption des ondes de longueurs différentes, que se soit de source émise, réfléchié ou transmise. Certaines images se font dans la partie visible du spectre (entre $0,4 \mu\text{m}$ (violet) et $0,7 \mu\text{m}$ (rouge)), d'autres, par contre, se captent dans d'autres intervalles telles que les images IRM (Image à Résonance Magnétique), les Ultra-son (contrôle non-destructif) et les TEP (positron émission tomographie).

Tout d'abord, nous commençons par la présentation des différents types de l'image.

1.3.2.2 Image niveaux de gris

Une image en niveaux de gris est une matrice dont chaque élément est un pixel qui représente une valeur d'intensité comprise entre 0 et 256 teintes de gris, du plus sombre au plus clair, chaque pixel est codé sur 8 bits. Mathématiquement, la luminosité d'un pixel dans une image est une application I , telle que:

$$I: [0, L - 1] \times [0, C - 1] \rightarrow [0, I_{max}]$$

$$I(i, j) = I \text{ (Intensité lumineuse du pixel)} \quad (I.6)$$

A titre illustratif, la figure suivante représente la structure d'une image en 2D :

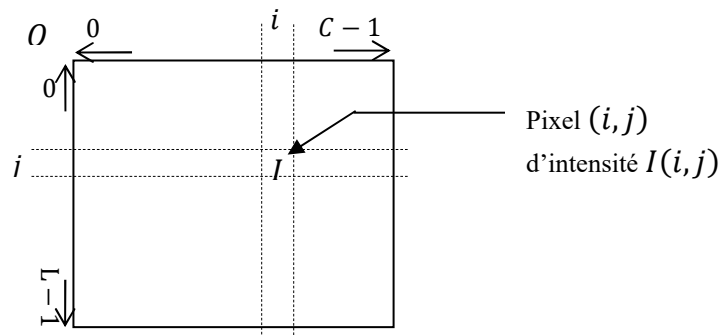


Figure I.7: Représentation d'une image niveaux de gris

1.3.2.3 Image couleur

1.3.2.3.1 Définition

Une image couleur peut être considérée comme une matrice de pixels caractérisés par leurs composantes trichromatiques. En effet, dans une image couleur chaque pixel représente un vecteur des intensités tridimensionnel ; les trois variables représentent respectivement les trois couleurs primaires rouge, vert et bleu. Ainsi une image couleur est codée principalement en RVB, Il existe d'autre espace de représentation 3D défini par d'autres bases tels que les espaces colorimétriques HSV, XYZ, $I_1 I_2 I_3, \dots$ etc.

I.3.2.3.2 Espaces de représentation de couleur

L'espace de représentation de couleur fait référence aux différentes façons d'utiliser les nombres pour décrire une couleur, il fait référence aussi à la manière dont on représente les différentes couleurs du spectre lumineux. La CIE (*Commission Internationale de l'Eclairage*) a alors montré qu'il suffit d'avoir trois variables indépendantes pour décrire presque toutes les couleurs du spectre visible et nous avons alors assisté à l'utilisation de systèmes de coordonnées en trois dimensions.

L'espace colorimétrique, étant qu'un espace vectoriel tridimensionnel permettant un codage rapide et simple des couleurs, distingue entre deux synthèses :

- *Synthèse additive* : consiste à combiner les lumières directement de sources colorées grâce à la somme des couleurs pondérées par des coefficients (cf. figure I.8(a)). Cette synthèse est adaptée aux dispositifs lumineux de restitution de la couleur tels que LCD, Plasma, ordinateur.
- *Synthèse soustractive* : il fait la combinaison d'absorption des couleurs pour obtenir des nouvelles couleurs par la soustraction d'une partie de spectre lumineux provenant d'une réflexion par un objet (cf. figure I.8(b)). Ce type de synthèse est utilisé dans les imprimantes et la peinture.

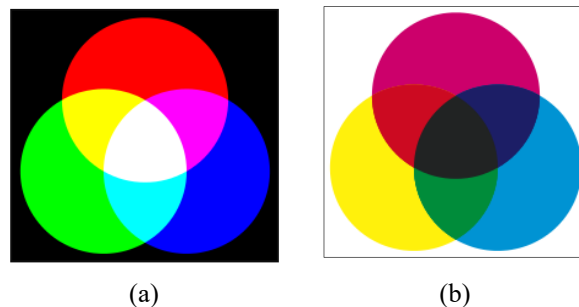


Figure I.8: (a) synthèse additive, (b) synthèse soustractive

Nous nous intéressons dans ce travail alors à la synthèse additive puisqu'elle est adaptée à l'informatique graphique. Un espace colorimétrique est une sorte de classement générique des familles d'espace colorimétriques. Une première façon de définir un espace de couleur consiste donc à choisir les trois couleurs primaires qui forment la base de l'espace vectoriel. On peut également créer un espace de couleur en définissant une matrice de changement de base définie sur un espace de couleur prédéfini. Dans la suite nous exposons quelques espaces colorimétriques les plus connus.

I.3.2.3.2.1 *Modèle RVB*

Le modèle RVB (en anglais RGB) est le plus célèbre parmi tous les espaces colorimétriques. Ce modèle, adopté dès 1931 par la CIE, est un système additif qui compose une large gamme de couleurs à partir des couleurs primaires rouge (R), vert (V) et bleu (B).

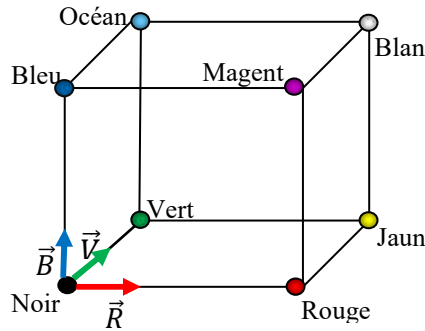


Figure I.9: Présentation spatiale du modèle RVB

L'espace RVB adapte le système de coordonnées cartésiennes (cf. figure I.9) où la diagonale allant du blanc (1, 1, 1) au noir (0, 0, 0) pour des valeurs de pixels incluant dans l'intervalle $[0,1] \subset \mathbb{R}$ appelé axe achromatique. Il reste l'espace de référence [LOZ-98] malgré qu'il présente quelques inconvénients comme dans certains cas des couleurs très saturées dont le matériel est incapable de produire, le modèle se rend à des composantes négatives pour les caractériser, ce qui peut provoquer deux stimulus de couleur de même chrominance tout en ayant des composantes trichromatiques différentes. Cela a été le sujet de plusieurs études [VAN-00][ZUP-01].

I.3.2.3.2.2 *Modèles XYZ et YIQ*

Ce modèle a été créé pour palier le problème des valeurs négatives dans l'espace RVB. La CIE a établi donc en 1931 un autre espace colorimétrique basé sur trois nouveaux primaires X, Y et Z [SKO-01]. L'espace de couleur induit par ces trois primaires permet de représenter toutes les couleurs par synthèse additive. Si une couleur est représentée dans l'espace RVB par :

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.166 & 0.125 & 0.093 \\ 0.060 & 0.327 & 0.005 \\ 0.000 & 0.004 & 0.46 \end{bmatrix} * \begin{bmatrix} R \\ V \\ B \end{bmatrix} \quad (I.7)$$

Le système XYZ est peu utilisé en tant que tel et sert souvent d'interface entre le système RVB et des systèmes plus évolués.

Le modèle YIQ est une variante du modèle RVB établi par le NTSC (*National Television System Committee*) [LUO-90] pour rendre plus efficace la transmission des signaux de télévision et la compatibilité avec les écrans noir et blanc. La composante Y contient l'information concernant la luminosité de l'image. L'information chromatique est quant à elle codée par les axes I et Q. la relation entre les modèles YIQ et RVB est la suivante :

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.30 & 0.59 & 0.11 \\ 0.060 & -0.27 & 0.32 \\ 0.11 & -0.52 & 0.31 \end{bmatrix} * \begin{bmatrix} R \\ V \\ B \end{bmatrix} \quad (I.8)$$

I.3.2.3.2.3 Modèles HSL et HSV

Le modèle TSL (HSV en anglais) est un modèle de représentation dit "naturel", ainsi que ses variantes (HSL, HSI). Ils sont basés sur la perception physiologique de la couleur par l'œil humain, en introduisant des notions de Teinte, de Saturation et de Luminance ou intensité.

I.3.2.3.2.4 Modèles $I_1I_2I_3$ et $H_1H_2H_3$

Le modèle $I_1I_2I_3$ proposé par Ohta et al. [OHT-80] répond à une approche totalement différente. C'est un espace de couleurs qui s'avèrent pertinents pour la segmentation d'image [GEV-99]. I_1 représente la luminance, I_2 et I_3 correspondent respectivement à une opposition bleu-rouge et magenta-vert. Le passage de l'espace RVB à ce système s'effectue ainsi par les équations suivantes :

$$\begin{cases} I_1 = \frac{R + V + B}{3} \\ I_2 = R - B \\ I_3 = \frac{2V - (R + B)}{2} \end{cases} \quad (I.9)$$

Outre ces modèles passés en revue, il existe d'autres espaces colorimétriques dédiés à d'autres objectifs, toute cette variété des modèles rend difficile de privilégier l'espace de représentation des couleurs le mieux adapté à un problème spécifique de segmentation.

Plusieurs recherches ont été faites afin de déterminer l'espace couleur le mieux adapté aux problématiques de segmentation de l'image couleur [TRE-04], et ils ont fini par inférer que cela dépend du type d'image et qu'il est impossible de définir un modèle au préalable pour toute sorte de segmentation [BUS-06].

1.3.2.4 Les images de TEP scan

Le TEP scan, tomographie par émission de positons (TEP), est un type d'imagerie en médecine nucléaire. C'est une technique d'imagerie fonctionnelle plus intéressée par le fonctionnement d'un organe que sa structure.

Cette technique d'imagerie est basée sur la détection des rayons *Gammas* émis par la rencontre des positrons, diffusés par le radiotracer, qui est injecté dans une région donnée du corps humain. Le traceur injecté est, généralement, à base de fluor lié à du glucose, le plus couramment utilisé est le 18 *fluoro-désoxy-glucose* ou 18-FDG, il est possible aussi d'utiliser l'oxygène 15, qui propage dans le corps, dont l'accumulation est due à l'augmentation locale du débit sanguin. Cette diffusion est détectée avec un scanner à hautes performances appelé le PET-Scan.



Figure I.10: équipement et résultat d'une machine TEP-scaner ; (a) Appareil physique de TEP-scan, (b) image TEP d'une coupe transversale du cerveau

Les images obtenues sont en couleurs et en trois dimensions, dont la coloration est en relation directe avec l'intensité de la réponse (cf. figure I.10(b)). Les grandes indications cliniques de la TEP scan sont en Oncologie, Neurologie, Psychiatrie, Cardiologie, Médecine Interne, Infectiologie et Rhumatologie

En comparaison par d'autres méthodes, l'imagerie par TEP scan est plus sensible que d'autres techniques de l'imagerie telles que la TDM (CT) et l'IRM [LAR-03]. En fait, les autres techniques d'imagerie montrent seulement la physiologie des parties du corps, tandis que l'image TEP montre la formation interne des tumeurs et des cellules cancéreuses, par exemple, au moyen du métabolisme des parties du corps [HAP-09].

Toutefois, ce genre des scans a des risques liés au produit de contraste tels que l'allergie et la neurotoxicité, à savoir qu'il est interdit aux diabétiques vue que le radio traceur est une sorte de glucose. La nouvelle norme dans la pratique clinique acquiert des images PET-CT au lieu

d'un PET scan unique pour tirer parti de l'information fonctionnelle et structurelle conjointement [FOS-14b].

En général, afin d'améliorer ou d'extraire des informations qu'une image peut contenir, un ensemble de techniques est utilisé pour traiter et modifier l'image, qui se manifeste dans une chaîne de traitement : l'acquisition, l'amélioration, la restauration, la compression, la segmentation et l'interprétation de l'image. Dans le cadre de notre étude nous nous intéressons à la phase de la segmentation.

1.3.3 Segmentation d'image numérique

La segmentation d'image est l'une des étapes de traitement les plus importantes vers la compréhension, le codage et la compression d'image.

1.3.3.1 Définition

La segmentation est un traitement de bas-niveau qui consiste à partitionner l'image en régions disjointes partageant une même propriété pour former des régions connexes. Selon *Horowitz* et *Pavlidis* [HOR-75], la segmentation d'une image I est définie comme une partition de I en n sous-ensembles non vides $R_i, i = 1, \dots, n$, expliqué par *Zucker* [ZUC-76] tels que :

- 1) Chaque élément de l'image appartient à une seule région : $I = \coprod_{i=1}^n R_i$,
- 2) Tous les éléments d'une même région sont connectés : $\forall i \in \{1, \dots, n\}, R_i$ est connexe,
- 3) Soit P une mesure d'homogénéité, Chaque région doit être homogène : $\forall i \in \{1, \dots, n\} P(R_i) = \text{Vrai}$,
- 4) La fusion de deux régions est hétérogène $\forall i \in \{1, \dots, n\}^2, R_i$ est adjacent à R_j et $i \neq j \Rightarrow P(R_i \cup R_j) = \text{Faux}$.

En d'autres termes, la segmentation est un processus qui permet d'extraire d'une image des objets sémantiques qui peuvent être identifiés par la couleur, l'intensité lumineuse ou par la texture, dont les techniques utilisées sont basées sur diverses méthodes.

1.3.3.2 Méthodes de segmentation

A l'instar de la classification, les résultats non uniques de la segmentation sont subjectifs et dépendent des buts finals, et donc il existe plusieurs catégorisations de ses méthodes où l'on trouve différentes approches :

- Méthodes basées sur le seuillage,
- Méthodes basées sur la détection des contours,
- Méthodes basées sur la détection des régions,
- Méthodes basées sur la classification,
- Méthodes basées sur la morphologie mathématique,
- Méthodes basées sur partitionnement de graphe,
- ...etc.

Certaines de ces approches sont incluses dans d'autres ; généralement, l'interprétation d'une image passe par la localisation et la reconnaissance des éléments primitifs tels que les régions et les contours. Nous pouvons ainsi distinguer deux grandes catégories de segmentations ; l'approche région et l'approche contour.

Ces deux approches sont duales du fait que chaque région possède un contour et qu'un contour délimite forcément une région [CLE-02].

I.3.3.2.1 Approche Région

Dans cette famille de méthodes, l'objectif est de trouver les régions supposées constituées de pixels connexes de l'image ayant des propriétés communes (couleur, intensité, ...), suivant un critère d'homogénéité. Cette approche inclut des méthodes de classification, méthodes de croissance de région et des méthodes basées sur le graphe.

I.3.3.2.1.1 Méthodes de classification

Similaire à ce qu'on a vu au début de ce chapitre, la classification supervisée ou non supervisée, consiste à partitionner l'image à des classes selon plusieurs critères : statistiques ou métriques, paramétriques ou non, où chaque pixel est considéré, dans la majorité des techniques, comme une observation. Parmi les méthodes de classification les plus utilisées, on trouve :

- ✓ La méthode de *K-Moyennes* et ses dérivées [DHA-15] [KAN-02], spécialement la méthode *FCM* (cf. §I.5), elle a été utilisée au début comme une deuxième segmentation pour un résultat plus fine, après avoir appliqué une première segmentation par multi-seuillage [LL-90]. La qualité de la solution dépend de l'initialisation du nombre de classes K , ils risquent donc de ne pas donner la solution optimale, mais la convergence de ces algorithmes est garantie.

- ✓ *Machine à vecteur de support (SVM)*, ce sont des méthodes probabilistes non paramétriques. Les méthodes de segmentation d'image basées sur la SVM peuvent toujours obtenir une meilleure performance de segmentation en raison de la capacité de généralisation du classificateur SVM, mais elles sont sensibles à l'initialisation des pixels [YYA-12], problème qui était résolu par l'application de la méthodologie de SVMs [YHU-04].
- ✓ *Les réseaux de neurones à apprentissage compétitif (RAAC)* (cf. §III.5) ont déjà été utilisées avec succès pour la segmentation d'image, tels que les réseaux multicouches et les carte de *Kohonen* [JIA-10]. L'identification des pixels de chaque classe se passe par la compétition entre les neurones de sortie qui représentent les prototypes des classes, afin d'affecter le pixel en entrée à la classe gagnante.

Ces algorithmes effectuent une étiquette à chaque pixel suivant les caractéristiques d'homogénéité de l'image correspondant à la classe à laquelle il appartient.

Par rapport aux méthodes supervisées, les méthodes de classification non supervisées ont moins de complexité de calcul; Cependant, ils sont sensibles au bruit et ne peuvent pas bien intégrer l'information spatiale en raison de l'incertitude inhérente des données [FOS-14b].

1.3.3.2.1.2 Méthodes de croissance en région (region growing)

C'est une technique largement appliquée à des fins de segmentation différentes [ADA-94] [XGH-12] [XBF-13]. Les méthodes de croissance en région reposent principalement sur l'hypothèse qui dit que les pixels voisins dans une région ont des valeurs similaires.

Ce sont des méthodes qui se basent sur l'homogénéité dans l'image, dont des régions distinctes se font croître progressivement à partir des pixels initiaux (germes) jusqu'au traitement de tous les pixels.

En effet, ces méthodes fonctionnent directement dans l'espace image, d'une manière ascendante, les régions grandissent en incorporant les pixels les plus similaires suivant un critère de ressemblance. A titre d'exemple les méthodes de *Ligne de Partage des Eaux* [ELM-13] issues de la morphologie mathématique. Quoique ces algorithmes soient sensibles au bruit et à l'initialisation, leurs performances restent fiables.

1.3.3.2.1.3 Méthodes basées sur le graphe

Les récentes recherches en segmentation d'image ont montré les avantages et la qualité des méthodes basées sur les graphes [BAGC-11]. Dans ce type des méthodes, l'image

est modélisée sous la forme d'un graphe pondéré et non dirigé dans lequel les nœuds représentent les pixels d'image et les arêtes lient les pixels voisins selon l'homogénéité dans l'image.

La segmentation est réalisée en minimisant le poids qui partitionne un graphe en sous-graphes. Chaque sous-graphe est considéré comme un objet dans l'image.

Il existe une large variété de méthodes de graphe utilisée pour la segmentation, les plus courantes sont *Graph Cut* [SHI-00] et *Random Walk* [GRA-06]. En outre, différentes techniques en été employées dans ce sens comme le champ de Markov [SIG-99].

I.3.3.2 Approche Contour

L'approche contour tente de chercher les pixels frontières correspondant aux variations locales significatives des caractéristiques des pixels, pour former un contour fermé autour d'un objet dans l'image. Au lieu de chercher les régions homogènes, il suppose que deux différentes régions adjacentes présentent des discontinuités locales de leur propriété à leurs frontières.

Les principales techniques liées à cette approche sont les méthodes basées sur le gradient [GEE-07] et la méthode de contour active [KAS-88].

L'inconvénient majeur de cette approche réside dans la connexion des lignes de contour cassées qui les rendent prédisposé à l'ambigüité [SEN-09].

Quel que soit la ramification de toutes ces méthodes de segmentation, elles ont une large gamme d'applications appartenant à divers domaines. Le paragraphe suivant met en évidence certains d'entre eux.

I.3.3.3 Applications de Segmentation

La segmentation d'image a une multitude d'application dans divers domaines utilisant l'image afin d'en extraire les informations utiles en regard de l'application considérée avant de les interpréter. Parmi ces applications on trouve la reconnaissance du visage, la télédiction, la reconnaissance d'iris et en particulier l'imagerie médicale qui fait le sujet d'application de notre travail.

I.3.3.3.1 Segmentation des images médicales : Image TEP

La segmentation des images médicales fournit un étiquetage pour chaque pixel (voxel), en particulier lorsque les formes des régions d'intérêt sont non convexes avec un fond

hétérogène. D'où son importance pour la recherche clinique, le diagnostic, le pronostic et le suivi.

D'autant plus que les structures et les pathologies à analyser sont très variables, ainsi que l'existence de diverses modalités de l'imagerie, basées sur différents types de rayonnements : champ magnétique, ultrasons, rayon x, rayon gamma, ..., etc, ce qui explique l'activité de la recherche dans ce sens, soutenue par plusieurs publications [FOS-14b] [LWE-15] [BCA-16]. Comme nous avons précisé au début de cette deuxième partie du chapitre, nous mettons l'accent sur la segmentation des images de type gamma spécialement la tomographie par émission de positons (TEP).

1.3.3.1.1 Segmentation de l'image TEP

Le TEP scan mesure la distribution et la concentration de l'accumulation du radiotracteur (glucose) dans toute la partie du corps du patient analysée. Ainsi l'image de l'activité métabolique du glucose est codée par couleur, de sorte que les régions les plus actives apparaissent en rouge ou blanc, et les régions les moins actives métaboliquement sont conçues par les couleurs foncées violet ou bleu.

Au cours de l'analyse de l'image TEP scan, une évaluation quantitative de l'absorption du radiotracteur dans l'image TEP est primordiale, la quantification est donc les valeurs proportionnelles à la concentration de radiotracteur dans l'image. Plusieurs méthodes ont en cours d'exploration pour être utilisées comme paramètres de quantification [BIL-15], telles que, de point de vue clinique, les techniques de régression non linéaires, SUV (*standardized uptake value*), TLE (*total lesion evaluation*) et TBR (*tumor-to-background ratio*).

Parmi ces paramètres, le SUV est l'indice de quantification le plus utilisé pour le TEP imagerie car elle donne une mesure physiologiquement pertinente du métabolisme cellulaire [BAS-11], à point que 90% des sites incluent le SUV dans le compte rendu de l'examen médical [BEY-11]. Le SUV (*standardized uptake value*) est une estimée du taux de métabolisme du glucose, calculé par la relation mathématique suivante [BON-12]:

$$SUV = \frac{\text{Concentration d'activité tissulaire (Mbg/ml)}}{\frac{\text{Dose injecté(Mbg)}}{\text{Poids du patient (g)}}} \quad ((I.10))$$

- Si $SUV > 1$, le traceur se distribuait uniformément dans l'organisme,
- Si $SUV > 1$, il y a une hyperfixation.

Un ensemble de méthodes ont été ensuite dérivées de SUV telles que SUV_{max} , SUV_{mean} , SUV_{peak} , $SUV_{40\%}$, SUV_{fit} , SUL et bien d'autres.

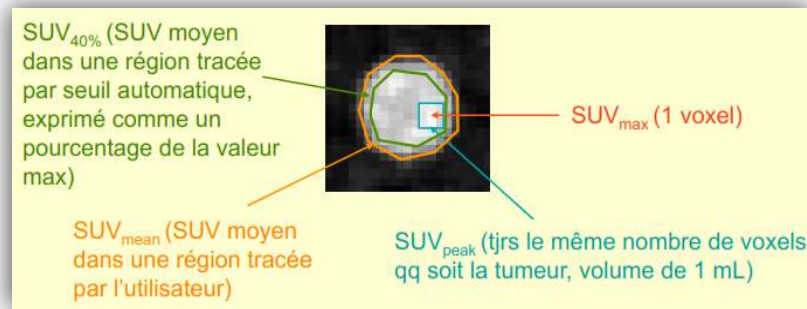


Figure I.11: présentation de quelques indices de quantification d'une image TEP [BUV-11]

La conversion des images en SUV permet de normaliser les images et de les rendre comparables d'un sujet à l'autre, et d'un examen à l'autre, puisque la valeur attendue est toujours 1, quel que soit l'activité injectée et la corpulence du sujet [BUV-11].

En l'occurrence, la délimitation des régions d'absorption à partir d'images TEP est nécessaire pour le calcul des VUS, par la définition des régions d'intérêt ROI (*region of interest*), qui demande une très grande précision, afin d'extraire des informations morphologiques et fonctionnelles. Par conséquent, la quantification et éventuellement les décisions diagnostiques exigent une segmentation bien précise [FOS-14b].

La segmentation des images TEP, étant décomposée de deux tâches : la reconnaissance de l'emplacement de l'objet et la délimitation spatiale de la région de l'objet, trouve certains obstacles tels que [FOS-14b]:

- Les questions relatives à la résolution,
- Une grande variabilité de la forme, de la texture et de l'emplacement des pathologies,
- Le bruit.

Ces facteurs augmentent la difficulté de la segmentation. Au cours des dernières décennies, une grande variété d'algorithmes a été proposée au niveau de toutes les approches de segmentation. En fait, il existe un intérêt croissant pour l'utilisation des méthodes basées sur la classification pour segmenter des régions de capture de formes complexes. En effet, la segmentation en termes de classification est très utile, notamment lorsque les formes des régions d'accrochage sont non convexes avec un fond hétérogène. Les régions non convexes sont très communes pour différentes maladies, en particulier les infections pulmonaires [FOS-13]. Parmi ces méthodes de classification on trouve le *K-moyennes flou* (FCM)[MER-13],

Machine à vecteur de support (SVM)[KER-09] et les réseaux de neurones en général [SAE-10].

La majorité des algorithmes de segmentation des images TEP utilisent des méthodes basées sur le seuillage pour leur simplicité mais elles restent moins optimales [FOS-14a]. Malgré qu'il existe de nombreux algorithmes de segmentation en TEP, mais il n'y a pas encore un algorithme idéal et performant de segmentation.

I.3.3.3.2 Segmentation de l'iris

L'iris, la partie colorée de l'œil humain, est un diaphragme mince circulaire contractile qui entoure la pupille et qui est entouré à son tour par le blanc de l'œil (*sclera*). Il est situé entre la cornée et le cristallin de l'œil humain ce qui fait de lui un organe interne visible et bien protégé de l'extérieur. Son diamètre en général varie entre 12 et 13mm. Son rôle est de contrôler la quantité de lumière qui entre dans la pupille [KRIC-07].

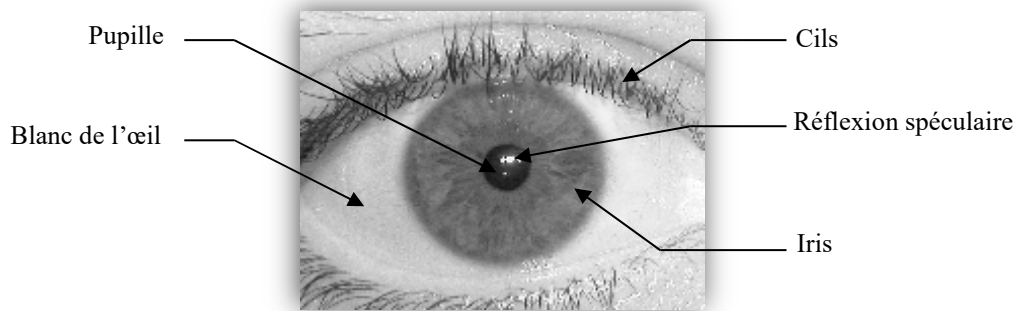


Figure I.12: Vue frontale d'un œil humain avec ses différentes caractéristiques

La reconnaissance de l'iris entre dans de nombreuses applications de sécurité comme les contrôles d'accès, elle est considérée parmi les diverses modalités biométriques qui permettent l'identification des personnes, l'iris est considéré comme très fiable, avec un taux d'erreur remarquablement faible [OTH-16] grâce à sa stabilité au cours de la vie [JAI-04]. Cependant, la qualité d'image acquise qui garantit cette fiabilité, a certaines contraintes non contrôlables comme l'utilisateur. Cela fait appel à un processus de traitement après l'acquisition, dont on trouve quatre étapes:

- **La segmentation** : Cette étape consiste à isoler l'iris des autres éléments de l'œil, mais l'élimination de certains d'eux forme un vrai défi pour la segmentation tels que les cils et la réflexion spéculaire. En général, un ensemble de prétraitement est appliqué à l'image dans le but d'améliorer sa qualité globale tels que le filtrage, l'amélioration du contraste et l'égalisation d'histogramme. Deux grandes approches existent dans la littérature séparant

entre un groupe de chercheurs qui a proposé de segmenter l'iris en utilisant des méthodes qui modélisent sa frontière par un cercle [HAN-07][YAH-16], et d'autres chercheurs qui ont approximés les frontières de l'iris par des contours elliptiques [GRAG-15].

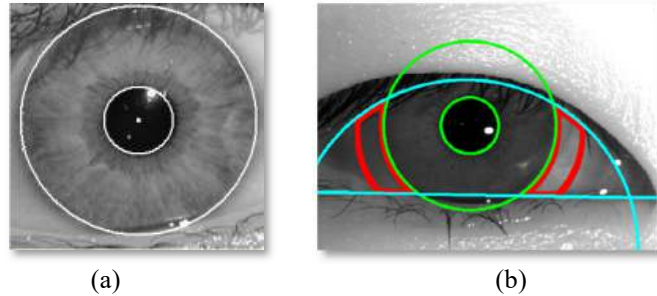


Figure I.13: exemple de deux images d'iris segmentées par les deux approches (a) méthode de modélisation par cercle [HAN-07], (b) méthode de modélisation par contours elliptiques [GRAG-15]

Cependant, quel que soit les circonstances de l'acquisition d'une image d'iris, des irrégularités peuvent apparaître que ça soit au niveau de l'iris ou au niveau de la frontière de la pupille. Par conséquent, les recherches récentes sur la segmentation d'iris tendent à développer des modèles qui décrivent les frontières de l'iris sans considération d'une forme prédéterminée [HIL-13], tel que les contours actifs [SHAR-09], les méthodes statistiques [VAT-08] et les méthodes de classification plus particulièrement l'apprentissage automatique [DEMA-16].

- **La normalisation** : L'objectif essentiel de cette phase est de résoudre le problème de la dilatation pupillaire. En outre, l'iris segmenté peut être une iris complète ou une fraction d'iris complet, d'où la nécessité de transformer la région d'iris à une région de dimensions fixe afin de pouvoir faire des comparaisons.

La variété des modalités de standardisation et de repères possibles a fait naissance à plusieurs approches de normalisation dont on cite les deux principales [HIL-13]; la première consiste à conserver la forme initiale de l'iris en appliquant les opérations de traitement d'image sur la région de l'iris dans le repère initial de l'image [GRAG-15]. La deuxième projette l'iris segmenté vers un autre repère standardisé [HAN-07].

- **L'extraction des caractéristiques** : Il s'agit de transformer l'iris en un code par l'extraction de texture caractéristique de l'iris qui contient toutes les informations et forme une base pour la comparaison entre les images d'iris. Le résultat obtenu est souvent quantifié pour être représenté par un code binaire composé de bits.
- **Classification** : Cette dernière étape effectue une correspondance entre l'iris analysé et les modèles existants dans la base des données en se basant sur une métrique de similarité

efficace conçue pour comparer entre deux codes binaires. Cette tâche permet de prendre la décision si deux images appartiennent à la même personne ou non.

I.4. Conclusion

Les objectifs finaux de la classification, la nature des données et les différentes mesures de ressemblance utilisées ont donné naissance à une large variété des méthodes pour résoudre le problème de la classification. Chacune d'elles possède ses points de force et de faiblesse.

Dans le monde réel, les données de forme aléatoire sont les plus fréquentes. Pour ce type de données les méthodes connexionnistes et celles basées sur la densité sont les plus adaptées. De plus, les travaux effectués dans ce sens pour la classification automatique, avaient comme point commun d'augmenter la robustesse de l'approche de détection des modes.

Parmi les applications les plus courantes de la classification automatique on trouve la segmentation d'image, qui a, à son tour, une variété de méthodes appliquées sur plusieurs domaines y compris la segmentation médicale et la segmentation d'iris.

Dans le cadre de notre étude, nous cherchons à classer des données dans un contexte non supervisé, c'est à dire que nous supposons ne connaître aucune information à priori sur les données à analyser, exceptées celles qui peuvent être extraites des données elles-mêmes. Ces données sont des ensembles d'observations multidimensionnelles des réels \mathbb{R}^N , où N représente le nombre d'attributs nécessaires à la caractérisation de chacun des objets à classer.

L'approche proposée utilise le réseau de neurones comme outil principal dans la procédure de détection des modes. Dans le chapitre suivant, nous mettons en évidence ces fameux réseaux.

CHAPITRE II : APPLICATION DES RESEAUX DE NEURONES A LA CLASSIFICATION AUTOMATIQUE DES DONNEES MULTIDIMENSIONNELLES

II.1. Introduction

Pendant des décennies, de nombreuses recherches ont été menées pour réaliser une modélisation et une simulation de l'intelligence humaine, tout en offrant aux machines plus de vitesse, d'efficacité et d'automatisme. Les modèles de réseaux neuronaux, étant des méthodes prometteuses pour contourner certaines failles des techniques classiques, ont réussi à se propager au-delà du domaine de modélisation biologique.

Les réseaux de neurones artificiels sont des outils bien adaptés, entre autres, à la classification, la segmentation et l'aide à la décision. Ils servent aujourd'hui à toutes sortes d'applications dans divers domaines, à savoir le traitement du signal, le traitement d'image et de parole et la reconnaissance de caractères. Leurs architectures parallèles donnent une puissance de calcul pouvant traiter des données multidimensionnelles non linéaires complexes en temps réel, avec une certaine tolérance aux erreurs.

En effet, un réseau de neurones artificiels est un assemblage d'unités de calcul simples [HAY-94], nommés "neurones", connectés entre eux avec une pondération des poids synaptiques en tant que paramètres du réseau, ces poids sont gérés et adaptés par des règles d'apprentissage qui permettent au réseau de reconnaître les objets qui lui sont présentés. Cela peut se passer dans un mode supervisé, qui suppose que l'on connaisse à priori la sortie désirée, ou non supervisé qui sert dans le cas des données non étiquetées où on ne possède aucune information sur la structure des données.

A l'instar des neurones biologiques, la communication entre les neurones formels revêt plusieurs formes, d'où la diversité des topologies des réseaux et leurs méthodes d'apprentissage. En l'occurrence, les recherches dans le domaine des réseaux de neurones ces dernières années ont une orientation vers les réseaux dynamiques qui se ressemblent dans leurs objectifs mais se différencient dans la façon dont ils travaillent.

Dans ce chapitre, nous ouvrons tout d'abord une fenêtre sur l'histoire des réseaux neuronaux, avant de faire un rappel du neurone biologique et formel. Ensuite, nous donnons un aperçu global sur les réseaux de neurones artificiels, leurs architectures, leurs objectifs visés, leurs fonctionnements et leurs types d'apprentissage.

Dans le cadre de notre étude, nous sommes intéressés aux réseaux de neurones à apprentissage compétitif dont nous présentons, dans le reste du chapitre, ses deux catégories principales, en montrant l'importance du choix de la fonction d'activation et de la métrique utilisées dans les différents modèles. Nous aborderons également certains points de forces et de faiblesses de ces méthodes.

II.2. Historique

Le mécanisme de la mémoire attirait la curiosité des savants depuis l'époque grecque. Pendant des siècles, la vision primitive du système nerveux restait sous la forme d'un réseau continu de nerfs, jusqu'à ce que *Waldeyer* (1891) trouva que le système nerveux est composé de cellules nerveuses indépendantes les unes des autres appelées "neurones" en visualisant ces cellules par la coloration au nitrate d'argent, méthode découverte par *Golgi* (1873, Nobel 1906).

Bien que *Tanzi* (1893) trouva que l'activation répétée des neurones est le responsable de la nature des liens entre ces neurones, le secret de l'apprentissage ne commença à être révélé qu'avec la théorie révolutionnaire de *Cajal* (Nobel 1906) en 1894, qui dit que l'apprentissage se fait grâce à des modifications anatomiques des connexions entre les neurones, nommées après "synapses" par *Sherrington* (Nobel 1932) en 1897, établissant ainsi les bases morphologiques du concept de plasticité.

Il fallait attendre jusqu'au progrès de la microscopie électronique grâce à *Max Knoll* et *Ernst Ruska* (1932, Nobel 1986) pour prouver et analyser le fonctionnement synaptique. Par conséquent, *Hebb* (1949), juste après *Konorski* (1948) qui donna naissance au terme de la plasticité, introduisit la théorie de Potentialisation/Inhibition synaptique, nommée à la suite Loi de Hubb. Cette théorie de *Hebb-Konorski* est encore utilisée jusqu'à aujourd'hui.

En parallèle des recherches neurobiologiques et depuis les années quarante, une nouvelle discipline inspirée de la neurobiologie a vu la lumière, exactement avec *McCulloch et Pitts* qui introduisirent en 1943 un modèle simplifié de neurone biologique (neurone formel). A partir des années cinquante, plusieurs modèles de réseau de neurones étaient inventés en se basant sur le neurone formel. Nous soulignons dans la suite ceux qui construisaient les bases de cette révolution :

- Donald Hebb (1949) demonstrated a simple updating rule for modifying the connection strengths between neurons. Hebb's Law, now called Hebbian learning. This theory is still used until today.
- En 1958, *Rosenblatt* développa le modèle du Perceptron, le premier modèle opérationnel du réseau de neurones,
- En 1960, *Windrow* et *Hoff* présentèrent le modèle de l'ADALINE (*ADaptive LINear Element*), le modèle de base des réseaux multicouches,
- En 1969, *Minsky* et *Papert* publièrent une étude mettant en évidence les limitations du Perceptron, qui résident dans l'impossibilité de classer les formes non linéairement séparables,
- En 1972, *Kohonen* présenta ses travaux sur les mémoires associatives et proposa des applications à la reconnaissance de formes,
- En 1982, *Hopfield* démontra l'intérêt d'utiliser les réseaux récurrents pour la compréhension et la modélisation des fonctions de mémorisation.
- En 1986, *Rumelhart* popularise l'algorithme du gradient qui permet la rétropropagation et l'entraînement par couches cachées des réseaux multicouches, permettant ainsi de dépasser les limites du Perceptron.

Et depuis, les réseaux de neurones ont été beaucoup étudiés, et ses nombreuses applications concernent divers domaines.

II.3. Du neurone biologique au neurone formel

II.3.1. Le neurone biologique

II.3.1.1. Définition

Le système nerveux du cerveau humain est constitué d'environ 10^{11} neurones ; cela veut dire qu'en moyenne, chaque neurone a de l'ordre de 10^4 connexions. A savoir qu'il existe quelques dizaines de catégories distinctes de neurones qui sont dénombrées jusqu'à présent [VIB-05].

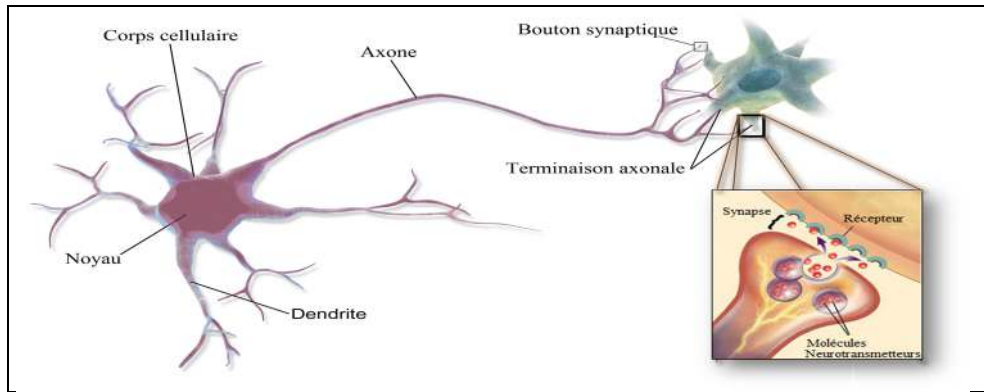


Figure II.1: Schéma d'une cellule neuronale

En générale, un neurone est composé de trois parties principales:

1. **Les dendrites:** Ce sont les récepteurs principaux du neurone pour capter les signaux qui lui parviennent.
2. **Le corps cellulaire:** Il contient un noyau et du cytoplasme, comme toutes les cellules. C'est un sommateur à seuil, il effectue une sommation des influx nerveux transmis par ces dendrites. Si la somme est supérieure au seuil, le neurone répond par un flux nerveux ou par un potentiel d'action qui se propage le long de son axone. Si la somme est inférieure au seuil, il reste inhibiteur.
3. **L'axone:** Il présente une membrane plasmique qui génère et transporte des potentiels d'action, synthétise les neurotransmetteurs, les stocke puis les libère à son extrémité qui est l'élément présynaptique des synapses [VIB-05].

II.3.1.2. Fonctionnement neurobiologique

Le cerveau est un ensemble des unités auto-organisées et dynamique qui stockent et traitent l'information de manière distribuée et parallèle. En effet, un neurone, en tant qu'une cellule nerveuse, possède des caractéristiques un peu spéciales par rapport aux autres cellules nerveuses.

A la réception d'un influx nerveux par les dendrites provenant des neurones voisins, le neurone intègre et évalue la somme de la stimulation reçue, si elle est inférieure à un certain seuil, il inhibe le signal, affaiblissant ainsi la connexion synaptique, sinon il s'excite et engendre un influx nerveux conduit à travers l'axone jusqu'aux boutons synaptiques (cf. fig.1) qui le transmet aux autres neurones voisins, augmentant ainsi la connexion synaptique (Loi de Hubb).

Le fonctionnement neuronal ne se restreint pas dans ces fonctions, il existe notamment une pertinence importante deux à deux entre les régions cérébrales éloignées manifestant par une synchronie d'activité [VAR-01]. En outre, en 2014 *O'Keefe, Mr et Mme Moser* ont eu un prix Nobel¹ pour leurs découvertes de cellules nerveuses dans le cerveau qui permettent d'avoir un sens du lieu et de la navigation [BJE-14].

Ces découvertes et beaucoup d'autres encore, continuent à tracer des nouveaux itinéraires pour les réseaux de neurones artificiels.

II.3.2. Le neurone formel

II.3.2.1. Définition

Le premier modèle de neurone formel, celui de *MacCulloch et Pitts* [MCC-43], est un automate avec des entrées et une sortie binaire. Ce neurone formel effectue une somme pondérée de ses entrées, qui peuvent être des signaux issus d'autres neurones ou des données externes, suivie d'un seuillage, comme le montre la figure II.2.

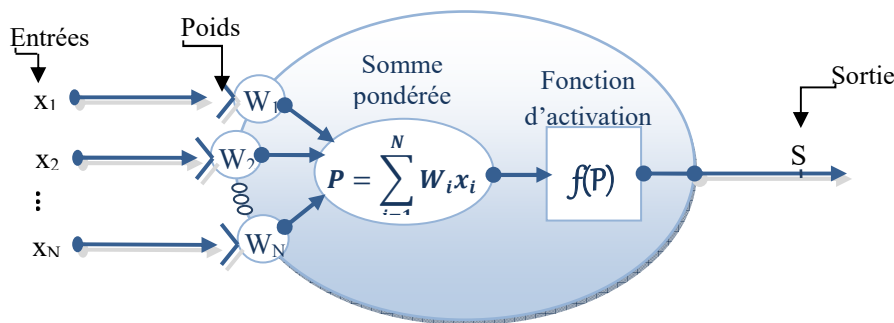


Figure II.2: Le modèle du neurone formel

Le neurone formel est caractérisé mathématiquement par :

1. **Les entrées** : définies par un vecteur $X = \{X_1, X_2, \dots, X_N\}$,
2. **Les poids synaptiques** $W = \{W_i, i = 1, 2, \dots, N\}$: appelés aussi coefficients de pondération. Ils caractérisent l'importance attribuée aux différents stimuli provenant d'autres neurones. Ainsi, chaque synapse a sa propre mémoire.
3. **L'entrée totale** P : le produit scalaire du vecteur des entrées et du vecteur des poids synaptiques. Elle est définie par : $P = \sum_{i=1}^N W_i x_i$.
4. **La fonction d'activation** f ou fonction de transfert qui calcule à partir du potentiel d'entrée P la réponse S du neurone artificiel $S = f(P)$. La fonction utilisée par le modèle *Mac Culloch et Pitts* est la fonction de Heaviside (tableau II.1).

¹ https://www.nobelprize.org/nobel_prizes/medicine/laureates/2014/press.html

II.3.2.2. Fonctionnement d'un neurone formel

Généralement le fonctionnement interne d'un neurone formel distingue entre deux phases : l'évaluation de la somme pondérée P et l'évaluation de son activation f. Mathématiquement,

soit deux algèbres A₁ et A₂, le neurone formel ξ est la composition des fonctions P et f :

$$\xi = P \circ f; \quad P : A_1^N \rightarrow A_2 \quad f : A_2 \rightarrow A_1$$

Ainsi un neurone formel est caractérisé par ses poids de connexion, le niveau d'activation reçu en entrée et son état qui peut être binaire, discrète ou continue. L'état de neurone est défini par sa fonction d'activation dont plusieurs formes existent. Le tableau II.1 énumère quelques fonctions d'activation selon l'état du neurone :

Etat de neurone	Fonction d'activation	Equation Mathématique	Présentation Graphique
Neurone binaire	Fonction Binaire à seuil	Si $P > \theta$ alors $f(P) = 1$; Sinon ; $f(P) = -1$;	
	Fonction Heaviside (tout ou rien)	Si $P > 0$ alors $f(P) = 1$; Sinon ; $f(P) = 0$;	
Neurone discret	Fonction Plancher	$f(P) = [P]$	
Neurone continu	Fonctions sigmoïdes	$f(P) = \frac{1}{1 + e^{-\lambda P}},]0,1[$	
	Fonction Linéaire à seuils	$f(P) = \min\{1, \max(-1, P)\} \in]-1,1[$	
	Fonction Gaussienne	$f(P) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(P-\mu)^2}{2\sigma^2}}$	

Tableau II.1 : les fonctions d'activation courantes selon l'état de neurone

- **Neurone binaire** : les entrées et la sortie de ce neurone sont booléennes. En fait, l'automate se met dans son état actif et prend un état égal à +1 si le potentiel d'entrée du neurone est positif. Dans le cas contraire, il est inactif et prend un état égal à 0 (fonction de Heaviside) ou -1.
- **Neurone discret** : sa fonction d'activation prend ses valeurs dans un intervalle de l'ensemble des entiers naturels \mathbb{N} .
- **Neurone continu** : ce type de neurone est plus adapté aux distributions des données non linéairement séparable, qui nécessite l'utilisation des fonctions non linéaires telles que les fonctions sigmoïdes (arc tangente, tangente hyperbolique,..), utilisées par exemple dans la règle d'apprentissage basée sur la rétropropagation du gradient de l'erreur [WERB-74], ou la fonction gaussienne employée par le modèle neuronale du réseau RBF (*Radial Basis Function*) [PARK-91].
- Neurone Distance

Comme son nom l'indique, le neurone distance utilise la distance pour calculer le potentiel P entre le vecteur d'entrée et le vecteur des poids synaptiques. Généralement, la fonction d'activation f du neurone est une fonction Identité ($f(P) = P$) ou gaussienne.

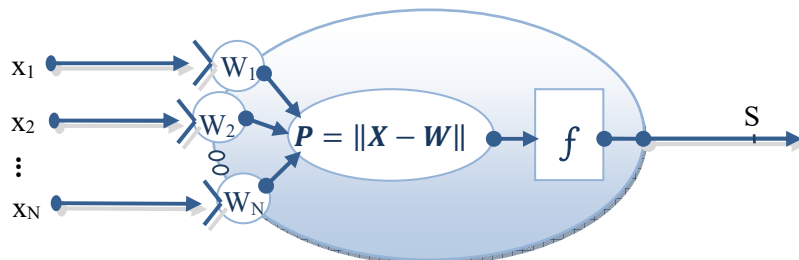


Figure II.3: neurone distance

En effet, le neurone distance compare le vecteur d'entrée X au vecteur poids W en calculant la distance entre eux, $P = D_{i=1}^N [x_i, w_i]$, la métrique choisie habituellement est la distance Euclidienne, vu sa simplicité et son utilité pour les données de dimensionnalité très élevées [XIA-15]. Il existe des neurones qui adoptent d'autres métriques comme la distance de Mahalanobis ou la distance de Manhattan (cf. §I.3.3).

- Exemple du fonctionnement de neurone

Pour analyser le fonctionnement d'un neurone formel, considérons un neurone binaire à deux entrées x_1 et x_2 et avec une fonction d'activation tout ou rien (cf. figure II.4).

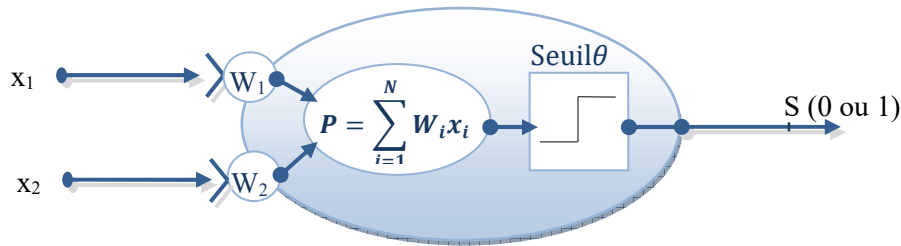


Figure II.4 : Neurone formel avec deux entrées

On suppose qu'on cherche à traiter une distribution bidimensionnelle de Q observations (cf. §I.1.2.1.1). Si on présente une observation $X_q = \begin{bmatrix} x_{1,q} \\ x_{2,q} \end{bmatrix}$, avec $q=1,2,\dots,Q$, aux entrées du neurone, ce dernier calcule la somme pondérée P des attributs du vecteur d'observation X_q et les composantes du vecteur poids :

$$W = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} : P = w_1 x_1 + w_2 x_2 \quad (\text{II.1})$$

Ensuite, il compare cette somme à θ pour définir la valeur de la sortie S :

$$\begin{cases} P = w_1 x_1 + w_2 x_2 > \theta, & S = 1 \\ P = w_1 x_1 + w_2 x_2 < \theta, & S = 0 \end{cases} \quad (\text{II.2})$$

Graphiquement, ces deux inéquations créent une droite d'équation : $w_1 x_1 + w_2 x_2 - \theta = 0$, qui sépare entre deux régions du plan de l'espace des observations (cf. fig II.5).

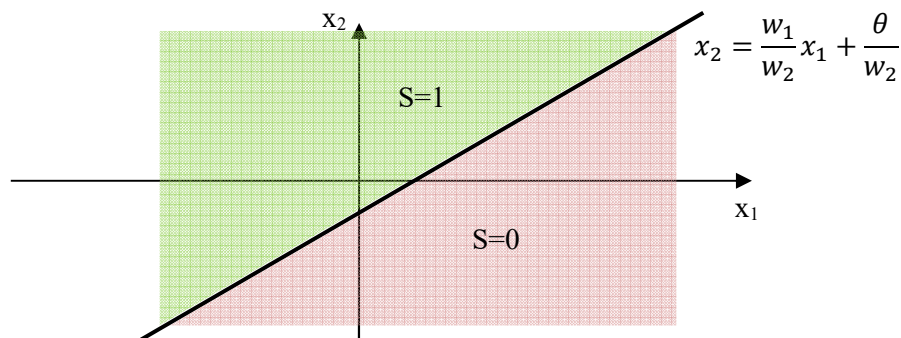


Figure II.5: Interprétation géométrique de la réponse d'un neurone formel

Ainsi nous pouvons dire qu'un neurone formel avec une fonction d'activation linéaire permet de séparer un ensemble d'observations multidimensionnelles en deux classes par un hyperplan.

De manière générale, le neurone formel est l'unité d'une architecture plus complexe où les neurones s'interconnectent pour former un réseau neuronal artificiel.

II.4. Réseaux de Neurones Artificiels (RNA)

- Définition

Un réseau de neurones artificiel est un processeur massivement distribué en parallèle qui a une propension naturelle pour stocker de la connaissance expérimentale et la rendre disponible à l'usage. Il ressemble au cerveau sur deux aspects [HAY-94]:

1. La connaissance est acquise par le réseau à travers un processus d'apprentissage,
2. Les connexions entre les neurones, connues sous le nom de poids synaptiques servent à stocker la connaissance.

- Intérêt de RNA

Ce qui explique l'attrait des RNA en tant que technique d'analyse de données, est leur disposition de certaines caractéristiques qui leurs distinctes des autres techniques telles que :

- Le parallélisme qui réduit le temps de calcul en permettant à un grand nombre des unités simples de fonctionner en parallèle.
- La capacité d'apprentissage qui offre une certaines auto-amélioration compte tenu du taux d'erreur,
- La mémoire est distribuée sur tout le réseau au lieu d'être localisée physiquement, la chose qui permet une tolérance des données manquantes et une évolution dynamique.

Toutes ces caractéristiques apportent de plus en plus au RNA une performance optimale.

En se basant sur les études biologiques, les chercheurs ont pu inférer divers modèles de réseaux de neurones artificiels caractérisés par leurs architectures, leurs méthodes d'apprentissage et surtout par leurs applications visées. Dans le cadre de notre étude nous visons l'application du réseaux de neurones sur la classification automatique et la segmentation d'image.

II.4.1. Architecture du réseau de neurones

Il n'est pas exagérant de dire que les modèles actuels des réseaux de neurones sont indénombrables, vues que chaque mois voire chaque semaine des variations de réseaux classiques sont inventées, afin d'arriver à la meilleure conception neuronale conduisant aux performances désirées selon les données du problème à traiter.

Il est primordial de savoir que la structure du graphe neuronal est configurée selon sa connectivité complète ou partielle. En effet, il existe deux topologies globales des réseaux de neurones ; les réseaux bouclés ou récurrents et les réseaux non bouclés.

II.4.1.1. Les réseaux récurrents (*Feedback*)

En général, comme son nom l'indique ce type des réseaux permet une récurrence de traitement au sens inverse, c'est-à-dire ; la propagation de l'information dans le réseau est au sens contraire, ce qui impose une topologie de réseau à connexions bouclées :

- **Totales** : de telle façon que chaque neurone est relié à tous les autres neurones et possède parfois même un retour sur lui.
- **Partielles** : dans lesquelles l'information circule de couche en couche tout en permettant sa rétropropagation.

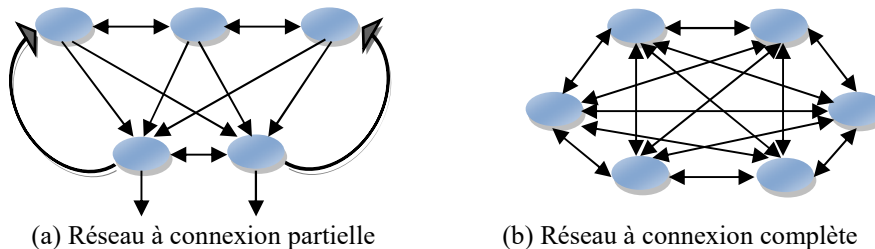


Figure II.6: topologie des réseaux récurrents

Là où toute connexion entre les neurones est possible, divers réseaux ont été conçus tels que les réseaux *Hopfield* [HOP-82], les *cartes auto organisatrices de Kohonen* [KOH-82] et les réseaux du *gaz neuronal* [MAR-91].

II.4.1.2. Les réseaux non bouclés (*Feed-forward*)

Ces réseaux sont structurés sous forme de couches de neurones ; en cas de plusieurs couches, la première et la dernière sont respectivement la couche d'entrée et la couche de sortie, avec des couches intermédiaires nommées couches cachées. Les neurones constituant la même couche ne sont pas connectés entre eux, par contre chaque neurone est connecté à tout ou à un nombre réduit de neurones de la couche suivante. L'information dans ces réseaux se propage des entrées vers les sorties en un sens unidirectionnel.

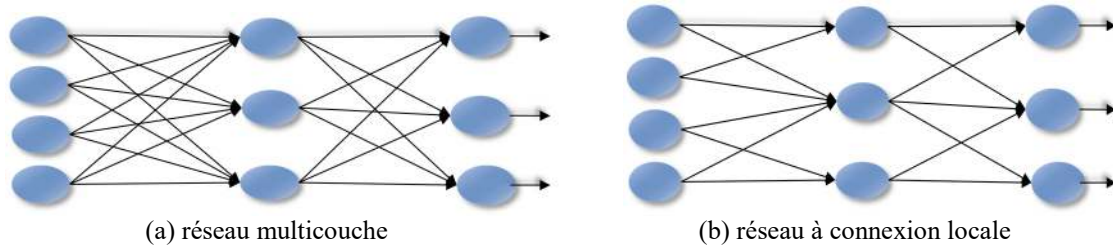


Figure II.7: topologie des réseaux non bouclés

Parmi les réseaux les plus célèbres de ce type on trouve le *perceptron monocouche* [ROS-58] et les *réseaux multicouches* dont les neurones de la même couche sont exclusivement reliés à l'ensemble des neurones de la couche suivante, tels que le *perceptron multicouche (PMC)* et le réseau *RBF* [BRO-88] qui est un PMC mais avec une fonction d'activation et un algorithme d'apprentissage différents.

En revanche, cette catégorisation des réseaux de neurones est fortement liée à la technique d'apprentissage adoptée par chacun de ces réseaux.

II.4.2. Apprentissage des réseaux de neurones

Après avoir défini l'architecture de réseau de neurones, il est inéluctable de lui affecter un algorithme d'apprentissage, malgré que la topologie de certains réseaux de neurones n'est créée qu'au cours de l'apprentissage (Ex: *Gaz Neuronal*).

De point de vue global, l'apprentissage est un processus itératif de développement améliorant la performance d'un réseau de neurones, durant lequel les poids sont modifiés dans le but de permettre à la sortie d'être aussi proche que possible du comportement désiré.

II.4.1.1. Types d'apprentissage

Les algorithmes d'apprentissage sont divisés en deux grandes classes, apprentissage supervisé et non supervisé.

II.4.1.1.1. Apprentissage supervisé

Les problèmes à traiter par les réseaux adoptant ce type d'apprentissage ont un objectif déjà déterminé. En effet, l'apprentissage supervisé consiste à présenter séquentiellement les observations et les classes auxquelles elles appartiennent. Le réseau adapte ses poids W de manière à minimiser l'erreur entre la sortie Y calculée par le réseau et celle désirée Y' connue à priori. De ce fait, il devient capable de classer une observation inconnue présentée à son entrée.

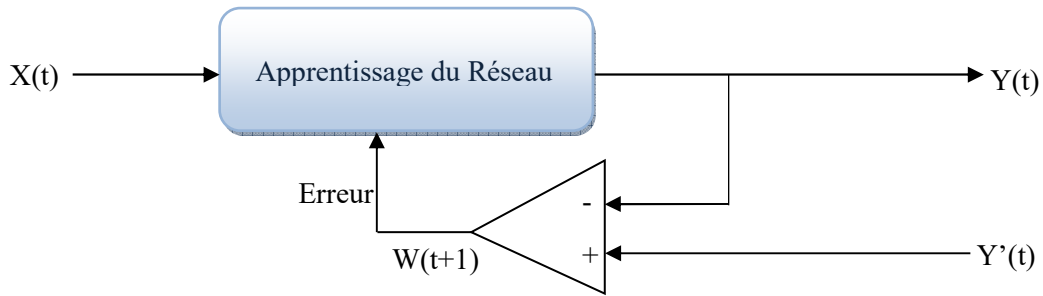


Figure II.8: Mode d'apprentissage supervisé

Cette technique d'apprentissage est basée sur l'idée de faire apprendre en donnant des exemples, en raison de profiter des connaissances acquises au futur. Parmi les réseaux de neurones à apprentissage supervisé on trouve les *Mémoires associatives* [SHL-76] qui fournit toute ou une partie de l'information mémorisée pour la reconstruction des données, l'annuaire électronique est une de ses applications. Cependant, la majorité des problèmes de classification n'ont pas une solution déterminée au paravent, d'où la nécessité des méthodes non supervisés.

II.4.1.1.2. Apprentissage non supervisé

L'apprentissage non supervisé est basé juste sur les données d'entrée sans avoir des données étiquetées qui approuvent la validité des données de sortie. En fait, c'est un apprentissage qui permet de découvrir la structure des données à partir des données elles-mêmes, sans information a priori quant à leur appartenance à différentes classes présentes dans l'échantillon analysé. Il consiste donc à assurer le maximum de similarité entre les observations pour grouper dans la même classe les observations semblables.

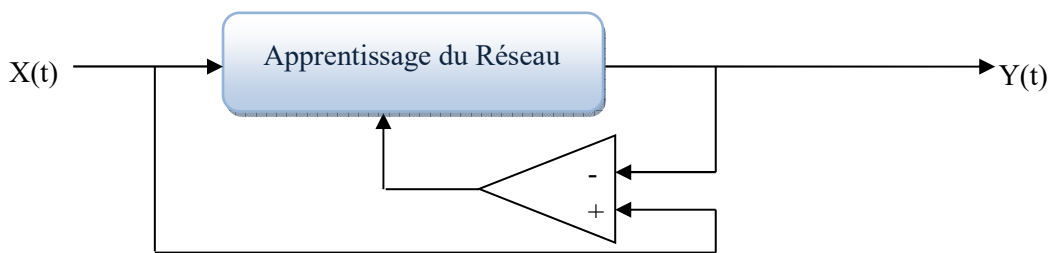


Figure II.9: Mode d'apprentissage non supervisé

Les données inexploitable sont les plus répondues dans la nature, ce qui explique l'orientation de la plupart des méthodes vers ce type d'apprentissage citant les algorithmes de *K-moyennes*, *cartes auto organisatrices (SOM)* et *la théorie de résonance adaptative (ART)*. Les réseaux de neurones à apprentissage non supervisé les plus utilisées sont souvent des réseaux compétitifs.

II.4.1.2. Règles d'apprentissage

La représentation de comportement dynamique d'un processus d'apprentissage est réalisée par l'obtention d'un modèle mathématique pour gérer les variations des poids synaptiques (cf. §II.3.2.1).

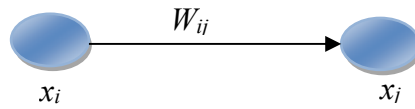


Figure II.10 : connexion entre deux neurones

Pour cela, il existe plusieurs règles de modification de poids dont nous citons les principales :

- **Règle de Hebb** : Parmi les lois fondamentales de l'apprentissage fondée par Donald Hebb en 1949, elle consiste à renforcer le poids de connexion entre deux neurones activés simultanément. C'est une loi corrélationnelle qui a fait naissance à divers modèles de réseaux de neurones, tels que le réseau de *Hopfield* et ses dérivés [HOP-07] et le *réseau BSB* (mémoire associative) qui utilisent des réseaux totalement connectés. Cette théorie est, jusqu'à nos jours, le sujet de plusieurs études [KUR-15].
- **Apprentissage par correction d'erreur** : c'est un apprentissage supervisé qui vise à minimiser le taux d'erreur entre la sortie théorique attendue et la sortie obtenue, à travers la modification des poids des neurones.
- **Apprentissage de Boltzmann** : c'est un apprentissage de type stochastique. Il ajuste les poids des connexions, pour que l'état des cellules visibles correspond à une distribution probabiliste souhaitée.
- **Apprentissage par compétition** : c'est un apprentissage issu de la méthode de k-Moyennes (cf. §I.4.1.2.2). Il permet d'estimer les noyaux des classes, chaque noyau est considéré comme un prototype d'une classe. Il est largement appliqué aux problèmes de classification.

Ce dernier type d'apprentissage fait l'objet de notre approche, dont nous allons mettre l'accent sur ses différentes procédures.

II.5. Apprentissage Compétitif (AC)

Cette procédure d'apprentissage permet à une partie du réseau d'apprendre uniquement des exemples qui le concernent : seul le ou les neurones de sortie gagnants d'une compétition subissent une mise à jour de leurs poids synaptiques, à l'opposé des autres règles

d'apprentissage utilisées jusqu'alors où, généralement, les mise à jour des connexions sont appliquées simultanément et de la même façon. Cette stratégie offre plus de flexibilité et de vitesse au réseau.

Cet apprentissage, étant une adaptation neuronale de l'algorithme des K-Moyennes, est basé sur un critère de ressemblance pour assurer le degré maximum de similarité entre les éléments du même groupe et le degré maximum de dissimilarité entre les éléments de différents groupes, ce qui invoque l'adoption des neurones distance.

Dans l'algorithme compétitif le neurone de sortie appelé aussi le neurone de compétition est caractérisé par un vecteur poids, qui n'est autre qu'une estimation du centre d'une classe, et aussi par une fonction d'activation sous forme d'une métrique qui gère la compétition entre les neurones lors de la présentation des observations à l'entrée de réseau.

L'apprentissage compétitif se situe au carrefour de plusieurs réseaux de neurones qu'on peut classifier selon la méthode d'apprentissage adoptée. Nous distinguons alors entre deux classes globales ; l'*apprentissage compétitif dur* et l'*apprentissage compétitif doux*.

II.5.1. Apprentissage compétitif dur

L'apprentissage compétitif dur est l'ensemble des méthodes compétitives standards qui permettent à un seul neurone de sortie à gagner la compétition, d'où son nom *Winner Take All* (WTA). Le neurone gagnant est celui qui a la probabilité d'appartenance maximum selon la règle de Bayes [STU-94].

Différentes méthodes spécifiques peuvent être obtenues en effectuant un des deux paradigmes dites *Batch* et *on-line* (K-moyennes) qui se distinguent comme suivant [FRI-97] :

- ***BaTch*** : toutes les entrées d'un ensemble fini possible sont évaluées d'abord avant que toute adaptation soit faite, dans un système itératif,
- ***On-line*** : effectue, par contre, une actualisation des poids après chaque entrée d'une observation.

L'apprentissage compétitif dur est implémenté sur des réseaux multicouches (cf. fig II.9(a)) ; supposons qu'on a un échantillon de Q observations estimées et groupées en K classes :

- La couche d'entrée d'un réseau composé de deux couches, est constituée de N neurones, correspondant au nombre N des attributs $x_{n,q}$ de l'observation X_q , $X_q = \{x_{1,q}, x_{2,q}, \dots, x_{N,q}\}$ avec $n=1, \dots, N$ et $q=1, \dots, Q$, ces neurones transforment juste

l'information présentée à l'entrée du réseau aux neurones de la couche suivante sans effectuer aucun traitement.

- La couche de sortie à son tour, est constituée de K neurones représentant les K classes menées d'un vecteur poids μ_k , $k=1, \dots, K$, $\mu_k = (\omega_{k,1}, \omega_{k,2}, \dots, \omega_{k,N})^t$ et d'une métrique $D(X_q, \mu_k)$ qui calcule la distance entre l'observation X_q et le vecteur poids μ_k du neurone

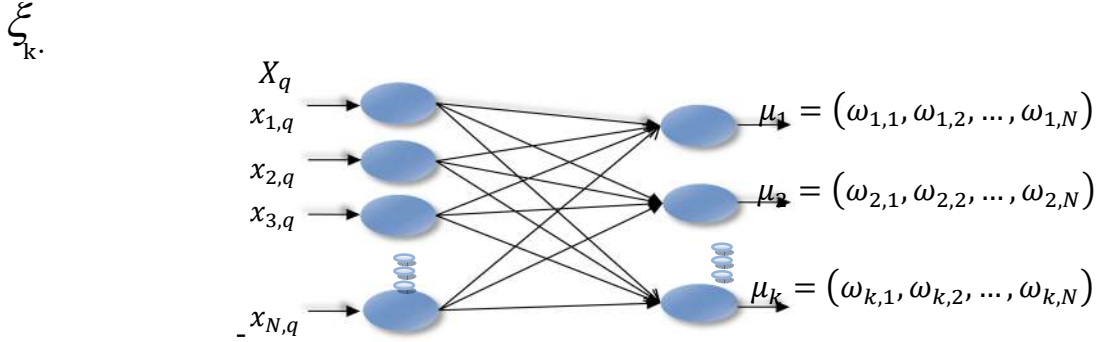


Figure II.11: Apprentissage de réseau à deux couches

II.5.1.1. Principe d'apprentissage compétitif du réseau multicouche

Dans le but de partitionner un échantillon des observations E en K classes, le principe de l'apprentissage est de modifier les paramètres de la fonction d'activation des neurones distance automatiquement à travers la démarche suivant :

On cherche à représenter chaque classe C_k par un neurone ξ_k dont le vecteur poids μ_k est une estimation du centre m_k de la classe C_k , i.e. le centre de gravité des observations de la classe C_k :

$$\text{Soit } m_k \in R^N, \quad m_k = \frac{1}{Q_k} \sum_{X_q \in C_k} X_q \quad (\text{II.3})$$

Q_k est le nombre estimé des observations dans la classe C_k .

- ✓ A l'état initiale, les vecteurs poids de neurones distance prennent leurs valeurs soit en leur affectons des vecteurs d'observations tirés aléatoirement parmi les observations de l'ensemble E , soit par des valeurs aléatoires.
- ✓ Lors de la présentation d'une observation $X_q(t)$ à l'itération t , chaque neurone calcule la distance $D_k(t)$ entre son vecteur poids $\mu_k(t-1)$ et l'observation $X_q(t)$, telle que :

$$D_k(t) = [X_q(t) - \mu_k(t-1)]^T [X_q(t) - \mu_k(t-1)] \quad (\text{II.4})$$

Le neurone dont la distance est minimale est déclaré gagnant et sa sortie S_g est mise à 1, alors que les sorties des autres neurones sont mises à 0.

$$\begin{cases} S_k = 1 \text{ si } k = g, g \text{ est l'indice du neurone gagnant } \xi_g \\ S_k = 0 \text{ si } k \neq g, \quad k = 1, 2, \dots, K \end{cases}$$

✓ Le vecteur poids du neurone gagnant ξ_g est modifié afin de minimiser l'erreur $e(t)$:

$$\begin{aligned} e_{AC}(t) &= \frac{1}{2} \sum_{k=1}^{k=K} S_k D_k(t) \\ &= \frac{1}{2} \sum_{k=1}^{k=K} S_k [X_q(t) - \mu_k(t-1)]^T [X_q(t) - \mu_k(t-1)] \end{aligned} \quad (\text{II.5})$$

L'équation (II.5) peut donc être simplifiée comme suit:

$$e_{AC}(t) = \frac{1}{2} \|X_q(t) - \mu_k(t-1)\|^2 \quad (\text{II.6})$$

L'apprentissage du réseau est effectué en calculant la sensibilité de l'erreur par rapport au vecteur poids μ_k .

$$\frac{\partial e_{AC}(t)}{\partial \mu_g} = - [X_q(t) - \mu_g(t-1)] \quad (\text{II.7})$$

L'adaptation des neurones est effectuée selon la méthode de descente du gradient telle que :

$$\mu_g(t) = \mu_g(t-1) - \alpha(t) \frac{\partial e_{AC}(t)}{\partial \mu_g} \quad (\text{II.8})$$

où $\alpha(t)$ est une fonction d'apprentissage que nous allons définir à la fin de ce paragraphe.

En remplaçant dans l'équation (II.8) $\frac{\partial e_{AC}(t)}{\partial \mu_g}$ par son expression de l'équation II.7, nous obtenons :

$$\mu_g(t) = \mu_g(t-1) + \alpha(t) [X_q(t) - \mu_k(t-1)] \quad (\text{II.9})$$

➤ **Interprétation géométrique :**

A chaque présentation d'une observation X_q , on rapproche le vecteur poids μ_g du neurone gagnant ξ_g d'une fraction de la distance qui le sépare de X_q , alors qu'on laisse les autres

neurones sans modification de leurs vecteurs poids de telle sorte que : $\forall k \neq g, k = 1, 2, \dots, K$
 $\mu_k(t) = \mu_k(t - 1)$.

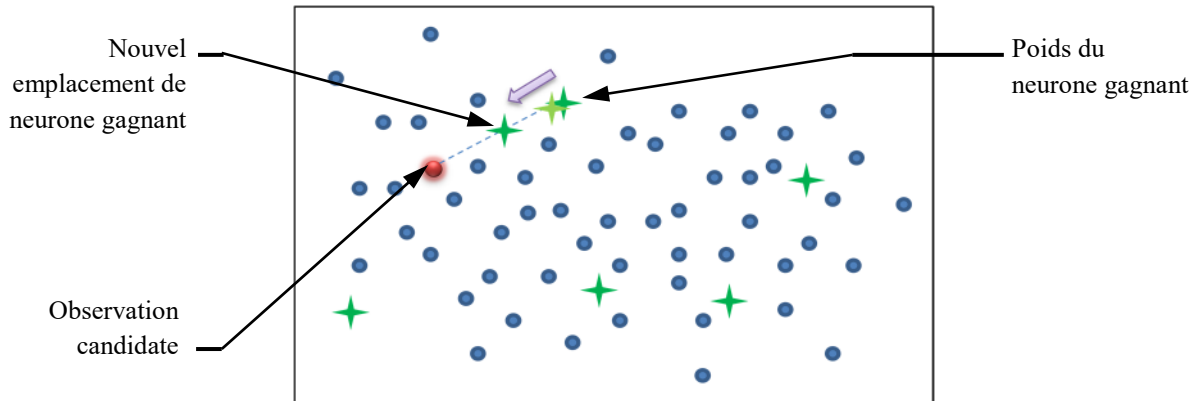


Figure II.12 : Principe de l'apprentissage compétitif standard

II.5.1.2. Algorithme d'apprentissage compétitif dur

➤ Phase d'initialisation :

- Fixer un nombre maximum d'itération T ,
- Fixer le nombre K de neurones de la couche de sortie,
- Prendre K observations de l'ensemble E comme vecteurs poids initiaux des neurones,
- Initialiser l'époque à 1 qui représente une période de Q itérations avant de reprendre à nouveau le traitement.

➤ Phase de traitement :

1. Tirage aléatoire d'une observation $X_q(t)$ à partir de l'échantillon E ,
2. Calcul de la distance $D[X_q(t), \mu_k(t)]$ pour chaque neurone de sortie $\xi_k, k=1, 2, \dots, K$

$$D[X_q(t), \mu_k(t)] = [X_q(t) - \mu_k(t - 1)]^T [X_q(t) - \mu_k(t - 1)] \quad (\text{II.10})$$

3. Recherche du neurone gagnant ξ_g , dont le vecteur poids est le plus proche de l'observation $X_q(t)$:

$$D[X_q(t), \mu_g(t)] = \min_{k=1} D[X_q(t), \mu_k(t)] \quad (\text{II.11})$$

4. Modification du vecteur poids du neurone gagnant :

$$\begin{cases} \mu_k(t) = \mu_k(t-1) + \alpha(t)[X_q(t) - \mu_k(t-1)] & \text{si } k = g \\ \mu_k(t) = \mu_k(t-1) & \text{si } k \neq g \end{cases} \quad (\text{II.12})$$

5. Si $t < T$ et $q < Q$ alors retour en 1,
6. Si $t < T$ et $q = Q$, alors $P = P + 1$ et retour en 1, si $t = T$ Alors fin.

II.5.1.3. Fonction d'apprentissage

D'après la relation (II.9) on peut déduire que la fonction d'apprentissage (*learning rate*) $\alpha(t)$ est une fonction qui détermine la mesure dans laquelle le neurone gagnant devrait s'approcher de l'observation en entrée à chaque itération. Cette fonction peut prendre plusieurs formes [DAR-91] qui doivent satisfaire aux conditions stochastiques de convergence [GRO-82] :

- ✓ Condition de plasticité : $\sum_{t=0}^{\infty} \alpha(t) = \infty$
- ✓ Condition de stabilité : $\sum_{t=0}^{\infty} [\alpha(t)]^2 < \infty$

La condition de plasticité garantit la convergence des poids vers les observations, tandis que la condition de stabilité permet d'éviter la divergence de l'algorithme.

La forme la plus simple de cette fonction d'apprentissage est une fonction constante :

$$\alpha(t) = \alpha_0 \quad ; \quad 0 < \alpha_0 \leq 1 \quad (\text{II.13})$$

Cependant, un taux d'apprentissage constant, représentant la valeur prise par cette fonction à l'itération t , ne permet pas aux vecteurs poids de converger lorsque le nombre d'itération tend vers l'infini. Les positions des vecteurs poids pourraient osciller autour des centres des classes sans jamais converger vers une position stable.

Pour permettre aux vecteurs poids de converger vers une solution stable, il était impératif d'ajouter une autre condition :

$$\lim_{t \rightarrow \infty} \alpha(t) = 0$$

La fonction appelée « *Running Average Schedule* » [DAR-91] répond à ces conditions, définie comme suit :

$$\alpha(t) = \frac{\alpha_0}{1+t} \quad (\text{II.14})$$

Cette fonction permet aux vecteurs poids de tendre vers une solution stable et donc on aura : $\mu_g(t) = \mu_g(t-1)$.

Néanmoins, le choix de la constante α_0 est crucial. Si la valeur de α_0 est mal choisie, $\alpha(t)$ tend alors rapidement vers zéro, et par conséquent, $\mu_g(t) = \mu_g(t-1)$ pour tout g, ce qui empêchent les vecteurs poids de converger vers les centres des classes.

Une autre fonction introduite par Darken, appelée « *Search then Converge Schedules* », facilite la tâche du choix des paramètres :

$$\alpha(t) = \frac{\alpha_0}{1 + \frac{\tau}{t}} \quad (\text{II.15})$$

Elle est caractérisée par une évolution en deux phases :

- ✓ Une première phase, dite de recherche de durée τ pendant laquelle le taux d'apprentissage est important, permet aux vecteurs poids de s'approcher rapidement des centres des classes, on peut écrire pour τ tend vers l'infini $\alpha(t) = \alpha_0$,
- ✓ Dans la deuxième phase, le nombre d'itérations t est très supérieur à τ et la fonction d'apprentissage décroît en $\alpha_0 \frac{\tau}{t}$ ce qui assure la convergence des vecteurs poids.

Le choix de α_0 et τ ne pose pas de problèmes. Il suffit de fixer τ à une valeur suffisamment grande pour permettre la recherche des centres des classes avant la convergence vers ces centres.

L'importance de la fonction d'apprentissage réside dans le fait que la précision de classification peut varier considérablement à cause d'elle. Ainsi, plusieurs recherches ont été dirigées pour optimiser cette fonction, certains chercheurs ont pensé à intégrer un modèle flou dans la fonction d'apprentissage [KAM-15], et d'autres ont essayé de trouver une fonction d'apprentissage optimale en utilisant un taux d'apprentissage adaptatif [ZEI-12] en conjonction avec le concept de partitionnement Laplacien pour faire varier les poids [CHA-16]. Bien que ces travaux ont prouvés leurs bonnes performances, les résultats obtenus n'ont pas été comparés à ceux utilisant la fonction « *Search then Converge Schedules* » comme fonction d'apprentissage.

II.5.1.4. Les extensions des modèles à apprentissage compétitif dur

En général, les algorithmes de l'apprentissage compétitif dur souffrent de plusieurs problèmes majeurs qui peuvent être causés par l'initialisation aléatoire des poids de neurones, à savoir :

- Les poids aberrants, connu sous le nom de « *dead units* », certains poids ne se déplacent vers aucun centre de classe ou encore plusieurs vecteurs poids se dirigent vers une solution sous optimale,
- Différentes initialisations des poids de neurones peuvent conduire à des résultats très différents,
- Une classe est présentée par plusieurs prototypes.

Pour palier certains de ces inconvénients, différentes variantes ont été développées dont nous analysons certains algorithmes:

- ✓ ***Apprentissage compétitif sensible à la fréquence (ACSF)***: proposé par Ahalt [AHA-90]. Ahalt songea à multiplier la distance entre une observation et un neurone par un terme de conscience ρ propre à chaque neurone :

$$\rho_g(t)D(X_q(t) - \mu_g(t)) = \min_{k=1} \rho_k(t)D(X_q(t) - \mu_k(t)) \quad (\text{II.16})$$

Avec une mise à jour de coefficient de conscience lors de l'activation du neurone gagnant :

$$\begin{cases} \rho_g(t) = \rho_g(t-1) + 1, & \text{si } k = g \\ \rho_k(t) = \rho_k(t-1) & , \quad \text{si } k \neq g \end{cases} \quad (\text{II.17})$$

Ce mécanisme de conscience permet de défavoriser les neurones qui gagnent souvent la compétition et permet aux neurones avec des vecteurs poids loin de nuage des observations de s'y approcher et par la suite de participer au processus de compétition induit par l'algorithme d'apprentissage. Il est parfois conseillé [DES-88] de fixer un terme de conscience maximum ρ_{max} , afin de terminer l'apprentissage par l'algorithme compétitif en mode standard. Cependant, dans une autre étude [BAL-96] l'algorithme de *k-Moyennes* a donné meilleure performance par rapport à l'ACSF.

- ✓ ***Apprentissage Compétitif avec Pénalisation du Rival (ACPR)***: proposé par Xu [XU-93], cette procédure d'apprentissage a été développé en se basant sur l'idée de faire converger le vecteur poids du neurone gagnant g vers le centre d'une classe, et de repousser le

vecteur poids de son rival r , le deuxième neurone gagnant, vers une autre classe. La remise à jour des vecteurs poids se fait donc en respectant ces équations :

$$\begin{cases} \mu_k(t) = \mu_k(t-1) + \alpha_g(t) [X_q(t) - \mu_g(t-1)], & \text{si } k = g \\ \mu_k(t) = \mu_k(t-1) - \alpha_r(t) [X_q(t) - \mu_g(t-1)], & \text{si } k \neq r \\ \mu_k(t) = \mu_k(t-1), & \text{si } k \neq g \text{ et } k \neq r \end{cases} \quad (\text{II.18})$$

Où $0 \leq \alpha_g(t) \leq 1$ et $0 \leq \alpha_r(t) \leq 1$ sont respectivement les coefficients d'apprentissage pour le neurone gagnant et son rival. En pratique, Xu propose d'utiliser des paramètres dépendants du rang d'itération, avec $\alpha_g \gg \alpha_r$.

L'algorithme possède de ce fait, l'avantage de déterminer automatiquement le nombre de classes. En effet, si le nombre de neurones fixés par l'utilisateur dans la phase d'initialisation, est supérieur au nombre de classes, les neurones en surnombre sont repoussés loin de toutes les observations. Par contre, si le nombre de neurone est inférieur au nombre de classes, l'algorithme ne converge pas vers une solution stable et l'initialisation doit être reprise, en augmentant le nombre de neurones. Cette méthode a connu plusieurs développements à la suite [XU-07], tel que *dynamique ACPR* [BUD-06] qui a prouvé que c'est une méthode prometteuse en la comparant avec la méthode de la *classification de minimisation d'erreur (CME)* [PAN-14].

- ✓ ***L'apprentissage compétitif généralisé*** : [PAL-93] qui consiste à modifier les fonctions d'activations de tous les neurones en fonction de leurs distances (Euclidienne) par rapport à l'observation présentée. Malgré que cette méthode arrive à limiter le phénomène « *dead units* », son inconvénient reste le temps de calcul qui est très élevé.
- ✓ ***K-Moyennes Globale (GKM)*** : [LIK-03] Elle consiste à effectuer une classification incrémentale en ajoutant itérativement un nouveau centre suivi par l'application du *k-Moyennes* jusqu'à la convergence. C'est algorithme a connu lui-même plusieurs variations telles que ***Fast GKM*** [XIE-10] et ***modified Fast GKM*** [BAG-11].

Un autre remède à ce problème d'initialisation est de faire un changement exhaustif au paradigme : 'le gagnant prend tout' –WTA- de l'apprentissage compétitif dure et le subroger par 'le gagnant prend la plupart' de l'apprentissage compétitif doux [FRI-97].

II.5.2. Apprentissage Compétitif Doux

Cependant que l'AC Dur traite les données comme des points discrets, l'AC doux les traite en tant que des régions continues géométriquement. Cette vision est inspirée du système biologique de cortex où les stimuli de même nature excitent une région de cerveau bien précise et d'une manière interprétable. La projection de ce concept sur les réseaux de neurones artificiels, a mené à chercher une organisation des neurones de sortie permettant une mise en correspondance de l'espace d'entrée avec l'espace du réseau, afin de préserver la carte topologique des données.

Au niveau algorithmique, l'apprentissage compétitif doux généralement, est basé sur l'idée de 'le gagnant prend la plupart' « *winner take most* » *WTM*. En effet, pas seulement le neurone gagnant qui actualise son vecteur poids mais aussi d'autres neurones ont l'opportunité d'être adaptées en s'appuyant sur leur proximité de l'observation d'entrée [BARA-99].

Les réseaux à apprentissage compétitif doux sont des réseaux récurrents où la structuration de leur topologie permet de distinguer entre deux types d'Apprentissage : *AC doux avec fixation de dimensionnalité de réseau* et *AC doux sans fixation de dimensionnalité de réseau*.

II.5.2.1. AC doux Avec Fixation de Dimensionnalité de Réseau

Il s'agit d'une construction préalable d'une carte sous forme de grille composée de neurones permettant à la suite une modélisation de l'espace des entrées au cours de l'apprentissage. Son point fort réside dans sa capacité de représenter les données de dimensionnalité très élevée sur un espace de faible dimension [KOH-95], ce qui rend ce genre des modèles très utile pour les problèmes de visualisation.

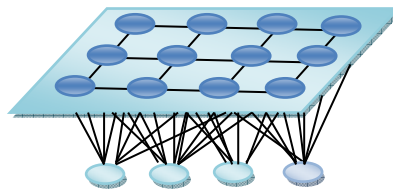


Figure II.13 : Carte auto organisatrice

Les réseaux qui adoptent ce type d'apprentissage sont particulièrement structurés, tels que le *réseau cellulaire croissant (Growing cell structure)* [FRI-94a], l'*incrémental grille croissant (Incremental Growing Grid)* [BLA-95] et les fameuses *cartes auto organisatrices de Kohonen*. Pour mieux comprendre, nous prenons le modèle des cartes de Kohonen comme exemple.

II.5.2.1.1 Les cartes auto organisatrices de Kohonen (SOM)

Ce modèle commença par *Van der Malsburg* en 1976 et élaborait par *Kohonen* en 1982 [KOH-82], son principe est de construire une carte à un ou deux dimensions et définir les relations de voisinage entre les neurones dans le but de représenter les observations voisines dans l'espace des entrées par des neurones voisins dans la carte de sortie. La carte peut être une grille rectangulaire ou hexagonale comme elle peut prendre aussi la forme d'un cylindre ou même une sphère [KIH-08].

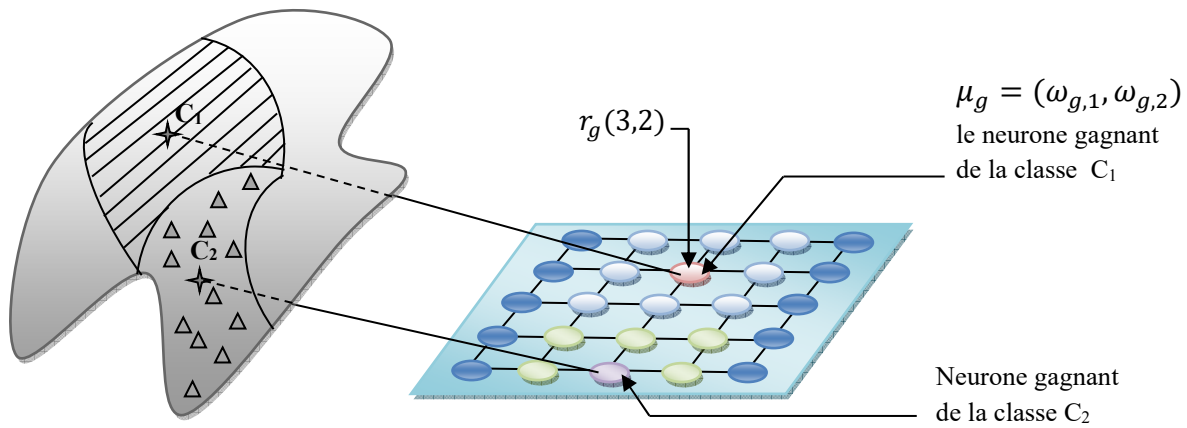


Figure II.14: apprentissage des cartes SOM de Kohonen

En tant qu'un réseau compétitif, les processus d'apprentissage d'une carte *SOM* pour trouver le neurone gagnant μ_g sont, en globale, pareils à ceux d'apprentissage standard, la différence apparaît dans l'influence de μ_g sur ses neurones voisins étroitement liés.

➤ **Pseudo-algorithme d'apprentissage**

1. Générer une carte A de $P = P_1 \times P_2$ grilles tel que $A = \{\mu_{1,N}, \mu_{2,N}, \dots, \mu_{P,N}\}$
2. Initialiser aléatoirement le réseau à partir de l'ensemble des observations E ,
3. Fixer un nombre maximum d'itération T ,
4. Calculer la distance $D[X_q(t), \mu_k(t)]$ pour chaque neurone de sortie ξ_k , $k=1,2,\dots,P$, pour chercher le neurone gagnant, dont le vecteur poids est le plus proche de l'observation $X_q(t)$:

$$D_g[X_q(t), \mu_g(t)] = \min_{k=1} D[X_q(t), \mu_k(t)]$$

5. Modifier le vecteur poids du neurone gagnant et ses voisins :

$$\mu_k(t) = \mu_k(t - 1) + \alpha(t)h(k, D_g)[X_q(t) - \mu_k(t - 1)] \text{ Pour tout } k=1,2,\dots,P \quad (\text{II.19})$$

$\alpha = \frac{\alpha_0}{1+t}$ est le taux d'apprentissage et h est le taux de diffusion défini par la fonction du 'Chapeau Mexicain' :

$$h(k, D_g) = \exp \left[-\frac{(r - r_g)^2}{2\sigma^2} \right] \quad (\text{II.20})$$

Avec r est l'emplacement du neurone et $\sigma = \frac{\sigma_0}{1+\frac{t}{\sigma}}$ une fonction qui représente le coefficient de voisinage.

6. Si $t < T$ alors retour en 1,

L'algorithme de Kohonen a subi plusieurs variations même par son propre hauteur [KOH-02], la plupart de ses extensions [COL-15] [UPA-15] ont été développées pour adapter cette méthode à des données précises, ce qui explique son utilisation sur une grande échelle. D'autres ont été développées pour combler certaines faiblesses et renforcer sa performance [KIA-01] [CHAU-14].

Cependant, il existe une ambiguïté dans les résultats de la classification par SOM, à cause de la difficulté de juger le degré d'appartenance à une classe particulière. Pour remédier à la situation, certains chercheurs ont pensé à appliquer la classification par la FCM sur les unités de grille de SOM permettant ainsi à chaque unité d'avoir une adhésion partielle dans toutes les classes enchevauchées [SAR-11].

II.5.2.2. AC doux Sans Fixation de Dimensionnalité de Réseau

Contrairement à tous les modèles vus jusqu'à l'instant, aucune topologie d'une fixe dimensionnalité n'est imposée [FRI-97], en fait, la dimensionnalité du réseau dépend de la dimension locale des données et peut varier selon l'espace d'entrée.

La structuration du réseau de neurones et son apprentissage se passe en même temps, de telle façon que chaque neurone a la possibilité d'être connecté à tout autre neurone suivant une certaines règles d'apprentissage.

En tant qu'une méthode de préservation de topologie des données, elle demande l'utilisation de deux outils géométriques pour permettre de localiser un objet rapidement ; le *diagramme de Voronoi* et son dual la *triangulation de Delaunay*, que nous donnons, par la suite, un aperçu sur eux :

➤ Diagramme de Voronoi (DV)

Le diagramme de Voronoi, formellement introduit par Voronoi [VOR-1908], est une méthode géométrique aussi appelée tessellation de Voronoi, ce nom donne une idée claire sur la forme de ce diagramme qui présente un partitionnement de l'espace des données, de telle sorte que chaque partition contient une seule unité de donnée (appelé germe), en décomposant l'espace à des régions convexes.

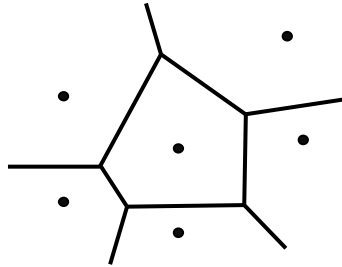


Figure II.15 : Diagramme de Voronoi représenté sur R^2

Le DV V_S d'un ensemble $S = (\omega_1, \omega_2, \dots, \omega_Q)$ de points $\omega_i \in R^N$ est donnée par Q polyèdres de N -dimension, le polyèdre de Voronoi V_i est défini comme suit: Le polyèdre de Voronoi V_i d'un point $\omega_i \in S$ est donnée par l'ensemble des points $v \in R^N$ qui sont plus proches de ω_i que tout autre $\omega_j \in S$ pour tout $i \neq j$ [MAR-94] :

$$V_i = \{v \in R^N \mid \|v - w_i\| \leq \|v - w_j\| \forall j\} \quad (II.21)$$

Le DV a été exploité par les réseaux neuronaux non bouclés et il a montré une utilité pour la dérivation des structures de réseaux de neurones alternatifs pour réaliser la classification désirée [BOS-93]. Au niveau algorithmique, il existe plusieurs algorithmes qui ont été conçus pour la construction du DV d'un ensemble de points, dont nous présentant les plus marquants :

- Le DV discret : Ces méthodes discrètes sont fondées sur la propagation de courbes de niveaux à partir des germes en utilisant des métriques discrètes. L'un des intérêts des méthodes discrètes est que le temps de calcul ne dépend pas du nombre de données. Cependant, elle a un inconvénient concernant l'élaboration de la structure de donnée telle qu'elle a été décrite auparavant [CHAS-91],
- Méthode incrémentale proposé par *Green* et *Sibson* [GRE-78] : Un algorithme qui construit le DV par insertion incrémentale avec un temps de calcul de $O(n^2)$ et il peut atteindre $O(n)$ pour des ensembles bien réparties de sites [AUR-98]. Cette méthode est

devenue populaire pour son spécifiques avantages-simplicité, son occupation économique de l'espace et sa commodité pour la mise à jour dynamique.

- Algorithme '*Divide and Conquer*' : cette méthode récursive, proposée par *Shamos* et *Hoy* en 1975 [SHA-75], est un autre algorithme itératif permettant de modifier localement un diagramme par insertion ou suppression de points, l'intérêt de cette approche concerne la complexité réduite à $O(n \log n)$ qui connue jusqu'à présent des optimisations [CHAN-15]. Cependant, son inconvénient majeur est son incapacité de faire des mises à jour, il exige ainsi une reconstruction complète du diagramme [CHAS-91].

➤ Triangulation de Delaunay (TD)

La triangulation de Delaunay ou tessellation de Delaunay (S) [DEL-34] est obtenue en connectant, avec un segment, chaque deux points quelconques p et q de S de telle façon qu'il existe un cercle C passant par p et q, et ne contient aucun autre site de S à son intérieure ou sur ses bornes. Les arêtes de TD (S) sont appelés arêtes de Delaunay [AUR-98].



(a) le cercle circonscrite de la triangulation de Delaunay

(b) la triangulation de Delaunay correspondante

Figure II.16: la triangulation de Delaunay d'un ensemble de six points

Le TD est aussi un graph où chaque deux sites de Voronoi V_i et V_j avec un arrêt en commun sont reliés par un arrêt de Delaunay, définit comme suivant :

$$DT(F) = \{i, j = 1, \dots, K \mid V_i \cap V_j \neq \emptyset\} \quad (\text{II.22})$$

De même, les sommets du diagramme de Voronoi sont les centres des cercles circonscrits des triangles de la TD. Les arêtes du DV sont sur les médiatrices des arêtes de la TD.

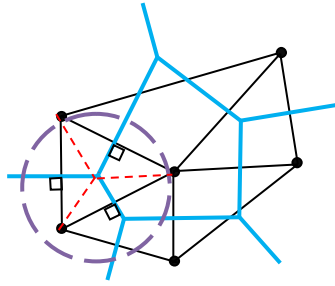


Figure II.17: topologie des réseaux non bouclés

D'un point de vue algorithmique, la TD a une variété d'algorithmes, qui est la même utilisé pour la création du DV, adapté à la création de chacun de ces graphes, tels que l'algorithme '*Divide and Conquer*', l'insertion incrémentale, l'algorithme '*sweep line*' [FORT-87] qui utilise une ligne de balayage et les méthodes utilisant l'enveloppe convexe [DWY-91]. Dans les dernières années, les chercheurs sont devenus de plus en plus attirés par les algorithmes de la TD basés sur le calcul parallèle [LO-12] [LI-13] [BI-14], grâce au développement rapide des ordinateurs. Malheureusement, l'exigence des processeurs multi-core est une limitation pour une large utilisation de parallèle triangulation de Delaunay [SU-16].

L'AC doux sans fixation de dimension inclus le réseau du *Gaz Neuronal* et l'*apprentissage compétitif Hebbien (Competitive Hebbian Learning (CHL))* qui ont fait l'inspiration de plusieurs d'autres réseaux tels que les méthodes *Neural Gas plus Competitive Hebbian Learning* et *Growing Neural Gas*.

Dans le but de mieux connaître le mécanisme de cette technique d'apprentissage, nous abordons, dans le reste du chapitre, les deux modèles de *Gaz Neuronal* et l'*apprentissage compétitif Hebbien*.

II.5.2.2.1 Gaz Neuronal (Neural Gas)

Dans le cas du réseau de *Gaz Neuronal (GN)*, proposé par Martinez [MAR-91], il n'existe absolument aucune topologie avant le commencement de l'apprentissage, durant lequel il se construit un graphe composé des sommets présentant les neurones et des arêtes indiquant les connexions entre ces neurones.

Le réseau de GN est considéré comme méthode de quantification vectorielle. Il est nommé par ce nom, 'gaz neuronal', parce que le déplacement des prototypes (dictionnaire (*codebook*)) des classes dans l'espace des données est semblable au mouvement Brownien des molécules de gaz dans un récipient fermé [PEN-08].

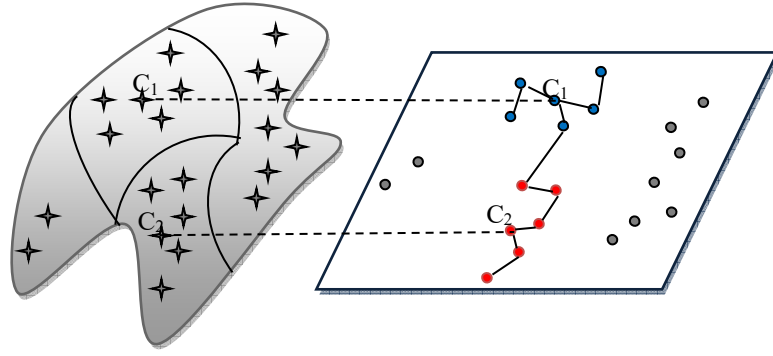


Figure II.18: la préservation de topologie des données par l'apprentissage du réseau de Gaz neuronal

L'algorithme de GN ne fait pas l'exception de la majorité des algorithmes d'apprentissage compétitif qui sont des extensions développées de k-Moyennes. La création du réseau est basée sur le classement des vecteurs poids selon leur voisinage avec le vecteur d'entrée dont la démarche se passe selon l'algorithme suivant:

➤ **Algorithme d'apprentissage**

1. Fixer un nombre maximum d'itération T ,
2. Initialiser aléatoirement N neurones $A = \{\mu_1, \mu_2, \dots, \mu_N\}$ dont les vecteurs de référence sont choisis parmi un ensemble de Q observations,
3. À l'itération t , calculer la distance Euclidienne $\|X_q(t), \mu_k(t)\|$, entre le vecteur de l'observation en entrée $X_q(t)$ et chaque neurone $\mu_k(t)$ $k=1,2,\dots,N$. Chaque prototype $\mu_k(t)$ est affecté à une classe $C_k(t)$ qui prend une valeur entière de $0, \dots, N-1$, de telle sorte que 0 est pour le prototype de minimum distance et $N-1$ est pour le maximum distance :

$$\|X_q - \mu_1\| < \|X_q - \mu_2\| < \dots < \|X_q - \mu_N\| \quad (\text{II.23})$$

4. Modifier les vecteurs poids des N neurones par :

$$\mu_k(t+1) = \mu_k(t) + \eta h(C_k(t)) [X_q(t) - \mu_k(t)] \quad (\text{II.24})$$

Où $\eta(t) = \left(\frac{\eta_f}{\eta_0}\right)^{\frac{t}{T}}$ et $h(C_k) = \exp\left[-\frac{C_k}{\lambda(t)}\right]$ réalise une compétition douce avec $\lambda(t) = \left(\frac{\lambda_f}{\lambda_0}\right)^{\frac{t}{T}}$ est le coefficient de voisinage, si $\lambda(t)$ tend vers 0, (II.23) se réduit à la fonction d'activation de K-moyennes. Les deux fonctions $\eta(t)$ et $\lambda(t)$ sont exponentiellement décroissantes, avec η_0 et λ_0 sont les paramètres de décroissance initiaux et η_f et λ_f sont les paramètres de décroissance finals.

Le Gaz Neuronal converge rapidement à des faibles erreurs de distorsion et atteint une erreur de distorsion inférieure à celle résultant de K-moyennes, de la classification d'entropie maximale et de la SOM [CHEN-06]. Outre son insensibilité de l'initialisation des neurones, le GN utilise une méthode d'apprentissage dynamique pour déterminer les relations de voisinage entre les neurones, contrairement à d'autres réseaux neuronaux tels que SOM qui utilise une méthode statique.

Néanmoins, Le tri des distances est la partie qui a la plus de consommation du temps dans cet algorithme, ce qui explique sa complexité de $O(n \log n)$. Une approche a été développée [ROV-99] pour diminuer le temps d'exécution en adoptant un trie partiel sans réduire sa performance.

Après avoir distribué un certain nombre de prototypes des classes, trouvés par le GN, il est temps de passer la tâche à l'*apprentissage compétitif Hebbien* pour générer un graphe préservant la topologie des données.

II.5.2.2.2 Apprentissage Compétitif Hebbien (CHL)

L'Apprentissage Compétitif Hebbien –*Competitive Hebbian Learning (CHL)*– formulé par *Hebb* [HEB-49] et développé à la suite par *Martinetz* [MAR-93], est basé principalement sur le postulat de *Hebb* qui dit que "si deux neurones sont simultanément actifs, la force de la synapse existant entre ces deux neurones est augmentée". Le but de CHL est de chercher la connexité entre les points d'un ensemble de données, de telle sorte que chaque ensemble des neurones connexes représente une classe.

Contrairement aux autres méthodes, CHL ne change pas les vecteurs poids des neurones, il reçoit un ensemble de neurones non connectés initialement et relie chacun au neurone le plus proche par une arête, en utilisant la distance Euclidienne. Le principe de cette méthode est élaboré dans l'algorithme suivant [FRI-97] :

➤ *Algorithme d'apprentissage*

1. Initialiser aléatoirement N neurones $A = \{\mu_1, \mu_2, \dots, \mu_N\}$ dont les vecteurs de référence sont choisis parmi un ensemble de Q observations,
2. Initialiser l'ensemble C des connexions entre les neurones, $C \subset A \times A$, par l'ensemble vide : $C = \emptyset$
3. À l'entrée d'une observation X_q , $q=1, \dots, Q$, déterminer les deux unités s_1 et s_2 tel que :

$$s_1 = \arg \min_{\mu \in A} \|X_q - \mu\| \quad (\text{II.25})$$

et

$$s_2 = \arg \min_{\mu \in A \setminus \{s_1\}} \|X_q - \mu\| \quad (\text{II.26})$$

4. Si la connexion entre s_1 et s_2 n'existe pas déjà, créer la :

$$C = C \cup \{(s_1, s_2)\} \quad (\text{II.27})$$

5. Revenir à l'étape 2, continuer jusqu'à le traitement de tous les Q observations.

En général, cet apprentissage n'est pas utilisé seul, mais en association avec d'autres méthodes. En fait, il exige souvent une méthode de quantification vectorielle comme le Gaz Neuronal, pour qu'il soit appliqué en tant que post-traitement. Dès la construction du diagramme de Voronoi et son dual la triangulation de Delaunay (fig II.21(a)), le résultat de l'application du CHL est une TD induite, en masquant la TD avec l'ensemble de données (fig.II.21(b)).

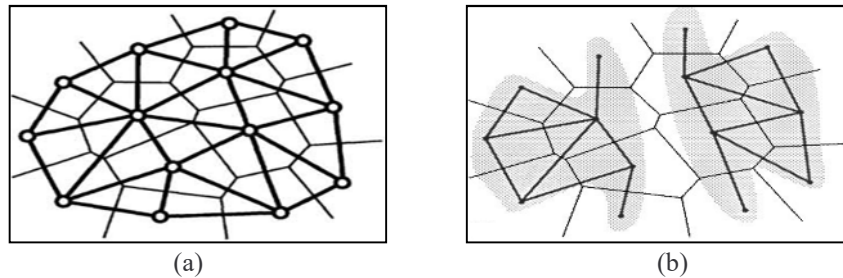


Figure II.19: (a) Le DV et la TD de 14 points de données 2D. (b) TD induite des 14 points de données

Grace à l'utilisation de la TD, l'algorithme de CHL n'a besoin de traiter la relation de voisinage d'un neurone qu'avec ses voisins déterminés par la TD. En plus de son taux d'erreur nul, le CHL a une complexité de $O(\log n)$ ce qui lui rend une méthode de préservation de topologie optimale.

Il est toutefois nécessaire de rappeler que le CHL ne fait qu'ajouter les connexions et ne pas les éliminer, Certains algorithmes d'apprentissage, tels que le Gaz neuronal incrémental '*Growing Neural Gas*' (GNG) [FRI-95] qui fusionne entre le NG et le CHL en utilisant une variable locale pour incrémenter 'l'âge' d'une connexion.

En effet, une connexion jeune implique une forte vraisemblance de la relation topologique, alors qu'à l'opposé, une connexion âgée signifie une vraisemblance faible de cette relation, lorsque l'âge d'une connexion dépassera un certain seuil, il sera donc supprimé. Cela donne au réseau une capacité de procéder à une réadaptation continue, même si de nouvelles observations sont ajoutées, sans avoir besoin de réinitialiser l'apprentissage [VIE-12].

Malheureusement, ce système de seuillage provoque une sensibilité envers le bruit. En l'occurrence, une amélioration du CHL a été développée [LIU-13] en affaiblissant la force de la connexion entre les neurones progressivement avec le temps.

II.6. Conclusion

Nous avons commencé notre chapitre par un exposé des notions de base des réseaux de neurones, un domaine prospère qui représente une alternative efficace aux méthodes classiques.

La nouvelle approche neuronale de classification automatique proposée dans cette thèse est issue de la procédure d'apprentissage compétitif. Ainsi, nous avons présenté dans ce chapitre les deux catégories des réseaux de neurones à apprentissage compétitif dur et doux conçus pour des tâches de classification et de reconnaissance de forme dans un contexte non supervisé, au sein des observations multidimensionnelles.

Dans le chapitre suivant, nous allons présenter une nouvelle procédure basée principalement sur des concepts compétitifs. Cette procédure d'apprentissage permet d'extraire les modes de la fonction de densité de probabilité sous-jacente à la distribution des données à analyser sous forme d'une composante connexe. Chacun de ces modes caractérise le noyau d'une classe au sein de la distribution des observations.

CHAPITRE III : NOUVELLE APPROCHE NEURONALE STATISTIQUE ET NON PARAMETRIQUE POUR LA CLASSIFICATION DES DONNEES MULTI DIMENSIONNELLES

III.1. Introduction

Nous avons introduit dans la première partie de ce mémoire un aperçu théorique de la classification automatique, son application sur l'image numérique et les méthodes neuronales compétitives dédiées à cet objectif. L'importance de la classification automatique s'augmente en parallèle avec la croissance et la variété de la nature et la construction des données multidimensionnelles de divers domaines, qui ont fait le sujet de plusieurs recherches.

Les approches statistiques de groupement postulent que chaque groupement correspond à un mode de la fonction de densité sous-jacente à la distribution des données à classer. Ainsi, le problème de la classification automatique se trouve posé en terme de détection des modes. La performance des méthodes classiques de classification dépend de plusieurs facteurs et nécessite le réglage de certains paramètres difficiles à fixer et certaines d'entre ces méthodes sont basées sur le seuillage.

Dans ce troisième chapitre, nous présentons l'approche neuromimétique [EDD-07a] qui est basée sur les concepts de détection des modes, pour la classification automatique non supervisé. Cette procédure qui est adaptée surtout aux distributions gaussiennes souffre toutefois de certains inconvénients à savoir qu'elle est non adaptée aux distributions non gaussiennes à degré d'enchevêtrement élevé, et elle nécessite un seuillage ce qui affaiblit son automaticité.

Notre objectif général est d'améliorer l'approche neuromimétique afin d'atteindre une automaticité et une performance qui lui permette de traiter une grande variété des données de dimensionnalité très élevée avec le maximum de précision, en surmontant ses inconvénients. Dans ce contexte, nous proposons, dans ce mémoire, une nouvelle approche neuronale, statistique et non paramétrique basée sur la détection des modes de la fdp et inspirée de l'approche neuromimétique.

Cette nouvelle approche est une procédure neuronale de classification automatique non supervisée, basée sur le concept compétitif, divisée en trois étapes de traitement. Elle

commence par une méthode non paramétrique de l'estimation de la fonction de densité de probabilité, suivie d'un réseau de neurones doté d'un apprentissage compétitif avec la distance de Mahalanobis comme une métrique de la fonction d'activation. Cette étape permet de détecter les maxima locaux du fdp sous-jacente à la distribution. Ensuite, nous utilisons l'apprentissage par CHL (*Competitive Hebbian Learning*) développé et adapté à notre contexte pour analyser la connectivité entre les maxima de la fdp détectés en utilisant la distance de Mahalanobis. Les classes des maxima ainsi détectées sont ensuite utilisées pour le classement du reste des observations de l'échantillon.

III.2. Estimation de la fonction de densité de probabilité

Les données à analyser par notre approche sont des données quantitatives multidimensionnelles où leurs attributs sont représentés par une matrice d'observation. Ainsi, un ensemble de Q observations est représenté par Q points dans un espace multidimensionnel R^N , en associant à chacune de ces observations un vecteur X_q , tel que $X_q = \{x_{q,1}, x_{q,2}, \dots, x_{q,N}\}$, avec $q=1, \dots, Q$.

Sous l'hypothèse que chaque classe d'observations corresponde à un mode de la fonction de densité de probabilité fdp, le problème de la classification se trouve donc posé en termes de détection des modes.

La méthode d'estimation de la fonction de densité de probabilité présentée au premier chapitre sera détailler dans la suite de ce chapitre.

III.2.1. Principe de l'estimation de densité

Dans un problème de classification non supervisée, pour pouvoir prendre une idée sur la façon dont les données sont distribuées, on est mené à étudier la densité P de chaque variable aléatoire X_q . Or, on ne dispose d'aucune information sur la forme de cette densité, ce qui incite à faire appel à une estimation non paramétrique.

Soit un échantillon $E = \{X_1, X_2, \dots, X_Q\}$ de Q observations indépendantes et identiquement distribuées suivant une loi de densité de probabilité $p(X)$. La probabilité P pour qu'une observation X soit située à l'intérieur d'un domaine $R(X_0)$ centré au point X_0 , est donnée par:

$$P = \int_{R(X_0)} p(X) dx \quad (\text{III.1})$$

La probabilité P_q pour que q des Q observations de l'échantillon soient situées dans le domaine $R(X_0)$ est donnée par la loi binomiale:

$$P_q = C_Q^q P^q (1 - P)^{Q-q} \quad (\text{III.2})$$

L'espérance mathématique de q est alors :

$$E(q) = Q.P \quad (\text{III.3})$$

Ce résultat indique que $\frac{q}{Q}$ est un estimateur non biaisé de P .

Si l'on suppose que $p(X)$ ne présente pas de variations importantes à l'intérieur du domaine $R(X_0)$, la probabilité P peut se mettre sous la forme:

$$P = p(X).V(X_0) \quad (\text{III.4})$$

Où $V(X_0)$ est le volume de $R(X_0)$.

L'estimateur de (X_0) , noté $p_e(X_0)$, s'obtient donc en combinant les équations (III.3) et (III.4). Ce qui donne :

$$p_e(X_0) = \frac{q}{Q} \quad (\text{III.5})$$

Le problème essentiel à résoudre est celui du comportement asymptotique de cet estimateur.

Pour assurer la convergence en moyenne quadratique de l'estimateur $p_e(X_0)$ vers $p(X_0)$, le volume $V(X_0)$ du domaine $R(X_0)$ ainsi que le nombre d'observations q_Q présentées dans ce domaine, doivent être liés au nombre total Q d'observations [DUD-73] [WEG-72] [BAN-77][FUKU-72].

Plus précisément, $p_e(X_0)$ converge vers $p(X_0)$ si :

$$1) \quad \lim_{Q \rightarrow \infty} V_Q(X_0) = 0 \quad (\text{III.6})$$

$$2) \quad \lim_{Q \rightarrow \infty} qQ = \infty \quad (\text{III.7})$$

$$3) \quad \lim_{Q \rightarrow \infty} \frac{q}{Q} = 0 \quad (\text{III.8})$$

- La condition 1 assure la convergence vers $p_e(X_0)$ comme indiqué par l'équation (III.5),
- La condition 2 assure la convergence en probabilité de $\frac{q^Q}{Q}$ vers P ,
- La condition 3 assure la convergence de $\frac{q^Q}{Q \cdot V_Q}$.

Il existe deux manières de construire des domaines $R(X_0)$ satisfaisant ces conditions auxquelles sont associées deux méthodes non paramétriques d'estimation de la fonction de densité de probabilité : la méthode du noyau de Parzen-Rosenbalt [ROS-56][PAR-62] et celle des K plus proches voisins de Cover et Hart [COV-67].

Dans notre approche, nous utilisons la méthode du noyau de Parzen [PAR-62] pour estimer la fonction de densité de probabilité sous-jacente.

III.2.2. Estimation par la méthode du noyau

Cette méthode consiste à réduire le volume du domaine $R(X_0)$ lorsque le nombre d'observations Q augmente en spécifiant V_Q comme fonction de Q satisfaisant la condition (III.6). Il faut vérifier que les variables q_Q et $\frac{q}{Q}$ répondent aux conditions (III.7) et (III.8). On peut utiliser des relations de type:

$$V_Q(X_0) = V_0 \cdot Q^{-\alpha} \quad (\text{III.9})$$

Avec $\alpha \in]0,1[$ où :

$$V_Q = \frac{V_0}{\log Q} \quad (\text{III.10})$$

Cette technique a été proposée par Rosenblatt [ROS-56] et approfondie par Parzen [PAR-62] dans le cas mono-variable, puis étendue au cas multivariables par Murty [MUR-66] et Cacoullos [CAC-62].

Dans cette approche d'estimation la région R_Q est un hypercube à N dimension. Si h_Q est le côté de cet hypercube alors son volume est donné par:

$$V_Q = h_Q^N \quad (\text{III.11})$$

L'estimateur de la fdp $p_e(X_0)$ en X_0 s'écrit donc :

$$p_e(X_0) = \frac{q_Q}{Q h_Q^N} \quad (\text{III.12})$$

q_Q représente le nombre d'observations appartenant à l'hypercube.

La fonction $\varphi(\cdot)$ dite noyau de Parzen-Rosenblat est définie par :

$$\varphi(v_i) = \begin{cases} 1 & \text{si } |v_i| < \frac{1}{2} \\ 0 & \text{Ailleurs} \end{cases} \quad i = 1, 2, \dots, N \quad (\text{III.13})$$

$\varphi(v_i)$ est différente de 0 si un hypercube de côté unité centré à l'origine donc :

$$\varphi\left(\frac{X_0 - X_i}{h_Q}\right) = \begin{cases} 1 & \text{si } X_i \in R_Q \\ 0 & \text{Ailleurs} \end{cases} \quad (\text{III.14})$$

En introduisant cette fonction, le nombre d'observations sera donné par :

$$q_Q = \sum_{i=1}^Q \frac{1}{V_Q} \varphi\left(\frac{X_0 - X_i}{h_Q}\right) \quad (\text{III.15})$$

L'estimateur dit estimateur de Parzen peut s'écrire sous la forme :

$$p_e^Q(X_0) = \frac{1}{Q} \sum_{i=1}^Q \frac{1}{h_Q^N} \varphi\left(\frac{X_0 - X_i}{h_Q}\right) \quad (\text{III.16})$$

La convergence de cet estimateur $p_e^Q(X_0)$ vers $p_Q(X_0)$ est assurée sous les conditions suivantes :

$$1) \quad \forall X, \exists s, \quad 0 \leq \varphi(X) \leq s \quad (\text{III.17})$$

$$2) \quad \lim_{|X| \rightarrow \infty} \varphi(X) \cdot \|X\|^N = 0 \quad (\text{III.18})$$

$$3) \quad \lim_{Q \rightarrow \infty} h_Q^{-1} = 0 \quad (\text{III.19})$$

$$4) \quad \lim_{Q \rightarrow \infty} Q \cdot h_Q^{-N} = \infty \quad (\text{III.20})$$

Cette procédure d'estimation reste limitée à des données de faible dimension. Lorsque la dimension de l'espace augmente, le dénombrement des observations situées dans chaque hypercube entraîne des calculs importants. Alors, pour contourner cette difficulté, un algorithme rapide d'estimation a été développé par Postaire et Vasseur [POS-82b].

III.2.3. Algorithme rapide d'estimation de la fdp

L'algorithme rapide de l'estimation adapte une procédure qui lui permette de déterminer les hypercubes non vides et donc supprimer les hypercubes vides. Ainsi, le temps de calcul, pour une taille d'un échantillon donné, est limité au temps nécessaire pour déterminer un nombre d'estimations qui sera toujours inférieur ou égal au nombre Q des observations.

Dans le cadre de ce travail, un échantillon à analyser est constitué d'un ensemble d'observations représentées par des points dans l'espace Euclidien R^N :

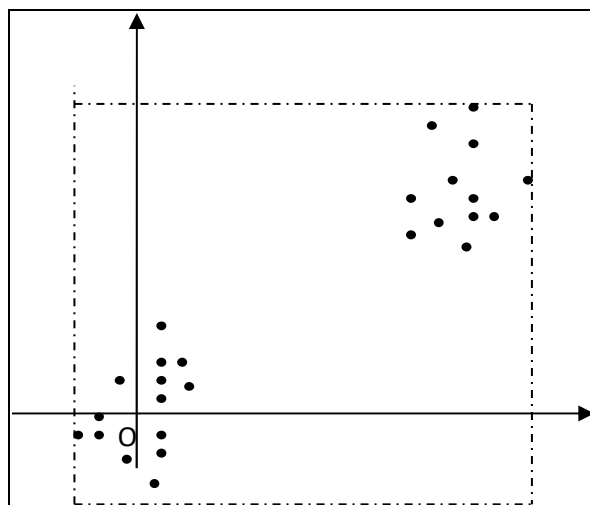


Figure III.1: Espace de représentation des observations bidimensionnelles

Avant tout calcul de densité, la procédure commence par une normalisation de l'espace puis une discrétisation qui permet de cerner les observations dans des hypercubes, pour calculer finalement la densité des données dans chaque hypercube.

III.2.3.1 Normalisation

On normalise, dans un premier temps, les plages de variations de chacune des composantes des observations [MEI-72] de telle sorte qu'elles soient situées à l'intérieur d'un hypercube de côté L. Ce nombre est ajusté par l'analyste. Ensuite, on divise chacune de ces plages en L intervalles adjacents et égaux à l'unité. On obtient alors un réseau d'échantillonnage formé d'hypercubes qui couvrent tout l'espace normalisé. La fdp est alors estimé au centre de ces hypercubes dont certains contiennent des observations alors que d'autres sont vides.

En effet, soit $E = \{X_1, X_2, \dots, X_Q\}$ un ensemble de Q observations à N dimensions d'une variable aléatoire X, telle que : $X_q = [x_{q,1}, x_{q,2}, \dots, x_{q,N}]^T$ avec $q=1, \dots, Q$, (cf. figure III.1)

L'origine O de l'espace R^N est d'abord translatée en O' tel que :

$$O' = \left[\min_q x_{q,1}, \min_q x_{q,2}, \dots, \min_q x_{q,n}, \dots, \min_q x_{q,N} \right]^T$$

La normalisation des plages de variation de chacune des coordonnées des observations est assurée par la transformation linéaire définie par une matrice diagonale telle que :

$$y_{q,n} = \frac{L \left(x_{q,n} - \min_{q'} x_{q',n} \right)}{\max_{q'} x_{q',n} - \min_{q'} x_{q',n}} \quad (\text{III.21})$$

Après cette normalisation, toutes les observations se trouvent situées à l'intérieur d'un hypercube de côté égal à L :

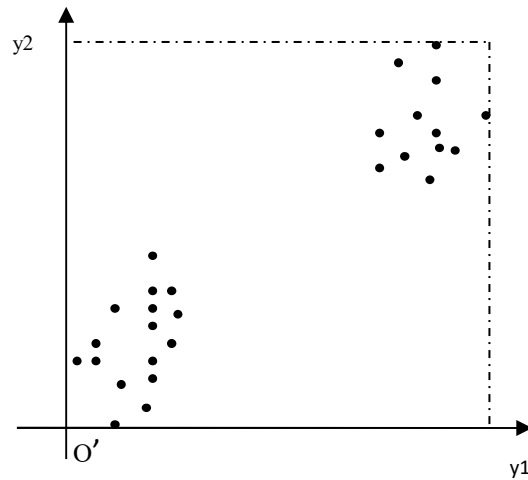


Figure III.2: Plage de variation des observations normalisée

III.2.3.2 Discrétisation

Chaque axe du nouvel espace de représentation des données est ensuite découpé en L intervalles adjacents et égaux à l'unité. Cette discrétisation définit un réseau de L^N hypercubes qui couvrent tout l'espace normalisé (Cf. figure III.3). Les centres de ces hypercubes constituent un réseau de points d'échantillonnage noté X . Chaque hypercube, noté $H(X)$ est repéré par les N coordonnées $x_1, x_2, \dots, x_n, \dots, x_N$ qui sont les parties entières des coordonnées de son centre X .

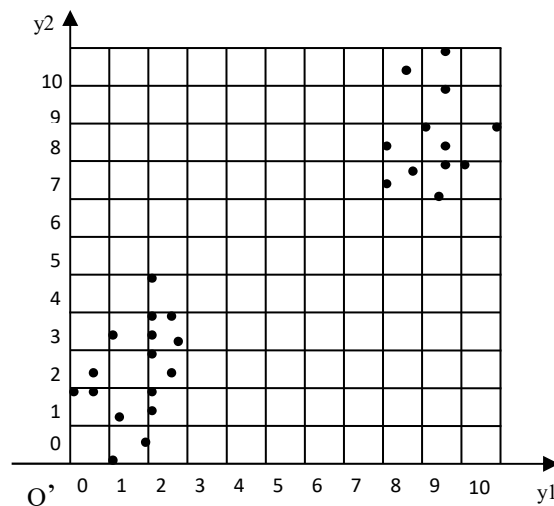


Figure III. 3: Réseau de points d'échantillonnage ($L=10$)

Soit $y_q = [y_{q,1}, y_{q,2}, \dots, y_{q,N}]^T$, $q=1, \dots, Q$ les observations dans l'espace normalisé. Chaque observation y_q se trouve à l'intérieur d'un hypercube non vide de coordonnées $X_n = \text{int}(y_{q,n})$

$n=1,2,\dots,N$ où $int(y_{q,n})$ désigne la partie entière du réel $y_{q,n}$. Si plusieurs observations se trouvent à l'intérieur du même hypercube celui-ci apparaît plusieurs fois dans la liste des hypercubes non vides.

III.2.3.3 Estimation de la densité locale

Ainsi, il est facile de connaître le nombre d'observations $q[H(X)]$ qui se trouvent dans chaque hypercube $H(X)$ non vide en comptant le nombre d'apparitions de cet hypercube dans cette liste.

Par conséquent, la valeur de l'estimateur de la fdp au point X est donnée par :

$$p_e(X) = \frac{q[H(X)]}{Q} \quad (\text{III.22})$$

Puisque le volume de $H(X)$ est égal à l'unité, $p_e(X)$ est donc un estimateur local (cf. figure III.4). Par soucis de simplicité, notons cette fonction $p_e(X)$ par $f(X)$ dans toute la suite de ce mémoire.

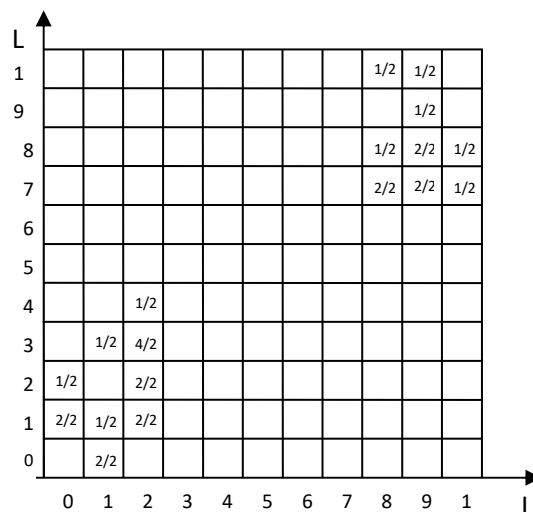


Figure III. 4: Estimation de la fdp dans l'espace discret

A l'issue de cette première phase de traitement des données, toute l'information dont on dispose pour assurer la classification des observations disponibles est contenue dans cette version discrète et la fdp sous-jacente $f(X)$ définie dans un sous-ensemble fini de l'espace $(Z^+)^N$ représenté par des lettres majuscules soulignées « X, Y,... ».

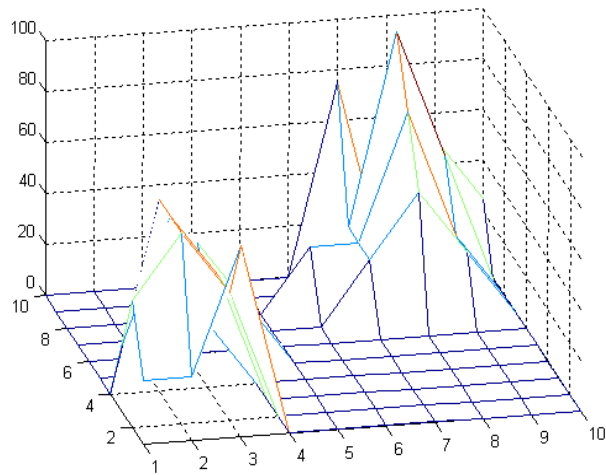


Figure III. 5: Présentation tridimensionnelles des valeurs de la fdp sous-jacente à la distribution

Le comportement de cette procédure d'estimation dépend de l'ajustement du paramètre L . si L est trop petit, tous les hypercubes du réseau d'échantillonnage seront connexes et un seul mode sera détecté. A l'opposé, s'il est trop grand, la procédure tendra à générer un grand nombre de petits modes non significatifs.

III.2.3.4 Exemple modèle

Pour une meilleure présentation de cette approche, nous allons nous servir d'un exemple modèle pour expliquer et visualiser les différentes étapes de notre procédure pour la détection des modes en classification non supervisée des données multidimensionnelles.

Notre modèle est un échantillon de 2000 observations composé de quatre classes gaussiennes bidimensionnelles de 500 observations chacune, générées artificiellement et dont les paramètres statistiques sont consignés dans le tableau suivant :

Données générées			
Distribution	Moyennes	Matrice de Covariance	Probabilité à Piori
Classe 1	0.0345 4.9841	2.9211 -0.0260 -0.0260 3.3739	0.25
Classe 2	-0.1029 -4.8435	2.8362 -0.0536 -0.0536 2.8707	0.25
Classe 3	5.0651 -0.1769	3.0612 -0.3191 -0.3191 2.9288	0.25
Classe 4	-4.9243 0.0526	3.0721 -0.0727 -0.0727 2.7638	0.25

Tableau III.1: Paramètres statistiques de l'exemple modèle

La distribution brute des observations est représentée sur la figure ci-dessous :

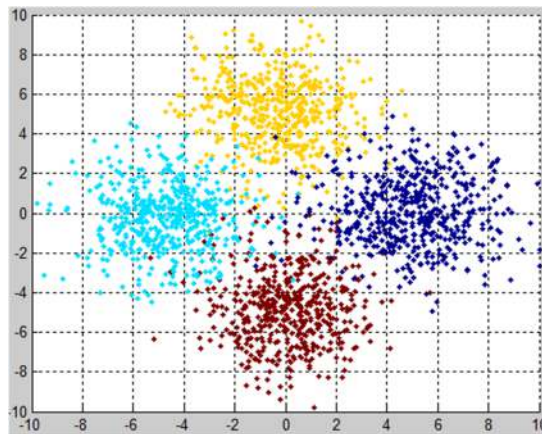


Figure III. 6: Ensemble des observations brutes

La procédure rapide d'estimation de la fdp est appliquée à la distribution (tableau III.1 et figure III.6) avec un pas de discrétisation $L=25$:

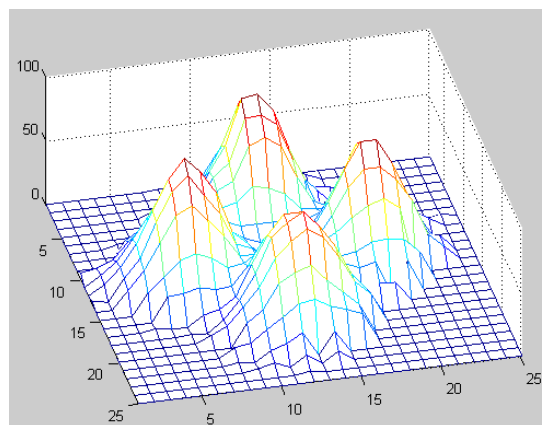


Figure III.7: Présentation de la fdp sous-jacente au quatre classes ($L=25$)

La figure III.7 représente la visualisation tridimensionnelle de la fdp sous-jacente. Dans cette représentation, à chaque point d'échantillonnage $X \in (Z^+)^N$ est associé une valeur qui correspond à la partie entière du pourcentage de $f(X)$ par rapport à la valeur maximale de la fdp dans un domaine de valeurs.

Cet algorithme d'estimation est une méthode rapide qui permet non seulement de construire un modèle qui nous fournit plus d'informations sur la structure statistique des données multidimensionnelles mais aussi de réduire le volume des données à traiter.

Le résultat de cette étape de l'approche dépend seulement du pas de discrétisation L . En fait, l'ajustement de ce paramètre dépend de la taille de l'échantillon analysé, de la dimension des données et de la structure de leur distribution. Dans un contexte non supervisé, ce paramètre est ajusté selon le concept de stabilité du nombre de modes [POS-81]. Il s'agit d'analyser les résultats obtenus pour une grande marge de variation de ce paramètre. Le choix de ce paramètre au milieu du plus grand intervalle à l'intérieur duquel le nombre des modes détectés reste constant et différent de 1 est une bonne procédure pour optimiser le pas L , ce qui a été prouvé par un grand nombre d'algorithmes de classification [TOU-88][EDD-07b].

III.3. Extraction des maxima locaux de la fdp par apprentissage compétitif

Dans cette deuxième phase de notre procédure, en assimilant les modes à des maxima locaux du pdf, l'approche de classification proposée utilise un réseau neuronal multicouche à apprentissage compétitif [EDD-07a], afin d'estimer les paramètres du réseau, qui sont les maxima locaux de la fdp et leurs sites dans l'ensemble \underline{X} .

Dans l'algorithme d'apprentissage, nous travaillons uniquement sur la fdp en présentant successivement, les centres des hypercubes non vides de l'ensemble \underline{X} au réseau, au lieu des Q observations brute comme dans d'autres réseaux de neurones à apprentissage compétitif [SOU-94]. Cela envisage une réduction importante du temps de convergence du réseau.

La procédure d'apprentissage va nous permettre d'estimer les paramètres du réseau adopté qui sont les maxima locaux de la fdp et leurs sites dans l'espace discrétisé.

III.3.1. Architecture du Réseau de Neurons

La couche d'entrée est constituée d'un nombre de neurones égal à la dimension de l'espace de représentation des données. Ces unités transmettent simplement les informations présentées à l'entrée du réseau à l'ensemble des neurones de la couche de sortie. Chaque neurone de sortie matérialise un hypercube représentant le site d'un maximum local de la fdp dans l'ensemble \underline{X} . Vu que la méthode de détection des maxima proposée est non supervisée, le nombre de neurones de la couche de sortie est alors initialisé arbitrairement lors de la phase d'apprentissage du réseau, il suffit de choisir un nombre de neurones de sortie du réseau un peu élevé pour éviter qu'il soit inférieur au nombre réel des classes et l'algorithme d'apprentissage converge ainsi rapidement vers les maxima locaux.

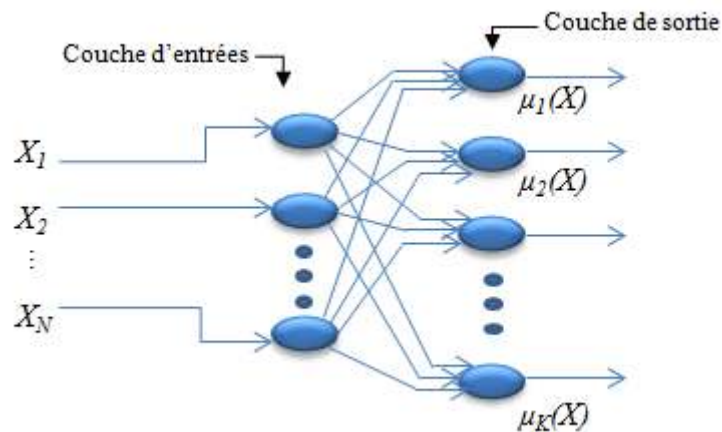


Figure III.8: Réseau compétitif à deux couches

A l'itération t , $H^t(X)$ un hypercube non vide dont les coordonnées de son centre $X = [x_1, x_2, \dots, x_n, \dots, x_N]$, $n=1,2,\dots,N$ sont présentées à l'entrée du réseau de neurones. A chaque itération, chaque neurone d'entrée I_n , $n=1,2,\dots,N$ est sollicité par l'attribut x_n du centre X de l'hypercube $H^t(X)$. Alors que chaque neurone de sortie ξ_k , $k=1,2,\dots, K$, définit un hypercube non vide matérialisé par le vecteur poids $\mu_k^t(X)$ de ce neurone, la fonction $f[\mu_k^t(X)]$ estimée dans ce hypercube et une fonction d'activation sous la forme d'une métrique.

Plus explicitement, chaque neurone de sortie ξ_k est défini à l'itération t par son vecteur poids $\mu_k^t(X)$, par la fonction $f[\mu_k^t(X)]$ estimée de ce vecteur poids et par la distance

$D[\mu_k^t(X), H^t(X)]$ séparant le vecteur poids $\mu_k^t(X)$ et l'hypercube $H^t(X)$ présenté à l'entrée du réseau (cf. §II.3.2.2).

Nous appelons un cycle ou époque, chaque présentation des tous les hypercubes non vides de l'ensemble \underline{X} au réseau.

III.3.2. Apprentissage du Réseau

Au cours de la phase d'apprentissage, les neurones de sortie entrent en compétition les uns avec les autres, en comparant la distance $D[\mu_k^t(X), H^t(X)]$, pour tout $k = 1, 2, \dots, K$, entre l'hypercube $H^t(X)$ en entrée et chaque neurone de sortie $\mu_k^t(X)$, le neurone gagnant est celui le plus proche de l'hypercube. Nous comparons ensuite la valeur de la fdp associée au neurone gagnant $\mu_g^t(X)$ avec celle de l'hypercube $H^t(X)$. Si la valeur de la fdp de ce dernier est supérieur de celle du neurone gagnant, alors le poids de ce neurone sera activé.

Pour notre exemple modèle (cf. figure III.6), la figure III.7 montre les dix-sept maxima locaux extraits par l'algorithme d'apprentissage pour un nombre des neurones de la couche de sortie initialisé arbitrairement à 17 dont chacun représente le site d'un maximum local de la fdp (cf. figure III.7) sous-jacente aux données de l'échantillon à analyser :

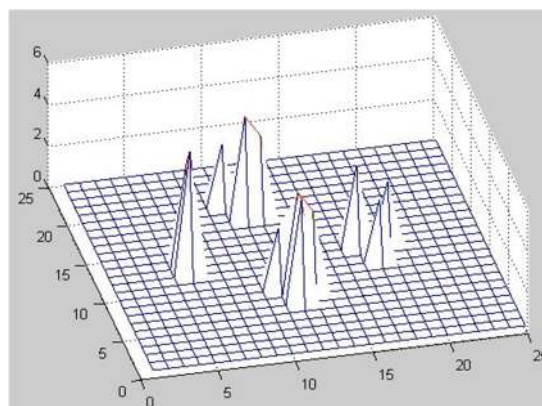


Figure III.9: Maxima locaux extraits par réseau de neurones pour $k=17$

La métrique utilisée dans ce réseau est la distance Euclidienne (DE) (cf. §I.3.3) qui s'adapte très bien avec les distributions gaussiennes. Pour vérifier la performance de l'approche sur les distributions non gaussiennes, nous allons lui adapter une autre métrique et nous allons appliquer les procédures avec les deux métriques sur un exemple non gaussien pour faire la comparaison des résultats.

III.3.3. Discussion

III.3.3.1 Choix de la distance

Parmi les objectifs de ce travail est l'adaptation de l'approche neuronale à la classification automatique des observations non gaussiennes.

Si on regarde la définition d'une classe selon d'Everitt (1974): « Une classe est une agrégation de points dans l'espace de test telle que la distance entre deux points dans la classe est inférieure à la distance entre un de ces deux point et tout point en dehors de la classe ». En utilisant la distance Euclidienne (DE) cette définition n'est valable que pour des distributions gaussiennes.

$$DE(\mu_k^t(\mathbf{X}), H^t(\mathbf{X})) = \sqrt{\sum_{n=1}^N (\mu_{k,n}^t(\mathbf{X}) - H^t(\mathbf{X}))^2} \quad (\text{III.23})$$

En effet, dans une distribution, le calcul de la distance DE ne prend pas en considération la dispersion des observations autour de leur centre. Cela apparaîtra plus clair dans l'exemple suivant :

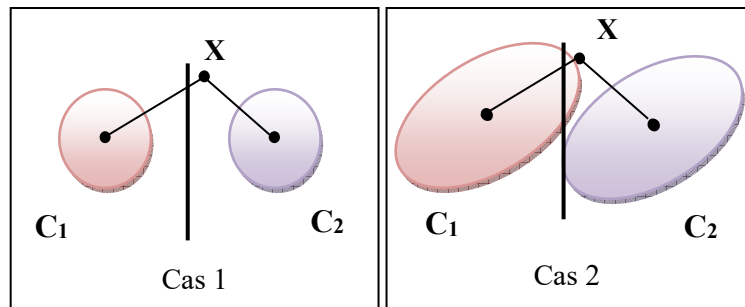


Figure III.10: analyse discriminante entre la distance DE et DM

Dans le premier cas, en calculant la distance DE qui éloigne X des centres des deux classes, il paraît trivial que X appartient à la classe C_2 , cependant, dans le deuxième cas, malgré qu'on a modifié la distribution des données autour de leurs centres et l'emplacement de X n'a pas changé, les distances Euclidiennes qui éloigne X des deux centres de classes C_1 et C_2 restent les mêmes. Ce qui nous mène à inférer que $X \in C_2$, la chose qui n'est plus juste dans ce cas.

A cette effet, nous avons fait appel à la distance de Mahalanobis (DM) (cf. §I.3.3) pour sa fiabilité par rapport à la distance euclidienne. En fait, cette métrique prend en considération

non seulement le noyau de la fdp, mais aussi la variation de cette dernière. Etant une métrique basée sur la corrélation entre les variables, la DM sera intégrée dans notre apprentissage sous la forme suivante :

$$DM = \sqrt{(\mu_k^t(\mathbf{X}) - H^t(\mathbf{X}))^t \Sigma^{-1} (\mu_k^t(\mathbf{X}) - H^t(\mathbf{X}))} \quad (\text{III.24})$$

Avec Σ^{-1} est l'inverse de la matrice de covariance.

En utilisant cette distance, un neurone distance ξ est redéfini aussi par sa matrice de covariance Σ que par ses paramètres. Il convient de mentionner que dans notre algorithme d'apprentissage, nous calculons la distance DM entre deux vecteurs et n'en pas entre un vecteur et un ensemble de données, nous sommes alors dans l'obligation de collecter les observations voisines pour obtenir la matrice de covariance sous-jacente.

III.3.3.2 Collecte des échantillons

La matrice de covariance contient toute l'information sur la façon dont se répartissent les formes de la classe dans l'espace des paramètres. Cependant, elle risque de ne pas être inversible dans les deux cas :

- La forte corrélation entre les variables d'une distribution,
- Le nombre des observations est insuffisant par rapport au nombre de variables.

Nous rappelons que une matrice de $N \times N$ n'est inversible que si elle est une matrice de rang plein, c'est à dire qu'il est carré et le rang de la matrice est égal à N .

Afin de s'assurer que la distance ne donnera pas des résultats erronés, nous nous inspirons du *Système de Mahalanobis de Tagushi (MTS)* [TAG-00] qui définit un groupe de référence pour construire une échelle de mesure qui calcule la distance des échantillons à ce groupe particulier. Ce groupe de référence est appelé « groupe normal », et les échantillons qui sont en dehors de ce groupe sont considérés comme « anormaux ».

Dans la littérature pertinente, plusieurs auteurs ont fourni différentes limites inférieures sur le nombre d'échantillons de groupe normal, citant quelques-unes :

- *Taguchi et Jugulum* (2000) [TAG-00]: affirment que le nombre des échantillons doit être supérieur au nombre des variables, qui est une exigence pour obtenir une matrice de

corrélation inversible des variables. Pourtant, cela peut ne pas être suffisant. Dans les cas où la matrice de corrélation de la population n'est pas connue et la matrice de corrélation de l'échantillon doit être utilisée, il y a une limite inférieure sur le nombre d'échantillons afin d'estimer une matrice de corrélation, dans un cas normal multivarié.

- *Hirohisa* et al. (2006) [HIR-06]: indiquent que la méthode MTS "exige de façon empirique que le nombre d'échantillons est supérieur à trois fois le nombre de variables.
- *Foley* (1972) [FOL-72]: avait travaillé sur cette question et avait montré que « le rapport entre le nombre d'échantillons et le nombre de variables doit être supérieur à 3 afin de minimiser les erreurs de classification pour déterminer si les échantillons sont de la même distribution multivariée ou non. La diminution de la taille de l'échantillon est montrée pour augmenter les faux négatifs ».

Par conséquent, nous proposons une méthode de collecte de données, basée sur l'hypothèse de *Hirohisa* [HIR-06] pour garantir l'inversibilité de la matrice de covariance, afin de calculer la distance de Mahalanobis, présenté par le diagramme suivant:

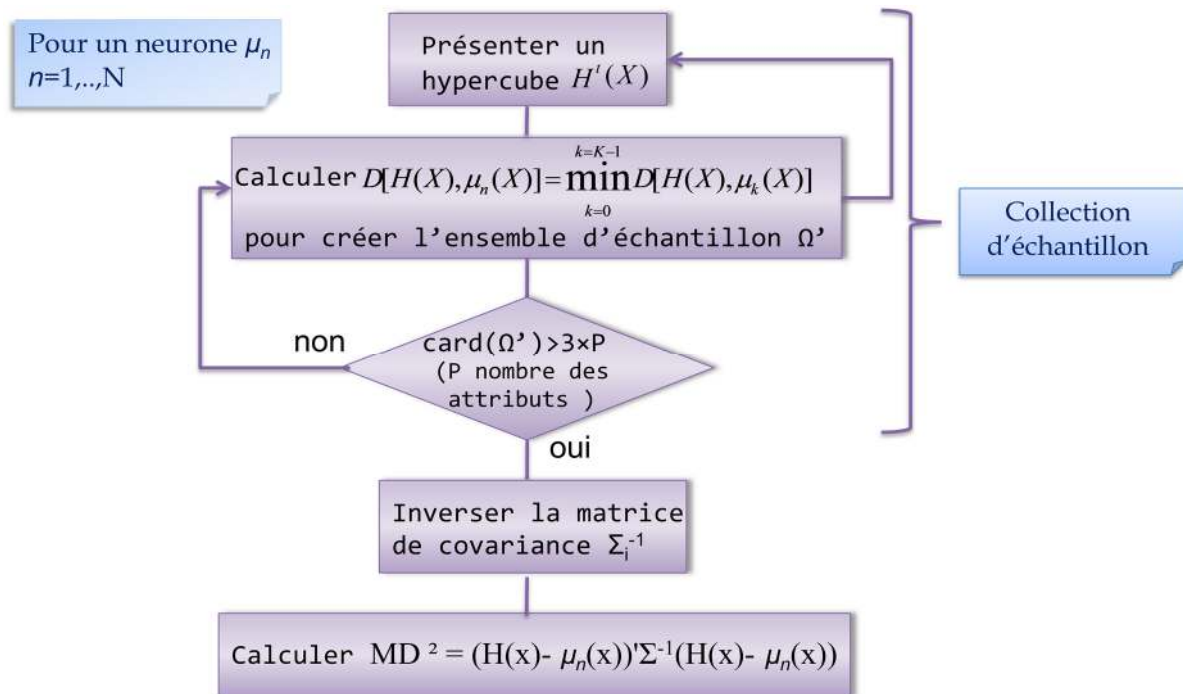


Figure III.11: diagramme de collecte de données pour le calcul de la DM

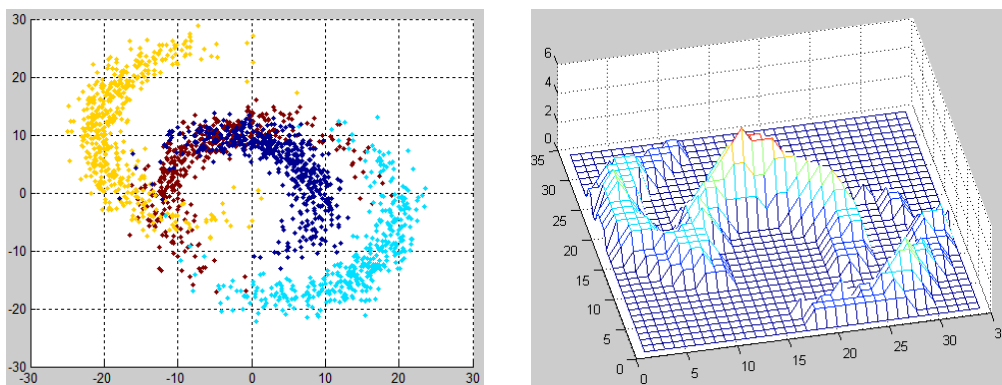
III.3.3.3 Comparaison entre les résultats de la distance euclidienne et la distance de Mahalanobis

Pour concrétiser la fiabilité de la méthode de détection des maxima locaux intégrant la distance de Mahalanobis, nous avons comparé ses performances à celles de la première version de l’algorithme (utilisant la distance euclidienne). Donc, nous avons testé les deux algorithmes sur le même échantillon.

Nombre de Population	θ	B1	B2	A1	Nombre d’observations
Classe 1	m=45 s=45	$\mu_1=-2$ $\sigma_1=3$	$\mu_2=-3$ $\sigma_2=3$	12	400
Classe 2	m=135 s=45	$\mu_1=-2$ $\sigma_1=3$	$\mu_2=4$ $\sigma_2=3$	16	500
Classe 3	m=270 s=45	$\mu_1=10$ $\sigma_1=3$	$\mu_2=-6$ $\sigma_2=3$	15	500
Classe 4	m=315 s=45	$\mu_1=-2$ $\sigma_1=3$	$\mu_2=2$ $\sigma_2=3$	14	400

Tableau III.2: Paramètres statistiques de l’exemple de comparaison

Notre exemple de test est un échantillon de 1800 observations composé de quatre classes non Gaussiennes bidimensionnelles générées artificiellement, le tableau ci-dessus montre ses paramètres statistiques. La distribution des 4 classes est présentée dans la figure III.12(a) suivant :



(a) Données brutes (b) estimation de la fdp
Figure III.12: Présentation des 4 classes non gaussiennes et leur estimation de densité

La figure III.12(b) représente l'estimation de la fdp sous-jacente à la distribution avec le paramètre de résolution $L=19$.

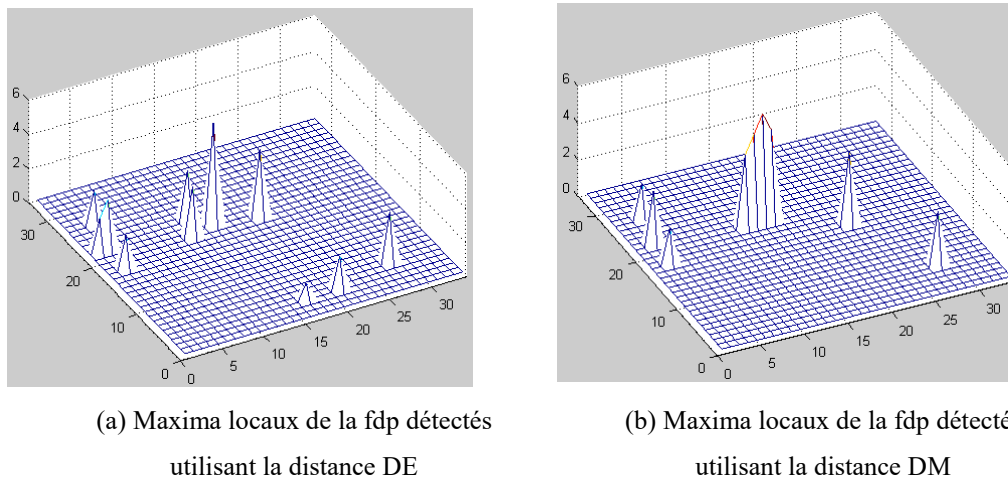


Figure III.13: Comparaison entre les résultats de l'apprentissage utilisant la DE et celui utilisant la DM

La figure III.13(a) montre que la méthode avec la DE n'a marqué que 3 classes, ce qui veut dire que ce procédé, donne des résultats ambigus lorsque les classe se chevauchent les uns aux autres pour les distributions non gaussiennes. Alors que l'algorithme avec la DM proposé a prouvé sa fiabilité en marquant clairement les 4 classes non Gaussiennes malgré l'enchevêtrement des observations de cet échantillon (cf. figure III.13), sans aucune information a priori sur le nombre réel des classes.

Néanmoins, le calcul de la distance de Mahalanobis provoque un temps d'exécution plus élevé. Par conséquent, dans le cas des données multidimensionnelles avec un nombre des attributs plus élevés que le nombre des observations ou des données déjà connues Gaussiennes, l'utilisation de la distance Euclidienne sera plus commode.

III.3.4. Algorithme d'apprentissage du réseau de neurones à apprentissage compétitif (RNAC)

L'algorithme suivant décrit les différentes étapes de la phase d'apprentissage proposée:

- *Phase d'initialisation:*
 1. Initialiser les vecteurs moyens; $k = 1, 2, K$, de la sortie de neurones K , avec un choix arbitraire de K hypercubes non vides de l'ensemble \underline{X} ,

2. Initialiser les coefficients de la fonction d'apprentissage α_0 et τ , Il suffit de donner une valeur très importante à τ pour que l'algorithme cherche les sites des maxima locaux avant sa convergence [EDD-07a].

- **Phase de traitement:**

1. Présenter au réseau, avec tirage arbitraire, un hypercube $H(X)$ non-vide,
2. Chercher le neurone gagnant g définie par le calcul de la distance qui sépare $\mu_k^t(X)$ et $H(X)$ de sorte que:

$$D[\mu_g(X), H(X)] = \min_{k=1} D[\mu_k(X) - H(X)] \quad (\text{III.25})$$

Avec:

$$D[\mu_k(X), H(X)] = \sqrt{(\mu_k(X) - H(X)) \sum^{-1} (\mu_k(X) - H(X))} \quad (\text{III.26})$$

\sum^{-1} est l'inverse de la matrice de covariance, les valeurs individuelle de covariance de \sum sont calculées à partir du produits de la somme donnée par:

$$\sum = \frac{1}{N} \sum_{n=1}^N (\mu_{kn}(X) - H(X)) (\mu_{kn}(X) - H(X))^t \quad (\text{III.27})$$

3. Comparer les deux valeurs $f dp(\mu_g(X))$ et $f dp(H(X))$:

➤ Si $f dp(\mu_g(X)) < f dp(H(X))$ alors modifier les paramètres du neurone gagnant, tel que :

$$\begin{cases} \mu_g(X) = \mu_g^{t-1}(X) - \alpha(t) (H(X) - \mu_g^{t-1}(X)) \\ f dp(\mu_g(X)) = f dp(H(X)) \end{cases} \quad (\text{III.28})$$

Où t est le nombre d'itérations et $\alpha(t)$ est la fonction d'apprentissage (cf. §II.5.1.3) définit comme suivant :

$$\alpha(t) = \frac{\alpha_0}{1 + \frac{t}{\tau}} \quad (\text{III.29})$$

- Sinon revenir à l'étape 1,
- 4. Aller à l'étape 1,
- 5. Critère d'arrêt : après le traitement de tous les hypercubes, comparer $\mu_k^c(X)$ de ce cycle avec $\mu_k^{c-1}(X)$ du cycle précédent pour tout $k=1,2,\dots,K$:
 - Si $\exists k, \mu_k^c(X) \neq \mu_k^{c-1}(X)$ alors passer au cycle suivant et aller à l'étape 1
 - Sinon, fin de l'apprentissage.

Le comportement de cet algorithme dépend des paramètres α_0 et τ de la fonction d'apprentissage $\alpha(t)$. Le choix des valeurs de ces paramètres ne pose pas de problèmes, il suffit de fixer α_0 et choisir τ assez important pour assurer la convergence de l'algorithme d'apprentissage vers les maxima locaux de la fdp (cf. §II.5.1.3).

III.4. Détection des modes significatifs de la fdp

Dans une troisième étape de cette procédure, nous cherchons à extraire automatiquement les modes significatifs de la fdp à partir des maxima locaux détectés et localisés durant la phase d'apprentissage.

Dans l'approche Neuromimétique, la procédure de détection des modes utilisée est inspirée de la technique des centres mobiles [DUD-73]. Elle permet de détecter les modes significatifs de la fdp parmi les maxima détectés dans la phase d'apprentissage mais son inconvénient majeur reste la nécessité d'un seuillage.

Étant donné que les maxima locaux qui sont détectés au cours de la phase d'apprentissage marquent parfaitement les régions modales et sont divisés en un nombre égal au nombre de classes existantes dans l'échantillon, nous avons pensé à relier chacun des groupes des maxima les plus proches de telle manière que nous obtenons une carte qui préserve la forme et la structure des classes.

III.4.1. La détection des modes par l'apprentissage compétitif Hebbien (CHL)

Une des méthodes parfaites qui forment des cartes préservant la topologie est l'apprentissage compétitif Hebbien "*Conpatitive Hebbien Learning*" (CHL) prouvé par Martinetz (1993) (cf. §II.5.2.2.2). Cet apprentissage nous permet d'analyser la connectivité

entre les maxima détectés de la fdp, de telle sorte que chaque composante connexe représente une classe.

Soit $F = \{H(X_1), H(X_2), \dots, H(X_K)\}$, l'ensemble des K hypercubes qui représentent les sites dans \underline{X} du K maxima locaux détectés ; qui sont la sortie du réseau multicouche à apprentissage compétitif. Dans cette phase, nous allons chercher la connexion inter-neuronal de ces maxima en utilisant une méthode de CHL adaptant la distance de Mahalanobis comme mesure de ressemblance.

Après avoir détecté les prototypes des classes, la classification par l'algorithme de CHL est basée sur deux outils géométriques ; le *diagramme de Voronoi* (DV) et son dual la *triangulation de Delaunay* (TD):

III.4.1.1 Diagramme de Voronoi

Le diagramme de Voronoï (DV) décompose l'espace de représentation des données à partir de sa structure métrique ; pour un ensemble K de points S appelés sites de l'espace Euclidien \mathbb{E}^N , chacun de ces sites s_i est associé à une région $V(s_i), i = 1, 2, \dots, K$ sous forme d'un polygone construit de l'intersection d'un nombre fini de demi-espace limités par les hyperplans médiateurs de segments $[s_i, s_j], i \neq j$.

Dans ce contexte, le diagramme de Voronoï V_F de l'ensemble F des hypercubes $H(X_i) \in R^N, i = 1, 2, \dots, K$, représentant les sites, est donné par K polyèdres N-dimensionnels. Le polyèdre de Voronoi V_i , est défini comme suit : Le polyèdre de Voronoi V_i d'un hypercube $H(X_i) \in F$ est donné par l'ensemble des hypercubes $v \in R^N$ qui sont près de $H(X_i)$ que tout autre $H(X_j) \in R^N, i \neq j$ [MAR-94] :

$$V_i = \{v \in R^N \mid \|v - H(X_i)\| \leq \|v - H(X_j)\| \forall j\} \quad (\text{III.30})$$

Le diagramme de Voronoi créé pour notre modèle est illustré dans la figure ci-dessous :

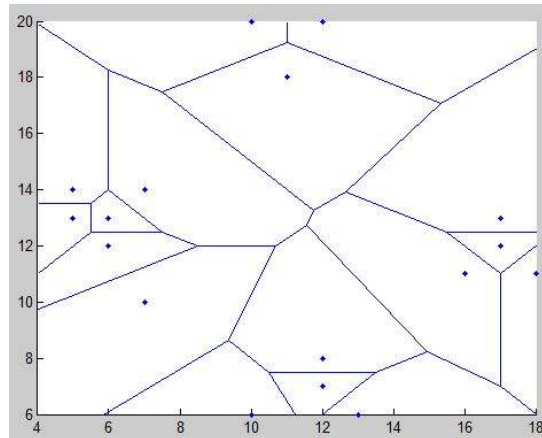


Figure III.14: Diagramme de Voronoï des 17 maxima locaux de l'exemple modèle

La figure montre la subdivision de l'espace de représentation des données dont les sites sont les maxima locaux détecté dans phase précédente.

III.4.1.2 Triangulation de Delaunay

Le graphe dual du diagramme de Voronoï est la triangulation de Delaunay [DEL-34] (cf. §II.5.2.2), cette dernière est créée en connectant tous les paires des hypercubes $H(X_i), H(X_j) \in F$ qui partagent une arête commune de deux polygones du DV sous-jacent respectivement aux V_i et V_j . Ainsi formés les circonscrips des triangles constitués de chaque 3 hypercubes de l'ensemble F , de telle sorte qu'aucun hypercube en F est à l'intérieur. En outre, chaque arête du triangle est une médiatrice de l'arête commune des deux polygones.

$$TD(F) = \{i, j = 1, \dots, K \mid V_i \cap V_j \neq \emptyset\} \quad (\text{III.31})$$

la figure ci-dessous illustre les deux graphes DV et TD créés pour notre modèle :

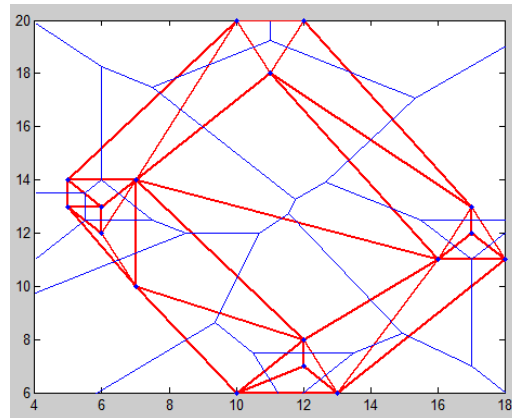


Figure III.15: la triangulation de Delaunay des 17 maxima locaux de l'exemple modèle

III.4.1.3 Triangulation de Delaunay induite

A chaque introduction d'un hypercube $H(X_i)$, $i=1,2,\dots,K$, le CHL le traite tant que nouveau neurone de son réseau. Sans modifier son poids, l'apprentissage cherche à connecter ce neurone, par le biais de la distance de Mahalanobis, avec les deux neurones les plus proches par une arête, en se basant sur l'ensemble d'arêtes produit $TD(F)$ par la DT. D'une façon indirecte, le CHL élimine les connexions non significatives. Ainsi, l'algorithme d'apprentissage génère des composantes connexes représentant les classes qui existent dans l'échantillon à classer.

Le nombre de composantes connexes engendrées par la méthode de CHL est limité aux 4 régions de l'exemple modèle. La figure suivante montre que les régions modales de l'exemple modèle ont été identifiées et qu'elles coïncident parfaitement avec les présentes quatre classes de la distribution:

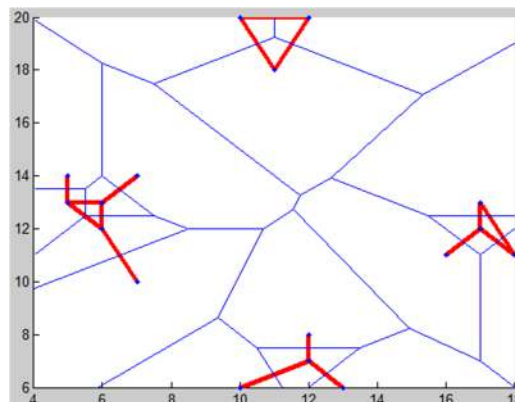


Figure III.16: la triangulation de Delaunay induite de l'exemple modèle

Cette technique ne nécessite pas de seuillage pour l'extraction des régions modales, à l'opposé à la technique probabiliste proposée par l'approche neuromimétique, ce qui rend l'approche proposée beaucoup plus performante.

III.4.1.4 Pseudo algorithme d'apprentissage

1. Initialiser l'ensemble C des connexions entre les neurones, $C \subset F \times F$, par l'ensemble vide : $C = \emptyset$
2. Calculer la triangulation de Delaunay, créer un ensemble des connexions tel que :

$$Td = \{(i, j) \mid \forall i \neq j \text{ indices de } H(X_i) \text{ et } H(X_j) \in F, \exists \text{ une connexion entre } i \text{ et } j\}$$

3. Choisir un neurone aléatoirement de l'ensemble $F = \{H(X_1), H(X_2), \dots, H(X_K)\}$ des maxima locaux,
4. À l'entrée d'un hypercube $H(X_i)$, $i = 1, 2, \dots, K$, déterminer les deux unités s_1 et s_2 tel que :

$$s_1 = \arg \min_{j \in A} \|H(X_i) - H(X_j)\| \quad (\text{III.32})$$

et

$$s_2 = \arg \min_{\mu \in A \setminus \{s_1\}} \|H(X_i) - H(X_j)\| \quad (\text{III.33})$$

Où A est l'ensemble des indices des voisins connectés à $H(X_i)$ selon Td .

5. Si la connexion entre s_1 et s_2 n'existe pas encore, créer la en l'ajoutant à C :

$$C = C \cup \{(s_1, s_2)\} \quad (\text{III.34})$$

6. Revenir à l'étape 3, continuer jusqu'au traitement de tous les hypercubes de F .

III.4.2. Discussion

Vue que l'apprentissage par CHL ne modifie pas les poids des neurones, il a un taux d'erreur nul, en outre, il n'utilise ni de seuillage ni des paramètres à déterminer avant l'apprentissage. En plus de tous les avantages des réseaux de neurones artificiels que possède notre méthode, elle est aussi adaptée aux différentes formes de distributions ; gaussiennes et non-gaussiennes, grâce à l'utilisation de la distance de Mahalanobis qui tient en compte la

dispersion des données. Cette méthode a prouvé ainsi son automaticité et son efficacité à la préservation de topologie des données optimale à travers la détection des régions modales.

Toutefois, au cours de traitement de la relation de voisinage d'un neurone avec ses voisins déterminés par la TD, Il se peut trouver une classe présentée par un seul neurone, alors qu'il est impossible qu'il soit isolé par la CHL (cf. fig.I.17) et donc il sera lié à d'autres prototypes qui présentent des classes différentes, comme dans les cas suivants :

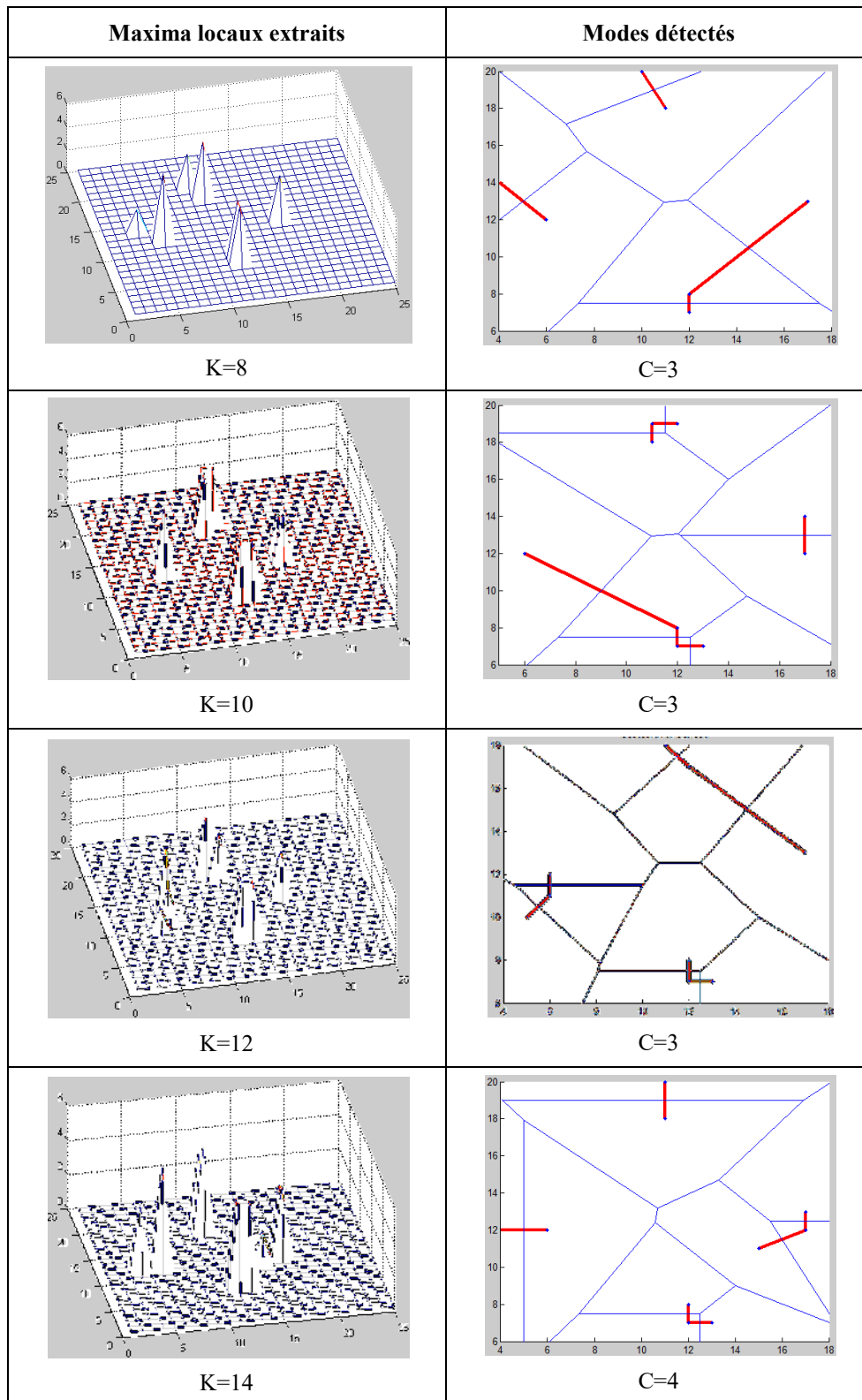


Figure III.17: présentation des triangulations de Delaunay induites de l'exemple modèle pour différents nombres K de neurones de sortie

Nous pouvons bien voir d'après cette figure que la méthode d'apprentissage CHL n'est pas capable de distinguer les classes présentées par un seul neurone qui serait souvent le cas des ensembles de données de petite taille.

Ainsi pour pallier à cet inconvénient du CHL, nous avons pensé à profiter de certaines propriétés de l'écart-type. En effet, l'écart-type, étant un paramètre statistique de dispersion absolue, permet de définir des pourcentages d'existence des observations autour de leurs centres selon les propriétés suivantes [ZAR-12a] :

- ✓ 68,27 % des cas sont compris entre $X - sd$ et $X + sd$ (un écart-type (sd) de part et d'autre de la moyenne)
- ✓ 95,45 % des cas sont compris entre $X - 2sd$ et $X + 2sd$ (deux fois l'écart-type de part et d'autre de la moyenne)
- ✓ 99,73 % des cas sont compris entre $X - 3sd$ et $X + 3sd$ (trois fois l'écart-type de part et d'autre de la moyenne)

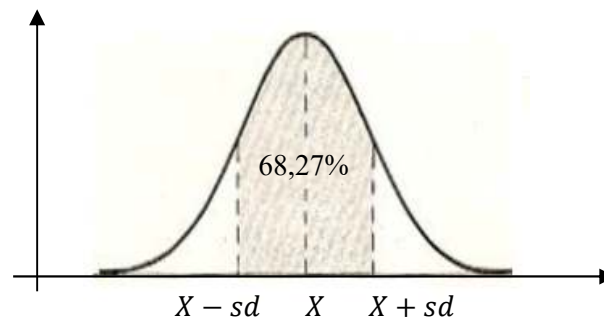


Figure III.18: présentation de la relation entre une distribution gaussienne et son écart-type

Ces pourcentages sont approximativement valables pour des distributions légèrement dissymétriques.

En appliquant la troisième propriété sur l'apprentissage de CHL (le pourcentage le plus élevé), nous pouvons déduire deux cas qui définissent la relation entre deux fonctions de densité de probabilité de n dimensions et de centre X_1 et X_2 , où $X_i = \{x_{1,i}, x_{2,i}, x_{3,i}, \dots, x_{n,i}\}$:

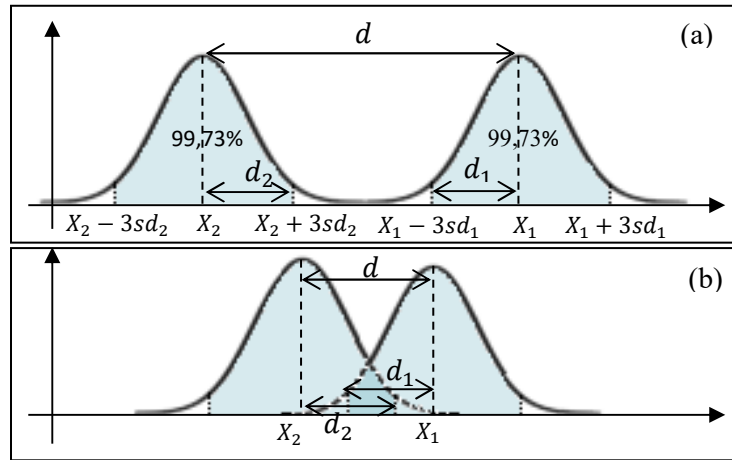


Figure III.19 : présentation de la relation entre deux distributions gaussiennes

Avec :
$$d_i = \|X_i - (X_i - 3sd_i)\| = 3sd_i \quad (III.35)$$

- Cas (a) :
$$\frac{d}{2} \geq d_2 \text{ ou } \frac{d}{2} \geq d_1 \quad (III.36)$$

N.B : le ‘OU’ n’est pas exclusive.

Cela signifie que la probabilité que X_2 appartienne à la distribution 1 ou X_1 appartienne à la distribution 2 est inférieure à 0,17%.

- Cas (b) :
$$\frac{d}{2} < d_2 \text{ et } \frac{d}{2} < d_1 \quad (III.37)$$

Dans ce cas, les deux distributions partagent un grand nombre d'observations, ce qui donne la possibilité de 99,73% que X_1 et X_2 appartiennent à la même classe.

Nous proposons alors d’ajouter une condition de vérification au cours de l’apprentissage :

- 1) À l’entrée d’un hypercube $H(X_i)$, $i = 1, 2, \dots, K$, calculer l’écart-type de chaque neurone et déterminer les deux unités s_1 et s_2 tel que :

$$sd(i) = \underset{j \in F}{sd}(H(X_i)) \quad (III.38)$$

et
$$D_1 = \min_{j \in F} \|H(X_i) - H(X_j)\|$$

$$D_2 = \min_{\mu \in F \setminus \{s_1\}} \|H(X_i) - H(X_j)\|$$

avec

$$s_1 = \arg(D_1)$$

et

$$s_2 = \arg(D_2)$$

Où A est l'ensemble des indices des voisins connectés à $H(X_i)$ selon Td .

- 2) Si la connexion entre s_1 et s_2 n'existe pas encore,
- 3) Si $\frac{D_2}{2} \leq 3 \times sd(1)$ et $\frac{D_2}{2} \leq 3 \times sd(2)$, créer là :

$$C = C \cup \{(s_1, s_2)\}$$

Si $\frac{D_2}{2} > 3 \times sd(1)$ ou $\frac{D_2}{2} > 3 \times sd(2)$, pas de connexion entre s_1 et s_2 .

A travers l'application de cette modification de l'algorithme sur l'exemple modèle, en utilisant les mêmes nombres de neurones qu'on a traités et illustrés dans la figure III.17, nous obtenons les résultats suivants :

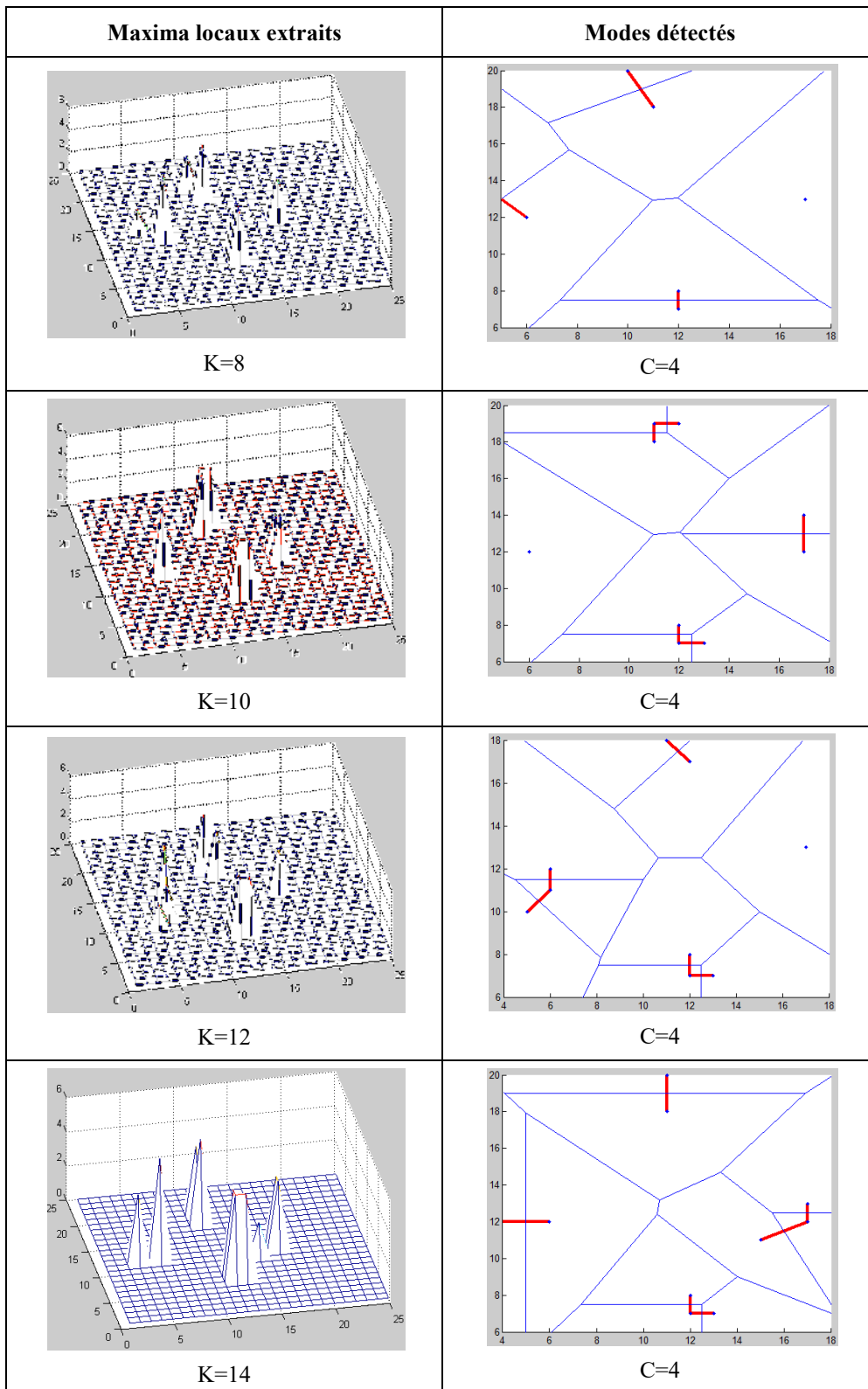


Figure III.20: présentation des triangulations de Delaunay induites de l'exemple modèle pour différents nombres K de neurones de sortie par la nouvelle méthode

A travers ces résultats nous pouvons déduire que nous avons pu réussir à surmonter cet obstacle de la classe représentée par un seul neurone.

En dépit de cela, la distance de Mahalanobis a un temps d'exécution plus élevé que la distance euclidienne à cause de sa complexité de traitement, mais la recherche du plus proche voisin peut être effectuée seulement dans $O(\log K)$ au lieu de $O(K)$ en exploitant la triangulation de Delaunay [KNU-73] qui permet de traiter la relation de voisinage d'un neurone qu'avec ses voisins déterminés par la TD. En outre, avec ce nombre ainsi réduit K de neurones représenté par les maxima locaux, la procédure de détection de modes proposée prouve un temps d'exécution réduit.

Les modes ainsi détectés sont ensuite utilisés comme des prototypes pour le processus de classification qui s'occupe de l'affectation du reste des données à leurs classes.

III.5. Classification des données

Les techniques de classification considérées dans la littérature sont simples. Elles consistent à utiliser les observations appartenant aux modes comme prototypes des classes pour réaliser ensuite la classification des observations restantes.

Une fois que les différentes régions modales sont identifiées, de nombreuses méthodes de classification peuvent être utilisées pour affecter le reste des observations aux classes qui leur sont rattachées.

La méthode de classification [EDD-02] que nous utilisons dans le présent travail consiste d'abord, à définir un ensemble dite interne des observations dans la composante connexe comme prototype d'une classe. Ensuite, les données restantes, qui ne tombent pas dans l'un des modes détectés, sont assignées aux ensembles attachés à leurs voisins les plus proches au moyen de la distance de Mahalanobis. La technique de classement est une méthode appelée Assignation Simple des Observations au Noyau le plus Proche (ASNP), explicité ci-après.

III.5.1. Assignation simple des observations au noyau le plus proche (ASNP)

L'objectif de cette technique de classification, comme mentionné avant, est d'affecter toute observation non encore classée à la classe dont son prototype est le plus proche.

L'observation concernée par chaque nouvelle affectation est considérée comme un nouveau prototype, et la procédure est itérée jusqu'à ce que toutes les observations aient été affectées.

Soit MC_i l'ensemble des modes interconnectés présentant la classe C_i , $i = 1, 2, \dots, NC$, avec NC est le nombre des composantes connexes trouvées par l'algorithme de détection des modes.

Cette technique peut être décomposée en deux étapes successives [BOTT-91]. La première étape consiste à affecter toutes les observations situées à l'intérieur de chaque région modale MC_i à la classe C_i associée à cette région. Les prototypes des classes ainsi définis constituent les noyaux des classes en présence dans l'échantillon. Le noyau N_i associé à la classe C_i est donc constitué de l'ensemble des prototypes situés à l'intérieur du MC_i .

La deuxième étape de la procédure de classification consiste à assigner chacune des observations situées à l'extérieur des noyaux N_i , $i = 1, 2, \dots, S$ à la classe la plus proche au sens de la distance de Mahalanobis. L'algorithme suivant numérise les différentes fonctions de décision à appliquer aux observations X_q d'un échantillon E dans cette méthode de classification.

III.5.2. Pseudo algorithme

- *Première étape :*

1. Entrer aléatoirement une observation X_q ,
2. Calculer la distance entre X_q et les modes dans chaque MC_i :

$$D_i(X_q) = \min_{j=1,2,\dots, NP_i} D(X_q, X_{i,j}), \quad i = 1, 2, \dots, NC \quad (\text{III.39})$$

Où NP_i le nombre d'observations situées à l'intérieur du mode MC_i formant les prototypes de la classe C_i , et $X_{i,j}$ est le jème prototype du noyau N_i .

3. Si $D_s(X_q) < D_i(X_q) \forall i = 1, 2, \dots, NC$ alors $X_q \in MC_s$ de la classe C_s et :
 $NP_s = NP_s + 1$
4. Si $q < Q$ Revenir à 1, si $q = Q$ terminer.

• *Deuxième partie :*

Après la détermination des NP_i prototypes de chaque classe C_i , $i = 1, 2, \dots, NC$:

1. Tirer sans remise d'une observation X_q non classée à partir de l'échantillon E .
2. Pour chaque région modale MC_i calculer la distance :

$$D_i(X_q) = \min_{j=1,2,\dots, NP_i} D(X_q, X_{i,j}), \quad i = 1, 2, \dots, NC \quad (\text{III.40})$$

4. Assigner X_q à la classe C_S vérifiant :

$$S = \underset{i=1,2,\dots,NC}{\text{Arg min}} D_i(X_q), \quad (\text{III.41})$$

5. Mise à jour des prototypes :

$$\begin{aligned} NP_S &= NP_S + 1 \\ NP_T &= NP_1 + NP_2 + \dots + NP_{NC} \end{aligned} \quad (\text{III.42})$$

6. Test du critère d'arrêt :

Si $NP_T = Q$ alors fin de l'algorithme

Sinon reprise du procédé à partir de l'étape 2

III.5.3. Discussion

La méthode de classification ASNP, a montré à travers des examens [EDD-07a] sa performance dans le cas de l'analyse de classes de forme sphérique, par rapport à plusieurs méthodes telles que, K-moyennes, nuée dynamique et l'ACDP (Assignation des observations au centre du mode le plus proche) qui consiste à déterminer le centre de gravité de chaque mode afin d'assigner chaque observation à la classe associée au centre de gravité le plus proche. En utilisant la distance de Mahalanobis comme critère de ressemblance pour affecter le reste des observations à leurs classes, la méthode sera capable de classifier différentes formes des données ; gaussienne et non gaussiennes. Néanmoins, l'ASNP effectuée dans ce cas un temps de calcul plus élevé que celle utilisant la distance Euclidienne, nous recommandons alors que l'usage de DM soit facultatif.

Les données classées de l'exemple de modèle sont présentées dans la figure suivante:

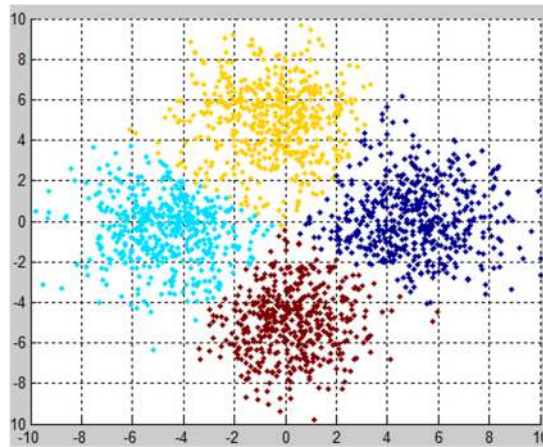


Figure III.21: Représentation de la distribution classée de l'exemple modèle

Le taux d'erreur obtenu par cette procédure est égal à 0,022, ce qui est légèrement inférieur à celui obtenu par le neuromimétique (0,028), et la K-Moyennes (0,0285). Donc, pour les distributions gaussiennes l'approche proposée reste comparable aux méthodes classiques mais son avantage majeur réside sur son automaticité.

III.6. Conclusion

Dans ce chapitre, nous avons présenté une nouvelle approche de détection des modes des fonctions de densité de probabilité sous-jacentes aux distributions des données multidimensionnelles. Cette approche est basée principalement sur les concepts compétitifs.

Nous avons commencé notre chapitre par une description de la nouvelle approche, puis nous avons expliqué les différentes étapes de la méthode d'estimation de la fdp sous-jacente à la distribution des données à analyser, par la suite, nous avons abordé les deux réseaux de neurones à apprentissage compétitif utilisés afin de détecter les modes de cette fonction, qui marquent parfaitement les classes de la distribution. Enfin, nous avons présenté la technique de regroupement qui permet d'assigner chacune des observations restantes au mode le plus proche.

L'avantage de cette approche est qu'elle ne nécessite aucune information a priori sur les données à analyser. En fait, elle ne nécessite ni de réglage compliqué des paramètres, ni de seuillage.

Dans l'objectif de présenter la performance de notre approche de classification automatique, nous allons appliquer cette procédure, dans le chapitre suivant, sur des exemples générés artificiellement ainsi que sur des images synthétiques et réels pour la segmentation d'images et plus particulièrement les images de la tomographie par émission de positons (TEP). Dans le reste du chapitre

CHAPITRE IV: PERFORMANCE ET RESULTATS EXPERIMENTAUX DE L'APPROCHE

IV.1. Introduction

Dans ce dernier chapitre, nous mettons en lumière la nouvelle approche de classification, que nous avons décortiquée dans le chapitre précédent. A travers la présentation et l'interprétation des résultats expérimentaux des différentes étapes de la procédure proposée, nous avons testé sa validité sur différents types de données.

Il convient de mentionner que notre approche est une méthode de classification automatique non supervisée, basée principalement sur les réseaux de neurones à apprentissage compétitif. Sans passer par un seuillage ni prétraitement, cette nouvelle procédure ne nécessite aucune information a priori sur le nombre des classes ni sur la structure des données classifiées.

Pour illustrer le comportement de la procédure ainsi que sa performance, nous avons divisé ce chapitre en trois parties ; la première vise à évaluer la performance de notre approche sur les données, en lui appliquant sur des échantillons multidimensionnels générés artificiellement, tout en comparant ses résultats avec les résultats d'autres méthodes de classification similaires, appliquées sur les mêmes échantillons.

La deuxième partie est consacrée à l'adaptation de la nouvelle procédure à la segmentation d'image couleur, testée sur plusieurs images synthétiques et réelles. Cette adaptation est éprouvée par des comparaisons avec d'autres méthodes et des évaluations qualitatives et quantitatives.

Dans la troisième partie, nous avons appliqué notre approche sur l'image TEP scan de la maladie Alzheimer, afin de prouver l'efficacité et l'utilité de l'approche. Les résultats de la nouvelle procédure sont ensuite comparés aux résultats des méthodes K-moyennes et FCM.

A la fin de ce chapitre nous avons développé une discussion sur la validité de notre approche de classification.

IV.2. Application pour la classification des données

IV.2.1. Exemple 1

L'exemple 1 est une distribution bidimensionnelle de 1800 observations arrangées en quatre classes constituées d'une classe gaussienne et trois classes non gaussiennes (cf. fig.IV.1), avec des paramètres statistiques donnés respectivement dans les tableaux IV.1 et IV.2.

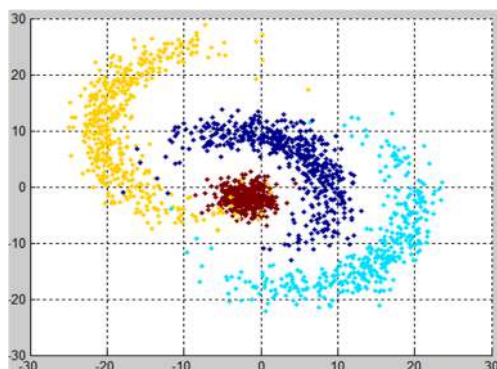


Figure IV.1: représentation des données brutes de l'exemple 1

Les attributs qui décrivent les observations des trois classes non gaussiennes sont définis par:

$$y_1 = A_1 \cos\theta + B_1 ; y_2 = A_2 \sin\theta + B_2 \quad (\text{IV.1})$$

Où θ est une variable aléatoire normale avec le vecteur moyenne m et l'écart type s . B_1 et B_2 sont des variables aléatoires normales avec la moyenne μ et la variance σ . Les valeurs de ces paramètres sont données dans le tableau IV.2 avec $A_1 = A_2$ dans cet exemple:

Classes	Nombre des observations	Moyenne	Variances
1	400	2.00234 -1.96618	2.80525 3.69549

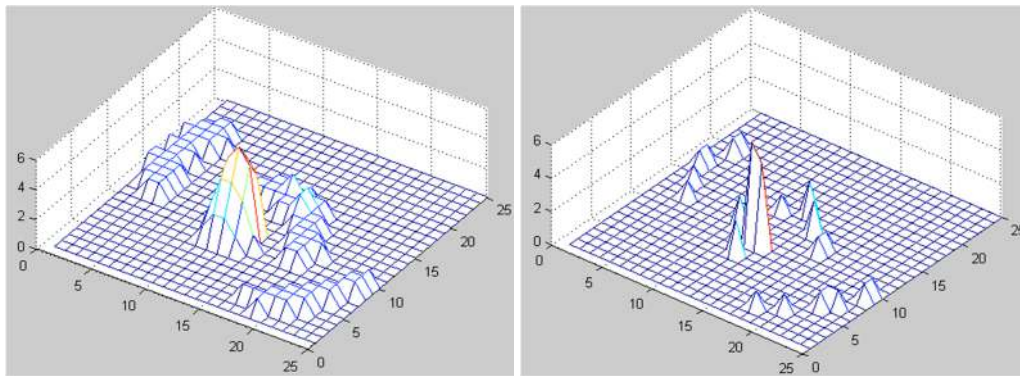
Tableau IV.1: paramètres statistiques de la classe gaussienne

Classe	$\Theta(\text{deg})$	B1	B2	A	Nombre des observations
1	$m=45$ $s=45$	$\mu_1=-2$ $\sigma_1=3$	$\mu_2=-3$ $\sigma_2=3$	12	500
2	$m=135$ $s=45$	$\mu_1=-2$ $\sigma_1=3$	$\mu_2=4$ $\sigma_2=3$	16	400
3	$m=270$ $s=45$	$\mu_1=10$ $\sigma_1=3$	$\mu_2=-6$ $\sigma_2=3$	15	500

Tableau IV.2: paramètres statistiques des 3 classes non gaussiennes

IV.2.1.1. Estimation des maxima locaux

Les observations subissent d'abord une estimation (cf. fig. IV.2(a)) de la fonction de densité de probabilité sous-jacente à la distribution, avec un pas de discrétisation $L = 25$. Cette valeur correspond au milieu de la plus longue plage stable où le nombre des modes détectés reste constant. La figure IV.2(b) montre le résultat de l'algorithme d'apprentissage, après 6 cycles (cf. §III.3.1), pour un nombre de neurones de sortie initialisé à 27 et avec des paramètres $\alpha_0 = 1$ et $\tau = 40$:

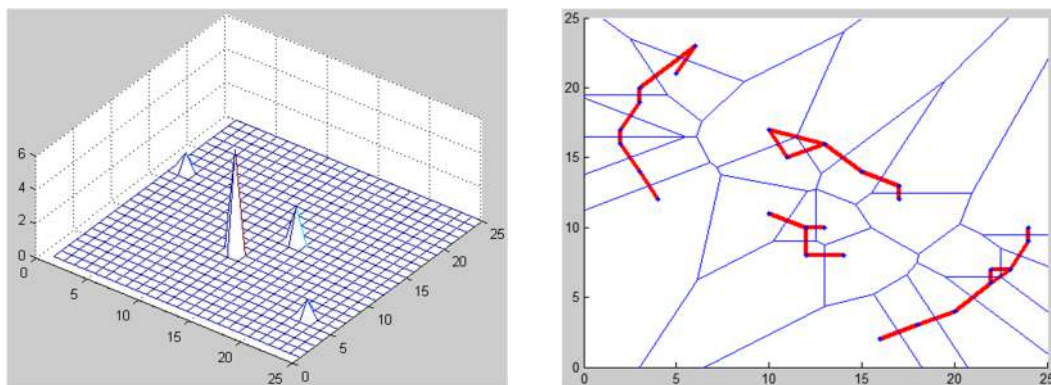


(a) Estimation de la fdp sous-jacent à la distribution (b) Les 27 Maxima locaux détectés après 6 cycles

Figure IV.2 : La détection des maxima après l'estimation de la fdp sous-jacente au modèle 1

IV.2.1.2. Détection des modes : Comparaison

Après avoir distingué les différents maxima locaux des classes, nous avons comparé la performance de la nouvelle technique de détection des modes par l'apprentissage compétitif Hebbien CHL, à celles de la procédure de détection des modes de l'approche neuromimétique [TIM-12], pour évaluer sa fiabilité. Donc, nous avons appliqué les deux approches sur le même échantillon.



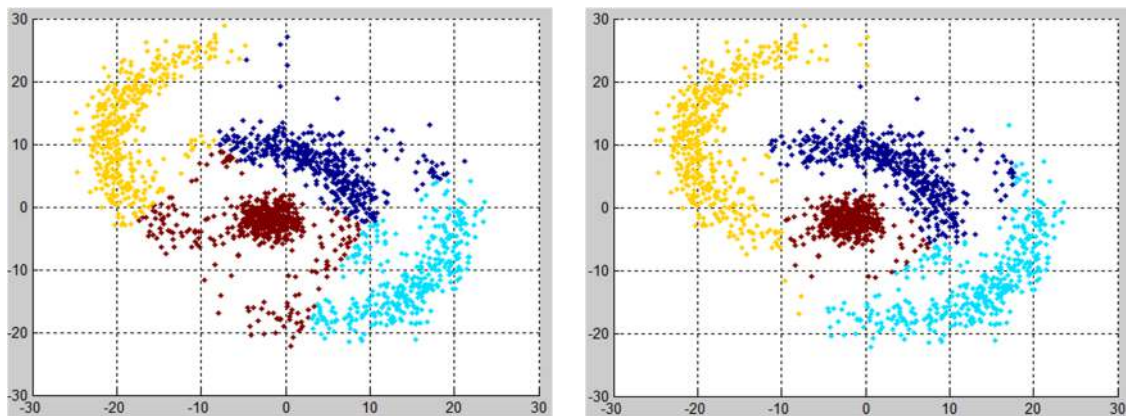
(a) Les modes détectés par l'approche neuromimétique (b) les modes détectés par l'approche proposée

Figure IV. 3: Les résultats des deux techniques de détection des modes de l'approche neuromimétique et l'approche proposée

La figure IV.3(a) montre les quatre modes détectés avec la technique d'analyse de proximité entre les maxima locaux résultant de la phase d'apprentissage dans l'approche neuromimétique. Alors que la figure IV.3(b) illustre la triangulation de Delaunay induite masquée par l'ensemble des maxima locaux; chacun des groupes de neurones reliés par des arêtes représente une classe de l'échantillon; cette méthode est une approche automatique qui ne nécessite pas de seuillage, comme le cas de l'apprentissage de GNG (*Growing Neural Gas*), et elle n'a pas besoin d'information ni sur le nombre des classes, ni sur la structure de leurs distributions dans l'échantillon. En fait, les résultats obtenus de la phase d'extraction des maxima locaux, rendent la tâche du CHL plus facile, ce qui explique l'efficacité de la nouvelle approche.

IV.2.1.3. Classement des données

Après avoir défini les modes qui préservent la topologie de la distribution, les observations qui ne font pas partie des modes estimés, sont ensuite affectées à la classe du prototype le plus proche au moyen de la distance de Mahalanobis comme critère de ressemblance. Cela est illustré dans la figure IV.4(a) en comparaison avec le résultat de cette phase utilisant la distance euclidienne présenté dans la figure IV.4(b):



(a) Observations classées par l'approche neuromimétique (b) Observations classées par l'approche proposée

Figure IV. 4: Comparaison entre les résultats de l'approche neuromimétique et ceux de la nouvelle procédure

Ces résultats montrent bien que cette nouvelle méthode de classification est plus performante que celle utilisant la distance Euclidienne et ceci grâce à l'utilisation de la distance de Mahalanobis qui tient en compte la dispersion de l'ensemble des données. Ainsi l'approche développée est adaptée à toute forme d'échantillon (gaussienne et non gaussienne).

En comparant les matrices de confusion de l'application des deux procédures et les taux d'erreur estimés, en tant que rapport entre le nombre d'observations mal classées et le nombre total des observations, nous obtenons les résultats suivant:

	Classification par l'approche neuromimétique	Classification par l'approche proposée
Matrice de Confusion	$\begin{pmatrix} 419 & 81 & 0 & 0 \\ 0 & 289 & 111 & 0 \\ 0 & 0 & 326 & 173 \\ 0 & 0 & 0 & 400 \end{pmatrix}$	$\begin{pmatrix} 469 & 31 & 0 & 0 \\ 0 & 372 & 28 & 0 \\ 0 & 0 & 459 & 40 \\ 0 & 0 & 0 & 400 \end{pmatrix}$
Taux d'erreur	0.1984	0.0478
Temps d'exécution	0.4410 s	2.2944 s

Tableau IV.3: Comparaison entre les matrices de confusion et les taux d'erreur des deux approches appliquées sur l'exemple1

Ces résultats montrent que l'approche proposée reste la plus efficace. Cela est dû à l'utilisation de l'apprentissage par CHL avec la distance de Mahalanobis, qui est également utilisé dans la procédure de regroupement comme un nouveau concept dans l'approche proposée, malgré que cela consomme plus de temps d'exécution par rapport à la distance euclidienne.

IV.2.2. Exemple 2

L'objectif de ce deuxième exemple est de démontrer la relativité et la stabilité de notre méthode de classification. L'exemple présente une distribution de 1200 observations bidimensionnelles, constituées de 2 classes gaussiennes et 2 classes non gaussiennes avec des paramètres statistiques donnés respectivement dans les tableaux IV.4 et IV.5.

Population	Nombre d'observations	Moyennes	Variances
Classe 1	200	0.0914	3.2064
		0.0914	2.7353
Classe 2	200	5.0488	3.3884
		-9.0809	3.9367

Tableau IV. 4: Paramètres statistiques des deux classes gaussiennes

Les observations des 2 classes non gaussiennes sont décrites par les mêmes attributs utilisés dans le premier exemple :

Population	$\Theta(\text{deg})$	B1	B2	A	Nombre des observations
Classe 3	$m=75$ $s=45$	$\mu_1=0$ $\sigma_1=3$	$\mu_2=-2$ $\sigma_2=3$	10	500
Classe 4	$m=225$ $s=30$	$\mu_1=4$ $\sigma_1=2$	$\mu_2=-5$ $\sigma_2=2$	10	300

Tableau IV. 5: paramètres statistiques des 2 classes non gaussiennes

La distribution de cet exemple est illustrée dans la figure IV.5(a). La figure IV.5(b) représente la fonction de densité de probabilité sous-jacente à la distribution, estimée avec le pas de discrétisation $L = 30$. Cette estimation a permis de réduire le nombre des observations de 1200 à 357 observations.

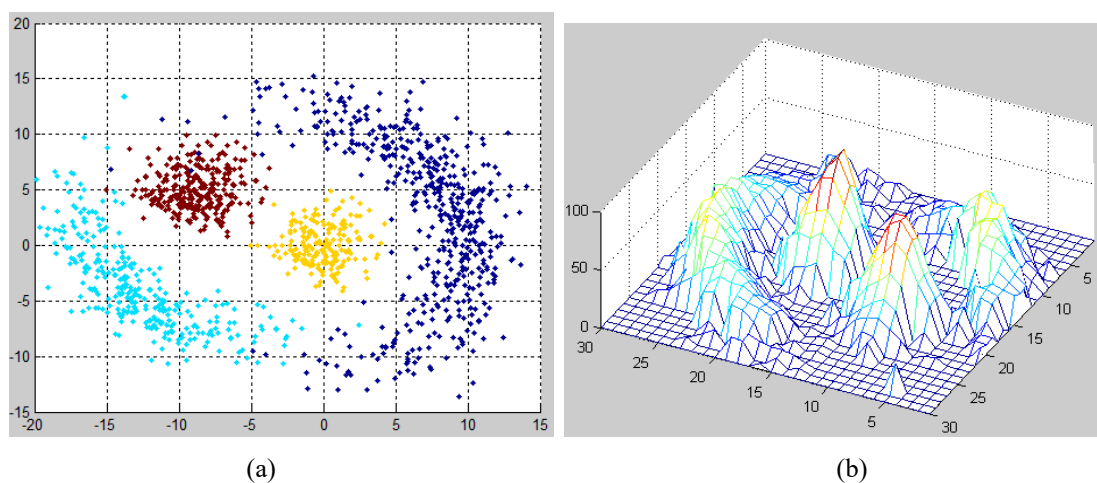


Figure IV.5: (a) représentation des données brute de l'exemple 2, (b) Estimation de la fdp sous-jacente à la distribution

La figure IV.6 montre le résultat de l'algorithme d'apprentissage, après 4 cycles, pour un nombre de neurones de sortie initialisé à 10 et avec les mêmes paramètres utilisés pour l'exemple 1 ($\alpha_0 = 1$ et $\tau = 40$) :

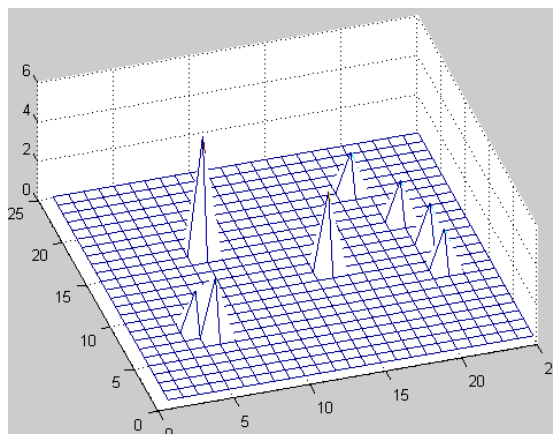


Figure IV.6: La détection des 10 maxima après l'estimation de la fdp sous-jacente du modèle 2

IV.2.2.1. Détection des modes

Bien que cette approche donne des résultats plus robustes que l'approche neuromimétique, l'utilisation de la méthode d'apprentissage compétitif Hebbien CHL, dans sa version usuelle, reste sensible aux irrégularités locales dans la répartition des données, en particulier pour les petits ensembles, souvent présentées par un seul neurone de sortie après l'apprentissage compétitif supervisé. Cela pose un problème des classes présentées par un seul prototype qui seront toujours liées à la classe voisine. Par conséquent, nous avons développé le CHL pour résoudre ce problème (cf. §III.4.2). Nous avons vu dans le chapitre précédent que cette modification du CHL est bien adaptée aux distributions gaussiennes, notre but de ce paragraphe est de vérifier son efficacité sur les distributions non gaussiennes. Voici maintenant, dans ce tableau, quelques cas d'initialisation de nombres k de neurones de sortie qui illustrent ce problème et sa résolution par notre nouvelle méthode :

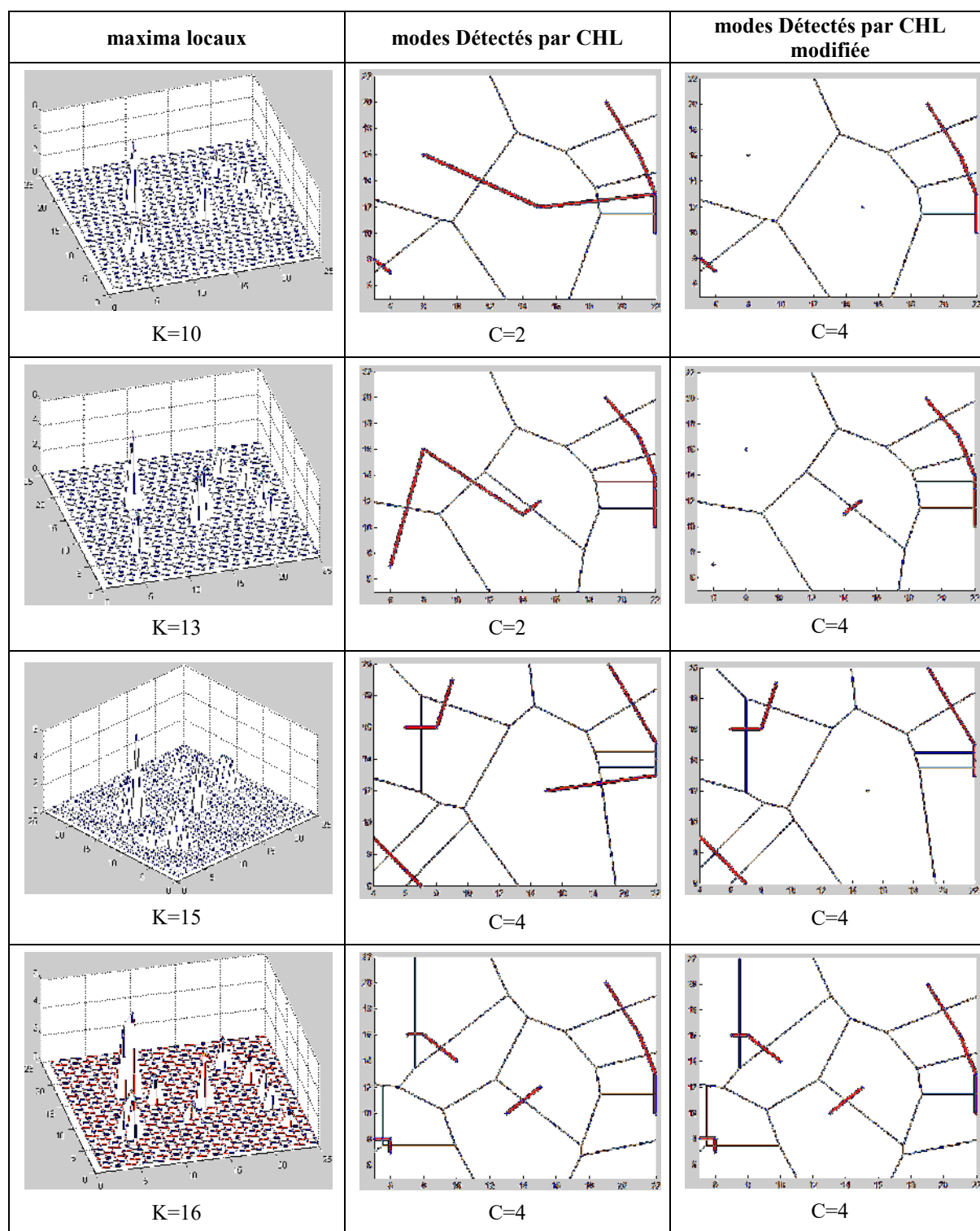


Tableau IV. 6: r esultats des deux phases d'apprentissage pour diff erents nombres de neurones de sortie

A travers ce tableau, et en comparant les r esultats de l'apprentissage de CHL avec les r esultats du CHL am elior e, nous pouvons bien conclure que ce nouvel algorithme a pu r esoudre ce

problème pour les données non sphériques aussi, grâce à l'utilisation de l'écart-type qui permet d'avoir une information sur la distribution des données, en plus de la métrique de Mahalanobis adoptée. Cette nouvelle méthode nous permet, en l'occurrence, d'abord de diminuer le nombre de neurones de sortie k et donc réduire le temps d'exécution, et aussi de fournir des résultats plus fiable en prenant en considération les regroupements des données de faibles taille.

IV.2.2.2. Classement des données

Le résultat final de la classification de ce deuxième exemple par notre approche est présenté dans la figure IV.7 suivante:

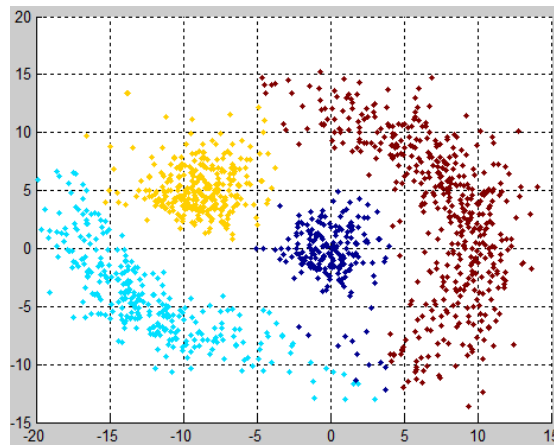


Figure IV.7: distribution classifiée par l'approche proposée

Comme le montre cette figure, les classes sont bien isolées en respectant la structure des données, ce qui prouve la faible variance intra-classes et la forte variance inter-classes des observations classées. Le taux d'erreur de cette classification par la procédure proposée est égal à 0.0216, approximativement supérieur à celui obtenu par l'approche *Neuromorphologique* [MUT-12a] qui vaut 0.0285, et meilleur que celui obtenu par la méthode classique Isodata qui vaut 0.2300.

IV.2.3. Exemple 3

Dans cet exemple, nous allons appliquer l'approche sur une distribution de 1800 observations tridimensionnelles composée de trois classes gaussiennes (cf. fig.IV.7(a)) avec les paramètres statistiques donnés dans le tableau suivant :

Classe	Nombre des observations	Moyennes	Variances
1	600	5.05576 3.81971 -2.06194	3.22752 4.17863 3.4554
2	600	-1.96581 6.84682 -2.29906	3.43273 3.47723 3.58331
3	600	-0.299437 0.369209 4.19357	3.25974 3.67446 2.9629

Tableau IV. 7: Les paramètres statistiques des trois distributions gaussiennes de l'exemple 3

La figure IV.9(b) montre les hypercubes générés après l'estimation de la fonction de densité de probabilité sous-jacente avec un pas de discrétisation $L = 25$. La figure IV.9(c) représente le résultat de l'algorithme d'apprentissage après 4 cycles pour un nombre de neurone de sorties égal à $k=12$ et avec les mêmes paramètres d'apprentissage de l'exemple 1.

IV.2.3.1. Classification de la distribution

Ces maxima détectés sont ensuite connectés par des arêtes (cf. fig. IV.8(d)) pour former trois composantes connexes représentant les trois classes de la distribution, en utilisant la méthode d'apprentissage CHL modifié. La phase finale est celle de l'affectation du reste des observations à leurs classes suivant la méthode de l'ASNP (cf. § III.6.1), présentée dans la figure IV.8(e) :

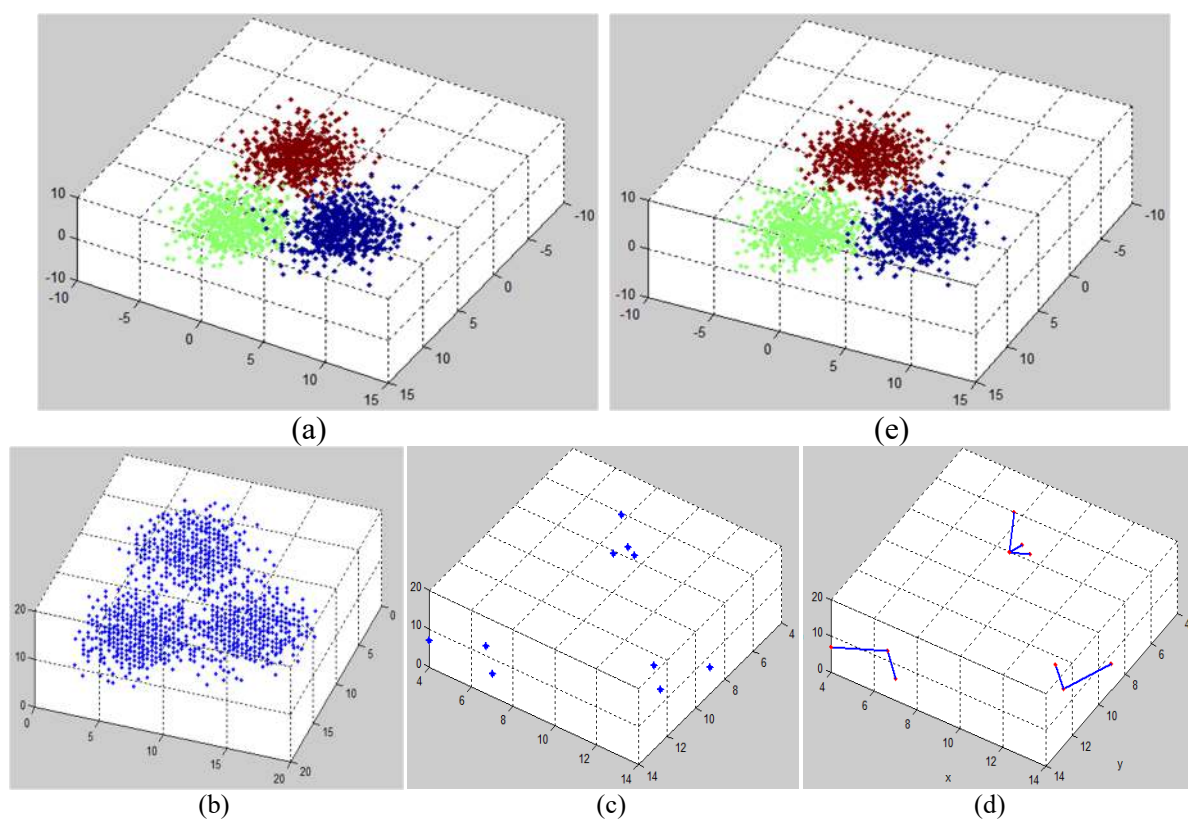


Figure IV. 8: Classification par l'approche proposée: (a) Données brut, (b) Estimation de la fdp sous-jacente, (c) 12 Maxima locaux détectés après 4 cycles; (d) Les composantes connexes des maxima représentant les 3 classes, (e) distribution classifiée

IV.2.3.2. Comparaison entre l'approche proposée et l'approche Neuromorphologique

La figure ci-dessous illustre les résultats de l'approche proposée appliquée sur ce troisième exemple en comparaison avec ceux de l'approche *Neuromorphologique* [MUT-12a] appliquée sur le même exemple :

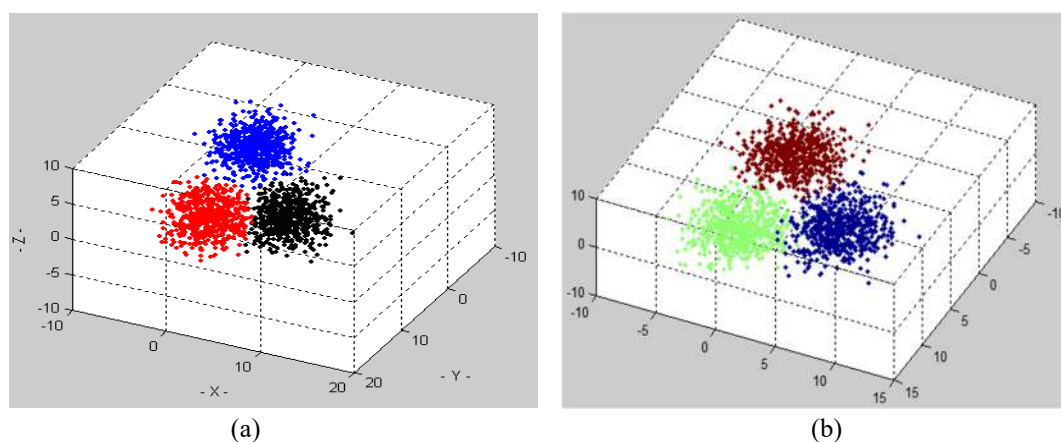


Figure IV.9: (a) Observations classifiées par l'approche Neuromorphologique, (b) Observations classifiées par la méthode proposée

Le tableau suivant expose les résultats statistiques de l'exécution des deux approches:

Résultats de la classification par l'approche neuromorphologique					Résultats de la classification par l'approche proposée					
Population	Vecteurs moyennes	Variances			Probabilité a priori	Vecteurs moyennes	Variances			Probabilité a priori
Classe 1	5.1142	2.9576	-0.0336	0.0023	0.3294	5.0282	3.2512	-0.0698	0.0012	0.3367
	3.7998	-0.0336	4.1001	0.0879		3.8164	-0.0698	4.0982	0.1057	
	-2.0834	0.0023	0.0879	3.3929		-2.0809	0.0012	0.1057	3.3943	
Classe 2	-1.9411	3.3989	0.0814	0.1582	0.3355	5.0593	3.1685	-0.1196	0.0454	0.3333
	6.8632	0.0814	3.3605	-0.0664		3.8344	-0.1196	4.0890	0.1330	
	-2.3055	0.1582	-0.0664	3.5134		-2.0984	0.0454	0.1330	3.3784	
Classe 3	-0.3084	3.3338	-0.0163	0.0342	0.3350	-2.0096	3.2734	0.2244	0.1308	0.3300
	0.3694	-0.0163	3.6568	0.0608		6.8922	0.2244	3.3338	-0.1185	
	4.1915	0.0342	0.0608	2.9228		-2.3000	0.1308	-0.1185	3.5426	
Taux d'erreur	0.01778				0.0067					
Matrice de confusion	$\begin{bmatrix} 584 & 5 & 4 \\ 12 & 590 & 2 \\ 4 & 5 & 594 \end{bmatrix}$				$\begin{bmatrix} 594 & 6 & 0 \\ 0 & 600 & 0 \\ 0 & 8 & 592 \end{bmatrix}$					

Tableau IV.8: Les résultats statistiques des deux approches appliquées sur des trois distributions gaussiennes de l'exemple 3

Le taux d'erreur estimé par la procédure proposée est de 0.0067, ce qui est mieux que les deux taux d'erreur obtenu par l'approche *Neuromimétique* [EDD-07a] avec 0,0895 et celle obtenue par l'approche *Neuromorphologique* utilisant l'épaississement morphologique comme concept pour détecter des régions modales, qui est égal à 0.01778. Cela montre l'adaptabilité de la procédure proposée aux distributions multidimensionnelles. En outre, si on observe plus en détail les matrices de confusion des deux approches, nous constatons que notre procédure obtient de très bonnes performances sur cet exemple.

Ces exemples, générés artificiellement, nous a permis d'évaluer les performances de notre approche de classification, en comparaison avec d'autres approches. Dans le reste de ce chapitre, nous allons employer cette approche dans la segmentation d'image par classification des pixels, dont le but est de prouver son efficacité et son utilité pour les données réelles.

IV.3. Application pour la segmentation des images

Toujours pour montrer l'intérêt de notre approche, nous avons cherché dans ce travail à l'appliquer sur l'image en vue de segmentation. Nous avons alors nous focalisée sur la classification automatique des pixels, en se basant sur le fait que les pixels appartenant à une même région possèdent des caractéristiques de similarités très proches et donc ils forment un nuage ou classe de points dans l'espace de représentation des couleurs ou d'intensité.

Ainsi, notre objectif est de classer les pixels d'une image selon leur distribution dans l'espace de représentation et de leur attribuer un étiquetage.

Dans ce paragraphe nous traitons plusieurs images modèles ; synthétiques et réelles, afin d'analyser les comportements de l'approche proposée et de prouver sa performance sur la segmentation d'image. Comme première phase de la procédure, l'estimation de la fdp dépend principalement du paramètre de discrétisation. Nous avons alors commencé par l'ajustement de ce paramètre pour les images.

IV.3.1. Ajustement de la procédure de détection des modes

L'ajustement du pas de discrétisation R (paramètre de résolution) a un effet majeur sur la fonction de densité fdp. En effet, une valeur de R trop petite, peut provoquer la connexion entre des hypercubes de classes différentes et donc un seul mode détecté présentera plusieurs classes.

A l'opposé, une valeur de R trop grande, permettra à la procédure de générer un grand nombre de petits modes non significatifs. En fait, l'ajustement de ce paramètre dépend de la taille de l'échantillon soumis à l'analyse, de la dimension des données et de la structure de leur distribution.

Afin de concrétiser cet effet, nous testons la procédure sur 3 images couleur pour un intervalle de R entre 10 et 68 (cf. fig. IV.10) :

- Image 1 est une image simple (au niveau de structure des objets dans l'image et d'enchevêtrement des voxels dans l'espace colorimétrique) de taille 128x128,
- Image 2 est une image compliquée de taille 128x128,
- Image 3 est une image moyenne de taille 256x256.

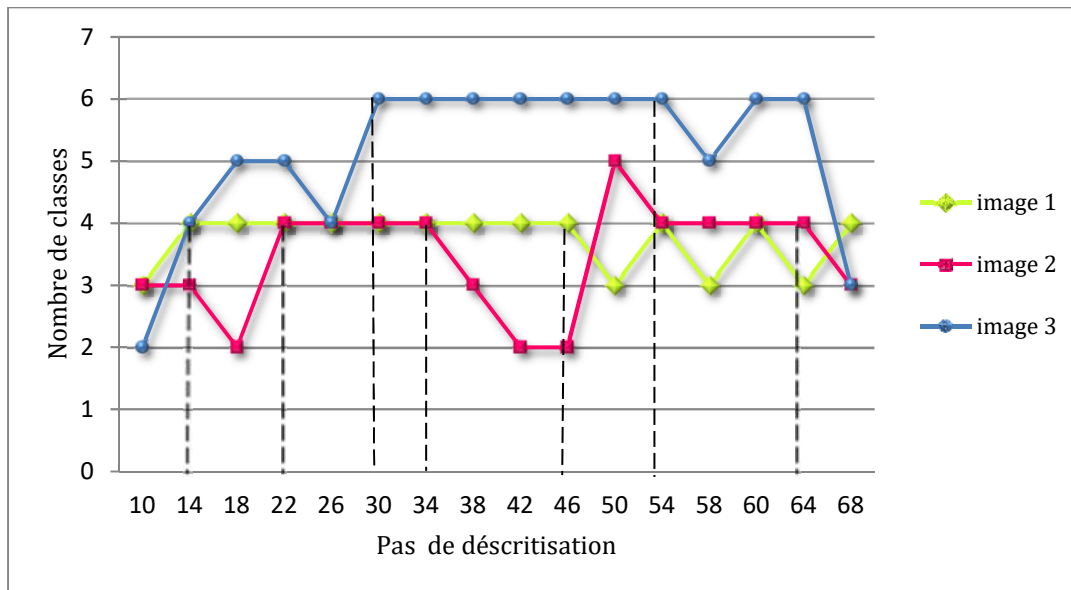


Figure IV.10: Effet du pas de discrétisation R sur le nombre de modes détectés

De ce graphe, nous pouvons constater que le choix du pas de discrétisation R dépend de la taille de l'image et de l'enchevêtrement des classes de pixels dans l'espace de représentation des couleurs (cf. §I.3.2.3). En fait, ce qu'on a tiré de l'exploitation de ces trois images et d'autres dans notre travail, est que plus l'image est grande et plus elle contient des détails, elle demande un pas de discrétisation élevé. Cela permet d'éviter la fusion des classes de petite taille.

En revanche, dans un contexte non supervisé, ce paramètre est ajusté selon le concept de stabilité du nombre de modes [POS-81]. Il s'agit d'analyser les résultats obtenus pour une grande marge de variation de ce paramètre. Il a été prouvé que le choix de ce paramètre au milieu du plus grand intervalle à l'intérieur duquel le nombre de modes détectés reste constant et différent de 1 est une bonne procédure pour optimiser un nombre important d'algorithmes de classification [TOU-88].

IV.3.2. Image couleur

Sous l'hypothèse que chaque mode détecté correspond à une région homogène de l'image, l'approche proposée affecte chaque pixel à la classe dont le centre de gravité correspond à la couleur la plus proche de celle du pixel à classer, en mesurant la distance dans l'espace colorimétrique. Les trois composantes colorimétriques de chaque pixel définissent les attributs des observations dans l'espace 3D du système de représentation de couleur RVB.

IV.3.2.1. Images synthèses

Pour illustrer le comportement de la procédure, nous présentons deux exemples modèles :

IV.3.2.1.1. Exemple modèle 1

Ce premier modèle est une image constituée de 4 régions de couleurs et de formes différentes (cf. fig.IV.11(a)). Les trois composantes RVB, codées sur 256 niveaux, constituent les axes des coordonnées de l'espace de représentation des pixels de l'image (cf. fig. IV.11(b)). L'approche proposée commence, tout d'abord, par l'organisation des pixels de l'image dans une matrice d'observation (cf. §I.2.1.1) pour pouvoir faire une estimation de la fdp avec un paramètre de résolution $R=30$ (cf. fig.IV.11.c), puis elle passe à l'extraction des maxima locaux de l'échantillon d'observation (cf. fig.IV.11.d) qui seront ensuite connectés par des arrêts pour former les modes des classes (cf. fig.IV.11.e) . Enfin, le reste des pixels sera affecté à la classe du mode le plus proche avant d'être étiqueter et reformer en image (cf. fig.IV.11.f).

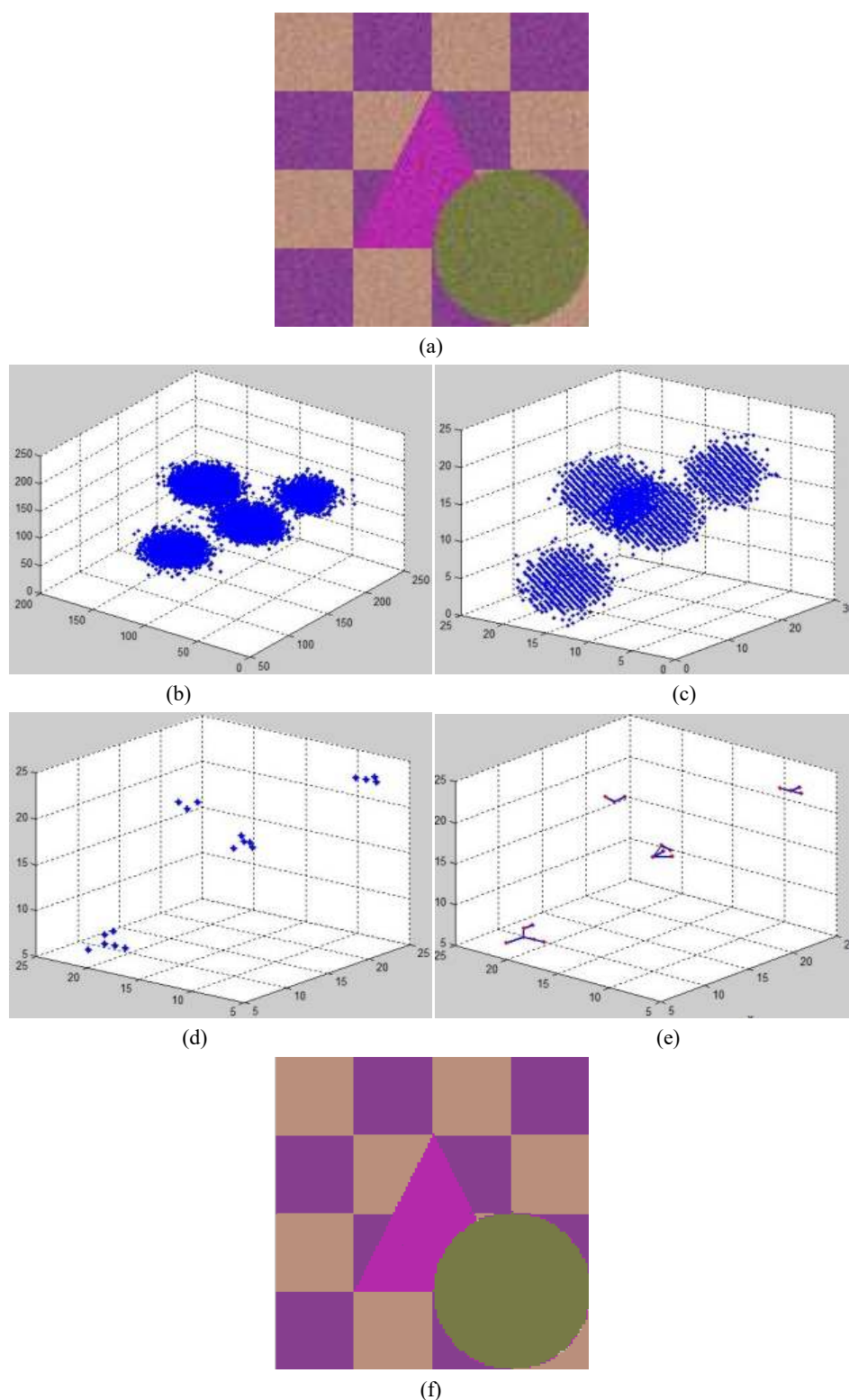


Figure IV. 11 : (a) image couleur originale; (b) présentation des pixels dans un espace RVB; (c) Estimation de la fdp sous-jacente; (d) Prototypes des classes dans l'espace RVB; (e) Prototypes connectés par des arrêtes ; (f) image couleur segmentée

Sous l'hypothèse que chaque mode de la fdp correspond à une région homogène de l'image, ce résultat montre que notre procédure arrive à extraire les quatre régions de cette image dans un nombre de cycles égal à 3, pour 18 neurones de sortie du réseau de neurones compétitif.

Ci-dessous nous illustrons séparément les 4 classes de l'image:

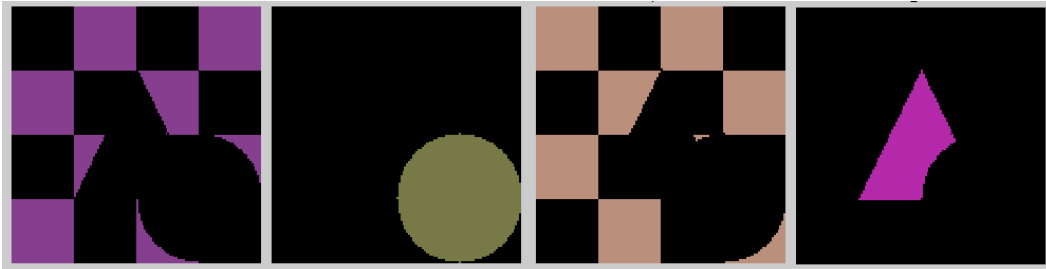


Figure IV. 12: Les quatre classes de l'image modèle

Cette figure montre que la technique proposée a réussi d'extraire parfaitement les quatre classes de l'image à travers les quatre groupes de neurones connectés qui ont été utilisés pour obtenir l'image segmentée.

IV.3.2.1.2. Exemple modèle 2

Nous exposons maintenant un exemple qui représente un certain degré de chevauchement [LAM-00]. Cette image couleur de synthèse (cf. fig.IV.14(a)) est composée de 6 régions de différentes formes ; les deux disques verts concentriques où les pixels associés à ces deux régions adjacentes constituent deux nuages qui se chevauchent dans l'espace RVB, ce qui rend difficile leur discrimination.

L'apprentissage compétitif dans l'approche proposée cherche généralement à détecter les centres de ces régions représentés par les valeurs maxima de la fdp sous-jacentes à la distribution des pixels (cf. fig.IV.14(d)).

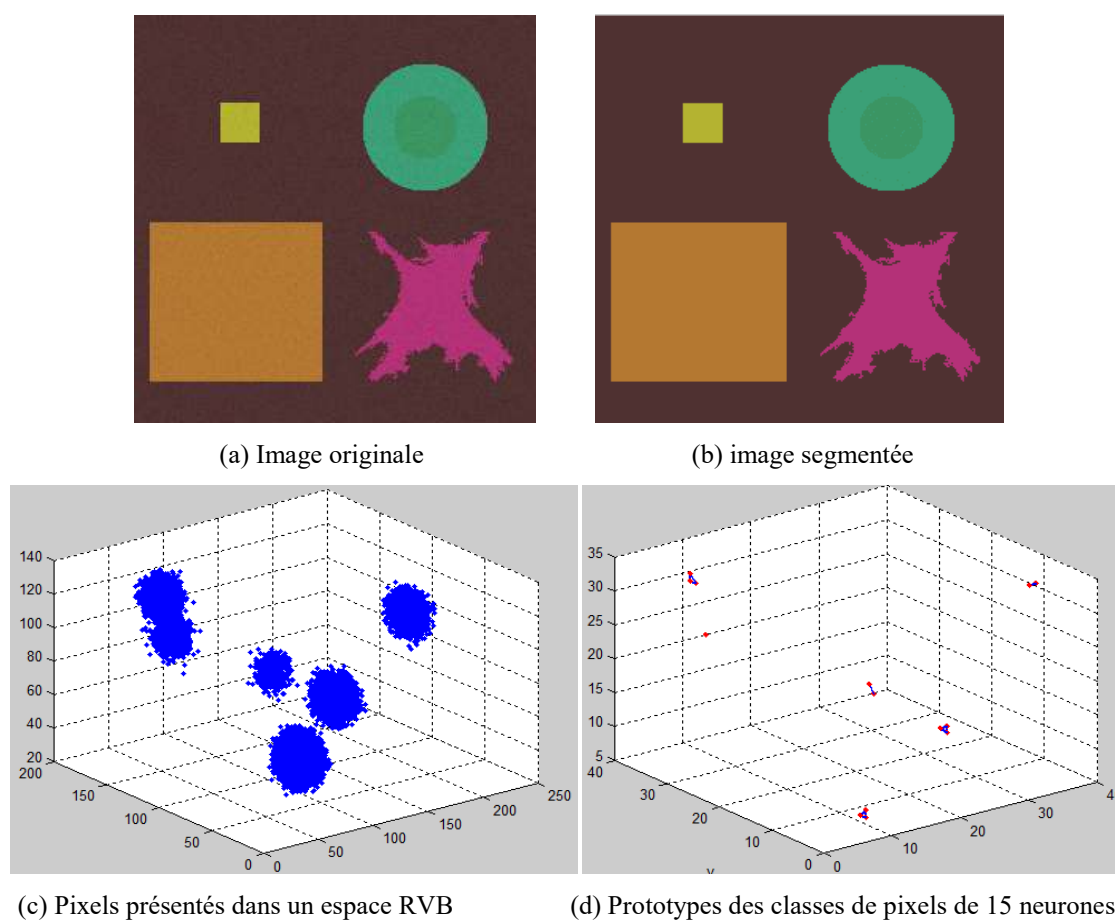


Figure IV.13 : segmentation de l'exemple 2 par l'approche proposée

Avec un paramètre de résolution $R = 40$, la procédure proposée pour la détection des modes de la fdp appliquée sur l'échantillon d'observations de la figure IV.13.c permet d'extraire les six régions associées à l'image.

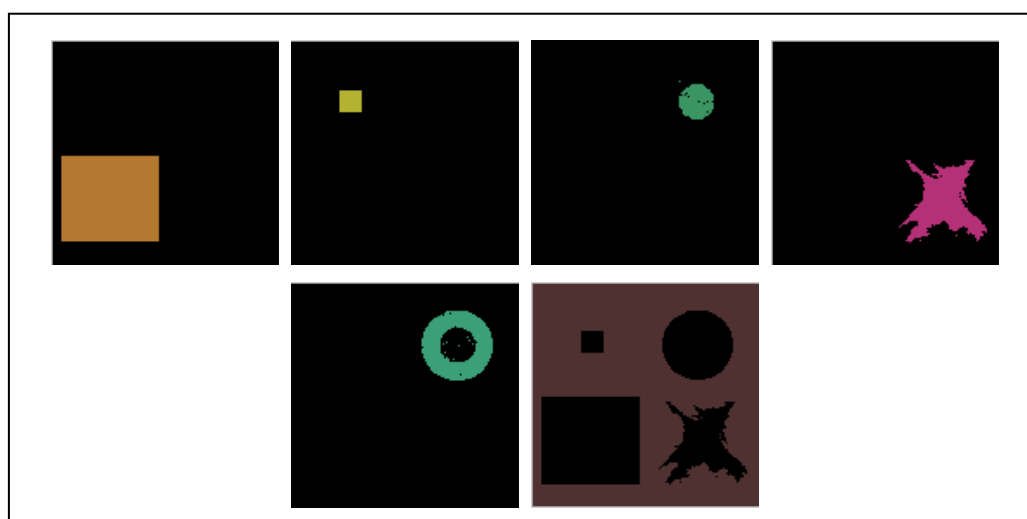


Figure IV.14: illustration des six régions séparées de l'image segmentée

A partir de la figure ci-dessus nous constatons que notre approche nous permet d'aboutir à une segmentation qui donne une bonne reconstruction de ces régions. En effet, après 5 cycles,

la procédure arrive à identifier les 6 régions parmi 15 neurones de sortie présentant des candidats de classes de l'image.

Ce résultat est comparable aux résultats obtenu, sur la même image, par d'autres approches de segmentation par classification, qui ont pu détecter les 6 classes de l'image, telles que l'approche *Neuromorphologique* [MUT-12a], *Isodata* [TAK-95], une méthode de classification qui prend en compte la disposition spatiale des pixels, selon une procédure beaucoup plus complexe [FON-01] et une autre méthode de classification spatio-colorimétrique des pixels, basée sur la détection des modes de l'histogramme 3D des couleurs[MACE-04].

IV.3.2.2. Images réelles

Maintenant, Nous proposons de segmenter, par notre approche, 3 images couleur réelles avec un degré d'enchevêtrement très important. La figure suivante illustre les 3 images modèles, leurs pixels dans l'espace colorimétrique RVB discrétisé, leurs prototypes détectés par la procédure proposée et leurs images segmentées :

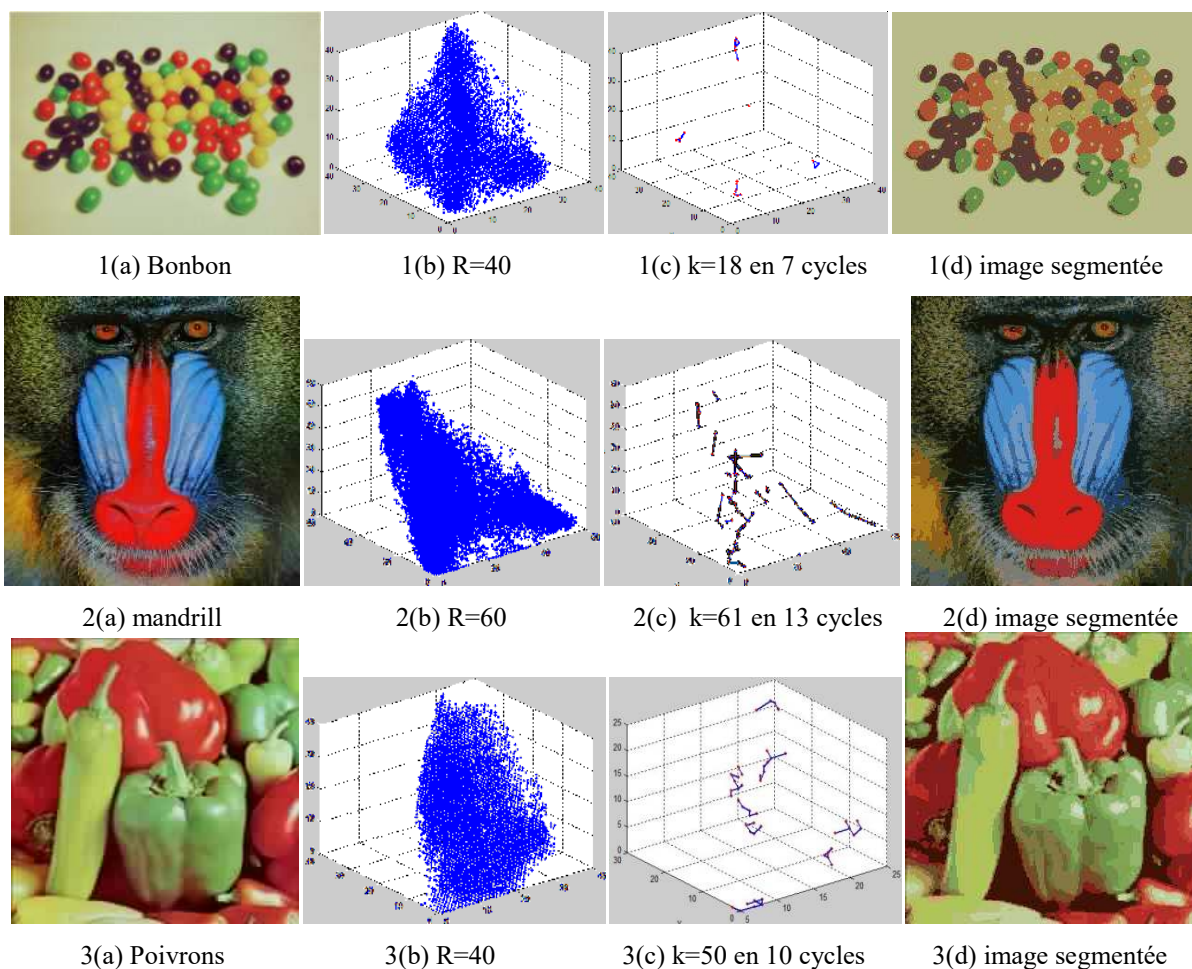


Figure IV. 15: (a) image Originale; (b) Hypercubes de pixels dans l'espace colorimétrique RVB discrétisé; (c) Modes détectés; (d) image segmentée

IV.3.2.2.1. Évaluation qualitative

Dans l'absence d'une segmentation de référence, nous évaluons la qualité de la segmentation par inspection visuelle de l'image segmentée par rapport à l'image originale. En examinant l'image 'Bonbon' on trouve que les classes correspondent effectivement aux régions de l'image. La technique parvient alors à discriminer les six classes de l'image 'Bonbon' qui sont les bonbons de couleurs rouge, jaune, noire et verte et le fond clair, en outre, le nombre de chaque couleur de bonbon dans l'image construite, correspond exactement à celui dans l'image originale.

En termes de pertinence de segmentation, il est à noter aussi que les voxels du fond sont considérés comme une seule classe alors que les objets de premier plan sont plus détaillés au niveau de degrés de connexité et d'homogénéité, comme le cas des deux images segmentées suivantes :

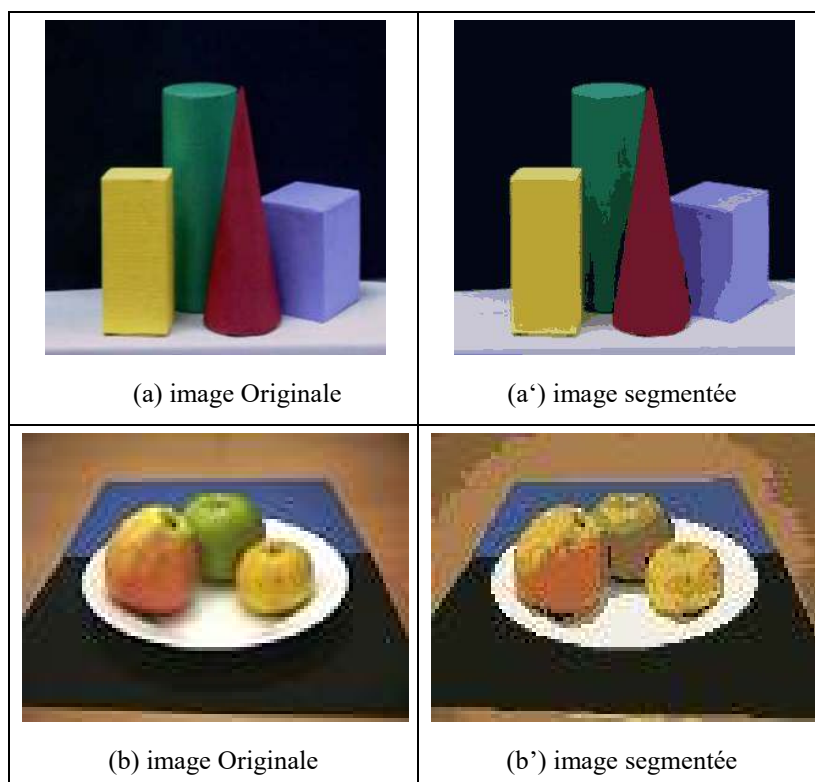


Figure IV.16: segmentation de deux images par l'approche proposée

De même, les deux images construites de 'mandrill' et 'Poivrons' (cf. fig.IV.15.2(d) et fig.IV.15.3(d)) montrent que notre procédure a pu mettre en exergue même les classes les plus petites.

Malgré la difficulté du traitement de ces images où les différents groupes présentent un degré de chevauchement significatif dans l'espace colorimétrique RVB, l'utilisation de la distance de Mahalanobis comme critère de ressemblance ainsi que la nouvelle méthode de détection des modes, permet à l'approche proposée, dans un contexte non supervisé, de donner des résultats plus performants par rapport à l'approche *neuromimétique*.

IV.3.2.2.2. Évaluation quantitative

Afin d'évaluer la performance de notre approche, nous procédons à une étude quantitative des résultats des 3 images (Figure IV.15.(a)), en évaluant l'homogénéité des images par rapport aux résultats de la méthode *K-moyennes* appliquée sur les mêmes images. Nous avons choisi d'utiliser l'intensité moyenne de voxels, l'écart-type et le coefficient de variance, comme critères quantitatifs, et nous avons exposé les résultats expérimentaux des images modèles dans le tableau suivant :

Paramètres	K-Moyennes			méthode Proposée		
	Bonbon	mandrill	Poivrons	Bonbon	mandrill	Poivrons
Intensité moyenne de voxels	39.235	38.343	45.764	59.654	30.4679	47.0573
Ecart-Type	14.502	22.553	21.6	14.9358	13.6251	19.3433
Coefficient de Variance	36.96%	58.89%	47.21%	25.04%	44.72%	41.106%

Tableau IV. 9: Paramètres statistiques des trois images segmentées par les deux méthodes

Ces résultats statistiques sont présentés plus explicitement dans les graphes suivants:

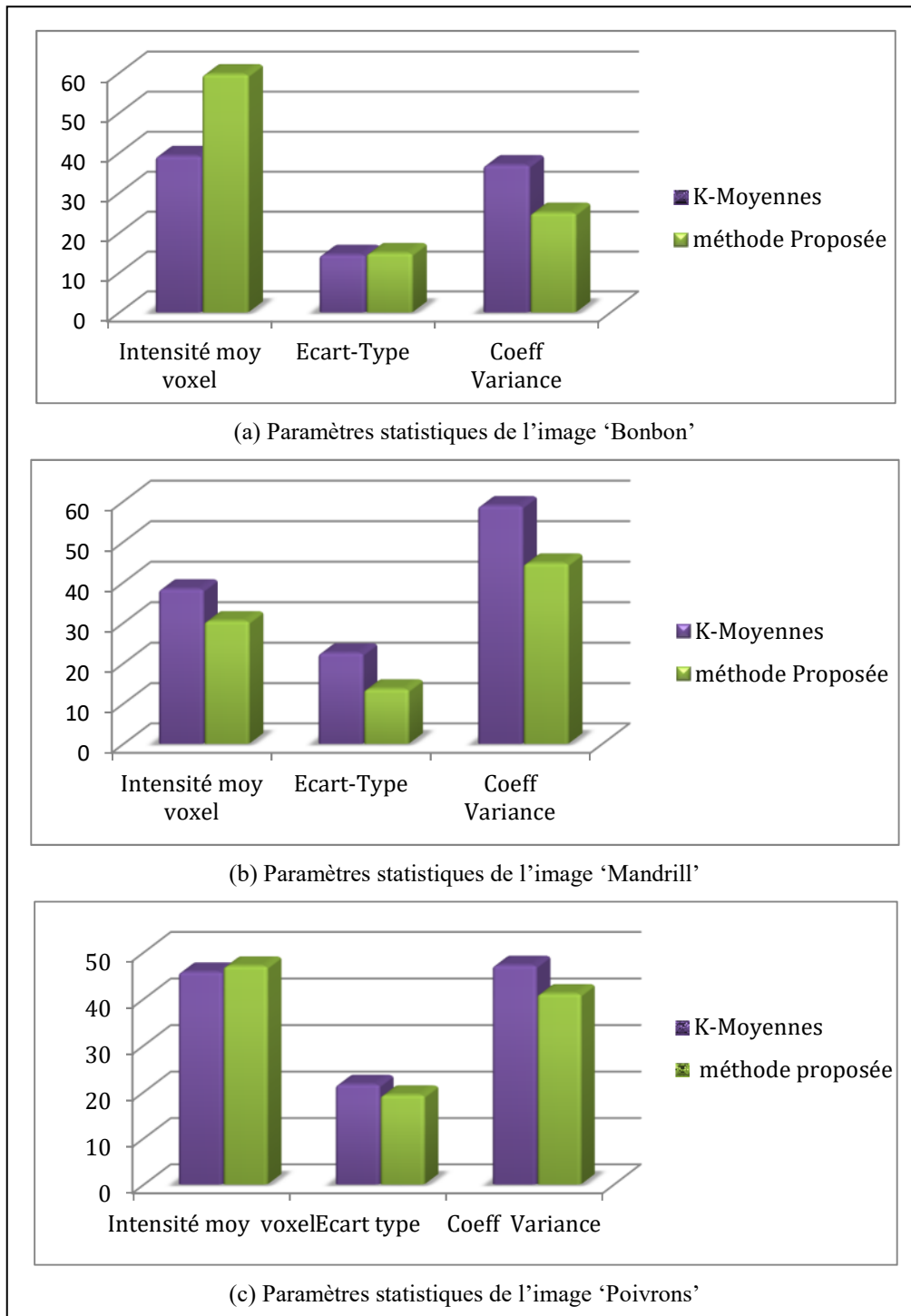


Figure IV.17: Comparaison entre les paramètres statistiques des trois images

Malgré que le résultat de la segmentation de l'image 'Bonbon' par notre méthode a donné un écart-type légèrement élevé et le résultat de l'image 'Mandrill' a donné une intensité moyenne de voxels faible par rapport au résultat de K-moyennes, les résultats de la segmentation des trois images par notre approche ont fourni des coefficients de variance plus faibles que ceux fournis par *K-moyennes*, ce qui est expliqué par une faible dispersion des pixels de même classe autour de leur moyenne dans l'espace de représentation RVB et donc une homogénéité

interclasses plus élevée. Par conséquent, nous pouvons constater que la procédure proposée démontre sa précision dans la segmentation d'image couleur.

IV.3.3. Images TEP scan de la maladie d'Alzheimer

Dans l'objectif d'analyser le comportement de l'approche de classification proposée, et de mettre en évidence son intérêt, nous avons choisi de l'appliquer à la segmentation d'image médicale, en particulier, *la tomographie par émission de positons (TEP)*, c'est une méthode d'imagerie très utilisée dans la recherche médicale, qui illustre le métabolisme de la radiotracer 18-FDG à haute sensibilité, en détectant les photons émis du radiotracer localisé vers des cellules anormales (cf. §I.3.3.3.1).

Avant toute analyse ou quantification (cf. § I.3.3.3.1), la segmentation des images TEP est une étape souvent nécessaire pour différencier le tissu anormal des zones normales environnantes dans l'image TEP, ce qui aide à la détection appropriée des maladies, le diagnostic, la planification du traitement et les suivis.

De nombreuses études récentes ont démontré l'intérêt de l'utilisation de l'imagerie cérébrale en se servant du TEP scan pour le diagnostic précoce de la maladie d'Alzheimer (MA) et pour le diagnostic différentiel de la démence ou troubles cognitifs légers (*Mild Cognitive Impairment (MCI)*). Ces études ont pu valider que la maladie d'Alzheimer montrera généralement une réduction de l'utilisation du glucose dans le cortex cérébral [MER-13] [MER-12].

Ainsi, la segmentation d'image médicale automatique a pour but de représenter le contenu de l'image en fonction de ses caractéristiques, dans l'espoir de cibler la maladie dans ses premiers stades, avant que des lésions cérébrales irréversibles ou un déclin mental ne se produisent.

IV.3.3.1. Image TEP synthétique

À cet effet, nous appliquons notre procédure sur une image synthétique de la maladie d'Alzheimer (Figure IV.18(a)), montrant un cerveau normal (cerveau supérieur) et un cerveau affecté par la maladie d'Alzheimer (MA) (cerveau inférieur).

Comme indiqué dans le premier chapitre bibliographique (cf. §I.3.3.3.1), l'image de l'activité métabolique du glucose est codée par couleur de sorte que les régions les plus actives

apparaissent en rouge ou blanc, et les régions les moins actives métaboliquement sont conçues par les couleurs foncées violet ou bleu².

La figure suivante montre les différentes phases de la procédure proposée appliquée à l'image MA (cf. fig.IV.18 (a)):

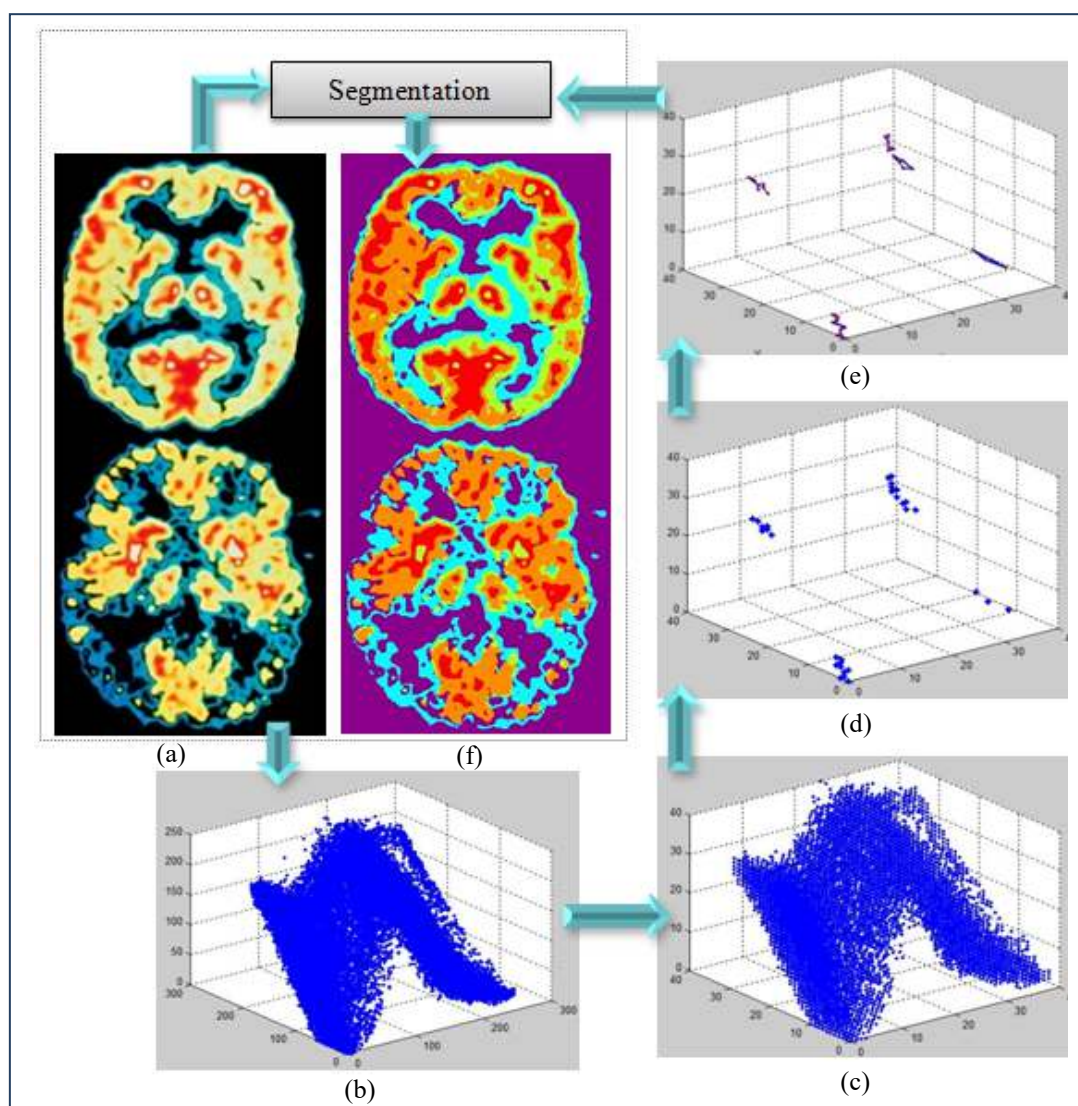


Figure IV. 18: Les différentes phases de la segmentation de l'image PET par la nouvelle approche : (a) Image MA originale ; (b) Pixels dans l'espace couleur RVB ; (c) Estimation de la fdp sous-jacente ; (d) 30 Maxima du pdf détecté après 7 cycles ; (e) Les modes détectés ; (f) Image MA segmentée avec des étiquettes de groupe colorées

Cette figure illustre clairement l'enchaînement des différentes étapes de la méthode proposée pour segmenter l'image (MA). En fait, après avoir réduire le nombre des observations par le calcul de l'estimation de la fdp (figure IV.18(c)) et l'extraction des maxima locaux de la fdp sous-jacente (figure IV.18(e)), la technique CHL modifiée détecte les modes (cf. fig.IV.18(d))

² <http://uuhsc.utah.edu/cacir/openhouse.html>

qui définissent des hypercubes des voxels prototypes des régions homogènes présentées dans l'image. Au lieu d'utiliser le grand nombre d'intensités de voxels [FOS-14a], la méthode rapide de l'estimateur de Parzen permet de réduire le nombre d'intensités de voxels à traiter ce qui rend la procédure de détection des modes plus rapide.

IV.3.3.1.1. Évaluation qualitative

Comme le montre la figure IV.18(b), nous pouvons bien voir le fort degré de chevauchement qui existe entre les voxels des différentes classes de l'image (MA), présentées dans l'espace RVB. Et pourtant, l'approche neuronale développée a permis de détecter 5 modes significatifs marquant parfaitement les 5 régions d'intérêts ROI (*Region Of Interest*) (figure IV.19(e)). Ces modes sont ensuite utilisés pour segmenter l'image (cf. fig.IV.18(f)). La figure ci-dessous visualise les différentes classes de l'image MA:

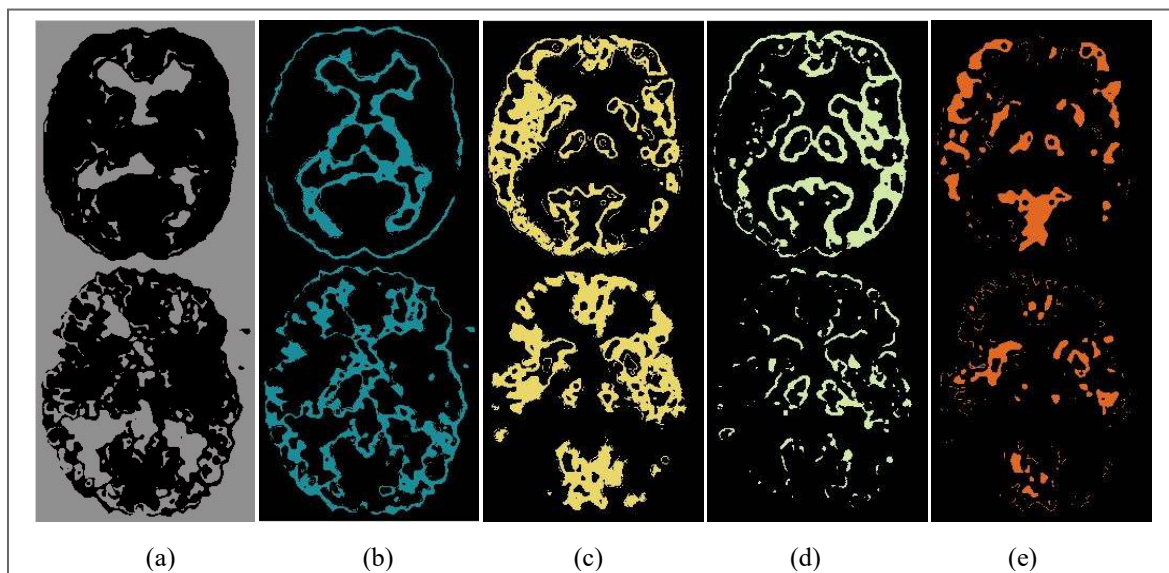


Figure IV.19: Visualisation des différentes régions de l'image MA séparées (a) arrière-plan de l'image MA en niveaux de gris; De (b) à (e) la délimitation des quatre classes de premier plan

La plupart des techniques de segmentation de l'image TEP sont bien adaptées à la détermination de l'absorption focale, mais donnent généralement de faibles résultats de segmentation pour les modèles d'absorption de radiotraceur diffusés et multifocales tels que ceux observés dans l'image MA. A l'exception d'une étude récente qui propose un cadre de segmentation du TEP adapté de manière unique à la segmentation de ces modèles d'absorption difficiles [THO-11][FOS-13][FOS-14a]. Néanmoins, cette méthode a besoin de choisir les seuils optimaux pour segmenter l'image en plusieurs régions, ce qui n'est pas adéquat, à cause de la limitation inévitable lors de l'utilisation de méthodes basées sur le seuil,

comme des incertitudes élevées dans les contours flous des objets, puisqu'il n'y a pas un accord général sur le choix d'un seuil [FOS-14b].

Comme le montre la figure IV.20, sans utiliser de seuil, notre approche permet de délimiter différentes régions du cerveau par la détermination de chaque classe d'intensité des voxels et permet ainsi d'identifier les régions d'activité ou d'atrophie provoquées par la maladie d'Alzheimer dans le cerveau.

IV.3.3.1.2. Évaluation quantitative

Nous allons comparer maintenant le résultat de notre approche aux résultats de K-moyennes et la c-moyennes floue FCM, couramment utilisés dans la segmentation de l'image TEP. Le tableau suivant (tableau IV.11) montre la comparaison entre les résultats statistiques des trois méthodes de segmentation appliquées sur la même image (MA) [MER-12] :

Paramètres	K-Moyennes	FCM	Procédure proposée
Moyen de Voxels	86.0916	76.4097	45.0215
Ecart-type	92.0758	42.8731	23.0798
Coefficient de Variance	106.951 %	56.109 %	51.264 %

Tableau IV.10: Paramètres Statistiques

La figure ci-dessous expose les paramètres expérimentaux d'évaluation de l'homogénéité de chaque méthode :

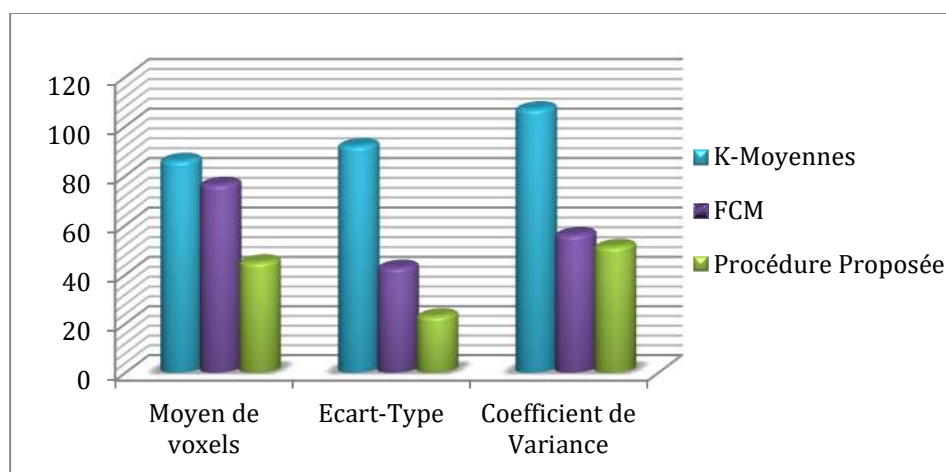


Figure IV.20: Comparaison des paramètres statistiques

A partir de ce graphe nous pouvons conclure que notre procédure a montré une homogénéité intra-classes plus élevée, grâce à sa faible écart-type par rapport aux deux méthodes K-moyennes et FCM, et en dépit de l'intensité de voxel moyenne la plus faible de la méthode

proposée, cette dernière a le plus faible coefficient de variation, ce qui prouve qu'elle a les résultats les plus cohérents.

Ces résultats montrent que la procédure proposée s'avère bien adaptée à la segmentation non supervisée des images TEP par classification des voxels.

IV.3.3.2. Images TEP réelles

Nous proposons de tester notre approche sur des images réelles collectées de la base de données numérique de *Harvard Medical School (the whole Brain Atlas)*³. Ce sont des images TEP de quatre coupes selon un plan axial du cerveau d'un homme de 70 ans qui a commencé à éprouver des difficultés avec la mémoire environ 9 mois avant l'imagerie. Les changements dans ce cerveau sont bilatéraux, mais l'hémisphère droit est légèrement plus affecté que la gauche. En vue de mesurer l'activité du radiotracer dans ces 4 coupes du cerveau, on se sert de la segmentation pour l'estimation des régions d'intérêt ROI. La figure ci-dessous illustre donc les résultats de la segmentation par notre procédure:

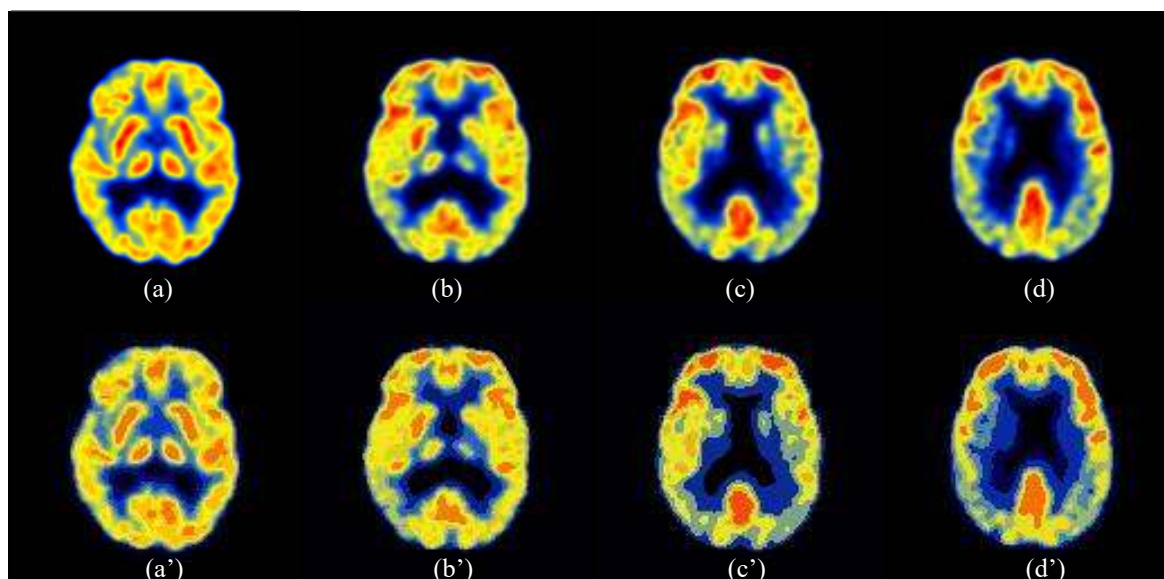


Figure IV.21: Segmentation des images TEP qui représentent 4 coupes axial du cerveau affecté d'MA par la nouvelle approche : de (a) à (d) Images TEP originales; de (a') à (d') Images TEP segmentées

En général, les méthodes de segmentation basées sur la région considèrent l'homogénéité de l'image en premier lieu pour la délimitation des régions d'intérêt ROI.

Dans les images segmentées, les régions sont colorées en calculant la moyenne des couleurs de chaque région dans les images originales. L'espace de représentation RVB de leurs voxels

³ <http://www.med.harvard.edu/aanlib/cases/caseNN1/mr1-dg1/015.htm>

a été discrétisé en $R=60$. Pour 9 cycles notre procédure a pu détecter 6 classes parmi 22 neurones candidats, nous avons extrait les régions de l'image fig.IV.21(c) et nous les avons illustré séparément dans la figure suivante :

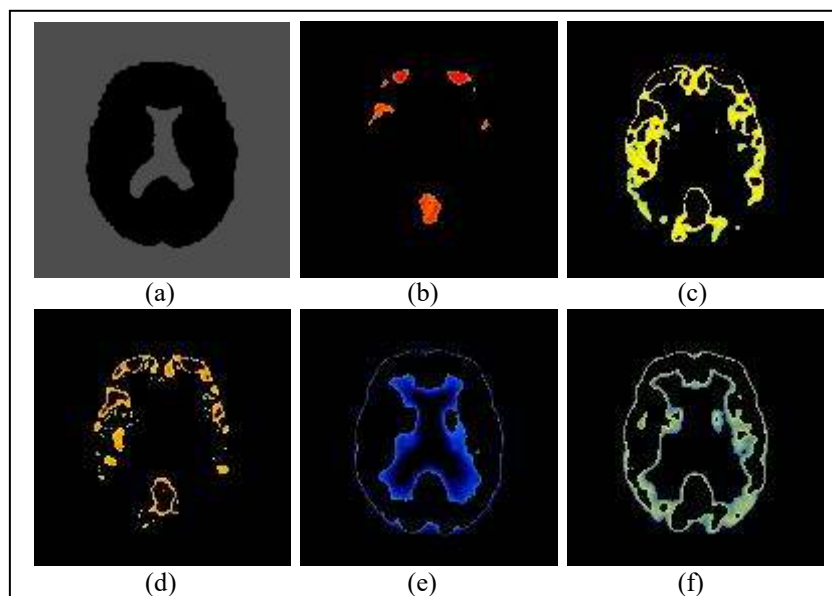


Figure IV.22: Visualisation des différentes régions séparées de l'image (IV.16.(c)): (a) arrière-plan de l'image MA en niveaux de gris; De (b) à (f) la délimitation des cinq classes de premier plan

Comme toutes les approches basées sur les graphes, l'approche proposée surpasse les autres méthodes de segmentation TEP en incorporant une reconnaissance efficace dans le processus de segmentation en utilisant des prototypes de premier plan et d'arrière-plan, comme des contraintes difficiles combinant des informations globales avec des similarités locales entre les paires des hypercubes, pour localiser les régions d'intérêt ROI dans l'image pour des résultats de segmentation optimale [FOS-14b].

L'avantage de notre approche réside dans le fait qu'il ne nécessite ni de prétraitement par exemple sous la forme d'un filtrage, ni post traitement comme le seuillage, par rapport à la FCM [MER-12] qui est sensible au bruit et à l'hétérogénéité d'intensité [BEL-10].

IV.4. Discussion

Tous ces exemples démontrent la performance de la procédure proposée, sachant que cette méthode présente tous les avantages du réseau de neurones artificiels (cf. §II.4.1) à savoir ; le parallélisme, la capacité d'apprentissage et la distribution de mémoire. En effet, les approches de l'intelligence artificielle (AI) peuvent fournir plus de précision et gagner du temps. Les réseaux de neurones artificiels (RNA), comme l'une des meilleures techniques

d'IA, ont la capacité de classer et de quantifier les lésions avec précision et de modéliser l'évaluation clinique d'un problème spécifique [SAE-10].

De même, l'utilisation de la distance de Mahalanobis, qui tient en compte les corrélations entre les variables, en tant que nouveau critère de ressemblance utilisée dans notre approche, permet de donner des résultats plus puissants par rapport à l'approche *neuromimétique* et même par rapport à l'approche *neuromorphologique*.

Néanmoins, le temps d'exécution que le calcul de la distance de Mahalanobis demande dans les deux phases d'extraction des maxima et de classement, reste important dans notre approche par rapport aux approches *Neuromemetique* et *Neuromorphologique*. Alors que la phase de détection des modes de notre approche est plus rapide que celles de ces dernières approches. Bien que, notre objectif dans le futur travail est de minimiser le temps d'exécution de la nouvelle approche en globale, vue que le calcul de la distance de mahalonobis peut poser des problèmes pour les données robustes ou/et de dimension très élevée. Dans l'attente d'une solution efficace, nous évitons ce problème en gardant provisoirement la possibilité d'utiliser la distance euclidienne dans les deux phases d'extraction des maxima et de classement de notre procédure dans des cas pareils.

Pour conclure, nous effectuons une évaluation de la procédure proposée, en vérifiant l'accomplissement des cinq indices de validation de la classification, par notre approche (cf. §I.2.4) :

- ***Validité relative :***

L'algorithme d'apprentissage compétitif supervisé de notre approche converge vers les maxima locaux de l'échantillon pour tout nombre de classes k supérieur au nombre exact des classes. Grâce à la phase de détection des modes, effectuée par la nouvelle technique d'*apprentissage compétitif Hebbien*, l'approche arrive à trouver les modes significatifs des classes existantes dans l'échantillon même celles présentées par un seul maxima local. Ainsi le choix du nombre k de classes n'a pas une grande influence sur les résultats de la procédure proposée.

- ***Validité interne :***

Notre approche, entant qu'une méthode de classification basée sur la région, le test de similarité est basé sur la distance entre les observations d'une distribution quelconque. Ainsi, nous avons vérifié la dispersion d'un ensemble des données autour de leurs

moyennes, au niveau qualitatif et quantitatif, et on a trouvé que les données similaires et les plus proches les uns des autres appartiennent à la même classe avec une marge très faible d'erreur.

- ***Validité externe :***

En comparaison avec d'autres méthodes similaires, telles que *k-Moyennes*, approche *neuromorphologique*, ISODATA et FCM, notre approche a fourni les meilleurs résultats pour les données numériques multidimensionnelles et les images couleurs. Ce qui prouve son adaptabilité pour ces types des données.

- ***Interprétabilité :*** tous les résultats de la classification sont interprétables et répondent aux objectifs de la classification.

- ***Validité de stabilité :***

Pour valider la stabilité de l'approche proposée par rapport à ses paramètres, qui sont d'abord, le pas de discrétisation que nous avons déjà discuté (cf. §IV.3.1), et le nombre k de neurones de sortie pour l'apprentissage compétitif qui, à son tour, prouvait qu'un changement léger ne cause plus le changement des résultats. Il reste les paramètres α_0 et τ de l'apprentissage compétitif supervisé. Ces derniers, d'après l'expérience de plusieurs échantillons, ne nécessitent pas forcément l'ajustement pour chaque classification, il suffit de leur affecter des valeurs qui assurent la convergence de l'algorithme d'apprentissage vers une solution stable (cf. §II.5.1.3).

IV.5. Conclusion

Dans l'objectif d'évaluer notre approche neuronale de la classification non supervisée, nous avons effectué, dans ce dernier chapitre, plusieurs expériences qualitatives et quantitatives sur une variété de données bidimensionnelles et tridimensionnelles, tout en comparant ses résultats avec d'autres approches similaires.

Les résultats montrent que cette approche, est adapté aux différentes formes de distributions; gaussiennes et non gaussiennes, et ne nécessite aucun seuillage ni détermination de plusieurs paramètres initiales.

L'application de notre approche à la segmentation des images couleur notamment l'image de TEP scan, donne des résultats plus fiables que celles obtenus par les techniques souvent utilisée actuellement pour les images TEP, et démontre son intérêt pour une image de TEP

réelle afin de cibler l'aide au traitement des maladies, comme la maladie d'Alzheimer, dans ses premiers stades, par le biais du diagnostic avant l'apparition des symptômes.

Grâce à toutes ces expériences, nous avons pu nous assurer que notre approche a répondu favorablement aux cinq indices de validation. La chose qui approuve la bonne performance de cette procédure.

CONCLUSION GENERALE

Dans ce mémoire, nous avons passé en revue quelques notions sur l'analyse des données, les diverses méthodes de classification automatique et leurs applications sur la segmentation d'images. Nous avons aussi exploré et examiné les possibilités offertes par les différents réseaux de neurones basés sur des concepts compétitifs et leurs dérivés dans le cadre de la classification non supervisée.

Le but de notre travail, est de surmonter certains obstacles qui jalonnent le chemin de la classification automatique vers la perfection. À cet égard, nous avons proposé une approche neuronale pour la classification non supervisée; son principe s'articule, en premier lieu, sur la détection des modes de la fonction de densité de probabilité sous-jacente à la distribution des données à traiter, en utilisant deux apprentissages compétitifs, et sur le classement des observations de la distribution selon les modes mis en évidence, en deuxième lieu.

Cette nouvelle procédure marche sans aucun seuillage et dans un contexte non supervisé, c'est à dire qu'à part les informations extraites des données elles-mêmes, elle ne nécessite aucune information a priori sur le nombre des classes, ni sur la structure de leurs distributions dans l'ensemble des données.

En fait, cette approche est constituée de quatre étapes :

- Elle commence par une estimation de la fonction de densité de probabilité. D'abord, l'espace de données est transformé en un espace discretisé, un algorithme rapide et simple approprié a été utilisé à cet effet. Ainsi, le traitement des données est réduit à celui des hypercubes non vides.
- Suivie d'un réseau de neurones compétitif adoptant la distance de Mahalanobis comme une métrique de ressemblance, pour mieux adapter la procédure aux données non sphériques. Ce réseau est appliqué uniquement à l'ensemble des hypercubes non vides, permettant ainsi d'extraire les maxima locaux de la pdf, ce qui permet de converger plus rapidement que s'il est appliqué à l'ensemble des observations de l'échantillon.
- Ensuite, nous utilisons une méthode développée de l'apprentissage compétitif Hebbien (CHL), afin de concevoir une procédure de classification non supervisée. Nous avons alors

cherché à relier ces maxima locaux détectés, de telle façon que chaque composante connexe de neurones représente un mode d'une classe.

- Le classement final consiste au début, à considérer des ensembles internes des observations les plus proches de chaque mode d'une classe comme des prototypes de cette classe. Enfin, les observations restantes sont assignées à leurs classes respectives au moyen d'une règle de voisinage utilisant la distance de Mahalanobis comme métrique.

Afin de donner un aperçu du comportement de l'approche proposée et de montrer sa performance, nous l'avons testé sur un ensemble des échantillons multidimensionnels générés artificiellement, puis nous l'avons appliqué à la segmentation de différentes images couleurs synthétiques et réelles, et éventuellement à la segmentation des images TEP. Les résultats ont été ensuite analysés qualitativement et quantitativement et comparés avec d'autres méthodes similaires. Ces résultats ont été satisfaisants d'après les indices de validation.

Par conséquent, nous envisageons, dans le cadre de nos travaux futurs, d'appliquer l'approche pour la segmentation des images TEP réelles de 2D ou 3D, ou encore plus récemment les images IRM-PET et d'effectuer une analyse approfondie des résultats.

Nous n'avons pas abordé l'ajustement du pas de discrétisation tant qu'un inconvénient qui affaiblit la stabilité de notre approche, puisqu'il reste bien géré et métrisable, quoique nous aspirons à trouver une solution cruciale pour la discrétisation. Parallèlement, nous nous attacherons à surmonter les inconvénients de la procédure proposée, qui se manifestent principalement dans le temps de calcul élevé à cause de l'utilisation de la distance de Mahalanobis dans plusieurs processus de l'approche.

Enfin, la perspective qui nous paraît plus intéressante et qui, peut-être, sera capable d'offrir des réponses à d'autres questions, est de développer un apprentissage capable de fournir l'automatisme, la vitesse et l'efficacité en s'inspirant des recherches les plus récentes dans le domaine de la neuroscience, pour des fins de classification.

LISTE DES PUBLICATIONS

Cette section regroupe par ordre chronologique, les publications que nous avons produites pendant cette thèse :

Communications dans des conférences internationales :

- 1) Timouyas M., Eddarouich S. and Hammouch A. (2012). «*A new approach of classification for non-Gaussian distribution upon competitive training*», IEEE Xplore, International Conference on Complex Systems, (ICCS'12), Agadir, Maroc, Pp.1-6. DOI: 10.1109/ICoCS.2012.6458572
- 2) Timouyas, M., Eddarouich, S., Hammouch, A., Touahni, R. and Sbihi, A. (2014). «*Unsupervised Neural-Morphological Colour Image Segmentation Using the Mahalanobis as Criteria of Resemblance*», ICMCS'14, IEEE conference, Marrakech, Morocco, ISBN: 978-1-4799-3824-7 April 14-16.
- 3) Timouyas M., Eddarouich S. and Hammouch A. (2016). «*Color Image Segmentation upon a New Unsupervised Approach using Amended Competitive Hebbian Learning*». In Proceedings of the 18th International Conference on Enterprise Information Systems - Volume 2: ICEIS, ISBN 978-989-758-187-8, pp. 205-210.

Publication acceptée dans un journal :

Timouyas M., Eddarouich S. and Hammouch A. (2017). «*A new neural unsupervised classification approach using amended competitive Hebbian learning: PET image segmentation insights*», International Journal of Bioinformatics Research and Applications, 2017 Volume 13, No.1, pp.1 – 21.

Timouyas M., Eddarouich S. and Hammouch A. (2017). “*Mode region detection using improved Competitive Hebbian Learning for unsupervised clustering*”, Engineering Science and Technology: An International Journal (ESTIJ), ISSN: 2250-3498 Vol.7, No.4, pp. 26-35.

BIBLIOGRAPHIE

- [ADA-94] Adams, R., Bischof, L. (1990). Seeded region growing, *IEEE Trans. Pattern Anal. Mach. Intell.* No 16 (6), 641–647.
- [AHA-90] Ahalt, S. C., Krishnamurty, A. K., Chen, P. et Melton, D. E. (1990). Competitive algorithms for vector quantization, *Neural Networks*, no. 3, 277–291.
- [ALI-07] Aliguliyev, R. M. (2007). Automatic document summarization by sentence extraction. *Journal of Computational Technologies*, 12, 5–15.
- [ALL-02] ALLAOUI, R. (2002). Généralisation des opérateurs récurrents hyperbolique et de Derache aux fonctions multivariables pour la classification automatique des données multidimensionnelles. Application à la classification des images couleur (Thèse de Doctorat), Université Ibn Tofaïl, Kenitra.
- [ALR-12] Alrashedi, M., Sbihi, M., Touahni, R., Moussa, A. et Sbihi, A. (2012). Mode Detection in Cluster Analysis using the EM and ICM Algorithms, *European Journal of Scientific Research* ISSN 1450-216X Vol.74 No.3, 456-468
- [AND-73] Anderberg, M. R. (1973), *Cluster Analysis for Applications*, New York: Academic Press.
- [AUR-98] Hammer, A. F. et Klein, R. (1998). Voronoi diagrams. North-Holland, Amsterdam. In J.-R. Sack and J. Urrutia, editors, *Handbook of Computational Geometry*, Elsevier Science Publishers B.V.
- [BAG-11] Bagirov, A. M., Ugon, J., Webb, D., Bagirov, A.M. et al. (2011). Fast modified global k-means algorithm for incremental cluster construction. *Pattern Recognition* 44, 866–876.
- [BAGC-11] Bağcı, U., Yao, J., Caban, J., Turkbey, E., Aras, O. et Mollura, D. J. (2011). A Graph-Theoretic Approach for Segmentation of PET Images. 33rd Annual International Conference of the IEEE EMBS Boston, Massachusetts USA. 8479-8482.
- [BAI-13] Baia, L., Lianga, J., Dang, C., Cao, F. (2013). A novel fuzzy clustering algorithm with between-cluster information for categorical data. *Fuzzy Sets and Systems*, Volume 215. 55–73.
- [BAL-96] (Sundar) Balakrishnan, P.V., Cooper, M. C., Jacob, V. S., Lewis, P. A. (1996). Comparative performance of the FSCL neural net and K-means algorithm for market segmentation. *European Journal of Operational Research*, Volume 93, Issue 2. 346-357.
- [BAN-77] Banon, G. (1977). Estimation Non Paramétrique de Densité de Probabilité pour le Processus de Markov. Thèse de Doctorat d'Etat, Université. P. Sabatier, Toulouse.
- [BAR-08] N. Barioni, M. C., L. Razente, H., J.M. Traina, A., Traina Jr., C. (2008). Accelerating k-medoid-based algorithms through metric access methods. *The Journal of Systems and Software* 81. 343–355.
- [BARA-99] Baraldi, A. et Blonda, P. (1999). A Survey of Fuzzy Clustering Algorithms for Pattern Recognition—Part I and II. *IEEE Transactions On Systems, Man, And Cybernetics—Part B: Cybernetics*, 29(26), 778-801.

- [BAS-11] Basu, S., Zaidi, H., Holm, S., Alavi, A. (2011). Quantitative techniques in PET-CT imaging. *Curr. Med. Imaging Rev.* 7 (3). 216–233.
- [BCA-16] Blaiotta, C., Cardoso, M. J., Ashburner, J. (2016). Variational inference for medical image segmentation, *Computer Vision and Image Understanding.* 81-92.
- [BEL-10] Belhassen, S. et Zaidi, H. (2010). A novel fuzzy C-means algorithm for unsupervised heterogeneous tumor quantification in PET. *Med. Phys.* Vol. 37, 1309-1324.
- [BER-14] Bernábe-Loranca, B., Gonzalez-Velázquez, R., Olivares-Benítez, E. J. Ruiz-Vanoye et Martínez-Flores, J. (2014). Extensions to K-Medoids with Balance Restrictions over the Cardinality of the Partitions. *Journal of Applied Research and Technology* Vol.12. 396-408.
- [BEY-11] Beyer, T., Czernin, J. et Freudenberg, L. S. (2011). Variations in Clinical PET/CT Operations: Results of an International Survey of Active PET/CT Users. *J Nucl Med* February 1, vol. 52 no. 2, 303-310.
- [BEZ-73] Bezdek, J. C. (1973). Fuzzy mathematics in pattern classification (dissertation). Cornell University. New York, USA.
- [BEZ-81] Bezdek, J. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, USA.
- [BI-14] Bi, S. B., Chen, D. Q., Yan, J. et Guo, Y. (2014). Planar Delaunay triangulation algorithm based on 2D convex hull. *Computer Science*, 41(10), 317–320.
- [BIE-99] Biela, P. (1999). *Classification Automatique D'observations Multidimensionnelles Par Reseaux De Neurons Competitifs (Thèse de doctorat)*. L'universite des Sciences et Technologies de Lille.
- [BIL-15] Bilgel, M., Carass, A., Resnick, S. M., Wong, D. F. et Prince, J. L. (2015). Deformation field correction for spatial normalization of PET images. *NeuroImage*, 119, 152–163.
- [BJE-14] Bjerknes, T. L., Moser, E. I. et Moser, M-B. (2014). Representation of Geometric Borders in the Developing Rat. *Report Neuron*, 82, 71–78.
- [BLA-95] Blackmore, J. (1995). *Visualizing High-dimensional Structure with the incremental Grid Growing Neural Network (Technical report)*. Departement of computer sciences of the university of Texas, Austin, USA.
- [BRO-88] Broomhead, D. S. et Lowe, D. (1988). Radial basis functions, multi-variable functional interpolation and adaptive networks (Technical report). RSRE, 4148.
- [BON-12] Bontemps X. et al., (2012). Comparaison de différentes méthodes de segmentation en TEP/TDM pour le ciblage en radiothérapie des carcinomes épidermoïdes des VADS. *Médecine Nucléaire*, 36, 303–312.
- [BOS-93] Bose N. K. et Garga, A. K. (1993). Neural Network Design Using Voronoi Diagrams, *IEEE Transactions On Neural Networks*, Vol. 4, No. 5. 778–787.
- [BOU-06] Boubou, M. (2006). *Contribution aux Classification non Supervisée via des Approches Prétopologiques Et d'Aggrégation d'Opinions (Thèse)*. Université Claude Bernard – Lyon I.

- [BOUG-15] Bouguettaya, A., Yu, Q., Liu, X., Zhou, X. et Song, A. (2015). Efficient agglomerative hierarchical clustering. *Expert Systems with Applications*, Volume 42, Issue 5, 2785–2797.
- [BUD-06] Budura, G., Botoca C. et Miclau, N. (2006). Competitive Learning Algorithms for Data Clustering. *FACTA UNIVERSITATIS (NIS) SER.: ELEC. ENERG.* vol. 19, no. 2, 261-269.
- [BUS-06] Busin. L. (2006). Sélection automatique d'espaces couleur pour la segmentation d'images (Thèse de Doctorat). Université des Sciences et Technologies de Lille.
- [BUV-11] Buvat, I. (2011). Outils de quantification pour l'imagerie TEP au FDG. *Innovations technologiques en Médecine Nucléaire*, <http://www.guillemet.org/irene/>.
- [CAB-12] Cabanes, G., Bennani, Y. et Fresneau, D. (2012). Enriched topological learning for cluster detection and visualization. *Neural Networks*, 32, 186–195.
- [CAR-12] Cardot, H., Cénac, P. et Monnez, J-M. (2012). A fast and recursive algorithm for clustering large datasets with k-medians. *Computational Statistics and Data Analysis*, 56, 1434–1449.
- [CEL-91] Celeux G. et Govaert, G. (1991). A classification EM algorithm for clustering and two stochastic versions. *Computational Statistics Quarterly*, vol. 2, no 1, 73–82.
- [CHA-16] Chandra, B. et Sharma, R. K. (2016). Deep Learning with Adaptive Learning Rate using Laplacian Score. *Expert Systems with Applications*, 63, 1–7.
- [CHAN-15] Chan T. M. et Tsakalidis, K.. (2015). Optimal Deterministic Algorithms for 2-d and 3-d Shallow Cuttings. *Proceedings of the 31st Annual Symposium on Computational Geometry (SoCG-15)*, 719–732.
- [CHAS-91] Chassery, J-M. et Melkemi, M. (1991). Diagramme de Voronoi appliqué à la segmentation d'images et à la détection d'évènements en imagerie multi-sources. *Traitement du Signal*, Vol. 8, N° 3, 155-164.
- [CHAU-14] Chaudhary, V., Bhatia, R.S. et Ahlawat, A. K. (2014). A novel Self-Organizing Map (SOM) learning algorithm with nearest and farthest neurons, *Alexandria Engineering Journal*, 53, 827–831.
- [CHE-02] Cheung, Y.M. (2002). Rival penalization controlled competitive learning for data clustering with unknown cluster number, in: *Proc. 9th Int. Conf. Neural Inf. Process*, Singapore.
- [CHEN-06] Cheng, G., Song, Z., Yang, J. et Gao, R. (2006). On Growing Self - Organizing Neural Networks without Fixed Dimensionality. *CIMCA, Computational Intelligence for Modelling, Control and Automation*, pp. 164, doi:10.1109/CIMCA.2006.158.
- [CLE-02] Clément A., (2002). Algorithmes et outils informatiques pour l'analyse d'images couleur application à l'étude de coupes histologiques de baies de raisin en microscopie optique (Thèse). université d'Angers, France.

- [CHO-11] Chouaib, H. (2011). Sélection de caractéristiques: méthodes et applications (Thèse), Université Paris Descartes, France.
- [CHR-14] Charrad, M. et Ghazzali, N. (2014). Véronique Boiteau, Azam Niknafs, NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, Vol 6(6). 1-36.
- [COL-15] Coleca, F. et State, A. (2015). Sascha Klement, Erhardt Barth, Thomas Martinetz, Self-organizing maps for hand and full body tracking, *Neurocomputing* 147, 174–184.
- [DAR-91] Darken, C. et Moody, J. (1991). Note on Learning Rate Schedules for Stochastic Optimization. *Advances in Neural Information Processing Systems* 3, Morgan Kaufmann Publishers.
- [DEL-34] Delaunay, B. (1934). Sur la Sphère vide. *Bulletin of the Academy of Sciences USSR*, VII, 793-800.
- [DEM-77] Dempster, A.P., Laird, N.M. et Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B* 39, 1–38.
- [DEMA-16] De Marsico, M., Petrosino, A. et Ricciardi, S. (2016). Iris Recognition through Machine Learning Techniques: a Survey. *Pattern Recognition Letters*, DOI: 10.1016/j.patrec.2016.02.001.
- [DES-88] DESIENO, D. (1988). Adding a Conscience to Competitive Learning Proceedings. *IEEE International Conference on Neural Networks*, San Diego, CA. 117-124.
- [DEV-82] Devijver, P.A. et Kittler, J. (1982). *Pattern Recognition: A statistical Approach*. Prentice-Hall, London, 109-119.
- [DHA-15] Dhanachandra, N., Mangle, K. et Jina Chanu, Y. (2015). Image Segmentation using K-means Clustering Algorithm and Subtractive Clustering Algorithm. *Procedia Computer Science*, 54, 764 – 77.
- [DID-86] Diday E. (1986). Une représentation visuelle des classes empiétantes : les pyramides, *Revue RAIRO APII*, 20(5), 475-526.
- [DUD-73] Duda, R. et Hart, P. E. (1973). *Pattern Classification and Scene Analysis*. Wiley New-York.
- [DUN, 1973] Dunn, J. C. (1973). A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of cybernetics*, 3, 32–57.
- [DU-10] Du, K.-L. (2010). Clustering: A neural network approach, *Neural Networks* 23, 89-107.
- [DWY-91] Dwyer, R. A. (1991). Higher-dimensional voronoi diagrams in linear expected time. *Discrete & Computational Geometry*, 6, 343–367.
- [EDD-07a] Eddarouich, S. et Sbihi, A. (2007). Neural Network for Modes Detection in Pattern Classification”, *ICTIS'07. Information and Communication Technologies International Symposium*, Morocco, Fez, 300-303.
- [EDD-07b] Eddarouich, S. (2007). Approche neuronale pour la detection des modes en classification automatique non supervisee (Thèse). Université Ibn Tofail Faculté des Sciences Kenitra.

- [ELG-15] Elgamel, M. S. et Dandoush, A. (2015). A modified Manhattan distance with application for localization algorithms in ad-hoc WSNs, *Ad Hoc Networks*, Volume 33, 168–189.
- [ELM-13] El Merabet, Y. (2013). Segmentation d'images couleur par combinaison LPE-régions/LPE-contours et fusion de régions -Application à la segmentation de toitures à partir d'orthophotoplans (Thèse). L'Université de Technologie de Belfort-Montbéliard.
- [ESS-97] Esson, F. (1997). Un Logiciel de Visualisation et de classification Interactives de Données Quantitatives Multidimensionnelles (Thèse). Université de sciences et de technologies de Lille. France. 6-9.
- [EST-00] Estivill-Castro, V. et Yang, J. A. (2000). Fast and robust general purpose clustering algorithm. *Pacific Rim International Conference on Artificial Intelligence*, 208-218,.
- [FAG-08] Esuli, A., Fagni, T. et Sebastiani, F. (2008). Boosting multi-label hierarchical text categorization. *Information Retrieval*, 11(4), 287–313.
- [FAR-13] Farahat, A.K., Ghodsi, A. et Kamel, M.S., (2013). Efficient greedy feature selection for unsupervised learning. *Knowl. Inf. Syst.* 35, 285–310.
- [FIH-15] Silva Filho, T. M., Pimentel, B. A. Souza, R. M.C.R., Oliveira, A. L.I. (2015). Hybrid methods for fuzzy clustering based on fuzzy c-means and improved particle swarm optimization. *Expert Systems with Applications*, Volume 42, Issues 17–18, 6315–6328.
- [FIL-08] Filippone, M., Camastra, F., Masulli, F., et Rovetta, S. (2008). A survey of kernel and spectral methods for clustering. *Pattern Recognition*, 41, 176–190.
- [FOL-72] Foley, D. (1972). Considerations of Sample and Feature Size. *IEEE TRANSACTIONS ON INFORMATION THEORY IT*, Vol(18), No. 5, 618 – 626.
- [FON-01] FONTAINE. M. (2001). Segmentation non supervisée d'images couleur par analyse de la connexité des pixels (Thèse de Doctorat). Université de Lille 1, France.
- [FOR-65] Forgy, E. W. (1965). Cluster Analysis of Multivariate data: efficiency vs. Interpretability of Classifications. *Biometrics*, Vol. 21, 768-769.
- [FORT-87] Fortune, S. (1987). A sweepline algorithm for voronoi diagrams. *Algorithmica*, 2153–174.
- [FOS-13] Foster, B., Bagci, U., Luna, B., Dey, B., Bishai, W., Jain, S., ... Xu, Z. (2013). Robust segmentation and accurate target definition for positron emission tomography images using affinity propagation. *IEEE , ISBI*, 1461–1464. DOI. 10.1109/ISBI.2013.6556810
- [FOS-14a] Foster, B., Bagci, U., Xu, Z., Dey, B., Luna, B., Bishai, W., ... J. Mollura, D. (2014a). Segmentation of PET Images for Computer-Aided Functional Quantification of Tuberculosis in Small Animal Models. *IEEE Transactions On Biomedical Engineering*, VOL. 61, NO. 3, 711-724.
- [FOS-14b] Foster, B., Bagci, U., Mansoor, A., Xu, Z., et J. Mollura, D. (2014b). A review on segmentation of positron emission tomography images. *Computers in Biology and Medicine*, VOL 50, 76–96.

- [FRI-95] Fritzsche, B. (1995). A growing neural gas network learns topologies. In G. Tesauro, D. S. Touretzky, & T. K. Leen (Eds.), *Advances in neural information processing systems: 7*. Cambridge, MA: MIT Press, 625-632.
- [FRI-97] Fritzsche, B. (1997). Some Competitive learning methods (Draft). Systems Biophysics, Institute for neural computation, Université de la Ruhr Bochum.
- [FUKU-72] Fukunaga, K. (1972). *Introduction to Statistical Pattern Recognition*. Academic Press, New-York.
- [GEE-07] Geets, X., Lee, J.A., Bol, A., Lonneux, M. et Grégoire, V. (2007). A gradient-based method for segmenting FDG-PET images: methodology and validation. *Eur. J. Nucl. Med. Mol. Imaging*, 34 (9), 1427–1438.
- [GEV-99] Gevers, T. et Smeulders, A. (1999). Color-based object recognition. *Pattern Recognition*, vol. 32, 453–464.
- [GHA-12] Ghanimi, F., Sbihi, A., Sbihi, M. et Touahni, R. (2012). Cluster Analysis by Binary Pretopology. *European Journal of Scientific Research*, Volume 74 Issue 3, 337-349.
- [GRA-06] Grady, L. (2006). Random Walks for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 11, 1768–1783.
- [GRAG-15] Gragnaniello, D., Poggi, G., Sansone, C. et Verdoliva, L. (2015). Using iris and sclera for detection and classification of contact lenses, *Pattern Recognition Letters* 000, 1–7.
- [GRE-78] Green, P. J. et Sibson, R. (1978). Computing dirichlet tessellation in the plane, *Computer Journal*, vol. 21, no 2, 168-173.
- [GRO-82] Grossberg, S. (1982). *Studies of Mind and Brain*. Eds Reidel.
- [HAL-01] Halkidi, M., Batistakis, Y. et Vazirgiannis, M. (2001). On Clustering Validation Techniques. *Journal of Intelligent Information Systems*, 17:2/3, 107–145.
- [HAM-06] Hammouche, K., Diaf, M. et Postaire, J.-G. (2006). Aclustering method based on multidimensional texture analysis, *Pattern Recognition*, Volume 39, Issue 7, 1265–1277.
- [HAM-09] Hammouche, K. et Postaire, J.-G. (2009). Multidimensional Texture Analysis for Unsupervised Pattern Classification (Chapitre). In book: *Pattern Recognition: Techniques, Technology and Applications*, Publisher: I-Tech book, Editors: Peng-Yeng Yin, 163-196.
- [HAMI-05] Hamid, A. Allaoui, R. et Sbihi, A. (2005). A New Unsupervised Color Image Segmentation Algorithm upon Statistical Multidimensional Data Analysis Approach, *ICGST-GVIP Journal*, Volume 5, Issue4, 33- 40.
- [HAMI-09] Hamid, A. (2009). Approche locale non paramétrique de classification des Images couleur à base de l'opérateur hyperbolique (Thèse de Doctorat). Univ. Mohammed V, Rabat.
- [HAMS-15] Hamse Y. Mussa, John B.O. Mitchell Avid M. Afzal, (2015). The Parzen Window method: In terms of two vectors and one matrix. *Pattern Recognition Letters* 63, 30–35.

- [HAN-07] Haniched, A. et Khenfer, N. (2007). Un Système Efficace D'identification Des Personnes Par Iris, 4th International Conference on Computer Integrated Manufacturing CIP'2007.
- [HAV-13] Háva, O., Skrbek, M. et Kordík, P. (2013). Supervised two-step feature extraction for structured representation of text data. *Simulation Modelling Practice and Theory*, Vol 33, 132–143.
- [HAP-09] Hapfelmeier, A., Schmidt, J., Muller, M., et Kramer, S. (2009). Interpreting PET scans by structured Patient Data: A Datamining case study in dementia Research. *IEEE Knowledge and Information Systems*, 213-222.
- [HEB-49] Hebb, D. (1949). *Organisation of Behavior*. Wiley, New York.
- [HER-14] Hernández, J., Sucar, L.E. et Morales, E.F. (2014). Multidimensional hierarchical classification, *Expert Systems with Applications*, 41, 7671–7677.
- [HIL-13] HILAL, A. (2013). *Système d'identification à partir de l'image d'iris et détermination de la localisation des informations* (Thèse de doctorat). Université de technologie de Troyes, France.
- [HIR-06] Hirohisa, A., Mochiduki, N. et Yamada, H. (2006). A Model For Detecting Cost-prone Classes Based on Mahalanobis-taguchi Method. *IEICE Transactions on Information and Systems* 89 - D, No. 4, 1347-1358. DOI : 10.1093/ietisy/e89-d.4.1347
- [HOR-75] Horowitz, S. et Pavlidis T. (1975). *Picture segmentation by a directed split-and-merge procedure* (Rapport de Recherche). Departement of Electrical Engineering, Princeton University, USA.
- [HOP-07] Hopfield, J. J. (1977). Hopfield network. *Scholarpedia*, 2 (5). 2007.
- [HOP-82] Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences* 79, 2554–2558.
- [HUA-98] Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge discovery*, vol. 2, 283-304.
- [ISM-11] Ismail T. S. E.-D., Ismail, A., Ramadan, M. A. et Samak, A. (2011). H. An efficient offline signature identification method based on fourier descriptor and chain codes. *International Journal of Biomedical Engineering and Technology* 5, 1–10.
- [IZE-08] Izenman, A. J. (2008). *Modern multivariate statistical techniques: regression, classification, and manifold learning*. Series: Springer Texts in Statistics, XXV, 733. Springer.
- [JAI-04] Jain, A.K., Ross, A. et Prabhakar, S. (2004). An introduction to biometric recognition. *IEEE Trans. on Circuits and Systems for Video Technology*, 14(1), 4–20.
- [JIA-10] Jiang, J., Trundle, P. et Ren, J. (2010). Medical image analysis with artificial neural networks, *Computerized Medical Imaging and Graphics*, 34, 617–631.
- [KAM-15] Kaminski, M., Orłowska-Kowalska, T. (2015). An on-line trained neural controller with a fuzzy learning rate of the Levenberg–Marquardt algorithm for speed control of an electrical drive with an elastic joint. *Applied Soft Computing*, Volume 32, July, 509–517.

- [KAN-02] Kanungo, T., Mount, D., Netanyahu, N., Piatko, C., Silverman, R., et Wu, A. (2002). An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, 881–892.
- [KAS-88] Kass, M., Witkin, A. et Terzopoulos, D. (1988). Snakes: active contour models. *Int. J. Comput. Vis.* 1 (4), 321–331.
- [KAU 90] Kaufman L., Rousseeuw P.J. (1990). Finding groups in data: an introduction to cluster analysis. John Wiley & sons.
- [KER-09] Kerhet, A., Small, C., Quon, H., Riauka, T., Greiner, R., McEwan, A. et Roa, W. (2009). Segmentation of lung tumours in positron emission tomography scans: a machine learning approach. *Artif. Intell. Med.* 146–155.
- [KIA-01] Kiang, M. Y. (2001). Extending the Kohonen self-organizing map networks for clustering analysis. *Computational Statistics & Data Analysis* 38, 161–180.
- [KIH-08] Kihato, P. K. Tokutaka, H. Ohkita, M. et al. (2008). Spherical and torus SOM approaches to metabolic syndrome evaluation. in *Proceedings of the ICONIP*, vol. 4985 of *Lecture Notes in Computer Science (LNCS)*, Springer, Heidelberg, Germany, 274–284.
- [KLA-15] Klawonn, F., Kruse, R. et Winkler, R. (2015). Fuzzy Clustering: More than just Fuzzification, *Fuzzy Sets and Systems*, 00, 1–8.
- [KNU-73] Knuth, D. E. (1973). *The art of computer programming*, Volume III: Sorting and searching. Reading, MA: Addison-Wesley.
- [KOH-82] Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43-59,
- [KOH-95] Kohonen, T. (1995). *Self-Organizing Maps*. Springer, Berlin.
- [KOH-02] Kohonen, T. et Somervuo, P. (2002). How to make large Self-Organizing Maps for Nonvectorial Data. *Neural Networks*, 15(8-9), 945-952.
- [KRI-08] Kriegel, H-P., Schubert, M. et Zimek, A. (2008). Angle-based outlier detection in high-dimensional data. *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, Las Vegas, 444-452.
- [KRIC-07] Krichen, E. (2007). *Reconnaissance des personnes par l’iris en mode dégradé (Thèse de Doctorat)*. Institut National des Télécommunications, France.
- [KUR-15] Kuriscek, E., Marsalek, P., Stroffek, J. et Toth, P. G. (2015). Biological context of Hebb learning in artificial neural networks (revue). *Neurocomputing*, Volume 152, 27–35.
- [LAI-11] Lai, Jim Z.C. et Huang, T-J. (2011). An agglomerative clustering algorithm using a dynamic k-nearest-neighbor list. *Information Sciences*, Volume 181, Issue 9, 1722–1734.
- [LAM-00] Lambert, P. et Macaire, L. (2000). Filtering and Segmentation: the Specificity of Color Images. *CGIP’2000 1st Int. Conf. on Color Image Process.*, St-Etienne, France, 57-64.

- [LAM-15] Lamberts, K. (2015). Connectionist Models of Concept Learning. *International Encyclopedia of the Social & Behavioral Sciences*, 2nd edition, Volume 4, 654–657.
- [LAR-03] Lardinois, D., Weder, W., Hany, T., Kamel, E., Korom, S., Seifert B., von Schulthess, G. et Steinert, H. (2003). Staging of non-small-cell lung cancer with integrated positron-emission tomography and computed tomography, *New Engl. J. Med.* 348 (25), 2500–2507.
- [LI-13] Li, J., Li, D. R. et Shao, Z. F. (2013). A streaming data Delaunay triangulation algorithm based on parallel computing. *Geomatics and Information Science of Wuhan University*, 38(7), 794–798.
- [LIK-03] Likasa, A., Vlassisb, N. et Verbeekb, J. (2003). The global k-means clustering algorithm. *Pattern Recognition* 36, 451–461.
- [LIU-06] Liu, Y., Zheng, Y. F., (2006). FS_SFS: a novel feature selection method for support vector machines. *Pattern Recognit.* 39, 1333–1345.
- [LIU-13] Liu, H., Kurihara, M., Oyama, S. et Sato, H. (2013). An incremental self-organizing neural network based on enhanced competitive Hebbian learning. *Neural Networks (IJCNN)*, IEEE, ISSN : 2161-4393, INSPEC Accession Number: 14027375, Dallas, TX, 1–8.
- [LL-90] Lim Y. W. et Lee, S. U. (1990). On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques. *Pattern Recognition*, 23(9), 935–952.
- [LO-12] Lo, S. H. (2012). Parallel Delaunay triangulation—application to two dimensions. *Finite Elements in Analysis and Design*, 62, 37–48.
- [LOZ-98] Lozano. V. (1998). Contribution de l’analyse d’images couleur au traitement des images textile (Thèse de doctorat). Université Jean Monnet de Saint-Etienne, France.
- [LPR-14] López-Rubio, E. (2014). A histogram transform for probability density function estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 36 (4), 644–656.
- [LPR-15] López-Rubio, E. et Muñoz-Pérez, J. (2015). Probability density function estimation with the frequency polygon transform, *Information Sciences* 298, 136–158
- [LU-16] Lu, Y., Hou, X. et Chen, X. (2016). A novel travel-time based similarity measure for hierarchical clustering. *Neurocomputing*, *Neurocomputing* 173, 3–8.
- [LUO-90] Luong, Q-T. (1990). La couleur en vision par ordinateur : 1-Une revue. Rapport de recherche INRIA n°1251.
- [LWE-15] Li, C., Wang, X., Eberl, S. et al., (2015). Supervised variational model with statistical inference and its application in medical image segmentation, *IEEE Trans. Biomed. Eng.* 62 (1), 196–207.
- [MAC-67] MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations, *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, University of California Press, 1, 281-297.

- [MACE-04] Macaire, L. (2004). Exploitation De La Couleur Pour La Segmentation Et L'analyse D'images (Mémoire). Habilitation à Diriger des Recherches, Université des Sciences et Technologies de Lille (USTL), France.
- [MAH-36] Mahalanobis, P. C. (1936). On the generalised distance in statistics », Proceedings of the National Institute of Sciences of India, vol. 2, no 1, 49–55.
- [MAL-09] Aliguliyev, R. M. (2009). Performance evaluation of density-based clustering methods. *Information Sciences* 179, 3583–3602.
- [MAO-94] Mao, J. (1994). Design and Analysis Networks for Pattern Recognition (Thèse de Doctorat). Michigan State University, USA.
- [MAR-91] Martinetz, T. et Schulten, K. J. (1991). A 'neural gas' network learns topologies. T. Kohonen, K. Mäkisara, O. Simula, and J. Kangas, editors, *Artificial Neural Networks*, North Holland, Amsterdam, 397-402.
- [MAR-93] Martinetz, T.M. (1993). Competitive Hebbian learning rule forms perfectly topology preserving maps. Proceedings of the (ICANN), Gielen S. and Kappen B. (eds), Springer, Heidelberg, 427-434.
- [MAR-94] Martinetz, T., et schulten K. (1994). Topology Representing Networks. *Neural Networks*. Vol. 7, No 3, 507-522.
- [MAS-05] Masson, M-H. (2005). Apports de la théorie des possibilités et des fonctions de croyance à l'analyse de données imprécise (Mémoire). Habilitation à Diriger des Recherches, Université de Technologie Compiègne, France.
- [MCC-43] McCulloch, W. et Pitts, W. (1943). A logical calculus of the ideas immanent in Nervous Activity". *Bulletin of Mathematical Biophysics*, vol. 5, 115-137.
- [MER-13] Meena, A. et Raja, K. (2013). A Study on Functional Brain Metabolism using PET Scan Image Datasets – An Analysis. *Journal of Computer Applications (JCA)* ISSN: 0974-1925, Volume VI, Issue 2, 28-30.
- [MER-12] Meena, A. et Raja, K. (2012). Segmentation Of Alzheimer's Disease In PET Scan Datasets Using Matlab. *International Journal of Computer Applications Journal* ,Volume 57(10), 15-19.
- [MEI-14] Mei, J-P., Chen, L. (2014). Proximity-based k-partitions clustering with ranking for document categorization and analysis. *Expert Systems with Applications*, Volume 41, Issue 16, 7095–7105.
- [MOE-14] Morales-Esteban, A., Martínez-Álvarez, F., Scitovski, S. et Scitovski. R. (2014). A fast partitioning algorithm using adaptive Mahalanobis clustering with application to seismic zoning. *Computers & Geosciences*, 73, 132–141.
- [MOK-11] MonirulKabir, Md., Shahjahan, Md. et Murase, K. 2011. A new local search based hybrid genetic algorithm for feature selection. *Neurocomputing*, 74, 2914–2928.
- [MOU-08] Moussa, A., Sbihi, A. et Postaire, J-G. (2008). A Markov random field model for mode detection in cluster analysis. *Pattern Recognition Letters*, 29, 1197–1207.

- [MOR-15] Moradi, P. et Rostami, M. (2015). A graph theoretic approach for unsupervised feature selection, *Engineering Applications of Artificial Intelligence*, 44, 33–45.
- [MUT-12a] Muthanna, H. A., Touahni, R., Sbihi, A., Eddarouich, S. et Messoussi, R. (2012). Unsupervised color image segmentation using new neural-morphological procedure. *GJCAT*, Vol 2 (1), 929-936.
- [MUT-12b] Muthanna H. A., (2012). Association de la Morphologie Mathématique aux Réseaux de Neurones Artificiels pour la Classification non Supervisée des Données Multidimensionnelles : Application à la Segmentation d’Images Couleur (Thèse de Doctorat). Université Ibn Tofail Faculte des sciences Kenitra, Maroc.
- [NIE-12] Nielsen, F. (2012). k-MLE: A fast algorithm for learning statistical mixture models. *IEEE, arxiv ICASSP’2012*, 869-872. DOI: 10.1109/ICASSP.2012.6288022
- [OHT-80] Ohta Y-I., Kanade T. et Sakai T. (1980). Color Information for regions segmentation. *Computer Graphics and Image Processing*, Vol. 13, 222-241.
- [OTH-16] Othman, N. (2016). Fusion techniques for iris recognition in degraded sequences (Thèse de doctorat). Image Processing. Université Paris-Saclay, France.
- [PAL-93] Pal, N.R., Bezdek, J. et Tsao, E.C.-K. (1993). Generalized Clustering Networks and Kohonen’s Self Organizing scheme. *I.E.E.E. Trans. On Neural Networks*, vol. N°4, 549-557.
- [PAN-14] Zaihu Pang, Shikui Tu, Xihong Wu, Lei Xu, comparative study of RPCL and MCE based discriminative training methods for LVCSR, *Neurocomputing* 134 (2014) 53–59.
- [PAR-62] Parzen, E. (1962). On the estimation of a probability density and the mode. *Ann. Math. Statist.*, 33, 1965–1976.
- [PARK_91] Park, J. et I. Sandberg (1991). Universal approximation using radial-basis-function networks. *Neural Computation* 3(2), 246–257.
- [PEN-08] Pena, M., Barbakh, W. et Fyfe, C. (2008). Topology-Preserving Mappings for Data Visualisation (Chapitre). Dans N. A Gorban. (dir.), *Lecture Notes in Computational Science and Enginee (LNCSE, Vol 120, 131-150)*, UK: Springer.
- [POS-89] Postaire, J-G. et Touzani, A. (1989). Mode boundary detection by relaxation for cluster analysis. *Pattern Recognition*, 22(5), 477-489.
- [POS-82] Postaire J.-G. et Vasseur C.-P.-A. (1982). A Fast Algorithm for Non Parametric Probability Density Function Estimation. *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 4, n° 6, 663-666.
- [QUO-15] Quost, B. et Denoeux, T. (2015). Clustering and classification of fuzzy data using the fuzzy EMalgorithm, *Fuzzy Sets and Systems*, 286, 134–156.
- [RAM-09] Aliguliyev, R. M. (2009). Clustering of document collection – A weighting approach. *Expert Systems with Applications*, 36, 7904–7916.

- [RAZ-13] Razavi Zadegan, S. M., Mirzaie, M. et Ranked S. F. (2013). k-medoids: A fast and accurate rank-based partitioning algorithm for clustering large datasets. *Knowledge-Based Systems*, 39, 133–143.
- [REZ-98] Rezaee, R., Lelieveldt, B.P.F., et Reiber, J.H.C. (1998). A New Cluster Validity Index for the Fuzzy c-Mean. *Pattern Recognition Letters*, 19, 237–246.
- [RIC-15] Richter, R. et Alexa, M. (2015). Mahalanobis centroidal Voronoi tessellations. *Computers & Graphics*, 4, 648–54.
- [ROS-58] Rosenblatt, F. (1958). The Perceptron : probabilistic model for information storage and organization in the brain. *Psychological Review*, 65, 386-408.
- [ROU-06] ROUX, M. (1985). *Algorithmes de Classification (Ouvrage)*. Université Paul Cézanne Marseille, version corrigée en juin 2006, France.
- [ROV-99] Rovetta, S., et Zunino, R. (1999). Efficient training of neural gas vector quantizers with analog circuit implementation. *IEEE Transactions on Circuits and Systems II*, 46(6), 688-698.
- [ROW-00] Roweis, S. T. et Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding, *SCIENCE*, 290, 2323-2326.
- [SAE-10] Saeed, S. M., Abbod, M., Amira, A. et Zaidi, H. (2010). Artificial neural network-based system for PET volume segmentation, *Int. J. Biomed. Imaging Volume 2010*, Article ID 105610, p. 11. <http://dx.doi.org/10.1155/2010/105610>
- [SAR-11] Sarlin, P. et Eklund, T. Fuzzy Clustering of the Self-Organizing Map(Chapitre). Dans P. Sarlin, T. Eklund (ind.). *Some Applications on Financial Time Series: Advances in Self-Organizing Maps (Volume 6731 of the series Lecture Notes in Computer Science, 40-50)*. Berlin/Heidelberg, Germany: Springer.
- [SBI-05] Sbihi, M., Moussa, A., Postaire, J-G. Sbihi, A. (2005). Approche Arkovienne pour la Classification Automatique non Supervisée de Données Multidimensionnelles. *Journal européen des systèmes automatisés*, 39 (9-10), 1133-1154.
- [SEN-09] Senthilkumaran, N. et Rajesh, R. (2009). Edge Detection Techniques for Image Segmentation – A Survey of Soft Computing Approaches. *International Journal of Recent Trends in Engineering*, Vol. 1, No. 2, 250-254.
- [SHA-15] Shatkay, H., Brady, S. et Wong, A. (2015). Text as data: Using text-based features for proteins representation and for computational prediction of their characteristics. *PubMed, MEDLINE*, 74, 54–64.
- [SHAM-75] Shamos, M. I. et Hoey, D. (1975). Closest-point problems », *Proceeding of 16th Annual IEEE Symposium on Foundations of Computer Science*, Los Angeles, IEEE Computer Society Press, 151-162.
- [SHAR-09] Shah, S. et Ross, A. (2009). Iris segmentation using geodesic active contours. *IEEE Transactions on Information Forensics Security*, Vol. 4, No. 4, 824-836.
- [SHI-00] Shi, J. et Malik J. (2000). Normalized Cuts and Image Segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 22, No. 8, 888-905.

- [SHL-90] Shiun Lin, C., Smith D. C. P. et Smith. J. M. (1976). The design of a rotating associative memory for relational database applications. *ACM Transactions on Database Systems*, Vol1(1), 53-65.
- [SIG-99] Sigelle M. et Tupin, F. (1999). Champs markoviens en Traitement d'Image (cours). Département Traitement de Signal et des Images, Université Mohammed V-ENST- Rabat, Maroc.
- [SKO-01] Skodras, A. Christopoulos C. et Ebrahimi, T. (2001). The jpeg 2000 still imageCompression standard. *IEEE Signal Processing Magazine*, 18, 36-58.
- [STU-94] A. Stuart, K. Ord, (1994). *Kendall's Advanced Theory of Statistics: Volume I—Distribution Theory*, Edward Arnold, §8.7.
- [SU-16] Su, T., Wang, W., Lv, Z., Wu, W. et Li, X. (2016). Rapid Delaunay Triangulation for Randomly Distributed Point Cloud Data Using Adaptive Hilbert Curve, *Computers & Graphics*, Vol 54, 65–74.
- [TAG-00] Taguchi, G. et Jugulum, R. (2000). New Trends In Multivariate Diagnosis. *Sankhya: The Indian Journal Of Statistics fuzzy Mathematical Morphology*, 62, 233-248.
- [TAK-95] Takahashi K. et Nakatani, H. (1995). Color Image Segmentation using ISODATA Clustering Method.2' Asian. Conf. on Computer Vision., Singapore, Vol. 1, 523-527.
- [TAN-06] Tan, P-N., Steinbach, M. et Kumar, V. (2006). *Cluster analysis: basic concepts and algorithms (Chapitre)*. In book: *Introduction to Data Mining (1 edition (May 12, 2005), 487-568)*. New York, USA: Pearson.
- [TEN-00] Tenenbaum, J. B., Silva, V. d. et Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500), 2319-2323.
- [THE-99] Theodoridis, S. and Koutroubas, K. (1999). *Pattern Recognition*. Academic Press.
- [THE-08] Theodoridis S, Koutroubas K (2008). *Pattern recognition*, Fourth edition. Academic Press.
- [THO-11] Thomas, B. A., Erlandsson, K., Modat, M., Thurfjell, L., Vandenberghe, R., Ourselin, S. et Hutton, B. F. (2011). The importance of appropriate partial volume correction for PET quantification in Alzheimer's disease. *Eur J Nucl Med Mol Imaging* 38, 1104–1119.
- [TIM-12] Timouyas, M., Eddarouich, S., et Hammouch, A. (2012). A new approach of classification for non-Gaussian distribution upon competitive training, *International IEEE Conference on Complex Systems (ICCS'12)*, Agadir, Morocco, ISBN: 978-1-4799-3824-7, 1-6.
- [TOD-13] Todeschini, R., Ballabio, D., Consonni, V., Sahigara, F. et Filzmoser, P. (2013). Locally centred Mahalanobis distance: A new distance measure with salient features towards outlier detection. *Analytica Chimica Acta*, 787, 1– 9.
- [TOR-12] Torra, V., et Narukawa, Y. (2012). On a comparison between Mahalanobis distance and Choquet integral: The Choquet–Mahalanobis operator. *Information Sciences*, 190, 56–63.
- [TRE-04] Trémeau, A., Fernandez-Maloigne, C. et Bonton, P. (2004). *Image numérique couleur de l'acquisition au traitement (cours)*. Master. Écoles d'ingénieurs, Collection : Sciences Sup, Dunod.

- [UPA-15] Upadhyay, P. et Chhabra, J. K. (2015). Modified Self Organizing Feature Map Neural Network (MSOFM NN) Based Gray Image Segmentation. *Procedia Computer Science*, 54, 671 – 675.
- [VAL-09] Valentini, G. (2009). True path rule hierarchical ensembles. In *Proceedings of the eighth international workshop on multiple classifier systems, MCS'09*, Springer-Verlag, 232–241.
- [VAN-00] Vandenbroucke, N. (2000). Segmentation d'images couleur par classification de pixels dans des espaces d'attributs colorimétriques adaptés. Application à l'analyse d'image de football (Sciences Sup). Université de Lille 1, France.
- [VAR-01] Varela, F., Lachaux, J.-P., Rodriguez, E., et Martinerie, J. (2001). The brainweb :Phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, 2, 229-239.
- [VAT-08] Vatsa, M., Singh, R. et Noore, A. (2008). Improving iris recognition performance using segmentation, quality enhancement, match score fusion, and indexing. *IEEE Transactions on Systems, Man, Cybernetics B*, Vol. 38, No. 4, 1021-1035.
- [VIB-05] Vibert, J-F., Sebille, A., Lavallard-Roussou, M-C., Boureau, F. et Mazières, L. (2005). *Neurophysiologie : de la physiologie à l'exploration fonctionnelle*. 2e édition, Elsevier Masson, Paris, France.
- [VIE-12] Viejo, D., Garcia, J., Cazorla, M., Gil, D. et Johnsson, M. (2012). Using GNG to improve 3D feature extraction—Application to 6DoF egomotion. *Neural Networks*, 32, 138-146.
- [VOR-1908] Voronoi, G. M. (1908). Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième Mémoire: Recherches sur les paralléloèdres primitifs. *J. Reine Angew. Math.*, 134, 198-287.
- [WEI-14] Wang, G-W., Zhang, C-X. et Zhuang, J. (2014). Clustering with Prim's sequential representation of minimum spanning tree. *Applied Mathematics and Computation*, 247, 521–534.
- [WEG-72] WEGWAN, E. J. (1972). Nonparametric Probability Density Estimation. I: A Summary, vol. 17, pp. 533-546.
- Werbos, P. 1974. *Beyond regression: New tools for prediction and analysis in the behavioral sciences,* Harvard University, 8.
- [WID-60] Widrow, C., Hoff, M.E., (1960). Adaptive switching circuits. Institute of Radio Engineers, Western Electronic Show and Convention, Convention Record 4, 96–104.
- [WU-15] Wu, X., Wu, B., Sun, J., Qiu, S. et Li, X. (2015). A hybrid fuzzy K-harmonic means clustering algorithm. *Applied Mathematical Modelling*, Volume 39, Issue 12, 3398–3409.
- [XBF-13] Xu, Z., Bagci, U., Foster, B. et Mollura, D.J. (2013). A hybrid multi-scale approach to automatic airway tree segmentation from CT scans, *IEEE 10th International Symposium on Biomedical Imaging (ISBI)*, 1308–1311.
- [XGH-12] Xu, Z., Gao, Z., Hoffman, E. et Saha, P. (2012). Tensor scale-based anisotropic region growing for segmentation of elongated biological structures, in: *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, 1032–1035.

- [XIA-15] Xia, S-Y., Xiong, Z-Y., Luo, Y-L. Xu, W. and Zhang, G-H. (2015). Effectiveness of the Euclidean distance in high dimensional spaces, *Optik - International Journal for Light and Electron Optics*. Volume 126, Issue 24, 5614-5619. DOI10.1016/j.ijleo.2015.09.093
- [XIE-10] Xie, J. et Jiang, S. (2010). A simple and fast algorithm for global K-means clustering. *Second International Workshop on Education Technology and Computer Science, IEEE*, Volume:2, 36 – 40.
- [XU-93] Xu, L. Krzyzak, A. et Oja, A. E. (1993). Rival penalized competitive learning for clustering analysis. rbf net and curve detection. *IEEE Trans. on Neural Networks*, no. 4, 636–64.
- [XU-07] Xu, L. (2007). Rival penalized competitive learning. *Scholarpedia*, 2 (8), 1810.
- [YAH-16] Yahiaoui, M., Monfrini, E. et Dorizzi, B. (2016). Markov Chains for unsupervised segmentation of degraded NIR iris images for person recognition. *Pattern Recognition Letters* 000, 1–8.
- [YHU-04] Yuan-Hui, Y. et Chang, C-C. (2004). Scenery image segmentation using support vector machines. *Fundamenta Informaticae* 61 (3–4), 379–388.
- [YUD-14] Yudong, Z., Shuihua, W., Preetha, P. et Ji, G. (2014). Binary PSO with mutation operator for feature selection using decision tree applied to spam detection. *Knowl. Based Syst.* 26, 22–31.
- [YYA-12] Yang, H-Y., Wang, X-Y., Wang, Q-Y. et Zhang, X-J. (2012). LS-SVM based image segmentation using color and texture information. *J. Vis. Commun. Image R.* 23, 1095–1112.
- [ZAD-65] Zadeh, L.A. (1965). Fuzzy sets, *Inform. Control* 8, 338–353.
- [ZAH-71] Zahn, C. T. (1971). Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE trans. Comput.* C-20 (Apr.), 68-86.
- [ZAR-12] Zarrouk F. (2012). Étude De La Relation Entre Deux Variables (le coefficient de corrélation), *Cours (9) de statistiques à distance, ISSEP Ksar-Said*.
- [ZAR-12a] Zarrouk F. (2012). Les Statistiques Descriptives, *Cours (6) de statistiques à distance, ISSEP Ksar-Said*.
- [ZEI-12] Zeiler, M. D. (2012). ADADELTA: An adaptive learning rate method. *arXiv preprint arXiv1212.5701*.
- [ZUC-76] Zucker, S. (1976). Region growing: Childhood and adolescence. *Computer Graphics, Image Processing*, Vol 5,382–399.
- [ZUP-01] Zuppiroli, L. et Bussac, M. N. (2001). *Traité des couleurs*. Presses Polytechniques Universitaires Romandes, PPUR (ind.). Collection : Physique - 1e édition, France.