République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université M'hamed BOUGARA de BOUMERDES

Faculté des Sciences

Département d'Informatique

MEMOIRE DE MAGISTER

Spécialité : Système informatique et génie des logiciels

Option : Spécification de Logiciel et Traitement de l'Information

Ecole Doctorale

Présenté par :

Zouaoui Hakima

Thème

Clustering par fusion floue de données appliqué à la segmentation d'images IRM

Devant le jury composé de:

Dr. MEZGHICHE Mohamed	Prof.	U.M.B. Boumerdes	*Président*
Dr. MOUSSAOUI Abdelouahab	M.C.	U.F.A. Sétif	*Rapporteur*
Dr. BABAHENINI Med Chawki	M.C.	U.M.K. Biskra	*Examinateur*
Dr. DJOUADI Yassine	M.C.	U.M.M.Tizi-Ouzou	*Examinateur*

Année Universitaire : 2007/2008

Résumé

Les données traitées en imagerie médicale sont souvent imprécises et/ou incertaines du fait du mode d'acquisition des images ou de la modélisation des connaissances des médecins. Lorsqu'un expert examine une ou plusieurs images médicales, il prend en compte simultanément ses propres connaissances théoriques ainsi que les informations fournies par les images afin d'effectuer son diagnostic. De même, la fusion de données agrège informations numériques et connaissances théoriques et contextuelles afin de fournir une information synthétique pour l'aide au clinicien.

L'objectif de ce mémoire consiste à développer une architecture de fusion de données basée sur la théorie possibiliste pour la segmentation d'une cible à partir de plusieurs sources d'images. Le processus de fusion est décomposé en trois phases fondamentales.

Nous modélisons tout d'abord les informations dans un cadre théorique commun. Le formalisme retenu consiste à faire la coopération entre l'algorithme FCM (C-moyennes floues) dont la contrainte d'appartenance d'un individu à une classe est gérée d'une manière relative et l'algorithme possibiliste PCM (C-means possibilistes) pour les points aberrants .

Nous agrégeons ensuite ces différentes informations par un opérateur de fusion. Celui-ci doit affirmer les redondances, gérer les complémentarités et prendre en compte les conflits soulignant souvent la présence d'une pathologie.

Nous construisons enfin une information synthétique permettant d'exploiter les résultats de la fusion.

Cette architecture développée est mise en oeuvre pour la segmentation des tumeurs cérébrales à partir des images IRM qui comprennent pour l'instant les séquences de base : T1, T2 et densité de protons (DP).

Mots-clés : Data mining, Imagerie médicale, Fusion de données, Segmentation, Cerveau, C-moyennes floues, C-moyennes possibiliste.

Abstract

Information treated in medical imaging is often inaccurate and uncertain due to image acquisition methods or to the modelling of clinician's knowledge. When analysing medical images to make his diagnosis, an expert takes into account his own knowledge as well as the information provided by each image. Data fusion aggregates in the same way numerical, theoretical and contextual knowledge to provide clinicians synthetic information.

The objective of this work consists of developing architecture of information fusion based on the possibilistic theory in order to segment a target from multiple sources of image. The fusion process is divided into three steps:

We first model the available information, numerical or symbolic, in a common theoretical frame. The formalism selected consists in cooperating between the algorithm FCM (Fuzzy C-Means) whose constraint of membership of an individual to a class is relative and the algorithm possibilist PCM (Possibilistic C-means) for the aberrant points.

Then we aggregate these information with a fusion operator. This operator has to affirm redundancy, manage the complementarities and also take into account conflicts that often underline the presence of pathology.

We finally propose a synthetic piece of information that allows to best represent the available data.

This developed architecture is performed for the segmentation of cerebral tumors from MRI images that presently include these routine sequences: T1, T2 and proton density.

Keywords: Data mining, Medical imaging, Tissue characterization, Data fusion, Clustering, Brain, Fuzzy c-means, possibilist c-means.

ملخص

إن المعطيات المعالجة في الصور الطبية غير دقيقة أو(و) غامضة وهذا راجع للصور المكتسبة أو كيفية معالجة الأطباء للمعرفة، عندما يفحص الخبير الصور الطبية يأخذ بعين الاعتبار هذه المعرفة النظرية و كذلك المعلومات المزودة من هذه الصور بهدف تشخيصها، بالإضافة إلى ذلك، المعطيات المحللة هي عبارة عن جمع المعلومات العددية و المعرفة النظرية و السياقية من اجل تزويد الطبيب المختص بالمعلومات التركيبية.

المذكرة تهدف إلى التركيز على تطوير هندسة المعطيات الملتحمة أساسها النظرية الممكنة، انطلاقا من مجموعة مصادر صورية لأجل تقسيم الهدف. معالجة الالتحام تنقسم إلى ثلاثة مراحل أساسية :

في البداية توضع المعلومات في أيطار نظري عام، بحيث الطريقة المتبعة تنسق بين خوارزمية المتوسطات الضبابية التي تحسب درجة الانتماء بطريقة نسبية و خوارزمية المتوسطات الممكنة للأفراد (البيكسال) المبعثرة .

وبعدها نقوم بتجميع مختلف المعلومات بواسطة عامل التجميع ،حيث هذا الأخير يؤكد على التكرار و يدير المتكاملات و يأخذ بعين الاعتبار التنازع الناتج في اغلب الأحيان عن وجود المرض.

و في الأخير نقدم معلومات تركيبية شاملة مستنبطة من خلال استغلالنا لنتائج الالتحام.

هذا التطور الهندسي يوضع من اجل تعيين الورم المخي بواسطة صور الرنين المغناطيسي للدماغ .

كلمات المفتاح: التنقيب عن البيانات، صور الرنين المغناطيسي, التحام المعطيات، التقسيم، الدماغ، بين خوارزمية المتوسطات الضبابية، و خوارزمية لمتوسطات الممكنة

DÉDICACES

A mes parents,

A mes frères et sœurs.

REMERCIEMENTS

Je remercie tout d'abord le bon dieu pour m'avoir donnée le courage et la santé pour accomplir ce travail.

Ce travail n'aurait pas pu aboutir à des résultats sans l'aide et les encouragements de plusieurs personnes que je remercie.

Mes vifs remerciements accompagnés de toute ma gratitude vont ensuite à mon promoteur MOUSSAOUI ABDELOUAHAB, maître de conférence à l'université de Sétif, pour ses conseils judicieux, sa grande disponibilité et pour m'avoir suivie et orientée.

Je remercie le président de jury Prof. MEZGHICHE MOHAMED, professeur à l'université de Boumerdes qui nous a fait l'honneur de présider le jury.

Je remercie également monsieur BABAHENINI MOHAMED CHAWKI, maître de conférence à l'université de Biskra pour m'avoir fait l'honneur d'évaluer mon travail.

Je remercie aussi monsieur DJOUADI YASSINE, maître de conférence à l'université de Tizi ouzou pour l'intérêt qu'il a bien voulu porter à ce travail en acceptant d'être examinateur.

Enfin, que tous ceux qui nous ont aidés et encouragés de prés ou de loin dans la concrétisation de ce projet, trouvent ici ma gratitude et mes sincères remerciements.

TABLE DES MATIÈRES

Introduction générale……..…1

Chapitre 1 : Extraction de connaissances et fouille de données

1.1 Introduction 4
1.2 Les étapes d'un processus d'extraction de connaissances à partir des données....4
1.2.1 Nettoyage et intégration des données 5
1.2.2 Pré-traitement des données....6
1.2.3 Fouille de données (Data Mining)....7
1.2.4 Evaluation et présentation 7
1.3 Fouille de données (Data mining)....9
1.3.1 Historique 9
1.3.2 Définition....11
1.3.3 Principales tâches de fouille de données 11
1.3.3.1 La classification 11
1.3.3.2 L'estimation....12
1.3.3.3 La prédiction 12
1.3.3.4 Les règles d'association....12
1.3.3.5 La segmentation 12
1.3.4 Les méthodes de data mining 13
A. Les méthodes classiques....13
B. Les méthodes sophistiquées....13
1.3.4.1 Segmentation (Clustering) 14
1.3.4.2 Règles d'association....16
1.3.4.3 Les plus proches voisins 20
1.3.4.4 Les arbres de décision....22
1.3.4.5 Les réseaux de neurones 25
1.4 Conclusion....29

Chapitre 2 : Méthodes de classification

2.1 Introduction 30
2.2 Méthodes supervisées....31
2.2.1 Méthodes bayesiennes 31
2.2.2 Champs de Markov....32
2.2.3 Algorithme des k plus proches voisins....33
2.2.4 Réseaux de Neurones....34
2.3 Méthodes non supervisées....36
2.3.1 Algorithmes de classification non flous 37
2.3.1.1 Algorithmes C-moyennes ("Hard C-Means" ou HCM)....37
2.3.2 Algorithmes de classification flous 38
2.3.2.1 Algorithmes C-moyennes foues ("Fuzzy C-Means" ou FCM)....38
2.3.2.2 Les variantes des C-moyennes floues....40
❖ L'approche de Kamel et Selim....41

❖ Méthode FCM semi-supervisée 41
❖ FCM et contraintes de voisinage 42
2.3.2.3 Algorithmes les C-moyennes possibilistes ("Possibilistic C-means" ou PCM) 43
2.4 Conclusion 49

Chapitre 3 : Fusion de données

3.1 Introduction 50
3.2 Divers contextes théoriques possibles pour la fusion de données 51
3.2.1 La théorie des probabilités 51
3.2.2 La théorie des croyances 52
3.2.3 La théorie des possibilités 54
3.2.3.1 Notion de sous -ensemble flou 54
3.2.3.2 Mesure et distribution de possibilité 57
3.2.3.3 Mesure de nécessiter 58
3.3 La Fusion de données 59
3.3.1 Définition 59
3.3.2 Intérêt de la fusion de données 60
3.3.3 Le processus du fusion de données 61
3.3.3.1 Représentation homogène et recalage des informations pertinentes 61
3.3.3.2 Modélisation des connaissances 61
3.3.3.3 Fusion 61
3.3.3.4 Décision par choix d'une stratégie 61
3.3.4 Classification des opérateurs de fusion 62
3.3.4.1 Opérateurs à comportement constant et indépendant du contexte 63
3.3.4.2 Opérateurs à comportement variable et indépendant du contexte 63
3.3.4.3 Opérateurs dépendants du contexte 63
3.3.4.4 Quelques propriétés 64
3.3.5 Fusion en théorie des probabilités 64
3.3.5.1 Combinaison 64
3.3.5.2 Règle de décision 65
3.3.6 Fusion en théorie des croyances 65
3.3.6.1 Combinaison 66
3.3.6.2 Règle de décision 66
3.3.7 Fusion en théorie des possibilistes 67
3.3.7.1 Combinaison 67
3.3.7.2 Règle de décision 68
3.4 Conclusion 68

Chapitre 4 : Contribution

4.1 Introduction 69
4.2 Modélisation 70
4.2.1 Choix de l'algorithme 70
4.2.1.1 Quel type de méthode ? 70
4.2.1.2 Classification floue ou non floue ? 71
4.2.1.3 C-moyennes floues ou algorithme possibiliste ? 71
1. Interprétation des degrés d'appartenance 72
2. Etude comparative entre FCM et PCM 74
4.2.2 Choix des paramètres de l'algorithme 75
4.2.2.1 Initialisation de l'algorithme 76
4.2.2.2 Détermination du nombre de classes 77
4.2.2.3 Choix du paramètre m 77
4.2.2.4 Choix de la distance 77
4.2.2.5 Détermination des paramètres de pondération η_i 79
4.2.2.6 Choix des vecteurs forme 80
4.2.3 Algorithmes utilisés dans notre approche 81
4.3 Fusion 83
4.3.1 Limitations de la fusion probabiliste 84
4.3.2 Comparaison des théories des possibilités et de l'évidence 84
4.3.3 Vers la théorie des possibilités 86
4.4 Décision 87
4.5 Conclusion 89

Chapitre 5 : Résultats et évaluation

5.1 Introduction 90
5.2 Images utilisées 90
5.2.1 Images réelles 90
5.2.2 Fantômes (images de synthèse) 91
5.2.3 Constructions des images simulées 92
5.3 Evaluation et étude comparative 94
5.3.1 Critères de validation 94
5.3.2 Le protocole d'évaluation 96
5.3.3 Évaluations des résultats 97
5.3.4 Comparaison aux C-moyennes floues et C-moyennes possibiliste 99
5.4 Analyse des résultats 101
5.4.1 Images de synthèse 101
5.4.2 Images réelles 105
5.5 Conclusion 108
Conclusion générale et perspectives 109
Annexe - Eléments d'anatomie cérébrale et IRM 112
Bibliographie 125

LISTE DES FIGURES

Figure 1. 1 : Processus d'extraction de connaissances à partir des données 8
Figure 1. 2 : L'extraction de connaissances à partir des données 9
Figure 1. 3 : Arbre de décision.....23
Figure 1. 4 : Nœud d'un réseau de neurone. 27
Figure 2. 1 : Distribution de possibilités des tissus dérivées de l'histogramme de l'image.....46
Figure 3. 1 : Exemple de sous-ensemble flou 55
Figure 3. 2 : Exemple d'intersection et d'union 56
Figure 4. 1 : Les étapes de la Fusion de données de notre proposition.....70
Figure 4. 2 : Démonstration de la mauvaise interprétation des degrés d'appartenance du FCM...72
Figure 4. 3 : Architecture générale des étapes de la fusion de données de l'approche proposée...88
Figure 5. 1 : Exemple de cartes floues du fantôme 92
Figure 5. 2 : Processus de construction des images simulées. 93
Figure 5. 3 : Processus d'évaluation sur des images fantômes 97
Figure 5. 4 : Coupes pondérées en T1, T2 et en densité de protons illustrant la fusion d'images. 97
Figure 5. 5 : Image T1 segmentée par FCM. 99
Figure 5. 6 : Image T1 segmentée par PCM. 100
Figure 5. 7 : Image T1 segmentée par l'approche proposée. 100
Figure 5. 8 : Comparaison des images segmentées (Images saines).....103
Figure 5. 9 : Comparaison des images segmentées (images pathologiques)……..104
Figure 5. 10 : Images originales : différentes coupes (axiales, sagittales, coronales) 106
Figure 5. 11 : Segmentation par approche proposée : différentes coupes.....107
Figure A. 1 : Structure générale de l'encéphale. 113
Figure A. 2 : Plans de coupe en imagerie médicale. 114
Figure A. 3 : Les trois axes de coupe pour la visualisation du cerveau. 114
Figure A. 4 : Structures anatomiques de la matière grise.....115
Figure A. 5 : Coupes IRM du cerveau. 117
Figure A. 6 : Différentes structures du cerveau. 117

Figure A. 7 : Appareil d'IRM 118
Figure A. 8 : Retour à l'équilibre du vecteur aimantation 119
Figure A. 9 : L'angle de basculement 121
Figure A. 10 : Images IRM. 122
Figure A. 11 : L'inhomogénéité RF. 124

Liste des tableaux

Tableau 1. 1 : La base de données avant le nettoyage.....5
Tableau 1. 2 : La base de données après le nettoyage.....6
Tableau 1. 3 : La base de données après le pré-traitement.....7
Tableau 1. 4 : Liste des achats.....17
Tableau 1. 5 : Tableau de co-occurrence.....17
Tableau 2. 1 : Comparaison entre méthodes de classification 48
Tableau 4. 1 : Degrés d'appartenance générés par le FCM 72
Tableau 4. 2 : Comparaison des degrés d'appartenance générés par FCM et PCM. 74
Tableau 4. 3 : Comparaison des théories de la possibilité et de l'évidence 85
Tableau 5. 1 : Evaluations de la segmentation par le système développé.....98
Tableau 5. 2 : Evaluations de la segmentation par approche coopérative.....98
Tableau 5. 3 : Comparaison des taux de recouvrement 101
Tableau 5. 4 : Propriétés des images utilisées. 105

INTRODUCTION GENERALE

Avec le développement des dossiers médicaux informatiques et la généralisation des techniques d'imagerie, il devient possible, pour une pathologie donnée, de disposer d'un grand nombre de données hétérogènes, complémentaires et parfois ambiguës. Le clinicien, analysant ces multiples informations, opère une agrégation de celles-ci, en fonction de jugements subjectifs et approximatifs fondés sur sa propre expérience. Le but de ce raisonnement est de synthétiser un état de la pathologie le plus complet possible, par exemple pour proposer un diagnostic, établir un pronostic ou même élaborer une aide à l'intervention chirurgicale.

Ces dernières années, des modélisations formelles de cette attitude ont été construites, fondées pour la plupart sur des approches prenant en compte les redondances, les complémentarités et les ambiguïtés inhérentes aux données médicales. Regroupées sous l'appellation "*fusion*", ces modèles ont pour but de gérer au mieux ces différents aspects pour faire converger les connaissances et proposer une information synthétique la plus exploitable possible.

Notre travail concerne essentiellement le développement de nouveaux outils issues des techniques de data mining pour l'extraction des connaissances par fusion floue de données. Il s'agit essentiellement de contribuer au développement de systèmes de classification guidés par les connaissances *a priori* où l'aspect flou et possibiliste sont pris en considération lors du processus de classification. Notre travail consiste à proposer une architecture de fusion de données guidée par ces connaissances *a priori*. Afin de valider les algorithmes développés, une application a été développée pour la segmentation des images IRM.

Le processus de fusion tel que nous l'envisageons ici est composé de trois étapes. Dans la première, les informations disponibles sont modélisées dans un cadre théorique commun, permettant de prendre en compte les connaissances vagues et ambiguës. Dans la seconde, les modèles d'informations sont agrégés, en tenant compte des redondances et des conflits exprimés. Dans la troisième, enfin, une décision est prise en fonction de toutes les informations précédemment fusionnées. Ce mémoire s'articule autour de ce schéma.

Le premier chapitre présente les concepts du data mining, où sont décrites les différentes étapes d'un processus d'extraction de connaissances à partir des données. Parmi ces étapes, nous détaillons la phase de fouille de données ainsi que les différentes approches de sa mise en œuvre.

Suivant le déroulement du processus de fusion, et après avoir introduit les données et les concepts théoriques de data mining, nous présentons dans un deuxième chapitre les différentes approches et stratégies généralement utilisées pour la classification des tissus cérébraux tout en évoquant les avantages et les inconvénients de chaque approche.

La fusion de données fait l'objet du chapitre trois. Ainsi après l'étape de modélisation, nous proposons tout naturellement une étude des méthodes d'agrégation des informations.

Les trois approches pour intégrer la représentation des connaissances *incertaines* et/ou *imprécises* à savoir : approche probabiliste, approche possibiliste et approche basée sur la théorie des croyance ainsi que l'étape de décision, dernière phase du processus de fusion, font également office de ce même chapitre.

Dans le quatrième chapitre nous décrivons d'abord, les différents algorithmes de fusion de données ainsi que leurs architectures et leurs différents paramètres. Nous présentons ensuite les étapes de la fusion de données de l'approche proposée en réponse à la problématique de l'extraction des connaissances par fusion floue de données.

Le cinquième chapitre présente les résultats obtenus dans le cadre de cette étude et illustre le fonctionnement du système sur des images médicales réelles et simulées.

La conclusion et les perspectives de ce travail seront présentées à la fin du mémoire.

En complément, nous proposons, en annexe, une présentation des modalités d'imagerie médicale utilisées dans ce travail de mémoire, en insistant sur quelques aspects de l'anatomie du cerveau et de son fonctionnement.

CHAPITRE 1

EXTRACTION DE CONNAISSANCES ET FOUILLE DE DONNÉES

1.1 Introduction

Les entreprises, mais aussi, dans une certaine mesure, les administrations, subissent aujourd'hui l'intensification de la concurrence ou la pression des administrés. Ces facteurs les poussent à porter une attention toujours plus grande aux clients (d'autant plus que leurs richesses aujourd'hui résident dans leurs clients), à améliorer constamment la qualité de leurs produits et à accélérer de manière générale leurs processus de mise sur le marché de nouveaux produits et services.

Pour répondre à ces besoins de découvertes, un ensemble d'architectures, de démarches et d'outils, certains nouveaux, d'autres existants depuis longtemps, a été regroupé sous le terme "*data mining*".

Ce premier chapitre est une introduction aux différents concepts de fouille de données (data mining en anglais), où les différentes étapes du processus d'extraction de connaissances à partir des données sont décrites. Nous insistons sur les différentes approches de mise en œuvre d'un modèle de fouille de données.

1.2 Les étapes d'un processus d'Extraction de Connaissances à partir des données

Définition

*«L'Extraction de Connaissances à partir des Données (ECD) est un processus itératif et interactif d'analyse d'un grand ensemble de données brutes afin d'en extraire des connaissances exploitables par un utilisateur- analyste qui y joue un rôle central»***[Zigh-01]**.

D'après **[Fayy-96]**, un processus d'ECD est constitué de quatre phases qui sont : *le nettoyage et intégration des données, le pré-traitement des données, la fouille de données* et enfin *l'évaluation et la présentation des connaissances.*

La figure 1.1 récapitule ces différentes phases ainsi que les enchaînements possibles entre ces phases. Cette séparation est théorique car en pratique, ce n'est pas toujours le cas. En effet, dans de nombreux systèmes, certaines de ces étapes sont fusionnées **[Kodr-98]**.

1.2.1 Nettoyage et intégration des données

Le nettoyage des données consiste à retravailler ces données bruitées, soit en les supprimant, soit en les modifiant de manière à tirer le meilleur profit.

L'intégration est la combinaison des données provenant de plusieurs sources (base de données, sources externes, etc.).

Le but de ces deux opérations est de générer des entrepôts de données et/ou des magasins de données spécialisés contenant les données retravaillées pour faciliter leurs exploitations futures.

Exemple :

Soit l'exemple suivant qui présente une base de données d'un éditeur qui vend 5 sortes de magazines : sport, voiture, maison, musique et BD. Il souhaite mieux étudier ses clients pour découvrir de nouveaux marchés ou vendre plus de magazines à ses clients habituels.

client	Nom	Adresse	Date d'abonnement	Magazine
23134	Bemol	Rue du moulin, Paris	07/10/1996	Voiture
23134	Bemol	Rue du moulin, Paris	12/05/1996	Musique
23134	Bemol	Rue du moulin, Paris	25/07/1995	BD
31435	Bodinoz	Rue verte, Nancy	11/11/2011	BD
43342	Airinaire	Rue de la source, Brest	30/05/1995	Sport
25312	Talonion	Rue du marché, Paris	25/02/1998	NULL
43241	Manvussa	NULL	14/04/1996	Sport
23130	Bemolle	Rue du moulin, Paris	11/11/2011	Maison

Tableau1.1 *: La base de données avant le nettoyage* **[Gill-00]**.

Intégrité de domaine : Dans notre exemple, la date d'abonnement des client 23130, 31435 (11/11/11) semble plutôt correspondre à une erreur de saisie ou encore à une valeur par défaut en remplacement d'une valeur manquante.

Informations manquantes : Dans notre exemple, nous n'avons pas le type de magazine pour le client 25312. Il sera écarté de notre ensemble. L'enregistrement du client 43241 sera conservé bien que l'adresse ne soit pas connue.

Après le nettoyage on aura la base de données suivante :

client	Nom	Adresse	Date d'abonnement	Magazine
23134	Bemol	Rue du moulin, Paris	07/10/1996	Voiture
23134	Bemol	Rue du moulin, Paris	12/05/1996	Musique
23134	Bemol	Rue du moulin, Paris	25/07/1995	BD
31435	Bodinoz	Rue verte, Nancy	NULL	BD
43342	Airinaire	Rue de la source, Brest	30/05/1995	Sport
43241	Manvussa	NULL	14/04/1996	Sport
23130	Bemolle	Rue du moulin, Paris	NULL	Maison

Tableau1.2 *: La base de données après le nettoyage* **[Gill-00]**.

1.2.2 Pré-traitement des données

Il peut arriver parfois que les bases de données contiennent à ce niveau un certain nombre de données incomplètes et/ou bruitées. Ces données erronées, manquantes ou inconsistantes doivent être retravaillées si cela n'a pas été fait précédemment. Dans le cas contraire, durant l'étape précédente, les données sont stockées dans un entrepôt. Cette étape permet de sélectionner et transformer des données de manière à les rendre exploitables par un outil de fouille de données.

Cette seconde étape du processus d'*ECD* permet d'affiner les données. Si l'entrepôt de données est bien construit, le pré-traitement de données peut permettre d'améliorer les résultats lors de l'interrogation dans la phase de fouille de données.

Exemple :

Soit la base de donnée nettoyée précédemment, le tableau1.3 présente le résultat de pré-traitement. Les clients qui ont des informations manquantes seront supprimés de la base.

client	Nom	Adresse	Date d'abonnement	Magazine
23134	Bemol	Rue du moulin, Paris	07/10/1996	Voiture
23134	Bemol	Rue du moulin, Paris	12/05/1996	Musique
23134	Bemol	Rue du moulin, Paris	25/07/1995	BD
43342	Airinaire	Rue de la source, Brest	30/05/1995	Sport

Tableau1.3 *: La base de données après le pré-traitement* **[Gill-00]**.

1.2.3 Fouille de données (Data Mining)

La fouille de données (*data mining* en anglais), est le cœur du processus d'ECD. Il s'agit à ce niveau de trouver des pépites de connaissances à partir des données. Tout le travail consiste à appliquer des méthodes intelligentes dans le but d'extraire cette connaissance. Il est possible de définir la qualité d'un modèle en fonction de critères comme les performances obtenus, la fiabilité, la compréhensibilité, la rapidité de construction et d'utilisation et enfin l'évolutivité. Tout le problème de la fouille de données réside dans le choix de la méthode adéquate à un problème donné. Il est possible de combiner plusieurs méthodes pour essayer d'obtenir une solution optimale globale.

Nous ne détaillerons pas d'avantage la fouille de données dans ce paragraphe car elle fera l'objet d'une section complète (*cf.*1.3)

1.2.4 Evaluation et présentation

Cette phase est constituée de l'évaluation, qui mesure l'intérêt des motifs extraits, et de la présentation des résultats à l'utilisateur grâce à différentes techniques de visualisation. Cette étape est dépendante de la tâche de fouille de données employée. En effet, bien que l'interaction avec l'expert soit importante quelle que soit cette tâche, les techniques ne sont pas les mêmes. Ce n'est qu'à partir de la phase de présentation que l'on peut employer le terme de *connaissance* à condition que ces motifs soient validés par les experts du domaine. Il y a principalement deux

techniques de validation qui sont la technique de validation statistique et la technique de validation par expertise.

La validation statistique consiste à utiliser des méthodes de base de statistique descriptive. L'objectif est d'obtenir des informations qui permettront de juger le résultat obtenu, ou d'estimer la qualité ou les biais des données d'apprentissage. Cette validation peut être obtenue par :

- le calcul des moyennes et variances des attributs,
- si possible, le calcul de la corrélation entre certains champs,
- ou la détermination de la classe majoritaire dans le cas de la classification.

La validation par expertise, est réalisée par un expert du domaine qui jugera la pertinence des résultats produits. Par exemple pour la recherche des règles d'association, c'est l'expert du domaine qui jugera la pertinence des règles.

Pour certains domaines d'application (le diagnostic médical, par exemple), le modèle présenté doit être compréhensible. Une première validation doit être effectuée par un expert qui juge la compréhensibilité du modèle. Cette validation peut être, éventuellement, accompagnée par une technique statistique.

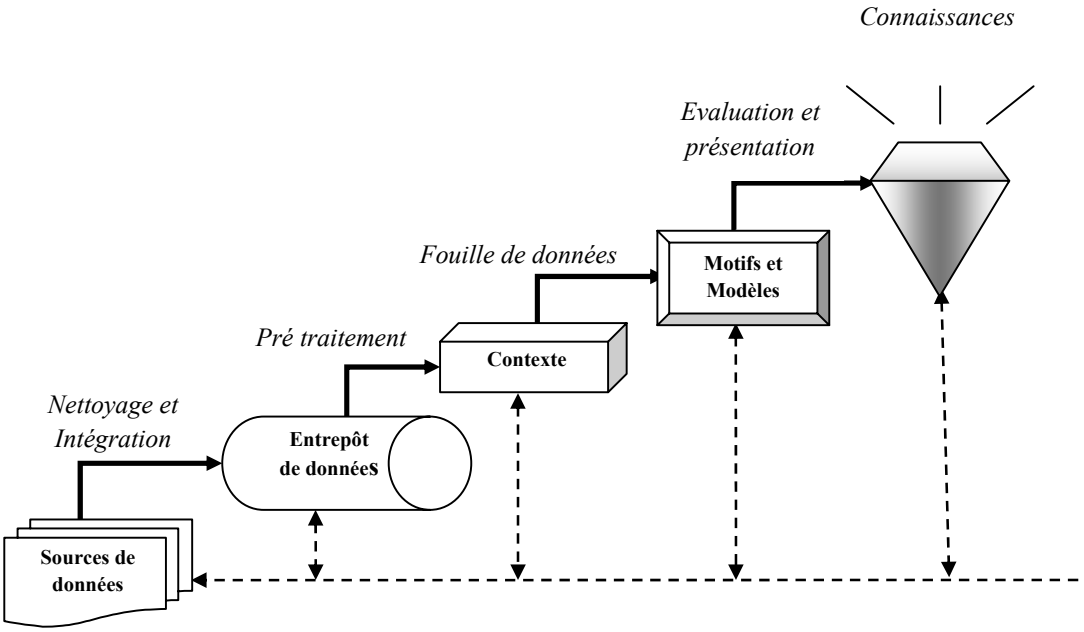


Figure 1.1 *: Processus d'extraction de connaissances à partir des données* **[Fayy-96]**.

Grâce aux techniques d'extraction de connaissances, les bases de données volumineuses sont devenues des sources riches et fiables pour la génération et la validation de connaissances. La fouille de données n'est qu'une phase du processus d'ECD, et consiste à appliquer des algorithmes d'apprentissage sur les données afin d'en extraire des modèles (motifs). L'extraction de connaissances à partir des données se situe à l'intersection de nombreuses discipline **[Kodr-98]**, comme l'apprentissage automatique, la reconnaissance de formes, les bases de données, les statistiques, la représentation des connaissances, l'intelligence artificielle, les systèmes experts, etc. *(cf*. Figure 1.2).

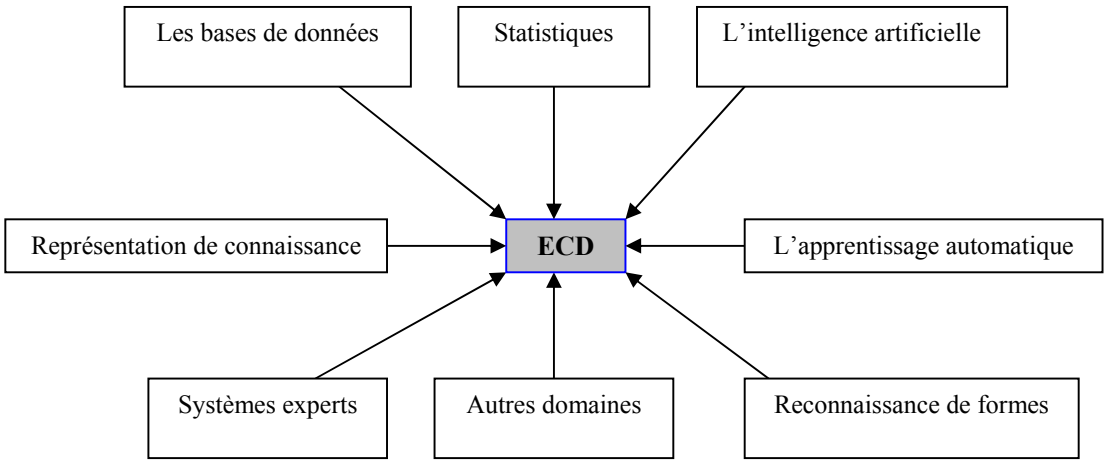


Figure 1. 2 : *L'Extraction de connaissances à partir des données à la confluence de nombreux domaines* [**Kodr-98]**.

1.3 Fouille de données (data mining)

Les concepts de fouille de données et d'extraction de connaissances à partir de données sont parfois confondus et considérés comme synonymes. Mais, formellement on considère la fouille de données comme une étape centrale du processus d'extraction de connaissances des bases de données (*ECBD* ou *KDD* pour *Knowledge Discovery in Databases* en anglais) **[Lieb-07]**.

1.3.1 Historique

L'expression "*data mining*" est apparue vers le début des années 1960 et avait, à cette époque, un sens péjoratif. En effet, les ordinateurs étaient de plus en plus utilisés pour toutes

sortes de calculs qu'il n'était pas envisageable d'effectuer manuellement jusque là. Certains chercheurs ont commencé à traiter sans *a priori* statistique les tableaux de données relatifs à des enquêtes ou des expériences dont ils disposaient. Comme ils constataient que les résultats obtenus, loin d'être aberrants, étaient tout au contraire prometteurs, ils furent incites à systématiser cette approche opportuniste. Les statisticiens officiels considéraient toutefois cette démarche comme peu scientifique et utilisèrent alors les termes " *data mining* " ou "*data fishing*" pour les critiquer.

Cette attitude opportuniste face aux données coïncida avec la diffusion dans le grand public de l'analyse de données dont les promoteurs, comme Jean-Paul Benzecri **[Zigh-00]**, ont également dû subir dans les premiers temps les critiques venant des membres de la communauté des statisticiens.

Le succès de cette démarche empirique ne s'est pas démenti malgré tout. L'analyse des données s'est développée et son intérêt grandissait en même temps que la taille des bases de données. Vers la fin des années 1980, des chercheurs en base de données, tel que Rakesh Agrawal **[Agra-93]**, ont commencé à travailler sur l'exploitation du contenu des bases de données volumineuses comme par exemple celles des tickets de caisses de grandes surfaces, convaincus de pouvoir valoriser ces masses de données dormantes. Ils utilisèrent l'expression "*database mining*" mais, celle-ci étant déjà déposée par une entreprise (Database mining workstation), ce fut "*data mining*" qui s'imposa. En mars 1989, Shapiro Piatetski **[Shap-91]** proposa le terme "*knowledge discovery*" à l'occasion d'un atelier sur la découverte des connaissances dans les bases de données. Actuellement, les termes data mining et knowledge discovery in data bases (*KDD*, ou *ECD* en français) sont utilisés plus ou moins indifféremment. Nous emploierons par conséquent l'expression "*data mining*", celle-ci étant la plus fréquemment employée dans la littérature.

La communauté de "*data mining* " a initié sa première conférence en 1995 à la suite de nombreux atelier (workshops) sur le *KDD* entre 1989 et 1994. La première revue du domaine " *Data mining and knowledge discovery journal* " publiée par "Kluwers " a été lancée en 1997.

1.3.2 Définition

« Le data mining, ou fouille de données, est l'ensemble des méthodes et techniques destinées à l'exploration et l'analyse de bases de données informatiques (souvent grandes), de façon automatique ou semi-automatique, en vue de détecter dans ces données des règles, des associations, des tendances inconnues ou cachées, des structures particulières restituant l'essentiel de l'information utile tout en réduisant la quantité de données» **[Stép-05**] **[Kant-03]**.

D'après **[Hadd-02]**, la définition la plus communément admise de Data Mining est celle de **[Fayy-98]** : *«Le Data mining est un processus non trivial qui consiste à identifier, dans des données, des schémas nouveaux, valides, potentiellement utiles et surtout compréhensibles et utilisables».*

En bref, le data mining est l'art d'extraire des informations (ou même des connaissances) à partir des données **[Tuff-05]**.

1.3.3 Principales tâches de fouille de données

On dispose de données structurées. Les objets sont représentés par des enregistrements (ou descriptions) qui sont constitués d'un ensemble de champs (ou attributs) prenant leurs valeurs dans un domaine. De nombreuses tâches peuvent être associées au Data Mining, parmi elles nous pouvons citer:

1.3.3.1 La classification

Elle consiste à examiner les caractéristiques d'un objet et lui attribuer une classe, la lasse est un champ particulier à valeurs discrètes. Des exemples de tâche de classification sont :

- attribuer ou non un prêt à un client,
- établir un diagnostic,
- accepter ou refuser un retrait dans un distributeur,
- attribuer un sujet principal à un article de presse,
- etc.

1.3.3.2 L'estimation

Elle consiste à estimer la valeur d'un champ à partir des caractéristiques d'un objet. Le champ à estimer est un champ à valeurs continues. L'estimation peut être utilisée dans un but de classification. Il suffit d'attribuer une classe particulière pour un intervalle de valeurs du champ estimé. Des exemples de tâche d'estimation sont :

- Estimer les revenus d'un client.

1.3.3.3 La prédiction

Cela consiste à estimer une valeur future. En général, les valeurs connues sont historisées. On cherche à prédire la valeur future d'un champ. Cette tâche est proche des précédentes. Les méthodes de classification et d'estimation peuvent être utilisées en prédiction. Des exemples de tâches de prédiction sont :

- Prédire les valeurs futures d'actions,
- prédire, au vu de leurs actions passées, les départs de clients.

1.3.3.4 Les règles d'association

Cette tâche, plus connue comme *l'analyse du panier de la ménagère*, consiste à déterminer les variables qui sont associées. L'exemple type est la détermination des articles (le pain et le lait, la tomate, les carottes et les oignons) qui se retrouvent ensemble sur un même ticket de supermarché. Cette tâche peut être effectuée pour identifier des opportunités de vente croisée et concevoir des groupements attractifs de produit.

1.3.3.5 La segmentation

Consiste à former des groupes (clusters) homogènes à l'intérieur d'une population. Pour cette tâche, il n'y a pas de classe à expliquer ou de valeur à prédire définie *a priori*, il s'agit de créer des groupes homogènes dans la population (l'ensemble des enregistrements). Il appartient ensuite à un expert du domaine de déterminer l'intérêt et la signification des groupes ainsi constitués. Cette tâche est souvent effectuée avant les précédentes pour construire des groupes sur lesquels on applique des tâches de classification ou d'estimation.

1.3.4 Les méthodes de data mining

Pour tout jeu de données et un problème spécifique, il existe plusieurs méthodes que l'on choisira en fonction de :

- La tâche à résoudre.
- La nature et de la disponibilité des données.
- L'ensemble des connaissances et des compétences disponibles.
- La finalité du modèle construit.
- L'environnement social, technique, philosophique de l'entreprise.
- Etc.

On peut dégager deux grandes catégories de méthodes d'analyse consacrées à la fouille de données [**Fiot-06]**. La frontière entre les deux peut être définie par la spécificité des techniques, et marque l'aire proprement dite du "*Data Mining*". On distingue donc :

A. Les méthodes classiques

On y retrouve des outils généralistes de l'informatique ou des mathématiques :

- ❖ Les requêtes dans les bases de données, simples ou multi-critères, dont la représentation est une vue,
- ❖ les requêtes d'analyse croisée, représentées par des tableaux croisés,
- ❖ les différents graphes, graphiques et représentations,
- ❖ les statistiques descriptives,
- ❖ l'analyse de données : analyse en composantes principales,
- ❖ etc.

B. Les méthodes sophistiquées

Elles ont été élaborées pour résoudre des tâches bien définies. Ce sont :

- ❖ Les algorithmes de segmentation,
- ❖ les règles d'association,
- ❖ les algorithmes de recherche du plus proche voisin,
- ❖ les arbres de décision,

- les réseaux de neurones,
- les algorithmes génétiques,
- etc.

La section suivante n'est pas une présentation exhaustive de l'ensemble des techniques de la fouille de données, mais une présentation de quelques méthodes sophistiquées pour fournir un aperçu du domaine.

1.3.4.1 Segmentation (Clustering)

La segmentation est l'opération qui consiste à regrouper les individus d'une population en un nombre limité de groupes, les segments (ou clusters, ou partition), qui ont deux propriétés : D'une part, ils ne sont pas prédéfinis, mais découverts automatiquement au cours de l'opération, contrairement aux classes de la classification. D'autre part, les segments regroupent les individus ayant des caractéristiques similaires et séparent les individus ayant des caractéristiques différentes (homogénéité interne et hétérogénéité externe).

La segmentation est une tâche d'apprentissage "*non supervisée*" car on ne dispose d'aucune autre information préalable que la description des exemples. Après application de l'algorithme et donc lorsque les groupes ont été construits, d'autres techniques ou une expertise doivent dégager leur signification et leur éventuel intérêt.

Nous présentons ici la méthode des *k*-moyennes car elle est très simple à mettre en œuvre et très utilisée. Elle comporte de nombreuses variantes et est souvent utilisée en combinaison avec d'autres algorithmes.

1. Méthode des k-moyennes

La méthode est basée sur une notion de similarité entre enregistrements. Nous allons pour introduire l'algorithme, considérer un espace géométrique muni d'une distance. Deux points sont similaires s'ils sont proches pour la distance considérée. Pour pouvoir visualiser le fonctionnement de l'algorithme, nous allons limiter le nombre de champs des enregistrements. Nous nous plaçons donc dans l'espace euclidien de dimension 2 et nous considérons la distance euclidienne classique. L'algorithme suppose choisi *a priori* un nombre *k* de groupes à constituer.

On choisit alors *k* enregistrements, soit *k* points de l'espace appelés centres. On constitue alors les *k* groupes initiaux en affectant chacun des enregistrements dans le groupe correspondant au centre le plus proche. Pour chaque groupe ainsi constitué, on calcule son nouveau centre en effectuant la moyenne des points du groupe et on réitère le procédé. Le critère d'arrêt est la stabilité, par lequel d'une itération à la suivante, aucun point n'a changé de groupe.

- **Description de l'algorithme**

On travaille avec des enregistrements qui sont des *n*-uplets de valeurs. On suppose définie une notion de similarité qui permet de comparer les distances aux centres. L'algorithme est paramétré par le nombre *k* de groupes que l'on souhaite constituer.

Algorithme des k-moyennes

Paramètre : le nombre *k* de groupes

Entrée : un échantillon de *m* objets $x_1,... x_m$

1-choisir *k* centre initiaux $c_1,... c_k$

2-pour chacun des *m* objets, l'affecter au groupe *i* dont le centre c_i est le plus proche

3-si aucun élément ne change de groupe alors arrêter et sortir les groupes

4-calculer les nouveaux centres : pour tout *i*, c_i est la moyenne des éléments du groupe *i*

5-aller en 2

2. Critiques de la méthode

a. *Avantages*

- *Apprentissage non supervisé* : la méthode des *k*-moyennes et ses variantes ne nécessitent aucune information sur les données. La segmentation peut être utile, pour découvrir une structure cachée qui permettra d'améliorer les résultats de méthodes d'apprentissage supervisé (classification, estimation, prédiction).
- *Applicable à tous type de données* : en choisissant une bonne notion de distance, la méthode peut s'appliquer à tout type de données (mêmes textuelles).

b. *Inconvénients*

- *Problème du choix de la distance* : les performances de la méthode (la qualité des groupes constitués) sont dépendantes du choix d'une bonne mesure de similarité ce qui est une tâche délicate surtout lorsque les données sont de types différents.
- *Le choix des bons paramètres* : la méthode est sensible au choix des bons paramètres, en particulier, le choix du nombre k de groupes à constituer. Un mauvais choix de k produit de mauvais résultats. Ce choix peut être fait en combinant différentes méthodes, mais la complexité de l'algorithme augmente.
- *L'interprétation des résultats* : il est difficile d'interpréter les résultats produits, en d'autres termes, d'attribuer une signification aux groupes constitués.

1.3.4.2 Règles d'association

Les règles d'association sont traditionnellement liées au secteur de la distribution car leur principale application est "*l'analyse du panier de la ménagère (market basket analysis)*" qui consiste en la recherche d'associations entre produits sur les tickets de caisse. Le but de la méthode est l'étude de ce que les clients achètent pour obtenir des informations sur "*qui*" sont les clients et "*pourquoi*" ils font certains achats. La méthode peut être appliquée à tout secteur d'activité pour lequel il est intéressant de rechercher des groupements potentiels de produits ou de services: services bancaires, services de télécommunications, par exemple. Elle peut être également utilisée dans le secteur médical pour la recherche de complications dues à des associations de médicaments ou à la recherche de fraudes en recherchant des associations inhabituelles.

Un attrait principal de la méthode est la clarté des résultats produits. En effet, le résultat de la méthode est un ensemble de *règles d'association*. Des exemples de règles d'association sont :

- si un client achète des plantes alors il achète du terreau,
- si un client achète une télévision, il achètera un magnétoscope dans un an.

Ces règles sont intuitivement faciles à interpréter car elles montrent comment des produits ou des services se situent les uns par rapport aux autres. Ces règles sont particulièrement utiles en marketing. Les *règles d'association* produites par la méthode peuvent être facilement utilisées

dans le système d'information de l'entreprise. Cependant, il faut noter que la méthode, si elle peut produire des règles intéressantes, peut aussi produire des règles triviales (déjà bien connues des intervenants du domaine) ou inutiles (provenant de particularités de l'ensemble d'apprentissage). La recherche de règles d'association est une méthode non supervisée car on ne dispose en entrée que de la description des achats.

1. Introduction à la méthode

On suppose avoir prédéfini une classification des articles. Les données d'entrée sont constituées d'une liste d'achats. Un achat est lui-même constitué d'une liste d'articles. On peut remarquer que, contrairement aux enregistrements d'une table, les achats peuvent être de longueur variable. Pour introduire la méthode, nous considérons l'exemple suivant :

	Produit A	Produit B	Produit C	Produit D	Produit E
Achat 1	X			X	
Achat 2	X	X	X		
Achat 3	X				X
Achat 4	X			X	X
Achat 5		X		X	

Tableau 1.4 *: Liste des achats*

À partir de ces données, si on recherche des associations entre deux produits, on construit le tableau de co-occurrence qui montre combien de fois deux produits ont été achetés ensemble :

	Produit A	Produit B	Produit C	Produit D	Produit E
Achat 1	4	1	1	2	1
Achat 2	1	2	1	1	0
Achat 3	1	1	1	0	0
Achat 4	2	1	0	3	1
Achat 5	1	0	0	1	2

Tableau 1. 5 *: Tableau de co-occurrence*

Un tel tableau permet de déterminer avec quelle fréquence deux produits se rencontrent dans un même achat. Par exemple, le produit A apparaît dans 80% des achats, le produit C n'apparaît jamais en même temps que le produit E, les produits A et D apparaissent

simultanément dans 40% des achats. Ces observations peuvent suggérer une règle de la forme " *si un client achète le produit A alors il achète le produit D*".

Ces idées se généralisent à toutes les combinaisons d'un nombre quelconque d'articles. Par exemple, pour trois articles, on cherche à générer des règles de la forme si X et Y alors Z.

Une *règle d'association* est une règle de la forme : Si **condition** alors **résultat**. Dans la pratique, on se limite, en général, à des règles où la *condition est une conjonction d'apparition d'articles* et le *résultat est constitué d'un seul article*. Par exemple, une règle à trois articles sera de la forme : Si **X** et **Y** alors **Z** ; règle dont la sémantique peu être énoncée : Si les articles X et Y apparaissent simultanément dans un achat alors l'article Z apparaît.

Pour choisir une règle d'association, il nous faut définir les quantités numériques qui vont servir à valider l'intérêt d'une telle règle.

❖ Le *support* d'une règle est la fréquence d'apparition simultanée des articles qui apparaissent dans la condition et dans le résultat dans la liste d'achats donnée en entrée, soit

$$Support=freq(condition\ et\ résultat)=d/m$$

où *d* est le nombre d'achats où les articles des parties condition et résultat apparaissent et *m* est le nombre total d'achats.

❖ La *confiance* est le rapport entre le nombre d'achats, où tous les articles figurant dans la règle apparaissent, et le nombre d'achats, où les articles de la partie condition apparaissent, soit

$$Confiance=freq(condition\ et\ résultat)/freq(condition)=d/c$$

où *c* est le nombre d'achats où les articles de la partie condition apparaissent.

La confiance ne dépend que des articles qui apparaissent dans la règle. Des règles, dont le support est suffisant, ayant été choisies, la règle de confiance maximale sera alors privilégiée Cependant, nous allons montrer sur l'exemple suivant que le support et la confiance ne sont pas toujours suffisants. Considérons trois articles A, B et C et leurs fréquences d'apparition :

Article(s)	A	B	C	A et B	A et C	B et C	A et B et C
fréquence	45%	42,50%	40%	25%	20%	15%	5%

Si on considère les règles à trois articles, elles ont le même niveau de support 5%. Le niveau de confiance des trois règles est :

Règle	Confiance
Si A et B et C	0.20
Si A et C et B	0.25
Si B et C et A	0.33

La règle si B et C alors A possède la plus grande confiance. Une confiance de 0.33 signifie que si B et C apparaissent simultanément dans un achat alors A y apparaît aussi avec une probabilité estimée de 33%. Mais, si on regarde le tableau des fréquences d'apparition des articles, on constate que A apparaît dans 45% des achats. Il vaut donc mieux prédire A sans autre information que de prédire A lorsque B et C apparaissent. C'est pourquoi il est intéressant d'introduire l'*amélioration* qui permet de comparer le résultat de la prédiction en utilisant la règle avec la prédiction sans la règle. Elle est définie par :

Amélioration=confiance/freq (résultat)

Une règle est intéressante lorsque l'amélioration est supérieure à 1. Pour les règles choisies, on trouve :

Règle	confiance	f (résultat)	amélioration
Si A et B alors C	0.20	40%	0.50
Si A et C alors B	0.25	42.5%	0.59
Si B et C alors A	0.33	45%	0.74

Par contre, la règle si A alors B possède un support de 25%, une confiance de 0.55 et une amélioration de 1.31, cette règle est donc la meilleure. En règle générale, la meilleure règle est celle qui contient le moins d'articles.

2. Critiques de la méthode

a. *Avantages*

- *Méthode non supervisée* à l'exception de la classification de différents articles en produits.
- *Clarté des résultats* : les règles faciles à interpréter.
- *Traite des données de taille variables* : le nombre de produits dans un achat n'est pas défini.
- *Simplicité de programmation* : même avec un tableur.

b. *Inconvénients*

- *Pertinence des résultats* : ils peuvent être triviaux ou inutiles.
- *Efficacité faible dans certains cas* : pour les produits rares.
- *Traitement préalable des données* : classement les articles en produits.

1.3.4.3 Les plus proches voisins

La méthode des plus proches voisins (*PPV* en bref, *nearest neighbor* en anglais ou *NN)* est une méthode dédiée à la classification qui peut être étendue à des tâches d'estimation. La méthode PPV est une méthode de raisonnement à partir de cas. Elle part de l'idée de prendre des décisions en recherchant un ou des cas similaires déjà résolus en mémoire. Contrairement aux autres méthodes de classification qui seront étudiées dans les sections suivantes (arbres de décision, réseaux de neurones, algorithmes génétiques), il n'y a pas d'étape d'apprentissage consistant en la construction d'un modèle à partir d'un échantillon d'apprentissage. C'est l'échantillon d'apprentissage, associé à une fonction de distance et d'une fonction de choix de la classe en fonction des classes des voisins les plus proches, qui constitue le modèle. L'algorithme générique de classification d'un nouvel exemple par la méthode *PPV* est :

Algorithme de classification par k-PPV

Paramètre : le nombre *k* de voisins

Donnée : un échantillon de *m* exemples et leurs classes

- La classe d'un exemple *X* est *c(X)*

Entrée : un enregistrement *Y*

1 Déterminer les *k* plus proches exemples de *Y* en calculant les distances

2 Combiner les classes de ces *k* exemples en une classe *c*

Sortie : la classe de *Y* est *c(Y)=c*

1. Critiques de la méthode

a. *Avantages*

- *Absence d'apprentissage* : c'est l'échantillon qui constitue le modèle. L'introduction de nouvelles données permet d'améliorer la qualité de la méthode sans nécessiter la reconstruction d'un modèle. C'est une différence majeure avec des méthodes telles que les arbres de décision et les réseaux de neurones.
- *Clarté des résultats* : bien que la méthode ne produise pas de règle explicite, la classe attribuée à un exemple peut être expliquée en exhibant les plus proches voisins qui ont amené à ce choix.
- *Données hétérogènes* : la méthode peut s'appliquer dès qu'il est possible de définir une distance sur les champs. Or, il est possible de définir des distances sur des champs complexes, tels que des informations géographiques, des textes, des images ou du son. C'est parfois un critère de choix de la méthode *PPV* car les autres méthodes traitent difficilement les données complexes. On peut noter, également, que la méthode est robuste face au bruit.
- *Grand nombre d'attributs* : la méthode permet de traiter des problèmes avec un grand nombre d'attributs. Cependant, plus le nombre d'attributs est important, plus le nombre d'exemples doit être grand.

b. *Inconvénients*

- *Sélection des attributs pertinents* : pour que la notion de proximité soit pertinente, il faut que les exemples couvrent bien l'espace et soient suffisamment proches les uns des autres. Si le nombre d'attributs pertinents est faible relativement au nombre total d'attributs, la méthode donnera de mauvais résultats car la proximité sur les attributs pertinents sera noyée par les distances sur les attributs non pertinents. Il est donc parfois utile de sélectionner tout d'abord les attributs pertinents.
- *Le temps de classification* : si la méthode ne nécessite pas d'apprentissage, tous les calculs doivent être effectués lors de la classification. Ceci est la contrepartie à payer par rapport aux méthodes qui nécessite un apprentissage (éventuellement long) mais qui sont rapides en classification (le modèle est créé, il suffit de l'appliquer à l'exemple à classifier). Certaines méthodes permettent de diminuer la taille de l'échantillon en ne conservant que les exemples pertinents pour la méthode *PPV*, mais il faut, de toute façon, un nombre d'exemple suffisamment grand relativement au nombre d'attributs.
- *Définir les distance et nombre de voisins* : les performances de la méthode dépendent du choix de la distance, du nombre de voisins et du mode de combinaison des réponses des voisins. En règle générale, les distances simples fonctionnent bien. Si les distances simples ne fonctionnent pour aucune valeur de k, il faut envisager le changement de distance, ou le changement de méthode.

1.3.4.4 Les arbres de décision

La méthodes des arbres de décision est l'une des plus intuitives et des plus populaires du data mining, d'autant plus qu'elle fournit des règles explicites de classement et supporte bien les données hétérogènes, manquantes et les effets non linéaires. Pour les applications relevant du marketing de bases de données, actuellement la seule grande concurrente de l'arbre de décision est la régression logistique, cette méthode étant préférée dans la prédiction du risque en raison de sa plus grande robustesse. Remarquons que les arbres de décision sont à la frontière entre les méthodes prédictives et descriptives, puisque leur classement s'opère en segmentant la

population à laquelle ils s'appliquent : ils ressortissent donc à la catégorie des classifications hiérarchiques descendantes supervisées.

1. Principale de l'arbre de décision

La technique de l'arbre de décision est employée en classement pour détecter des critères permettant de répartir les individus d'une population en n classes (souvent n=2) prédéfinies. On commence par choisir la variable qui, par ses modalités, sépare le mieux les individus de chaque classe, de façon à avoir des sous-populations, que l'on appelle nœuds, contenant chacune le plus possible d'individus d'une seule classe, puis on réitère la même opération sur chaque nouveau nœud obtenu jusqu'à ce que la séparation des individus ne soit plus possible ou plus souhaitable. Par construction, les nœuds terminaux (les feuilles) sont tous majoritairement constitués d'individus d'une seule classe avec une assez forte probabilité, quand il satisfait l'ensemble des règles permettant d'arriver à cette feuille. L'ensemble des règles de toutes les feuilles constitue le modèle de classement (*cf.* Figure 1.3.)

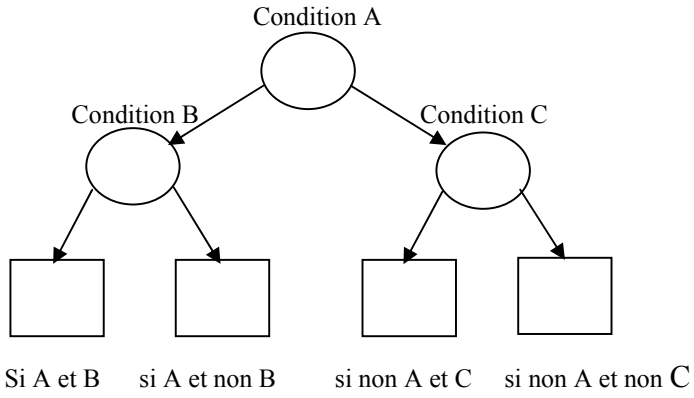


Figure 1.3 *: Arbre de décision.*

Algorithme d'apprentissage par arbre de décision

Donnée : un échantillon S de m enregistrements classés *(x, c(x))*
Initialisation : *A* :=arbre vide ;
noeud_courant := racine ;
échantillon_courant := *S*
Répéter
Décider si le noeud courant est terminal
Si (noeud_courant est terminal) **alors**
Étiqueter le noeud courant par une feuille
Sinon
Sélectionner un test :
Créer les fils
Définir les échantillons sortants du noeud
Finsi
noeud_courant := un noeud non encore étudié de *A*
échantillon_courant : échantillon atteignant noeud_courant
Jusque (noeud_courant = Ø)
Élaguer l'arbre de décision A obtenu
Sortie : l'arbre *A* élagué

2. Critiques de la méthode

a. *Avantages*

- *Adaptabilité aux attributs de valeurs manquantes* : les algorithmes peuvent traiter les valeurs manquantes (descriptions contenant des champs non renseignés) pour l'apprentissage, mais aussi pour la classification.
- *Bonne lisibilité du résultat* : un arbre de décision est facile à interpréter et est la représentation graphique d'un ensemble de règles. Si la taille de l'arbre est importante, il est difficile d'appréhender l'arbre dans sa globalité. Cependant, les outils actuels permettent une navigation aisée dans l'arbre (parcourir une branche, développer un noeud, élaguer une branche) et, le plus important, est certainement

de pouvoir expliquer comment est classé un exemple par l'arbre, ce qui peut être fait en montrant le chemin de la racine à la feuille pour l'exemple courant.

- *Traitement de tout type de données* : l'algorithme peut prendre en compte tous les types d'attributs et les valeurs manquantes. Il est robuste au bruit.
- *Sélectionne des variables pertinentes* : l'arbre contient les attributs utiles pour la classification. L'algorithme peut donc être utilisé comme pré-traitement qui permet de sélectionner l'ensemble des variables pertinentes pour ensuite appliquer une autre méthode.
- *Donne une classification efficace* : l'attribution d'une classe à un exemple à l'aide d'un arbre de décision est un processus très efficace (parcours d'un chemin dans un arbre).
- *Disponibilité des outils* : les algorithmes de génération d'arbres de décision sont disponibles dans tous les environnements de fouille de données.
- *Méthode extensible et modifiable* : la méthode peut être adaptée pour résoudre des tâches d'estimation et de prédiction. Des améliorations des performances des algorithmes de base sont possibles grâce aux techniques qui génèrent un ensemble d'arbres votant pour attribuer la classe.

b. *Inconvénients*

- *Méthode sensible au nombre de classes* : les performances tendent à se dégrader lorsque le nombre de classes devient trop important.
- *Manque d'évolutivité dans le temps* : l'algorithme n'est pas incrémental, c'est-à-dire, que si les données évoluent avec le temps, il est nécessaire de relancer une phase d'apprentissage sur l'échantillon complet (anciens exemples et nouveaux exemples).

1.3.4.5 Les réseaux de neurones

Les réseaux de neurones sont apparus dans les années cinquante avec les premiers perceptrons, et sont utilisés industriellement depuis les années quatre-vingt. Un réseau de neurone "*ou réseau neuronal*" a une architecture calquée sur celle du cerveau, organisée en neurones et

synapses, et se présente comme un ensemble de nœuds "ou neurones formels, ou unités" connectés entre eux, chaque variable prédictive continue correspondant à un nœud d'un premier niveau, appelé *couche d'entrée*, et chaque variable prédictive catégorique (ou chaque modalité d'une variable catégorique) correspondant également à un nœud de la couche d'entrée.

Le cas échéant, lorsque le réseau est utilisé dans une technique prédictive, il y a une ou plusieurs variables à expliquer ; elle correspondant alors chacune à un nœud (ou plusieurs dans le cas des variables catégorielles) d'un dernier niveau : la *couche sortie.* Les réseaux prédictifs sont dits "*à apprentissage supervisé*" et les réseaux descriptifs sont dits "*à apprentissage non supervisé*". Entre la couche d'entrée et la couche sortie sont parfois connectés à des nœuds appartenant à un niveau intermédiaire : *la couche cachée.* Il peut exister plusieurs couches cachées **[Tuff-02]**.

1. Définition

«Les réseaux de neurones sont des outils très utilisés pour la classification, l'estimation, la prédiction et la segmentation. Ils sont issus de modèles biologiques, sont constitués d'unités élémentaires (les neurones) organisées selon une architecture» **[Tuff-05]**.

Un nœud reçoit des valeurs en entrée et renvoie 0 à n valeurs en sortie. Toutes ces valeurs sont normalisées pour être comprises entre 0 et 1(ou parfois entre -1 et 1), selon les bornes de la fonction de transfert. Une fonction de combinaison calcule une première valeur à partir des nœuds connectés en entrée et poids des connexions. Dans les réseaux les plus courants, les perceptrons, il s'agit de la somme pondérée $\sum n_i p_i$ des valeurs des nœuds en entrée. Afin déterminer une valeur en sortie, une seconde fonction, appelée *fonction de transfert (ou d'activation)*, est appliquée à cette valeur. Les nœuds de la couche d'entrée sont triviaux, dans la mesure où ils ne combinent rien, et ne font que transmettre la valeur de la variable qui leur correspond.

Un nœud de perceptron se présente donc comme suit :

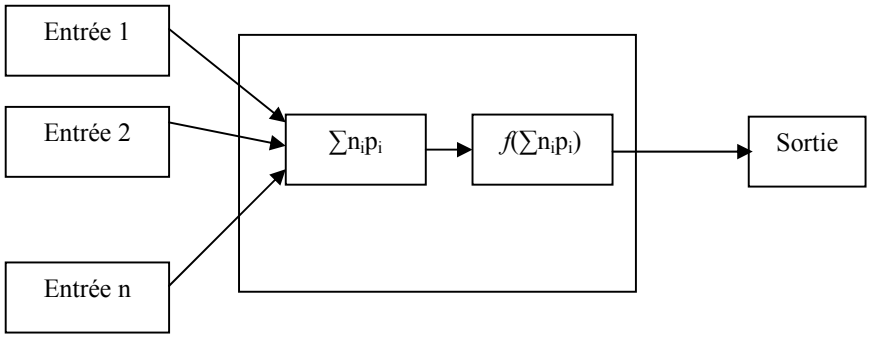


Figure 1.4 *: Nœud d'un réseau de neurone* **[Tuff-05]**.

Dans cette figure, on utilise les notions suivantes :

- n_i est la valeur du nœud i du niveau précédent (la sommation sur i correspond à l'ensemble des nœuds du niveau précédent connectés au nœud observé)
- p_i est le poids associé à la connexion entre i et le nœud observé.
- *F* est la fonction de transfert associée au nœud observé.

2. Mise en œuvre

De façon générale, les étapes dans la mise en œuvre d'un réseau de neurones pour la prédiction ou le classement sont :

1. l'identification des données en entrée et en sortie,
2. la normalisation de ces données,
3. la constitution d'un réseau avec une structure adaptée,
4. l'apprentissage du réseau,
5. le test du réseau,
6. l'application du modèle génère par l'apprentissage,
7. la dénormalisation des données en sortie.

3. Critiques de la méthode

a. *Avantages*

- *Lisibilité du résultat* : Le résultat de l'apprentissage est un réseau constitué de cellules organisées selon une architecture, définies par une fonction d'activation et un très grand nombre de poids à valeurs réelles.
- *Les données réelles* : les réseaux traitent facilement les données réelles "préalablement normalisées " et les algorithmes sont robustes au bruit. Ce sont, par conséquent, des outils bien adaptés pour le traitement de données complexes éventuellement bruitées comme la reconnaissance de formes (son, images sur une rétine, etc.).
- *Classification efficace* : le réseau étant construit, le calcul d'une sortie à partir d'un vecteur d'entrée est un calcul très rapide.
- *En combinaison avec d'autres méthodes* : pour des problèmes contenant un grand nombre d'attributs pour les entrées, il peut être très difficile de construire un réseau de neurones. On peut, dans ce cas, utiliser les arbres de décision pour sélectionner les variables pertinentes, puis générer un réseau de neurones en se restreignant à ces entrées.

b. *Inconvénients*

- *Temps d'apprentissage* : l'échantillon nécessaire à l'apprentissage doit être suffisamment grand et représentatif des sorties attendues. Il faut passer un grand nombre de fois tous les exemples de l'échantillon d'apprentissage avant de converger et donc le temps d'apprentissage peut être long.
- *Evolutivité dans le temps* : comme pour les arbres de décision, l'apprentissage n'est pas incrémental et, par conséquent, si les données évoluent avec le temps, il est nécessaire de relancer une phase d'apprentissage pour s'adapter à cette évolution.

1.4 Conclusion

Dans ce premier chapitre, nous avons présenté les principaux concepts de fouille de données (data mining), les processus, les tâches et les méthodes les plus utilisés en data mining ainsi que les avantages et les inconvénients de chaque méthode.

Dans notre travail nous nous intéressons aux techniques de la classification automatique, ou segmentation[1] (clustering). Nous avons vu que la segmentation (clustering) permet de regrouper des objets (individus ou variables) en un nombre limité de groupe ou de classes (segmentes, ou clusters). Les détails, font l'objets du chapitre suivant.

[1] Nous utiliserons dans ce manuscrit les deux terminologies.

CHAPITRE 2

MÉTHODES DE CLASSIFICATION

2.1 Introduction

D'un point de vue général, les méthodes de classification ont pour but de regrouper les éléments d'un ensemble $X = \{X^1,...,X^n,...,X^N\}$ en un nombre K optimal de classes selon leurs ressemblances **[Lure-03]**.

En effet, les objets sont souvent répertoriés par rapport à des catégories ou des classes auxquelles ils sont censés appartenir. Cette appartenance est, la plupart du temps, vague et/ou graduelle.

De manière générale, les problèmes de classification s'attachent à déterminer des procédures permettant d'associer une classe à un objet (individu). Ces problèmes se déclinent essentiellement en deux variantes selon Bezdek **[Bezd-93]** : la classification dite " *supervisée* " et la classification dite " *non supervisée* ".

La classification, supervisée ou non, en tant que discipline scientifique, n'a été automatisée et massivement appliquée **[Khod-97]**. Comme la plupart des activités scientifiques, l'essor des différentes techniques de classification a largement bénéficié de l'avènement et du perfectionnement des outils informatiques. De nos jours, la classification est une démarche qui est appliquée dans d'innombrables domaines. Un autre nom possible pour cette branche de la recherche est la typologie, et la science qui lui est associée est la taxinomie. Les méthodes de classification ont pour but de regrouper les éléments d'un ensemble *X*, de nature quelconque, en un nombre restreint de classes. La qualité de la classification peut être jugée sur la base des deux critères suivants :

- les classes générées doivent être les plus différentes possibles les unes des autres vis-à-vis de certaines caractéristiques,

- chaque classe doit être la plus homogène possible vis-à-vis de ces caractéristiques.

L'objectif de ce chapitre n'est pas de faire une revue exhaustive des méthodes existantes. Nous proposons d'expliciter quelques méthodes classiques, en utilisant pour cela la taxinomie introduite par Bezdek, Hall et Clarke dans **[Bezd-93]**. Ainsi, nous divisons la présentation de ce chapitre en deux parties : dans une première section, nous détaillons quelques méthodes supervisées, puis nous étudions quelques algorithmes non supervisés en insistant tout particulièrement sur les algorithmes de classification flous. Par la suite, nous présentons un tableau qui étudie les inconvénients et les avantages de chaque algorithme pour la caractérisation des tissus cérébraux dans les images anatomiques (**voir annexe**).

2.2 Méthodes supervisées

Ces sont des méthodes dans lesquelles les classes sont connues *a priori* avant d'effectuer l'opération d'identification des éléments de l'image. Elles demandent une **phase d'apprentissage** sur l'échantillon représentatif dans le but d'apprendre les caractéristiques de chaque classe et une autre phase pour décider l'appartenance d'un individu à telle ou telle classe.

Dans le cas qui nous intéresse, les données segmentées de l'ensemble d'apprentissage proviennent d'un étiquetage manuel des images ou des régions d'intérêt en C classes de tissus $C_1 \ldots C_c$ par un ou plusieurs experts. Chaque classe C_i se voit donc affecter un ensemble d'apprentissage E_i, et les données de l'ensemble de test sont segmentées en fonction des E_i. Puisque la structure anatomique d'un cerveau est différente d'un patient à l'autre, l'étiquetage doit être renouvelé pour chaque patient ou groupe de patients analysé, ce qui représente une tâche longue et fastidieuse pour les spécialistes.

2.2.1 Méthodes bayesiennes

La segmentation bayésienne consiste à calculer, pour chaque vecteur forme $\boldsymbol{x_j}$, la probabilité conditionnelle $P(C_i \setminus x_j)$ pour chacune des C classes $(C_1, ..., C_c)$ à l'aide de la règle Bayes :

$$(\forall i \in \{1..C\})\, P(C_i \setminus x_j) = \frac{P(C_i).P(x_j/C_i)}{\sum_{k=1}^{C} P(C_k).P(x_j/C_k)} \qquad 2.1$$

La segmentation bayésienne admet :

- soit une approche paramétrique, dans laquelle l'intensité d'un voxel est considérée comme la combinaison linéaire de probabilités d'appartenance attachées à chaque classe (en effectuant l'hypothèse que les données sont conformes à des distributions paramétriques). La probabilité conditionnelle $P(x_j/\mathrm{C_i})$ est modélisée par une fonction dépendant d'un vecteur de paramètre θ**.** Le problème est alors pour chaque classe C_i d'estimer le meilleur paramètre θ connaissant l'ensemble d'apprentissage E_i de cette classe.

- Soit une approche non paramétrique, dans ce type d'approche, les probabilités conditionnelles $P(x_j/\mathrm{C_i})$ sont supposées quelconques. La forme non paramétrique doit permettre de rendre compte le plus fidèlement possible de la réelle distribution statistique des niveaux de gris dans l'image.

2.2.2 Champs de Markov

Les *champs de Markov* et plusieurs méthodes de minimisation pour la classification sont présentés par Pony *et al* **[Pony-00]**.

La segmentation par *champ de Markov* étant un processus supervisé, les ensembles d'apprentissage E_i sont supposés connus pour chaque classe à segmenter. Held *et al* **[Held-97]** et Jaggi *et al* **[Jagg-97]** utilisent cette méthode pour obtenir les cartes d'appartenance aux tissus cérébraux. Le problème se ramène alors à l'estimation d'un processus label *L* (voxel[2] étiqueté) à partir d'un processus voxel connu *X*. La probabilité *a posteriori* d'un tel processus est définie par :

$$P(L = I \setminus \mathrm{X} = x) = \frac{P(X = x/\ \mathrm{L} = \mathrm{I})P(L = I)}{P(X = x)} \qquad 2.2$$

L'idée des champs de Markov est de rendre compte, à travers la formulation probabiliste, du fait que le niveau de gris d'un site donné dépend de la connaissance des états des sites voisins (leur état est supposé connu).

[2] Voxel est un point à trois dimensions

Dans ce type de méthodes, l'image est modélisée comme un ensemble de sites fini assimilable à un graphe et muni d'un système de voisinage. Une clique est alors soit un site isolé (singleton) soit un ensemble de sites tel que tous les couples de sites extraits de cette clique soient voisins. Dans un champ de Markov, la probabilité conditionnelle en un site n'est fonction que de la configuration de son voisinage. L'ensemble des valeurs possibles pour un site est l'ensemble des niveaux de gris de l'image et l'ensemble des sites est l'image elle-même (l'ensemble des voxels). Le système de voisinage utilisé est souvent la 4-connexité ou la 8-connexité. Une énergie est attribuée à chaque clique, et la somme des énergies des cliques donne l'énergie totale d'une configuration donnée. Une fois choisie le modèle d'énergie, le but est de trouver la configuration optimale qui minimise cette énergie.

Le modèle de Potts est le plus utilisé en segmentation d'images, il permet d'associer une fonction d'énergie à chaque ensemble de cliques tout en privilégiant, par exemple, le reclassement d'un voxel isolé dans la classe la plus présente dans son entourage.

La segmentation est donc un problème de minimisation de termes d'énergie, qui peut être facilement résolu par des algorithmes déterministes de type gradient ou par des algorithmes de type recuit simulé.

2.2.3 Algorithme des k plus proches voisins

L'algorithme des k plus proches voisins (*KPPV*) (*k-Nearest-Neighbors (kNN)* en anglais) est une méthode non paramétrique et supervisée de classification introduite dans **[Duda-73]**. Elle est largement utilisée en classification d'une manière générale et en segmentation d'images en particulier. Elle repose sur un principe simple et intuitif de regroupement d'individus en fonction de leur voisinage.

L'algorithme de *K Plus Proche Voisin* se base essentiellement sur les deux éléments principaux suivants :

1. le nombre de cas les plus proches (*K*) à utiliser et une métrique pour mesurer le plus proche voisin.
2. La valeur de *K* est spécifiée à chaque utilisation de l'algorithme puisqu'il détermine le nombre de cas existants qui sont considérés pour prédire un nouveau cas.

Le *K Plus Proche Voisin* est basé sur le concept de distances. Une métrique est nécessaire pour déterminer les distances, cette dernière est à la fois importante car le choix de métrique influe beaucoup sur la qualité des prédiction et arbitraire du fait qu'il n'existe pas de définition préalable sur ce qui constitue une bonne métrique.

La méthode des *k plus proches voisins* repose sur le regroupement des voxels en fonction de leur voisinage : chaque point est affecté à la classe la plus représentée parmi ses *k* plus proches voisins. Cette méthode nécessite l'établissement d'une règle de distance et la détermination du nombre de voisins à prendre en considération, ainsi qu'un ensemble d'apprentissage représentant les différentes classes.

Le processus de classement d'un voxel *v* peut être résumé comme suit :

1. Choisir nombre de voisins (un entier *k* compris entre 1 et *n*, le nombre maximum de voisins).
2. Calculer les distances $d(v,v_i)$, $i=1..n$ (où v_i est un des *n* voxels de l'image).
3. Retenir les *k* voxels pour lesquels ces distances sont les plus petites.
4. Compter le nombre de fois $k_1,...,k_m$ que ces *k* voxels apparaissent dans chacune des classes.
5. Attribuer *v* à la classe la plus représentée dans son entourage.

La probabilité *a posteriori* d'appartenance de x_j à une classe est estimée d'après la fréquence des classes présentes parmi les *k* plus proches voisins de ce vecteur. La règle de décision classique consiste à affecter x_j à la classe la plus représentée parmi ses *k* plus proches voisins. Cette règle est simple et ne nécessite pas, contrairement à la segmentation bayésienne paramétrique, d'hypothèse forte sur la loi statistique des niveaux de gris d'une classe. Seuls la distance et *k* doivent être déterminés. La première détermine la forme du voisinage de chaque vecteur x_j, et le choix de *k* est arrêté lorsqu'une valeur supérieure ne modifie pas la classification.

2.2.4 Réseaux de Neurones

Un *réseau de neurones* est un réseau d'unités élémentaires (les noeuds) interconnectées, à fonctions d'activation linéaires ou non linéaires. Ces noeuds sont regroupés pour les réseaux

multicouches en au moins deux sous-ensembles de neurones : un sous-ensemble d'entrée, un autre de sortie et éventuellement un ensemble de neurones cachés. De nombreux modèles de réseaux existent (réseaux de Hopfield, perceptrons multicouches, etc.) **[Magn-99]**, les différents nœuds étant complètement ou partiellement interconnectés aux autres. L'ensemble des liens convergeant vers un noeud constitue les connexions entrantes du noeud. Ceux qui divergent vers d'autres noeuds sont les connexions sortantes. À chaque connexion, entre des noeuds *i* et *j*, est associé un poids w_{ij} représentant la force de l'influence du noeud *i* sur le noeud *j*. L'ensemble des poids est regroupé dans un vecteur de poids synaptiques *w*. Un vecteur de scalaires ***a*** présenté à tous les noeuds d'entrée est appelé exemple. A cet exemple sont aussi associées les valeurs *y* (le vecteur de sortie) que l'on désire apprendre. Les poids des connexions sont éventuellement modifiés au cours d'un cycle d'apprentissage.

Modifier la sortie des noeuds à partir de leurs entrées consiste tout d'abord à calculer l'activation présente à l'entrée du noeud puis à calculer la sortie du noeud suivant la fonction d'activation qu'elle possède. Un réseau de neurones peut ainsi être défini pour chaque nœud par quatre éléments :

- La nature de ses entrées, qui peuvent être binaires ou réelles.
- La fonction d'entrée totale *e,* qui définit le pré-traitement *e (a)* effectué sur les entrées. Généralement, *e* est une combinaison linéaire des entrées pondérées par les poids synaptiques des connexions entrantes.
- La fonction d'activation *f* du noeud qui définit son état de sortie en fonction de la valeur de *e*. Toute fonction croissante et impaire convient et la fonction sigmoïde est souvent utilisée. La valeur de *f* en *e (a)* est redirigée vers l'extérieur ou vers d'autres noeuds où elle contribue à calculer leur état d'activation.
- La nature de ses sorties, qui peuvent être binaires ou réelles.

Deux éléments sont enfin nécessaires au bon fonctionnement du réseau : une fonction de coût et un algorithme d'apprentissage.

Les réseaux de neurones sont également utilisés pour obtenir une classification de l'image en tissus cérébraux. Ils sont organisés autour d'un ensemble de cellules (ou neurones) interconnectées. On dispose d'une base de connaissances constituée de couples (entrées, sorties)

et on utilise cette base pour entraîner une "mémoire" informatique à raisonner en prenant comme référence cette base empirique. Ainsi, en fonction de la base d'apprentissage, le réseau de neurones détermine la sortie pour chaque nouvelle entrée ; une mesure d'erreur est calculée pour chaque sortie obtenue, on cherche donc la meilleure valeur de sortie (c'est-à-dire celle qui minimise l'erreur). Ici encore, la base d'entraînement est constituée d'images segmentées par un expert, ou d'un échantillon de l'image à segmenter.

Les résultats de la classification tissulaire peuvent ensuite être utilisés comme point de départ pour des algorithmes de segmentation de structures cérébrales ou pour ajouter des contraintes au cours de l'exécution de ces derniers.

Les méthodes de segmentation supervisée offrent l'avantage d'être plus rapides et plus reproductibles que les méthodes manuelles. Toutefois, elles ont le désavantage de rester très dépendantes de la *base d'apprentissage*. C'est pourquoi il est intéressant de développer des algorithmes entièrement automatiques "*non supervisée*", qui présenteront l'avantage d'offrir un résultat reproductible et indépendant des actions de l'opérateur.

2.3 Méthodes non supervisées

L'intérêt des méthodes *non supervisées* est qu'elles ne nécessitent **aucune base d'apprentissage** et par là même aucune tâche préalable d'étiquetage manuel. Elles ont pour but de découper l'espace d'individus (pixel[3]) en zones homogènes selon un critère de ressemblance (critère de proximité de leurs vecteurs d'attributs dans l'espace de représentation entre les individus).

Les méthodes non supervisées peuvent être une simple automatisation des algorithmes proposés dans la section *(cf.* 2.2), l'ensemble d'apprentissage peut être remplacé par l'analyse d'une image apportant des informations complémentaires. Les auteurs utilisent l'analyse d'une image IRM en tenseurs de diffusion pour extraire des paramètres permettant d'identifier les tissus fcérébraux dans l'IRM classique.

[3] Pixel est un point à deux dimensions

Ici encore, nous ne prétendons pas être exhaustifs. Nous nous attachons plutôt à décrire les algorithmes qui nous ont paru utiles pour notre étude. En particulier, nous développons les méthodes de génération de fonctions d'appartenance des voxels aux classes de tissus. La présentation détaillée d'une de ces classes de méthodes, les algorithmes de clustering, permettra dans un second temps d'introduire la solution que nous avons développée.

2.3.1 Algorithmes de classification non flous

2.3.1.1 Algorithmes C-moyennes ("Hard C-Means" ou HCM)

Dans la méthode des C-moyennes (HCM) **[Dunn-74]**, **[Bezd-81]**, un élément de X est attribué à une classe et une seule parmi les C proposées. Dans ce cas, la fonctionnelle à minimiser est :

$$J(B,U,X)=\sum_{i=1}^{C}\sum_{j=1}^{N}u_{ij}d^2(x_j,b_i) \qquad 2.3$$

u_{ij} : Le degré d'appartenance d'un élément x_i à une classe i.

$d^2(x_j,b_i)$: C'est la distance euclidien entre (point) l'élément de la classe et le centre de la classe.

Les solutions au problème s'écrivent de la façon suivante :

$$u_{ij}=\begin{cases} 1 \text{ ssi} \quad d^2(x_j,b_i)<d^2(x_j,b_k)\ \forall k\neq i \\ 0 \text{ sinon} \end{cases} \qquad 2.4$$

$$\text{où} \qquad b_i=\frac{\sum_{j=1}^{N}u_{ij}x_j}{\sum_{j=1}^{N}u_{ij}} \qquad 2.5$$

La classification des éléments de X s'effectue de manière itérative en alternant l'étape de classification (2.4) et l'étape de mise à jour des centres (2.5), jusqu'à stabilisation de la segmentation ou de la fonction objectif. Dans une variante nommée ILSC (Iterative Least Squares Clustering) **[Velt-95]**, les vecteurs b_i sont recalculés à chaque ajout d'un élément dans une classe. Ces auteurs rapportent qu'un bon taux de reconnaissance des tissus cérébraux en IRM est obtenu (**voir annexe**), affirment également que cette modification entraîne à la fois un temps

de calcul plus long et une dépendance de la partition finale à l'ordre de parcours de l'image **[Pena-99]**.

Dans une méthode comme HCM, les éléments sont classés de façon certaine comme appartenant à une classe et une seule. Quelle que soit la modalité d'imagerie, cette assertion ne reflète pas la réalité physique de l'échantillon étudié (bruit, volume partiel, hétérogénéité de champ (**voir annexe**)). Les méthodes présentées dans les paragraphes suivants permettent d'obtenir une segmentation floue qui prend en compte ces aspects *imprécis* et *incertains*.

2.3.2 Algorithmes de classification flous

2.3.2.1 Algorithmes C-moyennes foues ("Fuzzy C-Means" ou FCM)

L'algorithme des C-moyennes floues **[Bezd-81]**, **[Bezd-87]** effectue une optimisation itérative en évaluant de façon approximative les minimums d'une fonction d'erreur. Il existe toute une famille de fonctions d'erreur associées à cet algorithme qui se distinguent par des valeurs différentes prises par un paramètre réglable, *m*, appelé indice de flou "*fuzzy index"* et qui détermine le degré de flou de la partition obtenue. Les FCM sont un cas particulier d'algorithmes basés sur la minimisation d'un critère ou d'une fonction objectif.

Dans ce cas, les x_j ne sont plus assignés à une unique classe, mais à plusieurs par l'intermédiaire de degrés d'appartenance u_{ij} du vecteur x_j à la classe *i*. Le but des algorithmes est alors non seulement de calculer les centres de classe *B* mais aussi l'ensemble des degrés d'appartenance des vecteurs aux classes.

a. Présentation de l'algorithme FCM

❖ *Formulation du problème*

Si u_{ij} est le degré d'appartenance de x_j à la classe *i*, la matrice *CxN*, *U*= $[u_{ij}]$ est appelée matrice de *C*-partition floue si et seulement si elle satisfait :

$$(\forall i \in \{1..C\})\ (\forall j \in \{1..N\})\ \ u_{ij} \in [0,1],$$

$$0 < \sum_{j=1}^{N} u_{ij} < N, \qquad 2.6$$

$$\forall j \in \{1..N\} \sum_{i=1}^{C} u_{ij} = 1. \qquad 2.7$$

Bezdek a montré **[Bezd-81]** que le problème de partition de X en C classes floues pouvait être formulé comme la minimisation d'une fonctionnelle $J(B,U,X)$ définie par :

$$J(B,U,X) = \sum_{i=1}^{C}\sum_{j=1}^{N} (u_{ij})^{m} d^2(x_j, b_i) \qquad 2.8$$

sous les contraintes (2.6) et (2.7) où :

- m>1 est un paramètre contrôlant le degré de flou de la partition résultante,

- $(x_j, b_i) \rightarrow d^2(x_j, b_i)$ est une distance du vecteur x_j au prototype b_i.

❖ *Obtention de U et B*

Les solutions au problème de minimisation précédent sont obtenues par une méthode lagrangienne. Plus précisément, si $H(x_j)$ est défini pour chaque vecteur x_j par :

$$H(x_j) = \sum_{i=1}^{C} u_{ij}^{m} d^2(x_j, b_i) - \alpha\left(\sum_{i=1}^{C} u_{ij} - 1\right), \ \alpha > 0,$$

l'annulation des dérivées partielles par rapport à u_{ij} et α donne :

$$\frac{\partial H(x_j)}{\partial \alpha} = 0 \ et \ \frac{\partial H(x_j)}{\partial u_{ij}} = m.u_{ij}^{m-1}.d^2(x_j, b_i) - \alpha = 0,$$

de telle sorte que, si $d^2(x_j, b_i) \neq 0$:

$$u_{ij} = \left(\frac{\alpha}{m.d^2(x_j, b_i)}\right)^{\frac{1}{m-1}}$$

avec

$$\sum_{i=1}^{c} u_{ij} = \left(\frac{\alpha}{m}\right)^{\frac{1}{m-1}} \sum_{i=1}^{C} \left(\frac{1}{d^2(x_j, b_i)}\right)^{\frac{1}{m-1}} = 1.$$

Et finalement

$$u_{ij} = \left[\sum_{k=1}^{C} \left(\frac{d^2(x_j, b_i)}{d^2(x_j, b_k)} \right)^{\frac{2}{(m-1)}} \right]^{-1} \qquad 2.9$$

Si $d^2(x_j, b_i)$=0, x_j est un prototype b_i et $u_{ij} = 1$ avec $u_{kj} = 0,\ k \neq i$.

Pour *C* et *X* fixés, l'annulation des dérivées partielles *H' (B,g)* par rapport à une direction quelconque g de *B* donne enfin :

$$b_i = \frac{\sum_{k=1}^{n} u_{ik}^m . x_k}{\sum_{k=1}^{n} u_{ik}^m} \qquad 2.10$$

L'algorithme des *C*-moyennes floues (FCM) consiste alors en l'application itérée de (2.9) et (2.10) jusqu'à stabilité des solutions. Le critère d'arrêt des itérations, définissant cette stabilité, peut par exemple consister en l'étude de la norme de la matrice *U* ou en la stabilité des centres de classe sur deux itérations successives.

L'algorithme FCM a été largement utilisé pour la segmentation des images de cerveau, quels que soient la modalité et le type d'acquisition (mono ou multi-spectrale). De nombreux travaux ont notamment été effectués en imagerie par résonance magnétique **[Barr-00]**. En IRM fonctionnelle. Baumgartner *et al.* **[Baum-98]** ont ainsi utilisé FCM pour segmenter les régions activées (simulation et aires motrices) du cerveau. Les voxels étant représentés par leur niveau de gris, les auteurs ont montré non seulement que l'algorithme avait des performances comparables à l'analyse de corrélation standard (avec l'avantage de ne nécessiter aucune connaissance *a priori* sur le paradigme), mais aussi que le FCM détectait des zones effectivement activées qui restaient silencieuses avec la corrélation **[Barr-99]**.

2.3.2.2 Les variantes des C-moyennes floues

De nombreuses variations de l'algorithme FCM sont possibles, en changeant la fonctionnelle à minimiser, la métrique d^2 ou les contraintes à appliquer. Nous présentons ici quelques variantes plus particulièrement utilisées pour la caractérisation des tissus cérébraux.

❖ L'approche de Kamel et Selim

Selon **[khod-97]**, Kamel et Selim. **[Kame-91]** ont proposé une généralisation de l'algorithme qui réduit le degré de flou de la partition finale. Lorsqu'un degré d'appartenance u_{ik} du vecteur x_k à la classe *i* est inférieur à une valeur seuil, τ, il est forcé à zéro et les autres degrés d'appartenance sont ensuite normalisés de telle sorte que leur somme soit égale à 1. Pratiquement, cette approche est mise en oeuvre en ajoutant les étapes suivantes à la fin de l'algorithme de Bezdek:

- calcul du seuil (global)

$$\tau = \min(\min_i \max_k u_{ik} , \min_k \max_i u_{ik}) \qquad 2.11$$

- seuillage :

$$(si(u_{ik} < \tau) \ alors \ u_{ik} = 0) \qquad 2.12$$

- normalisation, calcul de nouveaux degrés d'appartenance u_{ik}^* *:*

$$u_{ik}^* = \frac{u_{ik}}{\sum_{j=1}^{c} u_{ik}} \qquad 2.13$$

Cette approche est connue dans la littérature sous le nom de *C*-moyennes floues à seuil (TFCM, pour Thresholded FCM). Ces étapes supplémentaires peuvent être vues comme une tâche de pré-défuzzification.

❖ Méthode FCM semi-supervisée

Les algorithmes de minimisation utilisés en classification tendent selon certains auteurs **[Bens-96]**, **[Suck-99]** à privilégier les solutions dont les nuages de points ont des populations de taille sensiblement égales. Par conséquent, les amas de faible cardinalité sont intégrés aux nuages plus importants, d'où une classification qui risque d'être anatomiquement incorrecte. Le nombre de classes étant fixé *a priori*, l'absorption d'une classe de faible cardinalité par une classe plus importante résulte nécessairement dans la création d'une classe non significative ou de la partition d'une structure en deux sous-classes. Bensaïd *et al* **[Bens-96]**, ont alors proposé une version semi-supervisée de l'algorithme des *C*-moyennes floues. Un ensemble de vecteurs d'apprentissage *A* est défini par un expert dont chaque élément appartient de façon exclusive à l'une des classes. De

plus, une pondération est attribuée à ces vecteurs, qui permet de créer des copies des vecteurs d'apprentissage et ainsi de guider la segmentation vers un résultat anatomiquement correct. Le calcul des degrés d'appartenance des vecteurs n'appartenant pas à A reste inchangé par rapport à un FCM conventionnel. En revanche, la mise à jour des centres de classe nécessite d'intégrer les vecteurs d'apprentissage au calcul :

$$b_i = \frac{\sum_{x \in A} p_i u'^{m}_{ik} x + \sum_{x \notin A} u^{m}_{ik} x}{\sum_{x \in A} p_i u'^{m}_{ik} + \sum_{x \notin A} u^{m}_{ik}} \qquad 2.14$$

où p_i est un facteur de pondération attribué aux vecteurs d'apprentissage de la classe i et u'_{ik} est égal à 1 si l'élément k appartient à la classe d'apprentissage i et 0 sinon. Cet algorithme présente deux inconvénients majeurs : il nécessite une intervention humaine pour définir l'ensemble d'apprentissage et introduit les p_i comme paramètres supplémentaires. Dans certaines applications, l'intervention de l'utilisateur peut être évitée. Suckling *et al.* [**Suck-99]** initialisent ainsi les vecteurs d'apprentissage pour la matière grise et la matière blanche d'après l'histogramme de l'image. Ce type d'initialisation non supervisée est possible dans des cas simples où, malgré un chevauchement des paramètres caractéristiques, les structures sont identifiables sur l'histogramme. Si le facteur de pondération est trop fort, le centre du nuage restera égal au barycentre des vecteurs d'apprentissage de la classe tout au long des itérations. S'il est trop faible, le comportement de l'algorithme semi- supervisé sera similaire à celui d'un FCM conventionnel. L'algorithme semi-supervisé obtient selon ses auteurs de meilleurs résultats qu'un algorithme de *KPPV* utilisé avec le même ensemble d'apprentissage. Il est donc moins sensible au choix d'un ensemble d'apprentissage médiocre et produit dans ce cas de meilleures segmentations.

❖ FCM et contraintes de voisinage

Mohamed *et al.* **[Moha-98]** proposent une version modifiée de FCM pour la prise en compte du bruit dans la segmentation d'images. Les auteurs s'inspirent des relations de voisinage exprimées dans la modélisation par champ de Markov, et modifient dans (2.8) la distance entre le vecteur x_j et le prototype b_i en tenant compte des voisins du voxel j selon la formule :

$$d^2\left(x_j, b_i\right) = d^2\left(x_j, b_i\right)\left(1 - \alpha \frac{\sum_{k \in voi\,\mathrm{sin}\,age} u_{ik} h_{jk}}{\sum_{k \in voi\,\mathrm{sin}\,age} h_{jk}}\right) \qquad 2.15$$

où $\alpha \in [0,1]$ et h_{jk} est une mesure de la distance entre les voxels j et k. L'intérêt d'une telle approche est qu'elle permet d'utiliser les informations spatiales, grâce auxquelles chaque voxel tente d'attirer son voisin dans sa classe de plus grande appartenance, créant par là-même des régions homogènes. Cet effet peut cependant entraîner un biais dans les régions de fort volume partiel (zone d'ambiguïté entre deux tissus par exemple), où les voxels de la frontière sont alors classés dans l'une ou l'autre classe plutôt que partagés pour estimer le mélange des tissus **[Barr-00]**.

2.3.2.3 Algorithmes les C-moyennes possibilistes ("Possibilistic C-means" ou PCM)

Krishnapuram et Keller ont proposé une approche possibiliste des C-moyennes appelée *Possibilistic C-Means*, ou PCM **[Kris-93]** [**Bezd-97]**. Leur approche est censée conduire à une meilleure performance en présence de bruit. Mais leur travail est motivé essentiellement par le désir de remédier au caractère relatif des degrés d'appartenance générés par les FCM. En effet, ces derniers sont interprétés en tant que degrés de vérité relatifs décrivant l'appartenance d'un vecteur quelconque à chacune des classes possibles. Un élément à classer est donc, en quelque sorte, partagé entre ces différentes classes. A cette idée de partage Krishnapuram et Keller préfèrent substituer la notion de typicalité. En effet, le résultat d'un regroupement devrait décrire la parenté absolue entre un objet et chacune des c classes possibles, indépendamment du lien entre cet objet et les $(c-1)$ classes restantes.

Rappelons que l'algorithme des FCM minimise une somme pondérée des carrés des distances entre les vecteurs à regrouper et les centres des classes :

$$J_m(U, v) = \sum_{k=1}^{n} \sum_{i=1}^{c} \left(u_{ik}\right)^m \left(d_{ik}\right)^2 \qquad 2.16$$

Le degré d'appartenance d'un élément quelconque à une classe donnée doit naturellement être d'autant plus élevé que le vecteur en question est un élément typique de cette classe. La levée

de cette condition conduirait à des fonctions d'appartenance triviales identiquement nulles dans le cas des FCM. L'élimination de l'interférence entre les différents prototypes dans l'approche possibiliste, qui est beaucoup moins contraignante, et qui exige seulement qu'un, au moins, des différents degrés d'appartenance d'un vecteur quelconque ne soit pas nul, nécessite donc la définition d'un nouveau critère d'optimisation. Une nouvelle fonction erreur, à minimiser, a été proposée par Krishnapuram et Keller **[Kris-93]** :

$$J_m(U,v) = \sum_{i=1}^{c}\sum_{j=1}^{n} {u_{ij}}^m {d_{ij}}^2 + \sum_{i=1}^{c}\eta_i \sum_{j=1}^{n}\left(1-u_{ij}\right)^m , \qquad 2.17$$

où les paramètres η_i sont des nombres positifs homogènes à des distances carrées. Plus précisément, η_i est le carré de la distance séparant le centre de la classe *i* de l'ensemble des points dont le degré d'appartenance à cette même classe est égal à 1/2. La minimisation du deuxième terme implique des degrés d'appartenance les plus élevés possibles, évitant ainsi les solutions triviales. L'estimation des paramètres η_i est mise à part les cas les plus simples, la tâche la plus ingrate de cette approche. Comme les degrés d'appartenance aux différentes classes sont, à présent, décorrélés, un centre donné peut évoluer selon n'importe quelle trajectoire indépendamment de tous les autres centres. On se ramène alors, dans ce cas, à la minimisation de *c* critères partiels indépendants. Un tel critère, $J_{m,i}(U,v)$, correspondant à la classe *i* est de la forme :

$$J_{m,i}(U,v) = \sum_{j=1}^{n} {u_{ij}}^m {d_{ij}}^2 + \eta_i \sum_{j=1}^{n}\left(1-u_{ij}\right)^m \qquad 2.18$$

Signalons que (2.17) n'est, évidemment, qu'un exemple possible de critère pouvant servir à un regroupement de type possibiliste. Sauf indication contraire, c'est celui-ci qui sera discuté dans la suite. A titre d'exemple, Krishnapuram et Keller **[Kris-96]** donnent un autre critère possible :

$$J_m(U,v) = \sum_{i=1}^{c}\sum_{j=1}^{n} {u_{ij}}^m {d_{ij}}^2 + \sum_{i=1}^{c}\eta_i \sum_{j=1}^{n}\left(u_{ij}.\log\left(u_{ij}\right)-u_{ij}\right) \qquad 2.19$$

Cependant, c'est (2.17) qu'ils ont essentiellement appliqué et commenté.

La convergence de l'algorithme est encore assurée à la suite d'un certain nombre d'itérations dans lesquelles les degrés d'appartenance sont mis à jour à l'aide de la formule suivante :

$$\left(\forall i \in \{1..C\}\right)\left(\forall j \in \{1..N\}\right)\ u_i(x) = \frac{1}{1+\left(\frac{d^2(x_j, v_i)}{\eta_i}\right)^{\frac{1}{m-1}}} \qquad 2.20$$

Les centres des groupes sont toujours calculés de la même façon que dans le cas des FCM :

$$\left(\forall i \in \{1..C\}\right)\ v_i = \frac{\sum_{k=1}^{n} u_{ik}^m . x_k}{\sum_{k=1}^{n} u_{ik}^m} \qquad 2.21$$

Comme on peut le voir sur l'équation (2.20), à chaque itération la valeur de u_{ij} ne dépend donc plus que de la distance, d_{ij}, séparant le vecteur x_j du centre v_i, ce qui est conforme à l'esprit de l'approche possibiliste. La position de x_j relativement aux centres des classes autres que C_i n'interfère donc plus avec le calcul de u_{ij}.

L'algorithme résultant de l'application itérée de (2.20) et de (2.21) constitue l'algorithme de classification possibiliste, noté dans la suite PCM.

L'utilisation de l'algorithme de classification possibiliste en imagerie médicale d'une manière générale, et pour la caractérisation de tissus cérébraux en particulier, est peu fréquente. Masulli et Schenone. **[Masu-99]** ont proposé de combiner une approche par réseaux de neurones avec l'algorithme PCM pour segmenter les tissus cérébraux et des entités pathologiques (méningiomes). Barra et Boire. **[Boir-00]** ont également appliqué l'algorithme PCM en IRM sur des vecteurs forme x_j.

Certaines heuristiques floues utilisent l'histogramme de l'image pour extraire les degrés d'appartenance de chaque voxel aux classes de tissus. La segmentation par *"fuzzy classifiers"*, appliquée dans le cas où les voxels sont représentés par leurs niveaux de gris, permet de construire des fonctions d'appartenance à partir des points caractéristiques de l'histogramme[4] de

[4] Histogramme des occurrences de niveaux de gris dans

l'image. La génération des "*fuzzy classifiers*" repose sur l'analyse de la forme de l'histogramme **[Meda-98]**. Des fonctions caractéristiques (floues ou non) sont assignées à chaque classe en normalisant l'histogramme ou en fonction des points remarquables (maxima, points d'inflexion). Il est ainsi possible d'associer à chaque classe de voxels un ensemble flou, une masse ou une probabilité représentés par exemple par un trapèze.

Par exemple, Colin *et al* **[Coli-99]** calculent des distributions de possibilité trapézoïdales dont les sommets sont déterminés par les points d'inflexion de l'histogramme. Si les abscisses de ces points et du niveau de gris d'intensité maximale sont notées *A*, *B*, *C*, *D* et *E*, et si les abscisses des sommets des trapèzes sont numérotées p_1, p_2, p_3, p_4, les distributions de possibilités des classes LCS, de MB et MG peuvent être construites, sur une acquisition donnée, sur l'histogramme de la figure 2.1 par :

Classe LCS : $p_1 = 0,\ p_2 = 0,\ p_3 = \mathrm{A},\ p_4 = \mathrm{B},$

Classe MG : $p_1 = \mathrm{C} - (\mathrm{D}\text{-}\mathrm{C})/2,\ p_2 = \mathrm{C},\ p_3 = \mathrm{E},\ p_4 = \mathrm{E},$

Classe MB : $p_1 = \mathrm{B} - (\mathrm{C}\text{-}\mathrm{B})/2,\ p_2 = \mathrm{B},\ p_3 = \mathrm{C},\ p_4 = \mathrm{D}.$

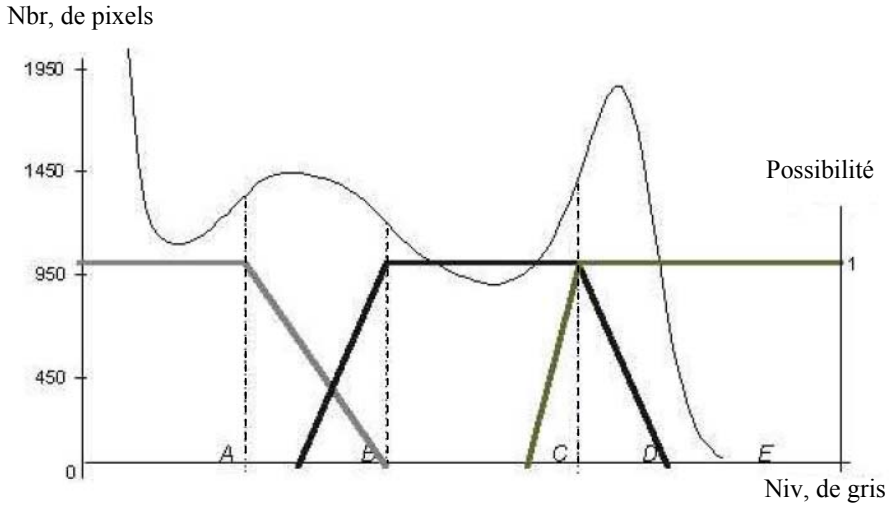


Figure 2. 1 *: Distribution de possibilités des tissus dérivées de l'histogramme de l'image* **[Barr-00]**, **[Fren-04]**.

Les valeurs p_i ont été sélectionnées par ajustements successifs des trapèzes sur les points d'inflexion jusqu'à ce que le modèle paraisse visuellement satisfaisant à l'auteur. Enfin, dans le cadre de la théorie des croyances, Bloch **[Bloc-96b]** construit des fonctions de masse sur des histogrammes d'images IRM multi-échos pour la caractérisation de l'adrénoleukodystrophie. Les hypothèses simples et composées sont déduites de connaissances à propos des informations contenues dans les deux images (contrastes parenchyme/ventricules et parenchyme/pathologie), et une fonction de masse trapézoïdale est assignée à chacune d'entre elles en fonction de l'histogramme des images. Le modèle est simple, mais se révèle suffisant pour la segmentation.

D'après l'étude des différentes méthodes citées dans ce chapitre, nous avons relevé (*cf.* Tableau 2.1) des inconvénients et des avantages pour chacune d'entre elles. Ce qui permis de dire donc qu'il n'existe pas de méthode de classification complète "*résultats optimaux"*.

		Approches	Structures à segmenté	Contraintes	inconvénients	Avantages
Statistiques (Classification)	Supervisées	**Champs de Markov [Held-97]**	*MG, MB, LCR, MB/MG* *LCS/MG*	*Base d'apprentissage*	*Sensibilité*	*Prennent* *En compte les relations spatiales*
		Kppv [Duda-73]	*MG, Mb, LCR*	*Le paramètre K*	*Temps de calcul élevé*	*Simplicité*
		Réseaux de neurones **[Magn-99]**	*Putamen, noyau caudé,* *Corps calleux*	*Base d'apprentissage*	*Intervention de opérateur*	*Apprentissage pour chaque image*
	Automatiques	**HCM [Bezd-81]**	*MG, MB, LCR*	*Centres des classes*	*Ne prend pas l'incertain et l'imprécis*	*Automatique*
		FCM [Bezd-81]	*MG, MB, LCR*	*Centres des classes*	*Degrés d'appartenance relatifs*	*Prend l'incertain et l'imprécis*
		PCM [Kris-96]	*MG, MB, LCR*	*Centres des classes*	*Choix des paramètres*	*Degrés d'appartenance absolus*

Tableau2. 1 *: Comparaison entre méthodes de classification* **[Semc-07]**.

2.4 Conclusion

La classification d'images cérébrales étant intrinsèquement limitée (effet de volume partiel, bruit, contraste insuffisant, etc.), aucun algorithme n'émerge dans la littérature comme supérieur aux autres. De plus, en l'absence de référence absolue, il est très difficile d'évaluer les performances des méthodes de classification.

Avec tous ces avantages et ces limites, nous avons opté pour la ***coopération*** de méthodes pour tirer partie des avantages de chacune, l'intérêt de telles approches est qu'elles exploitent la complémentarité d'informations en proposant un système de segmentation complet.

La *fusion de données* en IRM peut être effectuée à trois niveaux : au niveau pixel, au niveau caractéristique ou au niveau prise de décision. Au niveau pixel il y a les méthodes utilisant le recalage comme la fusion de données d'IRM-TEP (Tomographie d'Émission de Positron). Au niveau caractéristique il y a des techniques de segmentation à base de connaissances comme celles proposées par Clark [**Clar-98].** Enfin au niveau de la prise de décision la théorie des possibilités propose un modèle pour la fusion **[Barr-00] [Bloc-03]**.

Dans ce qui suit, nous allons présenter les principes de fusion d'informations et nous allons dresser un panorama des principales méthodes.

CHAPITRE 3

LA FUSION DE DONNÉES

3.1 Introduction

Dans beaucoup de nos décisions quotidiennes nous prenons en compte plusieurs informations. Un médecin, qui analyse des examens radioscopique et échographique, fait mentalement une fusion de données qu'il extrait de chacun des supports. On comprend donc tout l'intérêt de trouver des modèles mathématiques de fusion de données qui permettent de réaliser automatiquement ces combinaisons afin de faciliter les analyses, d'améliorer les prises de décision, voire de les rendre automatiques. Le concept de fusion est compréhensible, sa modélisation est plus complexe.

Les bases théoriques de la fusion de données ont été établies dans les années 1960, avec les recherches de Zadeh, Shafer et Dempster **[Leco-05]**. Le développement de la fusion s'est réalisé dans les années 1980. Elle est maintenant utilisée dans des domaines très variés comme le domaine médical, le domaine militaire, la technologie spatiale et la robotique.

Cette science récente, initiée aux Etats-Unis, connaît actuellement un fort développement. Cet essor est dû à la multiplication des sources d'informations et des données disponibles, et à une demande toujours plus forte de rapidité de traitement. La fusion répond aussi à une demande d'aide à la prise de décision **[Leco-05]**.

Nous présentons dans la section suivante trois théories permettant d'intégrer la représentation des connaissances *incertaines* et/ou *imprécises* : la théorie des probabilités, la théorie des possibilités et la théorie des croyances. Ceci nous permet d'introduire dans un second temps le concept de fusion de données, d'exprimer cette notion dans les trois cadres théoriques cités.

3.2 Divers contextes théoriques possibles pour la fusion de données

Divers cadres théoriques sont utilisés pour représenter les informations en fusion de données. Certains utilisent la théorie des croyances, la théorie des probabilités ou la théorie des possibilités. Le but de ce chapitre n'est pas de présenter de manière exhaustive chacun de ces contextes théoriques, mais de proposer une rapide présentation de chacun à titre d'exemple.

3.2.1 La théorie des probabilités

Les probabilités, du fait de leur ancienneté, développées au 19^{éme} siècle pour la mécanique classique **[Fémé-03]**, sont encore la base théorique la plus utilisée pour la représentation des données incertaines. La relation entre l'information des données et les hypothèses envisagées est représentée par une distribution de probabilité conditionnelle.

Plusieurs distributions de probabilités peuvent ensuite être combinées à l'aide de la règle de Bayes :

Soit $H = \{H_1, ..., H_N\}$, l'ensemble des hypothèses. Les H_i sont mutuellement exclusives. La probabilité *a posteriori* d'un événement H_i parmi les N hypothèses connaissant l'information I_j est donnée par

$$P(H_i \setminus \mathrm{I_j}) = \frac{P(H_i).P(I_j/\mathrm{H_i})}{\sum_{k=1}^{N} P(H_k).P(I_j/\mathrm{H_k})} \qquad 3.1$$

où $P(H_i)$ est la probabilité *a priori* de l'hypothèse H_i, *et* $P(I_j/\mathrm{H_i})$ représente la probabilité d'observer l'information I_j lorsque l'hypothèse H_i est réalisée.

Les lois $P(I_j/\mathrm{H_i})$ et $P(H_i)$ sont en pratique rarement connues. Elles sont souvent estimées à partir des données. Les probabilités $P(H_i)$ sont déterminées par l'expérience ou par une analyse d'exemples (base d'apprentissage) et les probabilités conditionnelles $P(I_j/\mathrm{H_i})$ sont estimées par des lois statistiques. Ces dernières peuvent être paramétriques. Dans ce cas une forme est choisie pour $P(I_j/\mathrm{H_i})$ et ses paramètres sont estimés (par maximum de vraisemblance par exemple), ou non paramétriques (par exemple fenêtres de Parzen **[Duda-73]**). Si l'hypothèse

d'une forme paramétrique simplifie grandement le problème, elle est souvent peu justifiée dans les cas réels et peut induire de nombreuses erreurs.

3.2.2 La théorie des croyances

D'après **[Abde-02]**, La théorie de l'évidence fut historiquement introduite par Shafer [**Shaf-76**]. Mais les origines de la théorie sont attribuables à Dempster **[Demp-67]**, **[Demp-68]**, par ses travaux sur les bornes inférieure et supérieure d'une famille de distributions de probabilités. À partir du formalisme mathématique développé, Shafer **[Shaf-76]** a montré l'intérêt des fonctions de croyance pour la modélisation de connaissances incertaines.

La théorie des croyances est une généralisation de la théorie bayésienne au traitement de grandeurs incertaines **[Appr-91]** et consiste en la quantification de la crédibilité attribuée aux faits observés.

Une fonction de masse est attribuée à chaque événement :
Soit X, un ensemble de N hypothèses H_i exclusives et exhaustives, 2^x désigne l'ensemble des 2^N sous-ensembles A_j de X.

Selon **[Bloc-03]**, une fonction de masse élémentaire, m, est définie de 2^X sur [0,1] par :

$$m(\phi) = 0 \qquad 3.2$$

$$\sum_{J=1}^{2^N} m(A_j) = 1 \qquad 3.3$$

Les *éléments focaux* sont les sous-ensembles dont la masse est non nulle. Lorsque ces derniers se réduisent aux singletons H_i, la notion de masse élémentaire est assimilable à celle de probabilité. L'apport de la théorie des croyances est donc de permettre l'évaluation conjointe d'ensembles quelconques d'hypothèses. Les événements considérés ne sont donc plus nécessairement exclusifs : par exemple, on pourra évaluer H_1, H_2 et H_3, mais également $H_1 \cup H_2$ et $H_2 \cup H_3$ de manière à prendre en compte une composante informative de la mesure propre à discriminer H_1 et H_3, mais insensible à H_2. Une masse $m(A_j)$ est représentative de la

vraisemblance attribuable à l'un des éléments du sous-ensemble A_j, sans aucun discernement possible entre les différents éléments de A_j. En particulier, $m(X)$ désigne le degré d'incertitude ou d'ignorance totale.

Une fonction de crédibilité, Cr, peut également être définie sur les mêmes ensembles par:

$$Cr(\phi) = 0 \qquad 3.4$$

$$Cr(X) = 1 \qquad 3.5$$

$$Cr(A_k) = \sum_{A_j \subset A_k} m(A_j) \qquad 3.6$$

Les fonctions de masse élémentaire et de crédibilité sont définies et utilisables de façon indépendante. Il existe donc une bijection entre l'ensemble des fonctions de masse élémentaire et l'ensemble des fonctions de crédibilité qui associe à chaque jeu de masses sur 2^X un jeu de crédibilités sur le même ensemble.

A partir de la notion de masse élémentaire, on peut également introduire une fonction de plausibilité, Pl :

$$Pl(A_k) = \sum_{A_j \cap A_k = \phi} m(A_j) \qquad 3.7$$

Cette fonction mesure à quel point les informations données par une source ne contredisent pas A_k. En fait, il a été démontré **[Shaf-76]** que la connaissance d'une des trois fonctions (m, Cr, Pl) sur 2^X était suffisante pour en déduire les deux autres.

De façon intuitive, la crédibilité peut être interprétée comme une mesure de vraisemblance minimale d'un événement, et la plausibilité comme une mesure de vraisemblance maximale.

Pour un événement A, l'intervalle $[Cr(A), Pl(A)]$ encadre la probabilité mal connue $P(A)$.

3.2.3 La théorie des possibilités

La théorie des possibilités a été introduite en 1978 par Zadeh **[Zade-78]** puis développée par Dubois et Prade en France **[Dubo-80]**, **[Dubo-88]**. De même que la théorie des croyances, elle constitue un cadre permettant de traiter des données à caractère *imprécis* ("cet arbre est très haut") et/ou *incertain* ("il va pleuvoir demain").

Elle a été bâtie à partir de la notion d'ensemble flou introduite par Zadeh **[Zade-65]** pour permettre de raisonner en fonction de l'incertitude et de l'imprécision **[Dubo-01]**. Dans un premier temps, nous introduisons donc les bases de logique floue permettant de bâtir la théorie des possibilités

3.2.3.1 Notion des sous -ensembles flous [Bouc-93] [Kauf-73]

- **Définitions**

Un sous-ensemble flou de X est une classe d'éléments de X dont les frontières ne sont pas binaires, comme par exemple la recherche des individus "plutôt blonds ou roux" dans une population. Un sous-ensemble flou peut être plus ou moins précis ou spécifique.

La notion de *sous-ensemble flou* permet de grader l'appartenance d'un élément à une classe, c'est-à-dire de permettre à cet élément d'appartenir plus ou moins fortement à cette classe. Par abus de langage, on parle généralement *d'ensemble flou* pour désigner un sous-ensemble flou.

Un *sous-ensemble flou* A de X est défini par une fonction d'appartenance qui associe à chaque élément x de X le degré $f_A(x)$, compris entre 0 et 1 avec lequel x appartient à A **[Zade-65] [Pedr-98]**:

$$f_A : X \to [0,1]$$

$f_A(x)$ est le degré d'appartenance de x à A.

Le support de A (supp(A)) est l'ensemble des éléments de X ayant un degré d'appartenance non nul à A : $supp\ (A) = \{x \in X / f_A(x) \neq 0\}$

La hauteur de A (*h (*A)) est le plus fort degré avec lequel un élément de X appartient à A. C'est la plus grande valeur prise par sa fonction d'appartenance : $h(A) = \sup_{x \in X} f_A(x)$

A est dit *normalisé* si sa hauteur est égale à 1 (il existe au moins un élément de X appartenant de façon absolue à A).

L'ensemble des sous éléments appartenant de façon absolue à A est appelé le *noyau* de A (noy(A)) : $noy\ (A) = \{x \in X / f_A(x) = 1\}$.

Ces différentes définitions sont illustrées dans la figure 3.1.

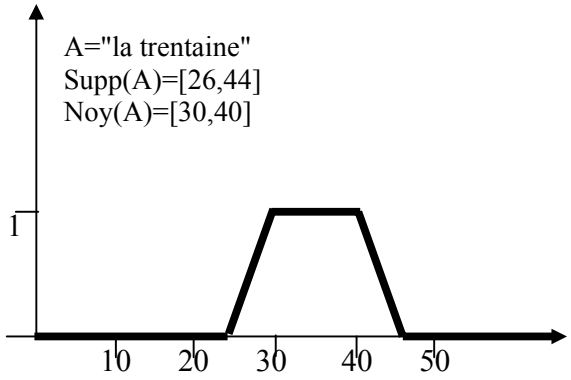


Figure 3. 1 *: Exemple de sous-ensemble flou* **[Fren-04]**

- **Opérations sur les sous-ensembles flous [Dou-06] [Fren-04]**

Soient deux sous ensembles flous, A et B d'un ensemble X.

A et B sont dits *égaux* si leurs fonctions d'appartenance prennent la même valeur en tout point de X :

$$\forall x \in X \qquad f_A(x) = f_B(x)$$

A est *inclus* dans B $(\mathrm{A} \subseteq \mathrm{B})$ si tout élément x de X appartenant à A appartient également à B avec un degré au moins aussi grand :

$$\forall x \in X \qquad f_A(x) \leq f_B(x)$$

L'*intersection* entre A et B est le sous-ensemble flou constitué des éléments de X affectés du plus petit de leurs deux degrés d'appartenance donnés par f_A et f_B :

$$\text{si } C = A \cap B, \forall x \in X \qquad f_C(x) = \min(f_A(x), f_B(x)) \quad (\textit{voir Figure 3.2})$$

L'*union* entre A et B est le sous-ensemble flou constitué des éléments de X affectés du plus grand des deux degrés d'appartenance donnés par f_A et f_B :

si $D = A \cup B, \forall x \in X$ $\qquad f_D(x) = \max(f_A(x), f_B(x))$ *(voir Figure 3.2)*

Le *complément* du sous-ensemble flou A, A^c, est défini comme le sous-ensemble de X de fonction d'appartenance $f_A{}^c$ définie par :

$\forall x \in X$ $\qquad f_A{}^c(x) = 1 - f_A(x)$

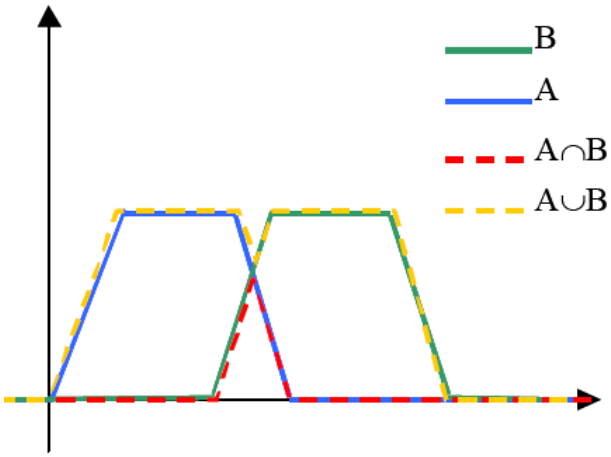


Figure 3. 2 *: Exemple d'intersection et d'union* **[Fren-04]**

- **Les α-coupes**

Il est souvent intéressant de pouvoir se référer au sous-ensemble ordinaire de X défini par un sous-ensemble flou A de X afin de pouvoir utiliser des critères de prise de décision ou des connaissances de la théorie des ensembles classique, par exemple. La façon la plus simple de créer un sous-ensemble flou correspondant à A est de fixer une limite inférieure, notée *a*, aux degrés d'appartenance pris en considération.

Pour toute valeur α de [0,1], on définit la α-coupe A_α d'un sous-ensemble flou A de X comme le sous-ensemble $A_\alpha = \{ x \in X \ / \ f_A(x) \geq \alpha \}$, de fonction caractéristique $\chi_{A\alpha}$ telle que :

$\chi_{A\alpha} = 1$ *si et seulement si* $f_A(x) \geq \alpha$

La théorie des sous-ensembles flous ne permet toutefois pas de prendre en compte les imprécisions et inexactitudes liées aux données. La théorie des possibilités a donc été introduite en 1978 par Zadeh **[Zade-78]**, en liaison avec la théorie des sous-ensembles flous, pour pallier cet inconvénient, en introduisant un moyen de prendre en compte les incertitudes sur les connaissances.

3.2.3.2 Mesure et distribution de possibilité

Soit un ensemble de référence fini X. Une *mesure de possibilité* Π est définie sur l'ensemble des parties de $X(P(X))$ et prend ses valeurs dans [0,1] telle que :

$$\Pi(\phi) = 0, \Pi(X) = 1 \qquad 3.8$$

$$\forall A_1 \in P(X), A_2 \in P(X), \ldots \qquad \Pi(\cup_{i=1,2,\ldots} A_i) = \sup_{i=1,2,\ldots} \Pi(A_i) \qquad 3.9$$

Dans le cas de deux éléments, la condition (3.9) s'écrit :

$$\forall A, B \in P(X)^2 \qquad \Pi(A \cup B) = \max(\Pi(A), \Pi(B)) \qquad 3.10$$

Ce qui exprime que la réalisation de l'un des deux événements, pris indifféremment, est affectée du même coefficient de possibilité que la réalisation de l'événement le plus possible.

Un événement est tout à fait possible si sa mesure de possibilité est égale à 1, et impossible si celle-ci est nulle.

Si une mesure de possibilité permet de déterminer le degré avec lequel l'union d'événements dont on sait à quel point ils sont possibles, sera elle-même un événement possible, elle ne permet pas de se prononcer sur l'intersection d'événements. Cependant, les conditions (3.8) et (3.9) imposent que le coefficient attribué à l'intersection des événements soit majoré par le plus petit des coefficients attribués à chacun d'entre eux.

Si l'on étudie un événement X et son contraire, l'un des deux au moins est tout à fait possible et :

$$\forall A \in P(X) \qquad \max(\Pi(A), \Pi(A^c)) = 1 \qquad 3.11$$

Une mesure de possibilité est totalement définie si un coefficient de possibilité est attribué à chaque partie de l'ensemble de référence X, ce qui suppose d'obtenir $2^{|X|}$ coefficients. C'est pourquoi, elle est définie plus simplement en indiquant les coefficients attribués seulement aux singletons de X, puisque leur union permet de constituer n'importe quel sous-ensemble de X et que l'on sait calculer les coefficients de possibilité de ces sous-ensembles à partir de la condition (3.9).

On introduit donc une fonction appelée *distribution de possibilité*, π, qui associe à tout événement de X, un coefficient compris entre 0 et 1 reflétant le degré avec lequel cet événement est possible. Cette fonction doit vérifier la condition de normalisation suivante :

$$\sup_{x \in X} \pi(x) = 1 \qquad 3.12$$

Une distribution de possibilité est un ensemble flou qui peut être associé à la mesure de possibilité bijectivement, et l'on a :

$$\forall x \in X \qquad \pi(x) = \Pi(\{x\}) \qquad 3.13$$

3.2.3.3 Mesure de nécessiter

Une mesure de possibilité fournit une information sur l'occurrence d'un événement *A* relatif à un ensemble de référence *X*, mais elle ne suffit pas pour décrire l'incertitude existante sur cet événement. Par exemple, si $\Pi(A) = 1$, il est tout à fait possible que A soit réalisé, mais on peut avoir en même temps $\Pi(A^c) = 1$, ce qui exprime une indétermination complète sur la réalisation de A, ou $\Pi(A^c) = 0$, c'est-à-dire que l'on est certain de la réalisation de A.

Pour compléter l'information sur A, on indique le degré avec lequel la réalisation de A est certaine par l'intermédiaire d'une *mesure de nécessité*.

Une mesure de nécessité, N, est une fonction définie sur l'ensemble $P(X)$ des parties de *X*, à valeurs dans [0,1], telle que :

$$N(\phi) = 0, N(X) = 1 \qquad 3.14$$

$$\forall A_1 \in P(X), A_2 \in P(X), \ldots \qquad N(\cap_{i=1,2,\ldots} A_i) = \inf_{i=1,2,\ldots} N(A_i) \qquad 3.15$$

Dans le cas de deux parties de X, (3.15) devient :

$$\forall(A, B) \in P(X)^2 \qquad N(A \cap B) = \min(N(A), N(B)) \qquad 3.16$$

Sur un ensemble de référence X, une mesure de nécessiter N peut être obtenue à partir de la mesure de possibilité Π correspondante, par l'intermédiaire du complémentaire A^c de toute partie A de X :

$$\forall A \in P(X) \qquad N(A) = 1 - \Pi(A^c) \qquad 3.17$$

Plus un événement A est affecté d'une grande nécessité, moins l'événement complémentaire A^c est possible, donc plus on est certain de la réalisation de A.

3.3 La Fusion de Données

3.3.1 Définition

D'après **[Leco-05]**, **[Arif-05]**, **[Dou-06]**, le concept de la fusion de données est facile à comprendre, mais il est difficile d'en trouver une définition qui rende compte de ce cadre formel et de ses multiples facettes.

La définition proposée par le Joint Directors of Laboratories (JDL), du ministère de la défense aux Etats-Unis d'Amérique **[Jdl-91]**. Elle est appelée "*le modèle JDL*". Ce modèle définit la fusion de données comme : «*un processus multi-niveaux à facettes multiples ayant pour objet la détection automatique, l'association, la corrélation, l'estimation et la combinaison d'informations de sources singulières et plurielles»*. Le modèle JDL est extrêmement populaire dans le domaine militaire **[Anto-90]**, **[Wald-02]**.

Le groupe européen SEE (Société d'électricité et d'électronique), la branche française de l'Institute of Electric and Electronics Engineers (IEEE), et la branche européenne de l'International Society for Photogrammetry and Remote Sensing (ISPRS) ont proposé la définition suivante : «la *fusion de données constitue un cadre formel dans lequel s'expriment les moyens et techniques permettant l'alliance des données provenant de sources diverses».* Cette définition met clairement l'accent sur le concept et non plus sur les méthodes, techniques ou stratégies **[Wald-02]**.

La définition qui nous sembles la plus adaptée dans notre travail est celle de **[Bloc-94]**, **[Bloc-03]** : «*la fusion de données consiste à combiner des informations issues de plusieurs sources afin d'améliorer la prise de décision»*.

3.3.2 Intérêt de la fusion de données

Les progrès permanents de la science et de la technologie et aussi de l'informatique permettent de disposer d'informations de plus en plus nombreuses et complexes, de nature et de fiabilité différentes. Il est de plus en plus demandé aux systèmes d'information et de communication ou de décision (au sens large) d'aider ou de coopérer avec les experts du domaine applicatif dans le but de décider. La fusion de données permet de répondre à ces demandes et de résoudre ces problèmes.

La fusion de données est un sujet de plus en plus actuel, car susceptible d'aider efficacement les scientifiques à extraire des informations de plus en plus pertinentes et précises.

D'après **[Walt-90]**, la fusion de données offre de nombreux avantages :

- Robustesse et fiabilité, le système est opérationnel même si une ou plusieurs sources d'informations sont défectueuses,
- augmentation de la couverture spatiale et temporelle de l'information et des déductions,
- accroissement du nombre de dimensions de l'espace des observations, menant à un accroissement de la qualité des déductions, et à une réduction de la vulnérabilité du système,
- réduction de l'ambiguïté des déductions, des informations plus complètes ou plus précises permettent un meilleur choix entre les différentes hypothèses,
- apport d'une solution à l'explosion de la quantité d'informations disponibles aujourd'hui.

La fusion de données permet de gérer une multitude d'informations, complémentaires, redondantes et incomplètes, issues de sources hétérogènes, afin d'obtenir la " **meilleure** " connaissance possible de l'environnement de décision étudié. Cette synergie n'est possible que si l'on est capable d'évaluer la connaissance ou l'information contenue dans chacune des sources de données, **[Faou-04]**.

3.3.3 Le processus du fusion de données

Dans un processus de fusion, quatre phases principales sont enchaînées successivement, chacune correspondant à un ou plusieurs traitements des données à fusionner **[Barr-05]** :

3.3.3.1 Représentation homogène et recalage des informations pertinentes

Les données à fusionner sont souvent hétérogènes, il est impossible de les combiner sous leur forme initiale. On est alors amené à rechercher un espace de représentation commun dans lequel les différentes informations pertinentes disponibles renseignent sur une même entité. Un premier traitement consiste donc à transformer certaines de ces informations initiales, en informations équivalentes dans un espace commun, dans lequel s'effectuera la fusion.

3.3.3.2 Modélisation des connaissances

Chaque jeu de données propre à chaque source n'est pas forcément exploitable en tant que tel, notamment si l'information fournie est très imparfaite, et ne donne qu'un aspect de la réalité. Cependant, même imparfaite, toute information peut apporter de la connaissance sur l'état du système. Donc, une *étape essentielle* du processus de fusion consiste à modéliser et à évaluer la connaissance apportée par chaque source. Elle est couplée au choix d'un cadre théorique adapté.

3.3.3.3 Fusion

C'est à ce niveau du processus que l'opération de fusion proprement dite est réalisée. Les informations *recalées* et *modélisées* sont combinées selon une règle de combinaison propre au cadre théorique choisi. Observons notamment que la fusion d'informations contradictoires doit permettre de gérer les conflits potentiels entre les diverses sources.

3.3.3.4 Décision par choix d'une stratégie

La fusion doit permettre de choisir l'information la plus vraisemblable, au sens d'un certain critère, parmi toutes les hypothèses possibles. En ce sens, la fusion de données aboutit bien souvent à une classification (affectation d'un ensemble de mesures aux hypothèses

possibles). Le critère de décision dépend du cadre théorique dans lequel le processus de fusion a été développé, et de l'objectif à atteindre.

3.3.4 Classification des opérateurs de fusion

La notion de fusion s'applique au cas de *n* sources **[Bloc-96a]**, **[Dubo-88]**. Pour simplifier l'exposé, nous avons choisi de la réduire au cas de deux sources.

On considère le problème de l'agrégation de deux informations, n_1 et n_2, issues de deux capteurs différents pour un même phénomène. On cherche à agréger les informations fournies par n_1 et n_2 en exploitant au mieux l'ambiguïté et la complémentarité des données.

L'agrégation des données est réalisée par un opérateur binaire, $F(.,.)$, auquel on impose une contrainte de fermeture (c'est-à-dire que la valeur retournée doit être de même nature que les informations d'entrée).

Bloch propose dans **[Bloc-96a]** la définition suivante pour le comportement de F :

- *Sévère* si $F(n_1, n_2) \leq \min(n_1, n_2)$.
- *Prudent* si $\min(n_1, n_2) \leq F(n_1, n_2) \leq \max(n_1, n_2)$.
- *Indulgent* si $F(n_1, n_2) \geq \max(n_1, n_2)$.

Un opérateur peut être de type conjonctif (sévère) ou de type disjonctif (indulgent) développées dans **[Dubo-85]**. Un opérateur prudent correspond à une position intermédiaire entre ces deux extrêmes.

Un opérateur *conjonctif* est utilisé dans le cas où les sources sont en accord. La conjonction limite le domaine de validité d'une propriété à l'intersection des domaines donnés par chacune des sources (la fusion ne conservera que les informations concordantes). Cette réduction du domaine augmente le risque d'éliminer de l'information pertinente mais donne un résultat plus précis, car seul la partie de l'information issue des différentes sources est conservé. Inversement, un opérateur *disjonctif* est utilisé plutôt dans le cas de sources en désaccord : le domaine de validité est alors l'union des deux domaines, le résultat obtenu risque donc d'être peu précis, mais plus certain (la fusion conservera la totalité des informations).

Il existe différents types d'opérateurs, suivant leur variabilité et leur dépendance au contexte. Les exemples cités dans la partie qui suit sont extraits de la classification effectuée dans **[Bloc-96a]**.

3.3.4.1 Opérateurs à comportement constant et indépendant du contexte

Ce sont des opérateurs ayant le même comportement quelles que soient les valeurs de n_1 et n_2 à agréger. Le résultat de la fusion est indépendant du contexte de l'agrégation. L'opérateur F est donc exclusivement sévère, indulgent ou prudent. Dans la suite, cette classe sera notée ***CCIC***.

3.3.4.2 Opérateurs à comportement variable et indépendant du contexte

Ce sont les opérateurs qui ne dépendent pas du contexte mais dont le résultat est fonction des valeurs de n_1 et n_2. Par exemple, les sommes symétriques sont des opérateurs à comportement Variable et indépendant du contexte, elles sont de la forme : $\sigma=\frac{g(x,y)}{g(x,y)+g(1-x,1-y)}$

où g est une fonction de $[0,1]\times[0,1]$ dans $[0,1]$, croissante, positive et continue, telle que $g(0,0)=0$. Cette classe sera notée ***CVIC*** dans la suite du manuscrit.

3.3.4.3 Opérateurs dépendants du contexte

La valeur retournée par F ne dépend plus seulement de n_1 et n_2 mais aussi d'une connaissance *a priori* sur le système de capteurs ou sur le phénomène étudié. Il est ainsi possible de construire des opérateurs dont le comportement sévère (resp. indulgent) est une fonction croissante (resp. décroissante) de l'accord entre les deux capteurs. On a ici à faire à un problème d'accord entre les sources. Cette classe d'opérateurs, qui sera notée ***CDC***, nous intéressera tout particulièrement dans la suite.

Lorsque les deux sources sont en conflit, le comportement de l'opérateur devra être conjonctif dans le cas où le conflit est faible, disjonctif s'il est très marqué et réalisé un compromis pour les cas intermédiaires.

Si l'on note π_1 et π_2 les deux distributions de possibilité à fusionner et π' la distribution résultante, les opérateurs peuvent être de la forme :

$$\pi'(s) = \max\left[\frac{i[\pi_1(s), \pi_2(s)]}{h(\pi_1, \pi_2)}, 1 - h(\pi_1, \pi_2)\right]$$

$$\pi'(s) = \min\left[1, \frac{i[\pi_1(s), \pi_2(s)]}{h(\pi_1, \pi_2)} + 1 - h(\pi_1, \pi_2)\right]$$

$$\pi'(s) = i[\pi_1(s), \pi_2(s)] + 1 - h(\pi_1, \pi_2)$$

où $h(\pi_1, \pi_2)$ est une mesure globale d'accord entre π_1 et π_2, i est un opérateur conjonctif.

3.3.4.4 Quelques propriétés

Un opérateur de fusion doit, si possible, être associatif et commutatif, afin d'être indépendant de l'ordre de présentation des informations, il doit également être continu, ce qui assure la robustesse de la combinaison pour des couples voisins, et enfin il doit être strictement croissant par rapport au couple (n_1, n_2).

3.3.5 Fusion en théorie des probabilités

Bien que la théorie des probabilités soit plus orientée vers un objectif d'estimation (recherche d'une valeur représentative à partir d'événements observés), il est possible d'envisager une fusion purement probabiliste. Le fait de manipuler des densités de probabilité étant plus familier que d'utiliser des notions de masse ou de distribution de possibilité, cette théorie est encore largement employée **[Drom-98]**. Nous détaillons dans la suite la méthode de fusion bayésienne **[Piat-96]**.

3.3.5.1 Combinaison

En reprenant les notations de la section (*cf.* 3.2.1), nous considérons ici que les capteurs fournissent deux observations n_1 et n_2 d'un événement H_i. La règle de Bayes (3.1) permet de calculer la probabilité d'obtenir l'événement H_i en ayant effectivement mesuré n_1 et n_2 :

$$P(H_i \backslash n_1, n_2) = \frac{P(H_i).P(n_1, n_2/H_i)}{\sum_{k=1}^{C} P(H_k).P(n_1, n_2/H_k)} \qquad 3.18$$

La fonction de vraisemblance $P(n_1,n_2/H_i)$ décrit la probabilité que le premier capteur observe n_1 et que le second observe n_2, étant donnée la vraie valeur de H_i. Si les mesures n_1 et n_2 sont issues de variables aléatoires indépendantes, cette fonction peut s'écrire comme un produit des deux mesures $P(n_1, n_2/H_i) = P(n_1/H_i).\ P(n_2/H_i)$, de sorte que (3.18) devient :

$$P(H_i \backslash \mathrm{n}_1, \mathrm{n}_2) = \frac{P(H_i).P(\mathrm{n}_1/\mathrm{H}_i)P(n_2/H_i)}{\sum_{k=1}^{C} P(H_k).P(\mathrm{n}_1/\mathrm{H}_k)P(n_2/H_k)} \qquad 3.19$$

L'équation (3.19) décrit un opérateur de fusion entre les données fournies par les deux capteurs. Cet opérateur est conjonctif et *CCIC*, comme produit de probabilités. Il est de plus associatif, commutatif et continu si les probabilités composant la règle le sont. Notons que d'autres opérateurs ont été introduits dans la littérature, dont le comportement est par exemple disjonctif et *CCIC* **[Dubo-99]**.

3.3.5.2 Règle de décision

Bien que de nombreux critères de décision bayésienne existent (les règles du maximum de vraisemblance, du maximum d'entropie, etc.), la règle la plus utilisée reste celle du maximum *a posteriori*. Elle consiste à privilégier l'événement ayant la plus forte probabilité *a posteriori* à l'issue de l'étape de fusion (3.19). Le résultat pour chaque couple de mesure est une affectation à l'événement H_i le plus probable.

3.3.6 Fusion en théorie des croyances [Bloc-03] [Lefe-00]

Nous considérons ici que les deux capteurs fournissent chacun une information sur un même cadre de discernement *X*. Pour chaque capteur $i \in \{1,2\}$ un jeu de masse m_i est construit, exprimant au mieux l'information apportée par chaque source. En particulier, un des intérêts de la

théorie des croyances est d'assigner des masses sur des hypothèses composées $H_j \cup H_k$ lorsque le capteur *i* ne peut distinguer entre ces deux hypothèses.

3.3.6.1 Combinaison

La combinaison de m_1 et m_2 en une masse résultante *m* est réalisée par la règle orthogonale de Dempster **[Shaf-76]** :

$$(\forall H \subset X, H \neq \phi)\ m(H) = (m_1 \oplus m_2)(H) \frac{\sum_{A_1 \cap A_2 = H} m_1(A_1) m_2(A_2)}{K} \qquad 3.20$$

$$m(\phi) = 0, K = 1 - \sum_{A_1 \cap A_2 \neq \phi} m_1(A_1) m_2(A_2) \qquad 3.21$$

K est une mesure d'accord entre les deux capteurs, si *K* est proche de 0, le conflit entre les sources est important et le fait même d'agréger les masses devient discutable. De plus, la règle est discontinue pour des valeurs de *K* proches de 0 **[Dubo-88]** et certains auteurs suggèrent même de ne pas normaliser **[Smet-90]**.

La règle de combinaison (3.20) implique une fusion *CCIC* conjonctive. L'opérateur est associatif, commutatif et peut être facilement généralisé au cas de la fusion de *n* sources. D'autres règles ont été proposées **[Smet-90]**, entraînant en particulier un comportement disjonctif (3.22), mais sont moins utilisées que la combinaison (3.20).

$$(\forall H \subset X, H \neq \phi) m(H) = (m_1 \oplus m_2)(H) = \sum_{A_1 \cup A_2 = H} m_1(A_1) m_2(A_2) \qquad 3.22$$

3.3.6.2 Règle de décision

La décision se fait en faveur de la classe qui a, soit la plus grande crédibilité (choix le plus pessimiste), soit la plus grande plausibilité (choix le plus optimiste), soit le maximum de crédibilité sans recouvrement des intervalles de confiance intervient le coefficient (sans risque d'erreur). Certains auteurs ont également proposé de prendre la décision en faveur d'une hypothèse composée (par exemple pour prendre en compte les effets de volume partiel en IRM **[Bloc-96b]**.

La théorie des croyances est également très utilisée pour la fusion d'images médicales **[Appr-91]**, [**Drom-98]**.

3.3.7 Fusion en théorie des possibilistes

L'information apportée par les deux capteurs est modélisée par deux distributions de possibilité. Tout l'intérêt de la théorie des possibilités s'exprime alors, puisqu'une gamme riche et variée d'opérateurs est disponible. L'idée générale derrière une approche possibiliste de la fusion d'informations est qu'il n'existe pas de mode unique de combinaison **[Barr-00]**, **[Fren-04]**. Tout dépend de la situation étudiée et de la confiance accordée aux capteurs.

3.3.7.1 Combinaison

De nombreux auteurs ont proposé une étude précise et détaillée des opérateurs de fusion dans le cadre possibiliste **[Bloc-96a]**, **[Dubo-99]**, **[Bouc-95]**, et nous rappelons ici brièvement les points essentiels.

Entre les comportements extrêmes conjonctif et disjonctif, il existe de nombreux opérateurs (familles paramétrées, formes explicites) *CCIC*, *CVIC* et *CDC* qui peuvent être classés à l'aide d'une relation d'ordre partiel :

$$(F_1 \le F_2) \Leftrightarrow \left((\forall n_1, n_2)\, F_1(n_1, n_2) \le F_2(n_1, n_2)\right) \qquad 3.23$$

Les opérateurs *CCIC* sont exclusivement sévères *(T*-normes, généralisant les intersections aux ensembles flous), prudents (moyennes, agissant comme un compromis entre les capteurs), ou indulgents (*T*-conormes, généralisant les unions aux ensembles flous). Il est de plus possible de construire des familles paramétrées décrivant l'ensemble des *T*-normes ou des *T*-conormes [**Dubo-88],** mais le caractère constant de leur comportement interdit de passer d'une classe à l'autre. Au contraire, il est possible d'élaborer des opérateurs paramétrés *CVIC* décrivant l'ensemble des comportements, du plus conjonctif au plus disjonctif, suivant les valeurs des mesures des deux capteurs .

Enfin, les opérateurs *CDC* prennent en compte une information contextuelle (conflit entre les sources, confiance accordée aux capteurs, contexte spatial, etc.) et adoptent un comportement dépendant de cette information. Sandri. **[Sand-91]** propose plus formellement une stratégie possibiliste de combinaison *CDC* se fondant sur la catégorisation des fiabilités des sources (une

ou deux sources fiables, connues ou non, conditionnement des informations d'une source par rapport à l'autre).

D'une manière générale, la construction d'un opérateur *F* peut être contrainte pour qu'il satisfasse les propriétés énoncées dans le paragraphe (*cf.* 3.3.4.4), et en particulier la contrainte de fermeture. L'opérateur doit en effet créer une distribution de possibilité à partir de deux distributions données. Dans certains cas (les *T*- normes notamment), il peut être nécessaire de modifier *F*, puisqu'une *T*-norme peut produire un résultat sous-normalisé $\left(\max\left(\pi(x)\right)<1\right)$, d'autant plus sous-normalisé que les sources sont en conflit. Le résultat n'est alors plus une distribution de possibilité $\left(\sup_{x\in X}\pi(x)=1\right)$ et deux hypothèses sont envisageables :

- Supposer que les deux sources sont fiables et normaliser le résultat par une mesure de conflit (élimination des informations affirmées par une source et rejetées par l'autre). Cette règle est souvent appliquée avec diverses mesures de conflit.

- Considérer que, dès qu'il y a conflit, une des deux sources se trompe et envisager une fusion à comportement disjonctif.

3.3.7.2 Règle de décision

La décision se fait en affectant à l'élément x la classe C_i pour laquelle le degré de possibilité est maximum, ou bien, pour le degré de nécessité est maximum, correspondant au critère le plus sévère.

3.4 Conclusion

Nous avons exposé dans ce chapitre trois modèles de représentation des connaissances ambiguës : les théories des probabilités, des possibilités et des croyances. Ces dernières nous ont ensuite guidé dans l'expression de la fusion dans les cadres formels correspondant, en terme d'analyse du comportement des agrégations et de prise de décision. Dans la suite de ce mémoire nous allons présenter notre contribution en justifiant le choix de méthodes de fusion de données et l'algorithme de classification adoptés, ainsi que les outils et les paramètres de chaque algorithme.

CHAPITRE 4

CONTRIBUTION

4.1 Introduction

Le concept de fusion de données, tel que nous l'envisageons, a été introduit dans la section (*cf.* 3.3). Il s'agit, pour nous, "*d'une agrégation d'informations ambiguës, conflictuelles, complémentaires et redondantes, autorisant une interprétation des données plus précise et/ou moins incertaine*". Cette définition exclut toute confusion avec des méthodes ne combinant pas réellement les informations et parfois qualifiée de fusion, comme le recalage ou la superposition d'images **[Bari-94]**, **[Cond-91]**, **[Levi-89]**, **[Stok-97]**.

Nous supposons dans la suite que les diverses images à fusionner sont déjà recalées**,** c'est-à-dire placé dans un même repère géométrique. La méthode de recalage se doit d'être suffisamment précise, d'une part pour ne pas accroître l'imprécision et l'incertitude présentes dans les images, et d'autre part pour pouvoir combiner des informations provenant effectivement de la même localisation anatomique.

La fusion d'images peut alors se décomposer en trois grandes étapes :

- Modélisation des informations dans un cadre théorique commun.
- Fusion des informations issues de la modélisation précédente.
- Prise de décision.

La figure 4.1 schématise les différentes étapes de fusion de données de notre approche.

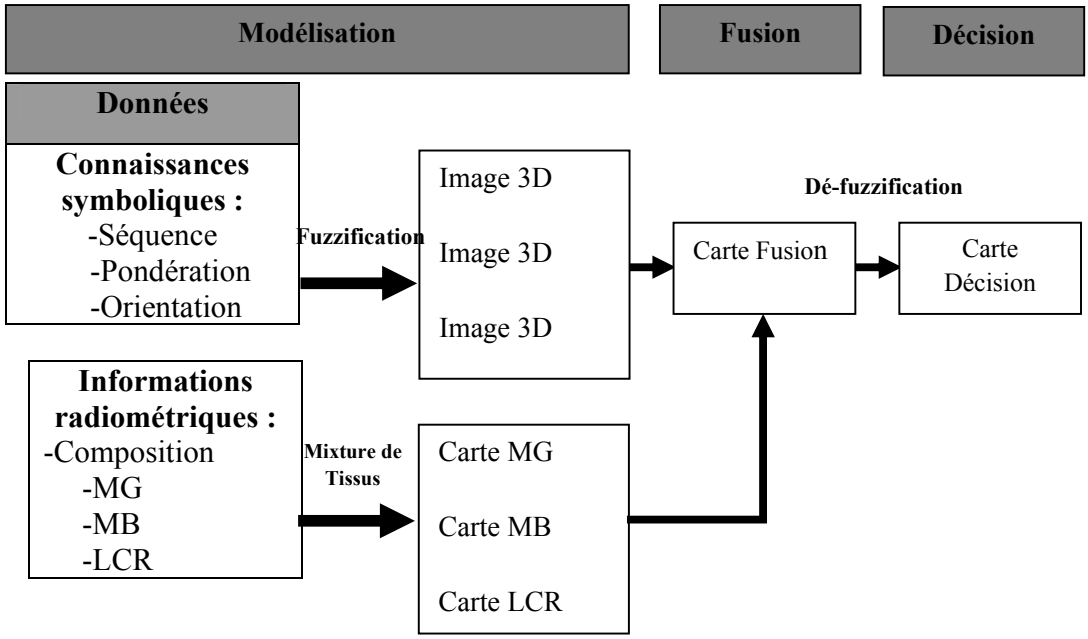


Figure 4. 1 *: Les étapes de la Fusion de données de notre proposition.*

4.2 Modélisation

L'étape de modélisation consiste en la représentation de l'information dans un cadre mathématique lié à une théorie particulière **[Bloc-94]**, **[Bloc-96a]**.

Notre but ici n'est pas de déterminer la meilleure méthode, mais la méthode la plus adaptée dans le contexte de la fusion des tissus cérébraux.

4.2.1 Choix de l'algorithme

4.2.1.1 Quel type de méthode ?

En reprenant la taxinomie de Bezdek *et al* **[Bezd-93]**. Nous devons tout d'abord préciser si l'algorithme doit être supervisé ou non. L'emploi d'un algorithme supervisé nécessite, comme nous l'avons vu dans la section (*cf.* 2.2), une base d'apprentissage pour chaque classe. Cela constitue un premier inconvénient de ce type de méthodes [**Barr-99]**. Pour notre approche, puisque nous aurons dans la suite à segmenter des images IRM, la création de cette base peut s'avérer fastidieuse pour les experts. De plus, les méthodes supervisées en segmentation d'images médicales peuvent être très dépendantes de la base d'apprentissage **[Barr-99]**. Clarke *et al*.

[Clar-93] ont par exemple comparé les méthodes de *KPPV* avec réseau de neurones sur la segmentation d'images IRM de cerveau, ont remarqué que les différences entre les segmentations sont faibles lorsque les tissus sont bien différenciés par leurs paramètres.

Plus généralement, de nombreux auteurs notent que des petites différences dans le jugement de l'expert lors de la phase d'apprentissage peuvent causer de grandes variations dans les résultats, ce qui rend les méthodes supervisées inadaptées pour des études quantitatives en particulier lorsque la base d'entraînement est mal adaptée **[Phil-95]**.

L'intérêt des méthodes non supervisées est qu'elles ne nécessitent aucune phase d'apprentissage ou d'étiquetage manuel préalable. La seule intervention de l'expert se situe à la fin du processus pour identifier les classes calculées avec les classes biologiques.

Notes ces raisons nous ont poussé plutôt à adopter une méthode non supervisée. Parmi cette classe d'algorithme, nous nous sommes intéressés aux algorithmes de clustering **[Barr-00]**, **[Fren-04]**.

4.2.1.2 Classification floue ou non floue ?

Di Gesu et Romeo. **[Gesu-94]** ont comparé la segmentation d'images IRM obtenue à l'aide de 4 algorithmes non flous (analyse d'histogramme, *k* plus proches voisins, méthode de division/fusion et partitionnement simple), ont déduit que la classification non floue est mal adaptée dans le cas où les nuages de points se chevauchent ou lorsque l'information disponible est vague et incertaine **[Barr-00]**. Les images de cerveau traitées dans ce mémoire présentent ces deux caractéristiques, que ce soit dans les zones de transition entre tissus (effet de volume partiel, information imprécise et vague) ou en raison du bruit présent dans l'image (information incertaine). Nous avons alors opté pour une approche floue pour segmenter les tissus cérébraux.

4.2.1.3 C-moyennes floues ou algorithme possibiliste ?

Dans cette partie, nous présentons notre approche, en expliquant les raisons qui nous ont amené à effectuer une **coopération** entre l'algorithme flou FCM et l'algorithme possibiliste PCM. Cela a pour but de rendre l'algorithme de classification plus robuste face aux imprécisions et aux données aberrantes. Mais avant, nous donnons d'abord les caractéristiques de chaque

algorithme, en faisant une étude comparative entre le FCM et le PCM (pour dégager les avantages et les limites de chaque algorithme).

1. Interprétation des degrés d'appartenance

❖ *FCM et degrés d'appartenance relatifs*

La contrainte de normalisation (2.7) utilisée pour la minimisation de la fonctionnelle (2.8) est source d'erreur dans l'interprétation des degrés d'appartenance issus du FCM. Krishnapuram et Keller dans **[Kris-93]** donnent une série d'exemples simples qui illustrent les problèmes associés à cette contrainte, que nous résumons dans la figure 4.2.

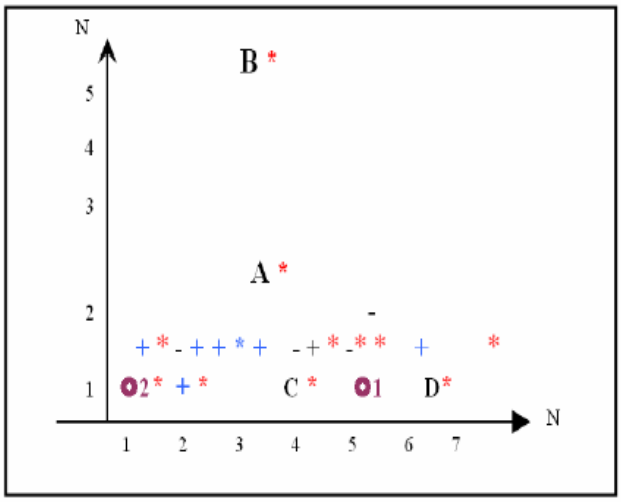


Figure 4. 2:*Démonstration de la mauvaise interprétation des degrés d'appartenance du FCM* **[Barr-00]**.

Cette figure présente deux nuages de points et deux points aberrants **A** et **B**. L'algorithme des C-moyennes floues appliqué à deux classes (***C*=2**) avec ***m*=2** et une distance euclidienne calcule les centres (triangles) et donne des degrés d'appartenance de chaque point aux deux classes présentés dans le tableau 4.1.

Points	**A**		**B**		**C**		**D**	
u_{ij} **FCM**	0.53	0.47	0.52	0.48	0.62	0.38	0.82	0.18

Tableau 4. 1 *: Degrés d'appartenance générés par le FCM sur l'exemple.*

Une analyse des résultats du tableau 4.1 montre que :

- Les points **A** et **B**, résultant par exemple d'une erreur de mesure ou d'un bruit, ne doivent pas avoir des degrés d'appartenance significatifs aux deux classes. De plus, les degrés d'appartenance de **B** doivent être plus faibles que ceux de **A**, ce point étant encore plus éloigné des classes. Néanmoins, en dépit de ces considérations intuitives, la contrainte (2.7) impose aux degrés d'appartenance de **A** et **B** aux deux classes d'être approximativement égaux à 0.5.

- Les degrés d'appartenance des points **C** et **D**, tels qu'ils sont calculés par FCM, sont significativement différents alors que ces deux points sont approximativement à la même distance du centre de la classe de droite. Cela est encore une fois dû à la contrainte de normalisation (2.7) qui force **C** à partager un peu de son appartenance à la classe de droite, dont il est plus proche que **D** ne l'est.

Ces exemples simples montrent que les degrés d'appartenance u_{ij} générés par les C-moyennes floues sont des degrés relatifs, dépendant non seulement de l'appartenance du vecteur j à la classe i, mais aussi à toutes les autres classes $k \in \{1 \ldots C\} \backslash \{i\}$. Cela s'observe d'ailleurs dans l'expression (2.9). Les réels u_{ij} générés par le FCM ne sont donc pas représentatifs, ils expriment plutôt un degré de partage des vecteurs dans toutes les classes. Ils ne peuvent pas, de plus, distinguer entre un point au bord des classes et un point aberrant (exemple des points **A** et **B**). Ce caractère relatif contraste avec l'idée de fonction d'appartenance énoncée par Zadeh **[Zade-65]**.

❖ ***PCM et degrés d'appartenance absolus***

La valeur de u_{ij} dans (2.20) ne dépend maintenant que de la distance du vecteur x_j à la classe i. Les degrés d'appartenance générés par PCM ne sont ainsi plus des degrés relatifs ou de partage, ils deviennent des valeurs absolues ("*typicality*" **[Kris-96])** reflétant la force avec laquelle chaque vecteur appartient à toutes les classes. L'application de l'algorithme de classification possibiliste sur le même ensemble de données de la figure 4.2 donne les résultats présentés dans le tableau 4.2.

Points	A		B		C		D	
u_{ij} FCM	0.53	0.47	0.52	0.48	0.62	0.38	0.82	0.18
u_{ij} PCM	0.26	0.18	0.09	0.06	0.77	0.13	0.76	0.07

Tableau4. 2 *: Comparaison des degrés d'appartenance générés par FCM et PCM.*

Les degrés d'appartenance générés par PCM reflètent de manière plus exacte la réalité de la distribution des points : **A** et **B** se voient affecter des degrés d'appartenance faibles aux deux classes. **C** et **D** ont un degré d'appartenance par rapport la première classe quasi identique.

D'après **[Lure-03]**, l'algorithme PCM permet la prise en compte des observations éloignées en leur associant une possibilité d'appartenance plus faible que le coefficient d'appartenance de nature probabiliste défini par l'algorithme FCM. D'autres algorithmes ont été développés afin d'améliorer la robustesse des algorithmes de regroupement floue aux points éloignés. Par exemple, **[Ohas-84]** considère une classe de points éloignés (classe de rejet) et modifie en conséquence la définition du critère d'optimisation.

Selon **[Khod-97]**, aujourd'hui encore, les problèmes liés à la validité de ces techniques, ou le traitement de données en présence de bruit n'ont pas trouvé de réponses satisfaisantes. Les FCM, comme toutes les approches de regroupement, sont sensibles au bruit. La version proposée par Krishnapuram et Keller **[Kris-93]** est censée entre autres améliorer les performances des FCM en présence de bruit.

2. Etude comparative entre FCM et PCM

- La différence principale, selon Krishnapuram et Keller [**Kris-96**], entre les FCM et les PCM réside dans le fait que les algorithmes du premier type cherchent avant tout une partition, alors que les algorithmes possibilistes identifient des régions denses de l'espace, indépendamment du nombre de classes présumé **[Khod-97]**.

- Le comportement des FCM face à une telle incertitude est plus déterministe : l'ensemble des données est avant tout divisé en exactement c régions, sans aucune

garantie sur la représentativité des centres obtenus. Ceci est inéluctablement achevé en fusionnant des classes différentes ou en subdivisant d'autres. Dans le cas d'un ensemble monoclasse, en imposant la valeur 2 pour ***c***, les FCM produisent deux classes différentes, quoique artificielles, alors qu'avec les PCM les deux classes obtenues seraient confondues. L'approche possibiliste semble donc moins sujette aux aléas dus à l'incertitude sur le nombre des classes, et pourraient éventuellement aider à trouver la bonne valeur de ***c***, résolvant ainsi un des problèmes de la validité de la classification **[Kris-94]**.

- Selon **[Khod-97]**, nous pouvons conclure que les PCM ne sont certainement pas robustes. Même lorsque la partition initiale est choisie extrêmement proche de la partition attendue, cette approche aboutit parfois à des résultats inacceptables, comme des centres très proches. Cette tendance à rapprocher les centres est confirmée sur l'ensemble IRIS (Indexation et Reconnaissance d'Images par la Sémantique), ainsi que lors de la segmentation d'images couleur. Contrairement à ce que soutiennent Krishnapuram et Keller, les PCM ne convergent pas toujours vers des centres de régions denses, même lorsqu'elles sont initialisées dans de bonnes conditions.

Malheureusement, tout comme le FCM, l'algorithme PCM souffre de quelques inconvénients. Pour cela nous proposons d'utiliser les résultats du FCM pour initialiser le PCM puisque ce dernier demande une initialisation plus précise que le FCM. C'est-à-dire, Le PCM peut être donc utilisé dans un deuxième passage pour les points aberrants, après l'application d'un autre algorithme de regroupement, tel que les FCM, qui fournit une partition initiale pour les PCM. Ce dernier améliore cette partition obtenue suite à la première étape.

4.2.2 Choix des paramètres de l'algorithme

Nous utilisons l'algorithme de classification possibiliste pour segmenter les tissus cérébraux dans les images IRM. Pour cela, il nous faut définir les différents paramètres gouvernant la méthode, à savoir les valeurs de m et C, le choix des poids η_i, la métrique utilisée et enfin les vecteurs forme représentant les pixels des images.

4.2.2.1 Initialisation de l'algorithme

Le problème de la classification par FCM ou PCM s'exprime comme une minimisation d'une fonctionnelle, sous certaines contraintes. Les algorithmes sous-jacents n'assurent pas l'optimalité de la solution, puisqu'un minimum local peut être trouvé qui stoppe les itérations.

L'étape d'initialisation, conditionnant la recherche du minimum, est donc fondamentale. Plusieurs stratégies ont été proposées dans la littérature. La plus simple consiste à demander à un expert de déterminer des régions d'intérêt représentatives des centres des classes. L'algorithme converge alors vers une solution acceptable, mais l'aspect non supervisé est perdu. Peña *et al*. **[Pena-99]** ont comparé quatre méthodes d'initialisation de la matrice de partition U (méthode totalement aléatoire, approches de Forgy, de MacQueen et de Kaufmann) et affirment que le meilleur schéma d'initialisation en termes de robustesse et de qualité de la partition est celui proposé par Kaufman et Rousseeuw **[Kauf-90]**. Ces derniers proposent de sélectionner itérativement les centres de classe jusqu'à ce que C vecteurs soit choisis. Le premier vecteur retenu est le plus proche du centre de gravité de l'ensemble des vecteurs. Le centre suivant est sélectionné selon la règle heuristique du choix d'un élément promettant d'avoir autour de lui un maximum de vecteurs non encore sélectionnés.

L'utilisation d'une initialisation non supervisée présente cependant l'inconvénient de ne pas nécessairement converger vers une solution correcte. Ainsi, Velthuizen *et al*. **[Velt-93**] rapportent des résultats décevants de taux de reconnaissance de tissus cérébraux et pathologiques (tumeurs) par l'algorithme FCM, le tissu tumoral n'étant reconnu par l'algorithme que dans quatre cas sur dix. Le résultat obtenu dépend en grande partie de l'initialisation et de l'algorithme choisi. D'autre part, les algorithmes de minimisation tendent à privilégier les solutions dont les nuages de points ont des effectifs sensiblement égaux **[Bens-96]**, **[Suck-99]**.

Les résultats de PCM vont évidemment dépendre de l'étape d'initialisation, comme tout algorithme de classification. Dans cette méthode, les nuages de points n'ont que peu de mobilité, puisque d'après (2.20) chaque vecteur x_j ne "*voit*" qu'un nuage à la fois plutôt que tous simultanément. Ainsi, l'initialisation, si elle doit exister, ne doit pas être aussi précise que dans d'autres algorithmes tel FCM **[Kell-93]**. Tout algorithme (flou ou non) peut donc être utilisé, et

FCM constitue une excellente manière d'initialiser les données, puisqu'il donne accès à une première estimation de U et de B .

L'algorithme de classification possibiliste est donc utilisé dans la suite en initialisant les données avec un FCM (m=2 pour une question de rapidité de calculs) et un critère d'arrêt n'autorisant que peu d'itérations (le PCM améliorera les résultats par la suite).

4.2.2.2 Détermination du nombre de classes

Nous nous plaçons ici dans une problématique de caractérisation de tissus cérébraux. Nous cherchons donc à segmenter la matière blanche, la matière grise, le liquide cérébrospinal et éventuellement le fond de l'image. Dans le reste de ce manuscrit, C sera donc pris égal à trois ou quatre. Cependant, suivant le cas étudié, rien n'empêche d'ajouter plusieurs classes pour détecter des éventuelles entités pathologiques (tumeurs, etc.) [**Phil-95]**.

4.2.2.3 Choix du paramètre *m*

Le paramètre m contrôle le degré de flou de la partition floue U. Si m est proche de 1, la partition résultante est quasiment non floue, chaque vecteur x_j est assigné à une classe i et une seule avec un degré d'appartenance $u_{ij} = 1$. Inversement, alors que la croissance de m dans le FCM tend à augmenter le degré de partage des vecteurs aux classes (les degrés d'appartenance de x_j à chacune des C classes sont égaux à $1/C$ lorsque m tend vers l'infini). Selon Bara dans [**Barr-99]**, il n'existe pas de méthode pour optimiser de manière générale ce paramètre, chaque problème appelle un choix dépendant de la nature des données. Une valeur comprise dans l'intervalle [1.5 ; 3] est généralement acceptée afin d'assurer la convergence de l'algorithme.

4.2.2.4 Choix de la distance

La métrique utilisée dans (2.17) conditionne la forme des nuages de points à séparer. D'une manière générale, la distance d^2 est définie par

$$\left(\forall i \in \{1..C\} \right)\left(\forall j \in \{1..N\}\right)\ d^2\left(x_j, b_i\right) = \left(x_j - b_i\right)^T . A . \left(x_j - b_i\right) \qquad 4.1$$

où ***A*** est une matrice définie positive. Lorsque ***A*** est la matrice identité, d^2 est la distance euclidienne et la structure des nuages de points est sphérique. D'autres choix sont possibles pour ***A***, permettant de détecter des nuages de forme plus complexe. Nous donnons ici trois exemples rencontrés dans la littérature **[Barr-99]** :

- Gath et Geva **[Gath-89]** ont proposé une distance exponentielle fondée sur l'estimation d'un maximum de vraisemblance. Cette distance permet de générer et de détecter des nuages hyper ellipsoïdaux variant par leur forme et leur densité. Si F_i est la matrice de covariance floue associée au nuage de points *i*, et si Q_i est la probabilité *a priori* de choisir la classe *i*, la distance est calculée par

$$d^2\left(x_j, b_i\right) = \frac{\left[Det\left(F_i\right)\right]^{\frac{1}{2}}}{Q_i} \exp\left[\frac{1}{2}\left(x_j - b_i\right)^T F_i^{-1}\left(x_j - b_i\right)\right], \qquad 4.2$$

où

$$F_i = \frac{\sum_{j=1}^{N} u_{ij}^m \left(x_j - b_i\right).\left(x_j - b_i\right)^T}{\sum_{j=1}^{N} u_{ij}^m} \; et \; Q_i = \frac{1}{N}\sum_{j=1}^{N} u_{ij}^m \qquad 4.3$$

- Bezdek a introduit dans [**Bezd-81**] une distance permettant de détecter des nuages de points linéaires ou planaires. Les *C* prototypes b_i ne sont plus de simples vecteurs, ce sont des variétés linéaires de dimension *r*, $0 \le r \le N-1$. La variété linéaire de dimension *r* passant par x_j, engendrée par les vecteurs $\left\{e_{j1},...,e_{jr}\right\}$ est donnée par :

$$L_{ri} = \left\{ y \in R^n \,/\, y = c_j + \sum_{k=1}^{r} t_k e_{jk}, t_k \in R \right\} \qquad 4.4$$

Si $\{s_{ij}\}$ est une base orthonormale de cet espace, la distance est calculée par :

$$d^2\left(x_j, b_i\right) = d^2\left(x_j, L_{ri}\right) = \left[\left\| x_j - c_i \right\|^2 - \sum_{k=1}^{r} \left(\left(x_j - c_i\right).S_{ik}\right)^2 \right] \qquad 4.5$$

- Davé **[Davé-90]** a recherché des nuages de points hypersphériques en modélisant chaque centre de classe b_i par une sphère de centre c_i et de rayon r_i. La distance associée est alors

$$d^2(x_j, b_i) = \left(\|x_j - c_i\| - r_i\right)^2 \qquad 4.6$$

Pour notre cas, nous avons donc utilisé l'algorithme PCM avec la distance la plus usuelle et la plus rapide à calculer, qui est la distance euclidienne (définition initiale de PCM **[Kris-93]**). Nous supposons donc que la forme des nuages de points représentant les classes de tissus est quasi-sphérique [**Barr-99]**.

4.2.2.5 Détermination des paramètres de pondération η_i

Les paramètres η_i déterminent la zone d'influence d'un vecteur x_j pour la répartition des centres de classe. Ce vecteur aura une influence d'autant plus faible sur l'estimation de b_i que la distance $d^2(x_j, b_i)$ sera grande devant η_i. En étudiant plus précisément les itérations du PCM, η_i peut être interprété comme le carré de la distance séparant b_i de l'ensemble des vecteurs dont le degré d'appartenance à la classe *i* est égal à *0.5*. Krishnapuram et Keller proposent dans **[Kris-96]** de choisir ce paramètre égal à la distance moyenne floue intra-classe, c'est-à-dire, pour tout *i* dans *{1..C}:*

$$\eta_i = \frac{\sum_{j=1}^{N} u_{ij}^m d^2(x_j, b_i)}{\sum_{j=1}^{N} u_{ij}^m} \qquad 4.7$$

D'autres choix sont bien sûr possibles, mais n'apportent pas d'amélioration significative dans la classification finale [**Kris-96]**.

L'algorithme PCM offre la possibilité de fixer, avant itérations, des valeurs pour ces paramètres, ou de faire recalculer les η_i au cours des itérations. Dans ce dernier cas, l'algorithme PCM peut évoluer vers des situations d'instabilité [**Kris-93]**. Les auteurs recommandent alors de

fixer des valeurs avant itérations puis, si c'est nécessaire, de recalculer les η_i après convergence avec les nouvelles valeurs de U et B pour appliquer une seconde fois PCM (dans le cas de données très bruitées par exemple).

4.2.2.6 Choix des vecteurs forme

Le choix des vecteurs forme est fondamental puisque leur pertinence va permettre de discriminer les pixels entre eux. Ce choix est défini suivant le type de modalité.

L'image anatomique que nous utilisons est l'IRM. L'imagerie par résonance magnétique est une modalité d'imagerie multi-spectrale donnant accès à un grand nombre de paramètres et donc de vecteurs forme. La première caractéristique qui peut être exploitée est le signal lui-même, principalement par l'intermédiaire d'images pondérées en T_1, T_2 et en densité de protons (**voir annexe**). Le vecteur forme x_j d'un pixel j est alors formé des niveaux de gris de ce pixel dans toutes les images. Cette information est très largement utilisée en segmentation d'images, en particulier dans un cadre flou **[Suck-99**] ou non flou **[Moha-99]**. Mais elle est dans ce cas très sensible aux variations du signal dues à l'instrumentation (hétérogénéité de champ).

Kiviniitty. [**Kivi-84]** affirme que les paramètres T_1 et T_2 suffisent à discriminer correctement les tissus sains dans des images IRM. Just *et al.* **[Just-88]** partagent cette opinion, mais notent que certaines entités pathologiques (tumeurs, oedèmes) sont caractérisées par un grand nombre de valeurs dans les images pondérées en T_1, T_2 et en densité de protons, et que cette variété peut affecter l'analyse de ces entités.

4.2.3 Algorithmes utilisés dans notre approche

Algorithme FCM

Etape 1 : Fixer les paramètres.

Les entrées : $X = \left(x_j, \mathrm{j}=1..\mathrm{N}\right)$ l'ensemble des vecteurs forme, C : nombre de classes

ε: Seuil représentant l'erreur de convergence (par exemple ε=0.001)

m :Degré de flou, $m \in [1.5,3]$.

Etape 2 : Initialiser la matrice degrés d'appartenances U par valeurs aléatoires dans l'intervalle [0,1].

Etape 3 : Mettre à jour la matrice prototype B par relation $b_i \leftarrow \dfrac{\sum_{k=1}^{n} u_{ik}^{m} x_k}{\sum_{k=1}^{n} u_{ik}^{m}}$

$$J^{Ancien} \leftarrow \sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij}^{m} d^2\left(x_j, b_i\right)$$

Etape 4 : Mettre à jour la matrice des degrés d'appartenance par la relation

$$u_{ij} \leftarrow \left[\sum_{k=1}^{C}\left(\frac{d^2\left(x_j, b_i\right)}{d^2\left(x_j, b_k\right)}\right)^{2/(m-1)}\right]^{-1}$$

$$J^{Nouveau} \leftarrow \sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij}^{m} d^2\left(x_j, b_i\right)$$

Etape 5 : Répéter les étapes 3 à 4 jusqu'à satisfaction du critère d'arrêt qui s'écrit :

$$\left\| J^{Ancien} - J^{Nouveau} \right\| \le \varepsilon$$

Les sorties : La matrice d'appartenance U et les centres des classes B.

Algorithme PCM

Etape 1 : Fixer les paramètres.

Les entrées : $X = \left(x_j, \text{j} = 1..\text{N}\right)$ l'ensemble des vecteurs forme, C : nombre de classes

ε: Seuil représentant l'erreur de convergence (par exemple ε=0.001)

m :Degré de flou, $m \in [1.5,3]$, η_i : Degré de pondération.

Etape 2 : Initialiser la matrice degrés d'appartenances U par valeurs aléatoires dans l'intervalle [0,1].

Etape 3 : Mettre à jour la matrice prototype B par relation $b_i \leftarrow \dfrac{\sum_{k=1}^{n} u_{ik}^{m} x_k}{\sum_{k=1}^{n} u_{ik}^{m}}$

$$J^{Ancien} \leftarrow \sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij}^{m} d^2\left(x_j, b_i\right) + \sum_{i=1}^{C} \eta_i \sum_{j=1}^{N} \left(1 - u_{ij}\right)^m$$

Etape 4 : Mettre à jour la matrice degrés d'appartenance par la relation

$$u_{ij} \leftarrow \frac{1}{1 + \left(\dfrac{d^2\left(x_j, b_i\right)}{\eta_i} \right)^{\frac{1}{m-1}}}$$

$$J^{Nouveau} \leftarrow \sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij}^{m} d^2\left(x_j, b_i\right) + \sum_{i=1}^{C} \eta_i \sum_{j=1}^{N} (1 - u_{ij})^m$$

Etape 5 : Répéter les étapes 3 à 4 jusqu'à satisfaction du critère d'arrêt qui s'écrit :

$$\left\| J^{Ancien} - J^{Nouveau} \right\| \leq \varepsilon$$

Les sorties : La matrice d'appartenance U et les centres des classes B.

Avec tous les avantages et les limites des algorithmes de classification automatique, nous avons opté pour la **coopération** de méthodes pour tirer partie des avantages de chacune, l'intérêt

de telle approche est qu'elle exploite la complémentarité d'informations en proposant un système de classification complet.

Algorithme général de l'approche proposée

Etape 1 : Fixer les paramètres.

Les entrées : $X = \left(x_j, \mathrm{j} = 1..\mathrm{N}\right)$ l'ensemble des vecteurs forme, C : nombre de classes

ε: Seuil représentant l'erreur de convergence (par exemple ε=0.001)

m: Degré de flou, $m \in [1.5,3]$, η_i : Degré de pondération.

Etape 2 : Lancer l'algorithme FCM.

Etape 3 : Lancer l'algorithme PCM pour les points aberrants.

Les sorties : La matrice d'appartenance *U*.

4.3 Fusion

Notre proposition n'est pas de comparer les théories de façon formelle, mais plutôt de choisir le cadre le plus adéquat à la représentation des informations.

Nous justifions donc ici l'orientation théorique du reste de ce manuscrit. En particulier, nous montrons pourquoi nous avons retenu la théorie des possibilités. Ce choix est motivé d'une part, par les limitations et les inadaptations des autres modèles, et d'autre part, par les qualités intrinsèques de la théorie des possibilités vis-à-vis des fusions que nous envisageons dans la suite.

De nombreux auteurs ont comparé les théories des probabilités, des possibilités et des croyances, et ont détaillé les transformations permettant de passer d'un formalisme à l'autre **[Bloc-95]**, [**Bouc-95]**, **[Dubo-99]**, **[Fabi-96]**, **[Lass-99]**, **[Sand-91]**.

4.3.1 Limitations de la fusion probabiliste

Notre choix ici est exclusivement orienté vers le domaine de l'imagerie. Dubois, Prade et Yager proposent par ailleurs une revue bibliographique plus générale de l'utilisation de la fusion probabiliste dans **[Dubo-99]**.

La théorie des probabilités, bayésienne essentiellement, a été souvent utilisée en fusion d'images. La fusion quant à elle est réalisée à l'aide de la règle de Bayes, et la décision est prise en fonction du maximum de vraisemblance. Cependant, cette théorie présente certains inconvénients qui limitent son utilisation dans le cas qui nous intéresse. Ces limites sont résumées ci-dessous :

- dans les images traitées dans le chapitre 5, les informations sont à la fois incertaines et imprécises. Or la théorie des probabilités ne peut pas bien prendre en compte l'imprécision des données,

- le formalisme, notamment introduit dans la règle de Bayes, requiert des connaissances *a priori* sur l'occurrence de chaque phénomène par l'intermédiaire des probabilités conditionnelles et des probabilités *a priori* des événements. Des modèles peuvent représenter ces connaissances mais imposent alors des hypothèses fortes sur les étapes de modélisation (densités, hypothèse d'indépendance, etc.) et de fusion,

- la connaissance qui n'est pas probabiliste par nature est difficile à inclure. Par exemple, nous pouvons savoir qu'un capteur a une dérive dans ses mesures, mais ne pas connaître la loi de dérive par rapport au temps ou à la position. Cependant, ce phénomène doit être pris en compte pour gérer l'incertitude des données.

4.3.2 Comparaison des théories des possibilités et de l'évidence

Les définitions des théories des possibilités et de l'évidence ont un formalisme assez proche. Dans le tableau 4.3, nous proposons un comparatif de ces deux théories.

Etape de fusion	Théorie des Possibilités	Théorie de l'Evidence
Cadre de discernement $D=(C_1,C_2,C_i,C_n)$	x appartient à C_i, élément de D	x appartient à A_k sous ensemble de D, $A_k = C_i$ ou $A_k = (C_i \cup C_j)$
Modélisation	Degré d'appartenance d'un élément x à une classe $C_i \pi(x)$	Masse donnant la confiance qu'un élément x appartienne à un sous ensemble A_k du cadre de discernement, $m_{Ak}(x)$ telle que
Estimation optimiste	Degré de possibilité $\Pi_{C_i}(x) = \sup\{\pi_{C_i}(a), a \in x\}$	Plausibilité que la vérité soit dans A $P_{Ak}(x) = \sum_{B \cap A_k \neq \Phi} m_B(x)$
Estimation pessimiste	Degré de nécessité, certitude de réalisation de l'hypothèse, $N_{C_i}(x) = \inf((1-\pi_{C_i}(a)), a \neq x)$	Crédibilité = croyance que la vérité est dans A_k $Cr_{Ak}(x) = \sum_{B \subset Ak} m_B(x)$
Relation	$\Pi_{C_i}(x) = N_{C_i}(x)$ $\Pi_{C_i}(x) = 1 - N_{C_i}(\bar{x})$	$P_{A_K}(x) \geq Cr_{A_K}(x)$ $P_{A_K}(x) = 1 - Cr_{\overline{A_K}}(x)$
Combinaison	Une fonction à choisir, par exemple : $\Pi_{Ci}(x) = \min(\Pi_{Ci}(x), ..., \Pi_{Ci}(x))$ $\Pi_{Ci}(x) = \max(\Pi_{Ci}(x), ..., \Pi_{Ci}(x))$	La somme orthogonale $M_{A_K} = \frac{\sum_{j=1-I} \oplus M_{A_K}}{1-K}$
Décision	$Max\ \Pi$, $Max\ N$	$Max\ Pl$, $Max\ Cr$

Tableau4. 3 *: Comparaison des théories de la possibilité et de l'évidence* **[Leco-05]**.

4.3.3 Vers la théorie des possibilités

Nous avons comparé les théories des croyances et des possibilités par rapport au problème spécifique de la fusion d'images médicales, cette comparaison permet de se rendre compte que des différences importantes existent, ces différences sont :

- La théorie de l'évidence prend en compte les ensembles composés de plusieurs classes, ce qui permet de considérer le doute entre les classes. Il n'est pas toujours évident de choisir l'appartenance d'un élément *x* entre deux classes **[Leco-05]**.
- La richesse de l'étape de modélisation des données en théorie des croyances est indéniable et sa flexibilité en fonction des contraintes peut être particulièrement adaptée à la représentation des connaissances en imagerie médicale (choix d'hypothèses composées pour représenter le volume partiel, ou l'incomplétude de l'information). Dans les applications d'images IRM , nous n'aurons cependant pas l'opportunité d'utiliser des hypothèses composées, puisque les informations que nous extrairons de ces images pourront directement être modélisées par des hypothèses simples ou par des distributions de possibilité **[Barr-00]**.
- L'étape de combinaison en théorie des croyances se résume la plupart du temps à l'application de l'opérateur orthogonal, dans le cas où toutes les ambiguïtés peuvent être introduites à l'étape de modélisation. Au contraire, la théorie des possibilités offre une grande variété d'opérateurs ayant des comportements différents suivant la situation présentée (hétérogénéité des sources, conflit, information contextuelle). Dans les informations que nous aurons à traiter, deux images pourront être tantôt concordantes (par exemple dans les zones saines) et tantôt en conflit (par exemple dans les zones pathologiques). De plus, certaines images seront plus susceptibles que d'autres d'extraire une information pertinente sur un événement (par exemple l'IRM pour la localisation du liquide cérébrospinal). Le fait de prendre en compte dans la combinaison le conflit et la fiabilité des sources nous a paru fondamental, en particulier pour préserver l'information pertinente pour le diagnostic **[Barr-00]**. Nous nous sommes donc plutôt orientés vers la construction d'**opérateurs Dépendants du Contexte (CDC)** et avons ainsi privilégié la représentation possibiliste des données.

L'analyse des ces modèles pour les fusions envisagées dans la suite a révélé tout d'abord que la théorie des probabilités était mal adaptée, principalement en raison des effectifs d'expérimentation faibles dont nous disposons. La relative simplicité des informations à extraire dans les images et la pauvreté des modes de combinaison en théorie des croyances ont été ensuite pour nous des éléments décisifs dans le choix d'un cadre formel et nous ont fait préférer l'approche possibiliste à la théorie des croyances.

La fusion des cartes floues 3D produit une carte floue de fusion 3D exprimant la distribution de possibilité fusionnée. La carte fusionnée a ses valeurs dans l'intervalle [0,1]. Cette méthode peut être étendue à la fusion de N cartes floues par combinaison de opérateurs présentés ci-dessus.

4.4 Décision

La dernière étape consiste à prendre la décision quant à l'appartenance d'un voxel de l'image I à une classe C_i. La règle de décision adoptée consiste à prendre une coupe de la carte de fusion en choisissant un seuil d'étiquetage.

$$\forall v \in I,\ v \in C_i \ \ Si \ \ \pi_{Fusionné}(v) \geq seuil$$

Nous pouvons résumer notre proposition d'approche de fusion floue de données appliquée à la segmentation d'images IRM, par l'architecture représentée dans la figure suivante.

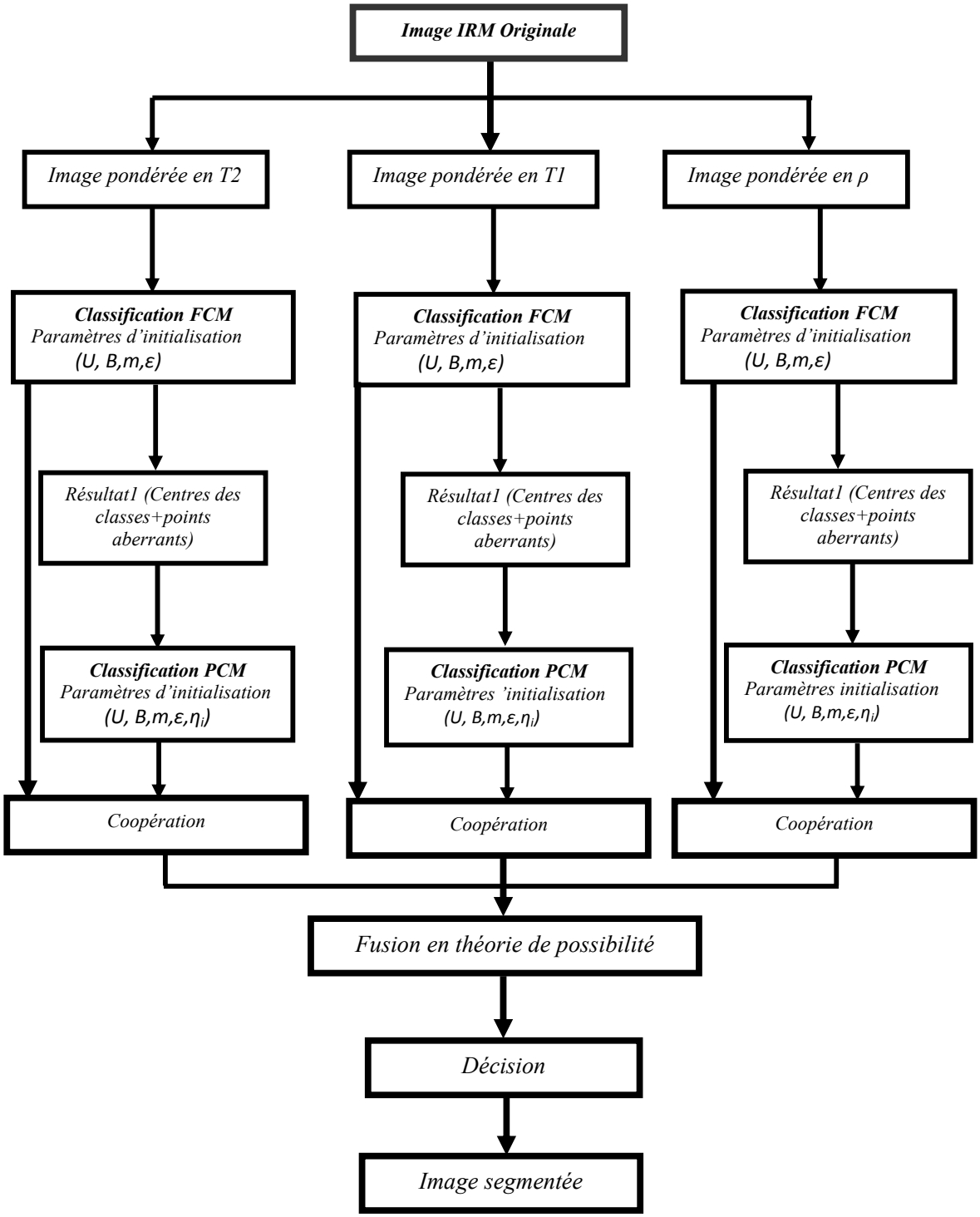


Figure 4. 3*: Architecture générale des étapes de la fusion de données de l'approche proposée.*

4.5 Conclusion

Dans ce chapitre nous avons décrit les étapes de la fusion de données de l'approche proposée pour répondre à la problématique de l'extraction des connaissances par fusion floue de données, avec la présentation de l'architecture générale et des paramètres utilisés par chaque algorithme.

D'abord, nous avons fait une étude comparative entre les algorithmes de classification automatique (clustering) et entre les théories de fusion de données. Nous avons remarqué à partir de cette étude que le problème majeur de l'algorithme FCM est sa sensibilité aux points aberrants (éloignés) qui peuvent avoir des valeurs d'appartenance élevées et ils peuvent affecter de façon significative l'estimation des centres des classes. Le formalisme retenu consiste à faire la **coopération** entre l'algorithme FCM (C-moyennes floues) dont la contrainte d'appartenance d'un individu à une classe est gérée d'une manière relative et l'algorithme possibiliste PCM (C-means possibilistes) pour les points aberrants. Enfin, nous avons retenu la théorie de possibiliste comme contexte théorique pour la fusion de données en imagerie médicale.

Dans le chapitre suivant nous allons présenter les résultats, afin d'évaluer l'efficacité du système développé.

CHAPITRE 5

RÉSULTATS ET EVALUATIONS

5.1 Introduction

Ce dernier chapitre est consacré à l'application des méthodes proposées dans les chapitres précédents et à leurs évaluations quantitativement et qualitativement à partir des différents critères proposés. Les méthodes proposées ont été appliquées sur des images IRM : simulées et réelles. Les images de référence représentent des images d'un cerveau. L'évaluation a été tout d'abord effectuée pour la segmentation de tissus sains, c'est à dire la matière blanche, la matière grise, et le LCR.

Pour valider la méthode développée, nous nous sommes basés sur la base de données Brainweb[5]. Cette base a été choisie dans la mesure où elle est très fréquemment utilisée dans la littérature et permet donc de pouvoir fournir un point de comparaison plus aisé avec des validations proposées dans d'autres documents. Le site Web de Brainweb permet de simuler des IRM cérébrales avec différents niveaux de bruit et d'inhomogénéités **[Aït-06]**.

5.2 Images utilisées

5.2.1 Images réelles

Une image réelle est obtenue à partir d'un signal continu bidimensionnel comme par exemple un appareil photo ou une caméra, etc. Sur un ordinateur, on ne peut pas représenter de signaux continus, on travaille donc sur des valeurs discrètes.

[5] http://www.bic.mni.mcgill.ca/brainweb/

Les images réelles sur lesquelles nous avons travaillé ont été acquises dans le cadre de la collaboration entre le laboratoire LSI (Laboratoire Systèmes Intelligents : équipe image et signaux) de l'université Ferhat Abbas de Sétif et le Centre de Recherche en Sciences et Technologies de l'Information et de la Communication (CReSTIC) équipe LAM (Traitement d'images) IUT de Troyes,

Il s'agit d'images pondérées en T_1, T_2 et en densité de proton pour des patients de différents âges (taille pixel =1mm, taille de matrice 256 × 256). Les images sont en format DICOM [6](*Digital Imaging and Communications in Medecine*).

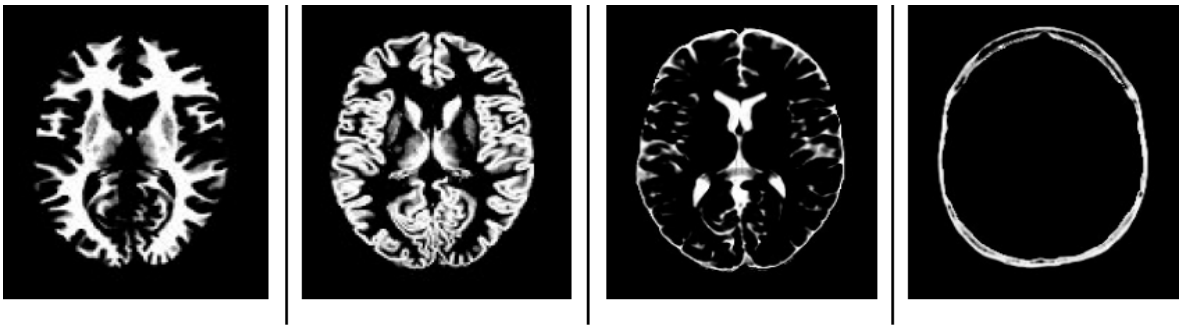
5.2.2 Fantômes (images de synthèse)

Nous avons utilisé un fantôme qui est ici une base de données synthétique qui permet de construire des données IRM.

Pour valider les méthodes de segmentation du cerveau, les chercheurs ont proposé divers fantômes imitant le cerveau. Cependant, ces fantômes sont relativement simples par rapport à la complexité du cerveau. Depuis quelques années, le centre d'imagerie cérébrale de l'institut Neurologique de l'université McGill à Montreal, met à la disposition de la communauté des chercheurs le fantôme dénommé Brainweb qui est devenu une référence très utilisée pour valider les algorithmes de segmentation du cerveau. Le modèle anatomique du fantôme consiste en un ensemble de volumes flous représentant des degrés d'appartenance aux différents tissus constituant l'image IRM (matière branche, matière grise, LCR, peau, crâne, graisse, etc.).

La construction du fantôme est fondée sur un ensemble de 27 volumes IRM de taille 181 × 127 × 181 voxels par volume et de haute résolution (1*mm* / voxel). Ces images ont été recalées et un volume IRM a été crée par moyennage de ces 27 volumes recales. Les voxels du volume moyen, ont été étiquetés par un neuroradiologue en MG, MB, LCR, graisse, etc. Les méthodes de classification floue ont été effectuées sur ce volume. Apres corrections manuelles des résultats obtenus, les cartes floues de tissu sont finalement construites. En illustration, figure 5.1 présente une coupe des cartes floues de MG, MB, LCR et du crâne.

[6] DICOM 3.0 : The ACR/NEMA Standard Home Page : http://www.xray.hmc.psghs.edu/dicom/



MB *MG* *LCR* *Crâne*

Figure 5. 1 *: Exemple de cartes floues du fantôme* **[Kwan-96]**.

5.2.3 Constructions des images simulées

A partir des cartes floues de tissu du fantôme et à l'aide des équations de Bloch **[Bloc-46**], des images IRM peuvent êtres simulées en différentes pondérations, telles que T1, T2, et aussi être bruitées à volonté afin d'obtenir des images réalistes. Le bruit dans l'image est un bruit blanc gaussien qui est exprimé comme un pourcentage de son écart type à l'intensité moyenne du signal. Chaque volume de données est constitue de 181 images de taille 181×217 pixels tout comme le fantôme.

D'après **[Semc-07]**, les principales étapes mises en oeuvre afin de transformer le volume initial en un fantôme permettant de générer des simulations sont les suivantes :

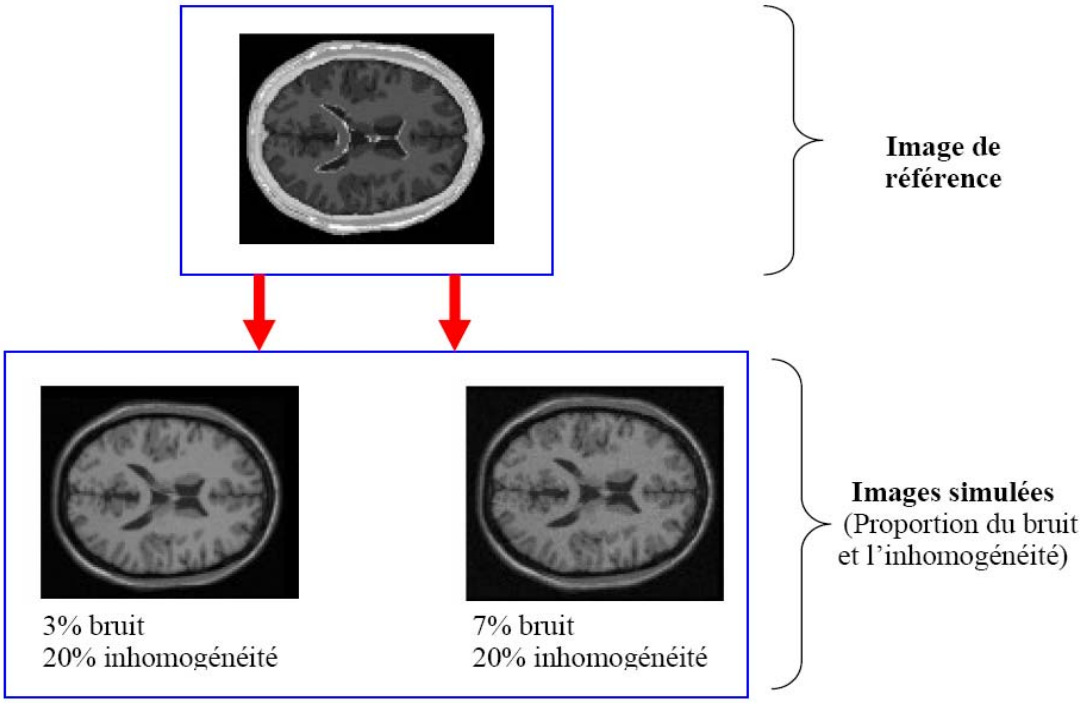


Figure 5. 2: *Processus de construction des images simulées.*

- *La correction des non-uniformités* : Avant tout traitement du volume initial, un algorithme de correction des non-uniformités des niveaux de gris a été appliqué afin de réduire au maximum les risques d'erreurs de classification.

- *La classification* : La classification a été réalisée à partir d'un ensemble d'apprentissage décrivant des exemples de pixels appartenant aux différents tissus recherchés et marqués par un expert. Plusieurs algorithmes de classification ont été testés et le principe des plus proches moyennes a été retenu comme fournissant le meilleur résultat **[Germ-99]**. Afin de construire un fantôme réaliste, les effets de volume partiel sont pris en compte. Ainsi pour chaque pixel du volume, un vecteur décrit la proportion de chacun des tissus qui le constituent. Le résultat de la classification est finalement constitué par 9 volumes décrivant chacun l'un des tissus recherchés (matière blanche et grise, liquide céphalo-rachidien, graisse, muscles, crâne, air, etc.). Au sein de chaque volume, l'intensité d'un pixel représente sa fraction pour le tissu correspondant.

- *Les corrections manuelles* : Un ensemble de corrections manuelles ou semi-automatiques a été réalisé afin d'améliorer le résultat de la segmentation. Ainsi, le fantôme final est

composé de 9 volumes de tissus corrigés et la carte de référence qui lui est associée correspond en chaque pixel au tissu le plus représenté. Ce fantôme permet de décrire une géométrie réaliste de cerveau humain.

- *Les simulations* : Les niveaux de gris simulés pour chaque pixel du volume ont été obtenus par résolution des équations de Bloch **[Germ-99]**.

Nous avons choisi des volumes avec une épaisseur de coupe de 1 mm, taille couramment utilisée pour générer des images anatomiques. Le choix du bruit proposé sur le site BrainWeb est compris entre 0% et 9% et celui des valeurs du paramétrage d'hétérogénéité entre 0% et 40 %. Pour des raisons de réalisme, la valeur de 0% n'a pas été retenue. De même, la valeur de 9% produit des images à l'allure très artificielle et nous l'avons écartée. Nous avons utilisé dans nos simulations la valeur de 3%. Les valeurs du paramètre d'inhomogénéité ont été choisies entre 0% et 20%. La valeur de 40% a été écartée car elle perturbe trop les distributions des niveaux de gris dans les images et que dans la réalité l'inhomogénéité se situe plutôt aux environs de 10%.

5.3 Evaluation et étude comparative

Les performances de nos algorithmes nous ont conduits à réfléchir sur la validité de la segmentation obtenue. Il n'existe pas de "*bonne*" segmentation. Seule l'appréciation de l'utilisateur (qualité visuelle) et le but recherché permettent de définir une bonne segmentation pour un type de données. C'est pour cette raison que nous avons mesuré et quantifié les performances de notre segmentation de l'ensemble de l'encéphale.

5.3.1 Critères de validation

Pour tester la méthode développée de façon pertinente, nous jugeons la qualité de la segmentation obtenue par rapport à plusieurs estimateurs souvent utilises dans la littérature **[Rich-04]** :

- *Sensibilité (SE)* : elle correspond à la proportion de vrais positifs par rapport à l'ensemble des structures qui devraient être segmentées :

$$SE = \frac{TP}{TP + FN}$$

La sensibilité tend vers 1 (resp. 0) s'il y a peu (resp. beaucoup) de faux négatifs. Cet indicateur permet d'évaluer dans quelle mesure l'intégralité d'une structure recherchée est segmentée.

- *Spécificité (SP*) : elle correspond à la proportion de vrais négatifs par rapport à l'ensemble des structures qui ne devraient pas être segmentées :

$$SP = \frac{TN}{TN + FP}.$$

La spécificité tend vers 1 (resp. 0) s'il y a peu (resp. beaucoup) de faux positifs. Cet indicateur permet d'évaluer dans quelle mesure l'intégralité du complémentaire d'une structure recherchée n'est pas segmentée.

- *Recouvrement (RE*) : il correspond la proportion de vrais positifs par rapport à l'ensemble des structures qui ont été ou devraient avoir été segmentées :

$$RE = \frac{TP}{TP + FP + FN}$$

Le recouvrement tend vers 1 (resp. 0) s'il y a peu (resp. beaucoup) de faux positifs et de faux négatifs. Cet indicateur permet d'évaluer dans quelle mesure la structure recherchée correspond quantitativement et qualitativement `a la segmentation.

- *Similarité (SI)* : elle correspond à la proportion de vrais positifs par rapport à l'ensemble des structures qui ont été et devraient avoir été segmentées :

$$SI = \frac{2.TP}{2.TP + FP + FN}$$

La similarité tend vers 1 (resp. 0) s'il y a peu (resp. beaucoup) de faux positifs et de faux négatifs. A l'instar du recouvrement, cet indicateur permet d'évaluer dans quelle mesure la structure recherchée correspond quantitativement et qualitativement à la segmentation.

- *Le Branching Factor (BF)* : elle correspond à la proportion de faux positifs par rapport à la vrais positives des structures qui devraient être segmentées :

$$BF = \frac{FP}{TP}$$

Quantifie la sur détection de pixels n'appartenant pas au tissu recherché dans la carte de référence. Dans le cas idéal ce coefficient vaut 0.

Notations employées :

- TP : vrais positifs (true positive), quand la méthode de segmentation a trouve des pixels qui ne figurent pas dans l'image de référence.
- FP : faux positifs (false positive), où l'image référence a indique des pixels que la méthode de segmentation n'a pas trouve.
- TN : vrais négatifs (true negative), où l'image de référence possède des pixels que la méthode de segmentation indique aussi.
- FN : faux négatifs (false negative), où l'image de référence n'a pas indique des pixels et la méthode de segmentation non plus.

5.3.2 Le protocole d'évaluation

L'évaluation objective et quantitative des résultats joue un rôle important dans la segmentation d'image **[Zhan-94]**. L'utilisation d'images de synthèse permet de comparer quantitativement la segmentation obtenue par rapport à un référentiel (Talairach[7]) (Figure 5.3). La robustesse des systèmes de segmentation peut ainsi être étudiée et comparée pour différents niveaux de bruit et d'hétérogénéité d'intensité des images simulées. Il est également possible pour un même système de segmentation de tester quantitativement l'influence du choix de certains paramètres sur les résultats de la segmentation.

[7] Le référentiel de **Talairach** est un système de coordonnées permettant de repérer la position de n'importe quel point dans le cerveau d'un individu quelconque en référence à un atlas publié par les médecins Jean Talairach.

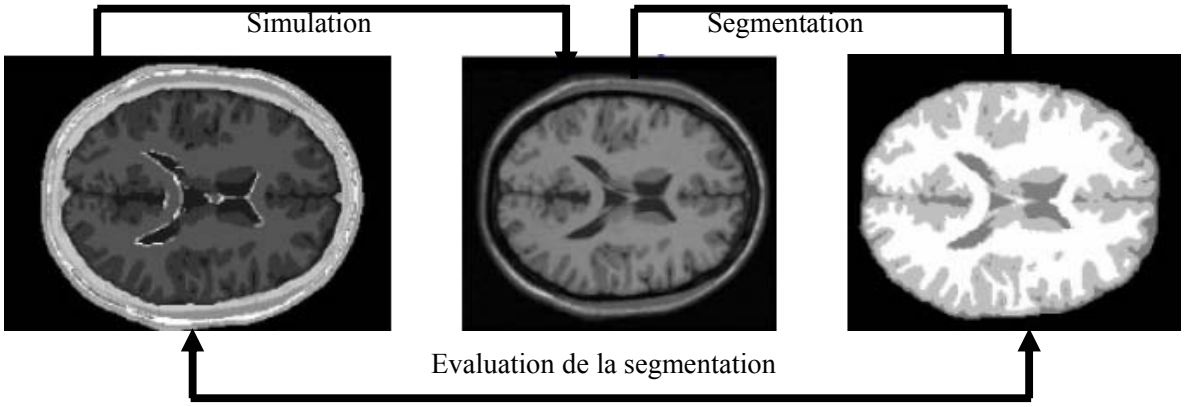


Figure 5. 3: *Processus d'évaluation sur des images fantômes*

5.3.3 Évaluations de la segmentation

Dans ce qui suit, nous présentons les expérimentations que nous avons réalisées sur 10 coupes successives du volume Brain 130 pour évaluer l'approche proposée. Les résultats obtenus sont comparés avec ceux des approches classiques. Ces résultats sont présentés dans les tableaux 5.1 et 5. 2.

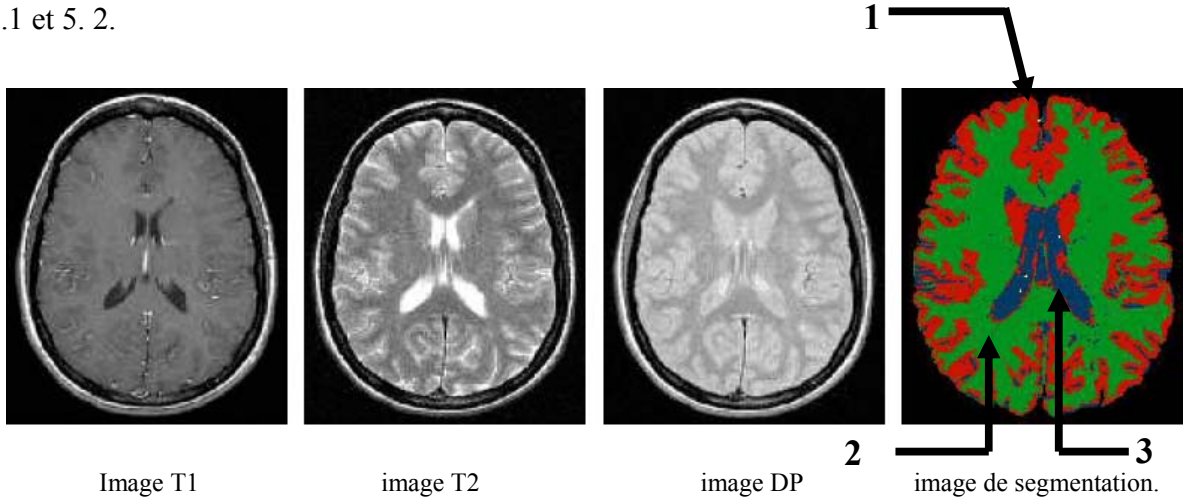


Figure 5. 4 : Coupes *pondérées en T1, T2 et en densité de protons illustrant la fusion d'images.*

Dans cette image, les tissus sont étiquetés : par (1), la matière grise ; par (2), la matière blanche ; et par (3) le liquide céphalo-rachidien.

Une image étiquetée a été créée en assignant chaque voxel au tissu pour lequel il a le plus grand degré d'appartenance (règle du maximum de possibilité).

	Matière grise		Matière blanche		Liquide céphalorachidien		
Coupe	***BF***	***SE***	***BF***	***SE***	***BF***	***SE***	***RE***
1	0,02	0,94	0,02	0,94	0,01	0,97	96,65
2	0,03	0,93	0,03	0,94	0,01	0 ,93	96,96
3	0,02	0,93	0,02	0,97	0,01	0 ,95	96,71
4	0,04	0,96	0,02	0,97	0,01	0 ,95	96,98
5	0,09	0,96	0,01	0,92	0,01	0 ,97	96,38
6	0,09	0,95	0,02	0,93	0,02	0 ,94	96,65
7	0,07	0,97	0,01	0,93	0,02	0 ,94	96,74
8	0,06	0,97	0,01	0,95	0,02	0 ,96	96,82
9	0,05	0,95	0,01	0,93	0,01	0 ,96	96,65
10	0,05	0,96	0,02	0,92	0,02	0 ,96	96,55
Moyenne	0,052	0,952	0,017	0,94	0,014	0,953	**96,709**

Tableau 5. 1 *: Evaluations de la segmentation sur 10 coupes successives du volume Brain130, par le système développé.*

	Matière grise		Matière blanche		Liquide céphalorachidien		
Coupe	***BF***	***SE***	***BF***	***SE***	***BF***	***SE***	***RE***
1	0,04	0,9	0,02	0,93	0,03	0 ,92	96,43
2	0,12	0,91	0,02	0,92	0,02	0 ,94	96,94
3	0,11	0,96	0,02	0,93	0,01	0 ,97	96,64
4	0,11	0,96	0,05	0,93	0,01	0 ,91	96,5
5	0,04	0,96	0,05	0,92	0,01	0 ,92	96,3
6	0,05	0,97	0,01	0,93	0,02	0 ,89	96,6
7	0,08	0,97	0,01	0,96	0,01	0 ,95	96,47
8	0,08	0,96	0,01	0,96	0,01	0 ,96	96,54
9	0,09	0,92	0,02	0,92	0,02	0 ,94	96,25
10	0,12	0,92	0,01	0,92	0,01	0 ,90	96,34
Moyenne	0,084	0,943	0,022	0,932	0,015	0,93	**96,501**

Tableau 5. 2 *:* *Evaluations de la segmentation sur 10 coupes successives du volume Brain130, par l'approche proposée.*

En comparant les résultats des deux systèmes, nous constatons que :

- La similarité des résultats obtenus pour les valeurs du SE *Gris* et du SE *Blanc*. Pour toutes les images, la différence entre les valeurs moyennes de ces coefficients est inférieure à 0.05, ce qui justifié notre choix d'écarter les pixels ambigus dans le processus de classification (entre MG, MB)
- La différence qui existe systématiquement entre le *BF Gris* et le *BF Blanc* des deux approches, ce facteur est un coefficient qui permet d'évaluer les sur détections de pixels. Dans le cas de notre application, ce coefficient est faible à cause du rejet des pixels aberrants par rapport aux centres des classes.

5.3.4 Comparaison aux C-moyennes floues et C-moyennes possibiliste

L'algorithme de l'approche proposée a ensuite été comparé à l'algorithme des C-moyennes floues (FCM) et l'algorithme C-moyenne possibiliste (PCM) sur les tissus cérébraux.

Les figures 5.5, 5.6 et 5.7 montrent les résultats de segmentation par les algorithmes FCM, PCM et l'approche proposée.

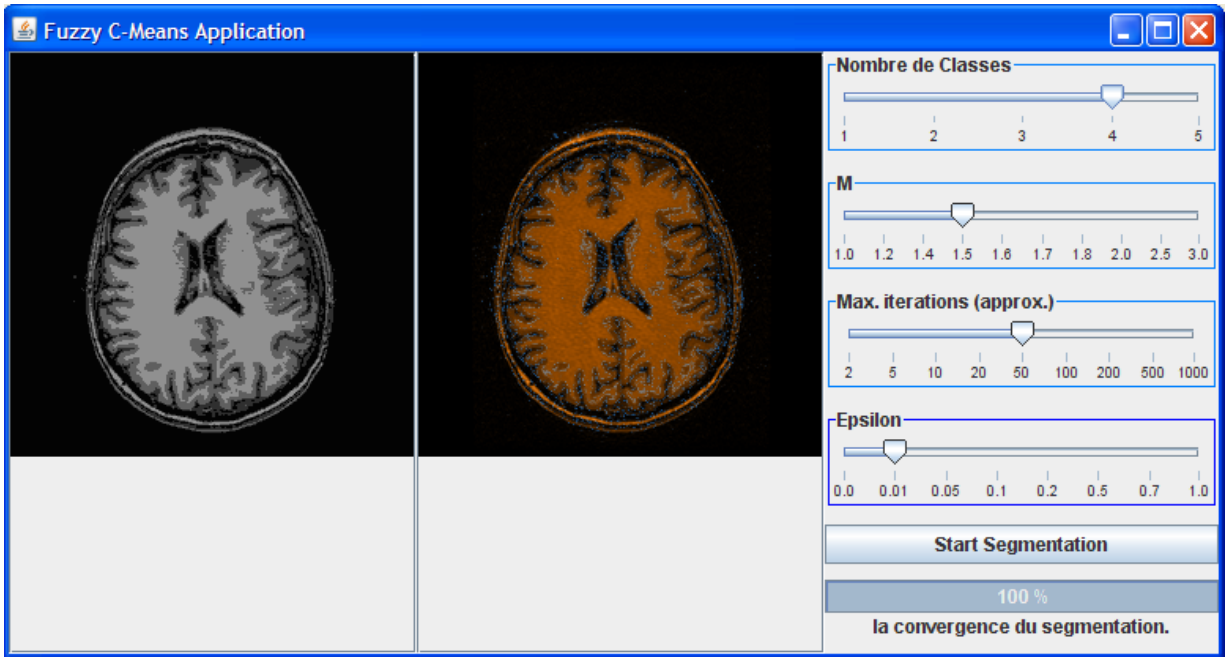


Figure 5.5 *: Image T1 segmentée par FCM.*

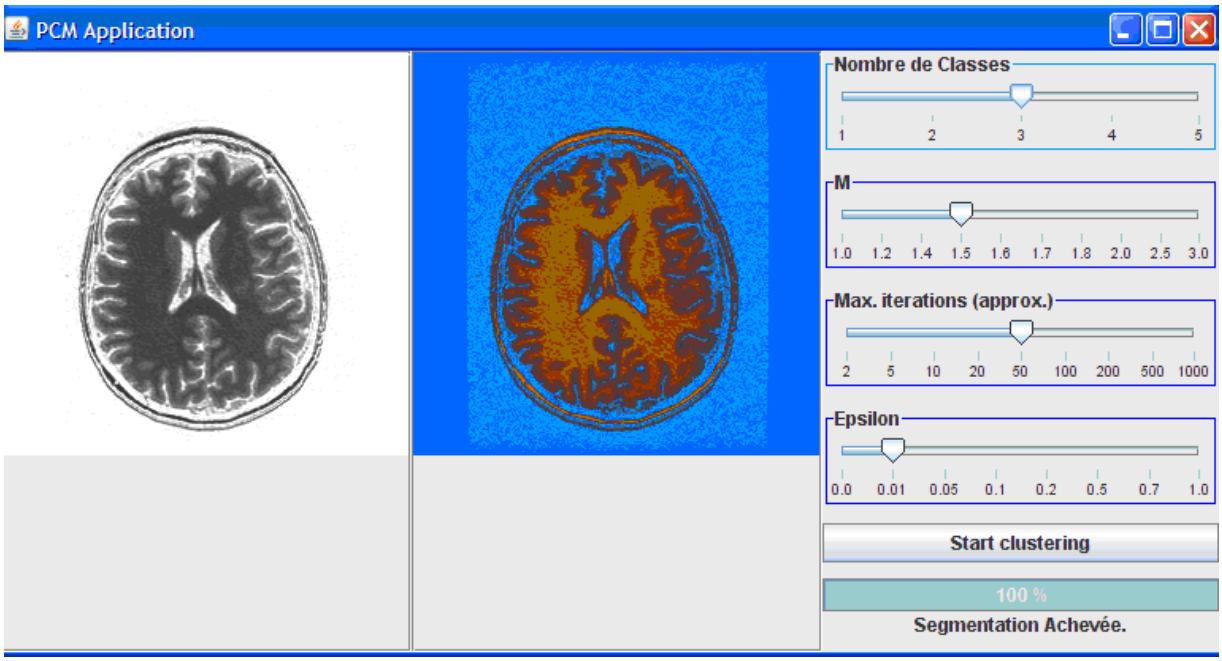


Figure 5.6 : *Image T1 segmentée par PCM.*

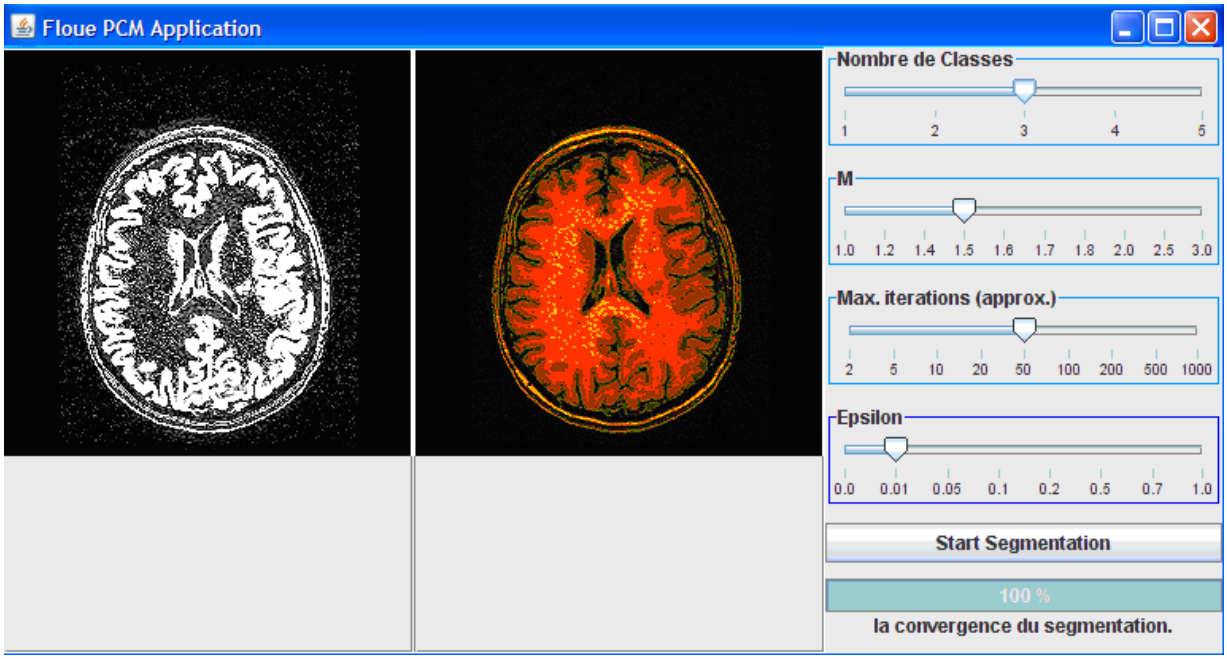


Figure 5.7 : *Image T1 segmentée par l'approche proposée.*

Le tableau 5.3 présente les taux de recouvrements obtenus sur ces volumes IRM pour les trois algorithmes flous.

	FCM	PCM	**Approche proposée**
MG	0,85	0,64	**0,9**
MB	0,9	0,52	**0,92**
LCR	0,63	0,65	**0,9**

Tableau 5. 3 *: Comparaison des taux de recouvrement obtenus par différents algorithmes.*

Les résultats du tableau 5.3 confirment l'intérêt de l'approche de fusion par rapport aux approches prises indépendamment.

5.4 Analyse des résultats

Les résultats de chaque étape de la fusion sont présentés sur un niveau de coupe (Figure 5.4) dont la localisation permet de distinguer les trois classes de tissus à séparer :

- MG (pallidum, putamen, noyaux caudés, thalamus et cortex).
- MB (parenchyme cérébral).
- LCS (espaces sous-arachnoïdiens, ventricules latéraux et V3).

L'interprétation de nos résultats est faite par un médecin spécialiste (CHU de Constantine) sur des images simulées et réelles.

5.4.1 Images de synthèse

En analysant les images de la figure 5.8 et la figure 5.9, le médecin a établit le bilan suivant :

- **Image *(e)*** : la classe LCS est très semblable à celle issue de l'image originale, la classe LCS n'apporte en effet aucune information supplémentaire quant à la localisation du liquide.

- **Image *(h)*** : la distinction entre les différentes classes segmentées ne s'exprime pas complètement (problèmes d'initialisations des centres de classes).
- **Image *(k)*** : l'interprétation des classes est complètement améliorée par rapport aux (FCM, PCM), on remarque la distinction entre les 3 classes du cerveau.
- **Image *(f)*** : la classe LCS n'apporte en effet aucune information supplémentaire quant à la localisation du liquide. La classe MB est fortement améliorée par rapport à l'image originale.
- **Image *(i)*** : le PCM n'apporte presque pas de grande chose par rapport au FCM.
- **Image (*l*)** : l'approche proposée apporte une grande performance à la segmentation pour la classe MB mais pour la classe LCS, il reste toujours pas complètement visible c'est-à-dire la classe LCS n'apporte en effet aucune information supplémentaire.

Remarque

Pour l'image *(l)* issue de la segmentation par le système développé l'expert a signalé la présence de nouvelle classe (entourer par le rectangle), cette classe n'a pas été détectée dans l'approche floue et possibiliste.

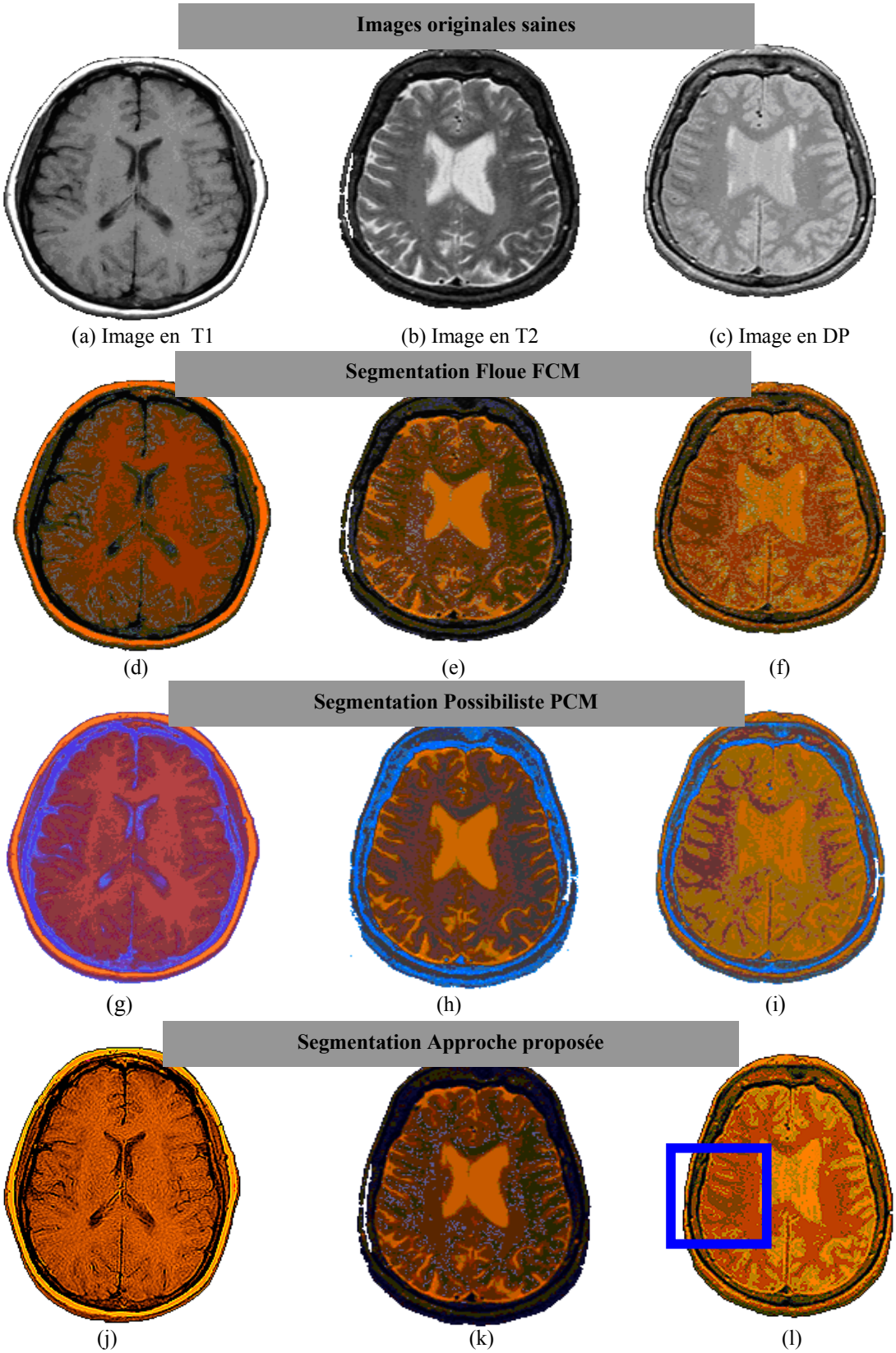


Figure 5.8 *: Comparaison des images segmentées par différentes approches de classification (Images saines).*

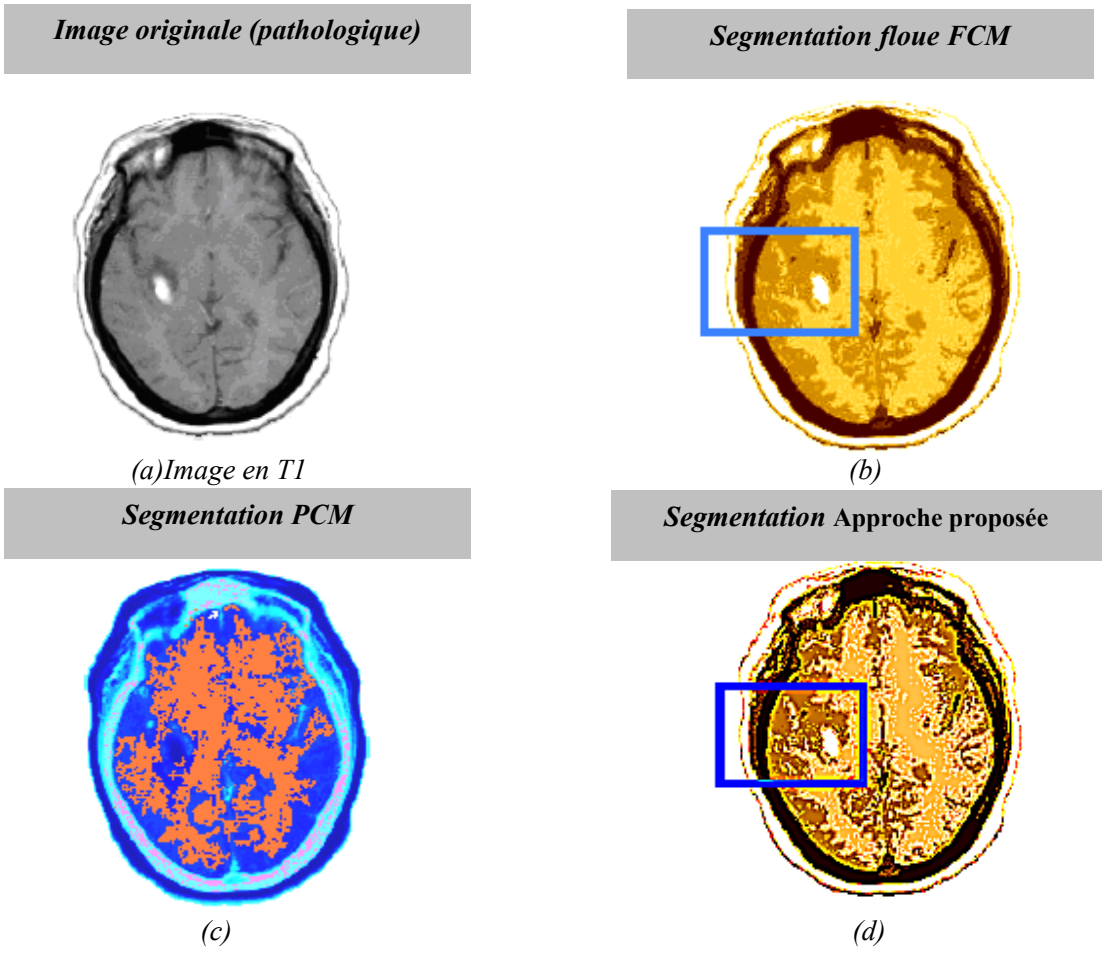


Figure 5. 9 *: Comparaison des images segmentées par différentes approches de classification (images pathologiques).*

- **Image (b*)*** : la classe LCS n'est pas conforme à celle de l'image originale. Le manque d'informations sur les petits sillons (image (a)) et la mauvaise discrimination LCS/MG font que la classe LCS segmentée représente mal la distribution du liquide. Les distributions de MB et MG quant à elles se rapprochent de celles fournies par l'image originale. La détection de la pathologie est signalée selon l'expert mais les détails sont mal exprimés

- **Image (c*)*** : PCM est mal adapté dans cette segmentation par rapport à l'image (d).

- **Image (d)** :l'approche proposée apporte une grande performance à la segmentation pour les trois classes et surtout pour la quatrième classe qui est la pathologie qui spécifie bien la taille et les détails sur cette dernière.

5.4.2 Images réelles

La segmentation du cerveau a été appliquée avec succès sur quelques images réelles (format DICOM). Les résultats sont illustrés sur la figure 5.11.

Nature	Images IRM Cérébrales
Pondérations	T1, T2, densité de proton
Coupes	Axiales, sagittales, coronales
Taille	256*256
Patients	Enfants, Hommes
Formats	DICOM

Tableau 5. 4 *: Propriétés des images utilisées.*

On a préféré présenter les images segmentées sur des coupes différentes afin de prouver l'efficacité de l'approche développée.

Pendant tout le processus de segmentation des images réelles, nous avons remarqué que le temps nécessaire pour segmenter une image (nombre d'itérations de chaque algorithme ≥ 50, taille 256*256) est très élevé. C'est pour cette raison qu'on a choisi $\varepsilon = 0{,}02$

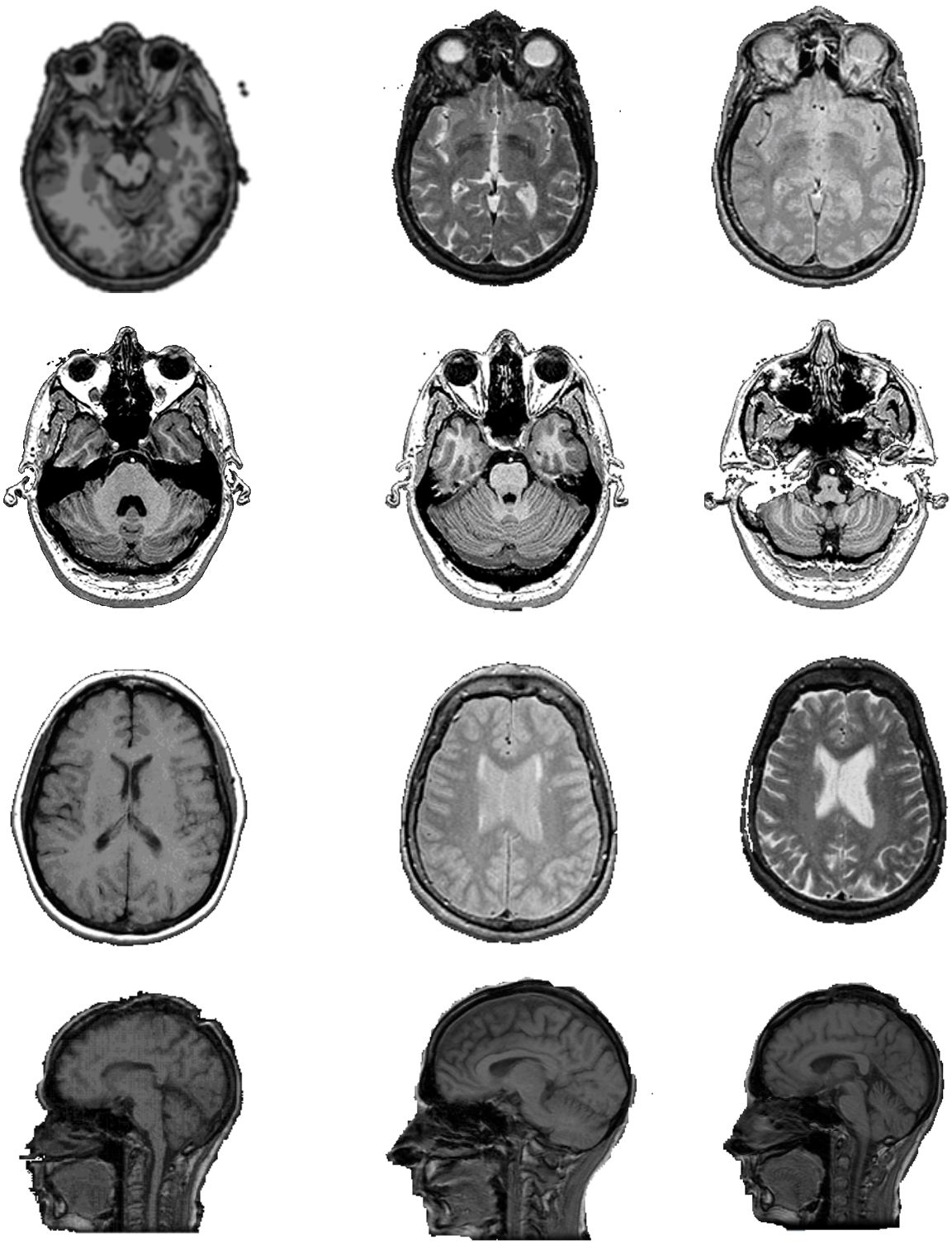


Figure 5.10 : *Images originales : différentes coupes (axiales, sagittales, coronales) et différentes pondérations (T1, T2, DP).*

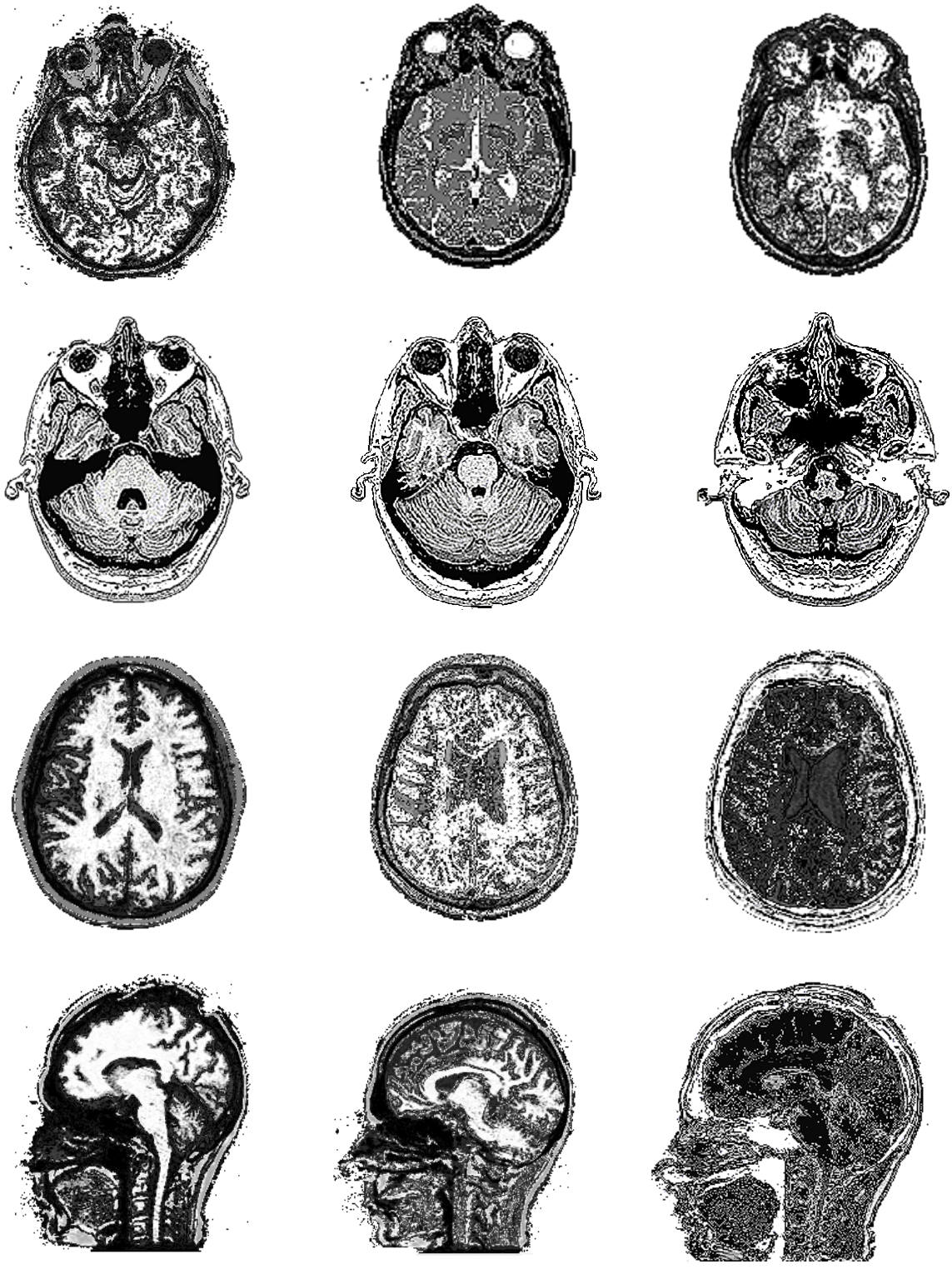


Figure 5.11 : *Segmentation par l'approche proposée : différentes coupes (axiales, sagittales, coronales) et différentes pondérations (T1, T2, DP).*

Les cartes segmentées de la figure 5.11 ont été créées par la règle du maximum d'appartenance.

5.5 Conclusion

Dans ce chapitre, nous avons tout d'abord évalué la méthode de segmentation des tissus cérébraux proposée. Nous avons appliqué notre méthode sur des images de synthèses et des images réelles pour la segmentation des classes normales (MG, MB, LCR) et des classes pathologiques (tumeur, œdème, etc.).

Les avantages de notre méthode sont les suivants :

1. Elle est totalement automatique.
2. Elle satisfait aux contraintes de l'application.
3. Elle est robuste en présence des points aberrants.
4. Sa performance est meilleure que celle de la méthode supervisée **[Jaya-96]**.
5. C'est un système à base de la théorie floue et possibiliste.
6. Elle est efficace sur au moins 2 types de tissus.

CONCLUSION GÉNÉRALE ET PERSPECTIVES

La localisation et la quantification de structures cérébrales fait appel à la synthèse de diverses informations issues non seulement de l'imagerie, mais également des connaissances anatomiques et fonctionnelles concernant ces structures et la pathologie qui les affecte. Le clinicien réalise la synthèse de ces différentes informations pour obtenir un diagnostic fiable et précis. Le but de ce travail était, à travers la modélisation de ce processus par des méthodes de fusion de données, de fournir un outil de segmentation et de quantification volumique et fonctionnelle des structures cérébrales.

Après avoir effectue une revue bibliographique sur la fusion de données dans les chapitres précédents, nous avons démontré l'importance et les difficultés de ces travaux de segmentation d'images IRM. Deux raisons expliquent ces difficultés :

- La première est qu'il existe une très grande variété de tissus anormaux qui différent par leur taille, leur forme, leur position et leur composition (nature et homogénéité). En outre, dans certains cas, les formes des structures sont difficiles à délimiter même par des experts.
- La seconde raison vient de ce que la donnée issue de l'acquisition IRM est sensible au bruit de fond et à l'échantillonnage. Ainsi un voxel peut appartenir à des tissus de différentes natures ce qui cause l'effet dit de volume partiel.

Nous avons tout d'abord décomposé le processus de fusion de données en trois phases fondamentales.

En premier lieu, nous avons modélisé des informations, numériques ou symboliques, dans un cadre commun permettant de prendre en compte les ambiguïtés, les imprécisions et les incertitudes. Nous nous sommes pour cela placés dans le cadre de la **coopération** entre l'algorithme flou FCM et l'algorithme possibiliste PCM afin de rendre l'algorithme plus robuste face aux imprécisions et aux données aberrantes. Dans un premier temps l'initialisation (la matrice des degrés d'appartenances et les centres de classes) est faite par l'algorithme FCM. Dans un second temps l'algorithme PCM pour les points aberrants, cet algorithme permet entre autre de générer un degré d'appartenance absolu reflétant de manière exacte la réalité de distribution des pixels.

En second lieu, nous avons fusionné différentes données. Cette agrégation a été réalisée par des opérateurs de fusion qui modélisent l'analyse quotidienne du médecin confronté à des données cliniques hétérogènes. Nous avons explicité l'opérateur qui nous a semblé le plus adapté, en fonction de propriétés mathématiques souhaitées, de considérations intuitives sur la nature des données et de tests expérimentaux.

En dernier lieu, nous avons présenté ces informations fusionnées au clinicien. Ici encore, nous avons proposé pour chaque type de fusion une solution, soit sous la forme d'une image, soit sous la forme d'une nouvelle image de synthèse.

Nous avons présenté les résultats de notre travail qui consiste à utiliser plus d'un algorithme pour segmenter des images médicales en vue d'améliorer la qualité de la segmentation. La segmentation a été réalisée sur des images IRM cérébrales bidimensionnelles, les résultats obtenus après segmentation sur un ensemble de données sont satisfaisants et comparés de façon favorable avec les résultats d'autres méthodes comme les algorithmes FCM et PCM sur le même ensemble de données, ce qui nous permet de dire, que l'utilisation combinée de plusieurs algorithmes de segmentation travaillant en coopération permet de pallier aux problèmes rencontrés par l'utilisation d'un seul algorithme,

Les perspectives d'amélioration et de développement de ce travail sont multiples.

Dans le domaine de la modélisation : nous nous sommes restreints aux cas où les informations étaient portées par les distributions de tissus cérébraux ou par des connaissances

symboliques expertes. Puisque la philosophie de la fusion indique qu'il faut réunir le maximum de connaissances avant de prendre une décision, nous souhaiterions pouvoir intégrer d'autres informations afin d'augmenter la masse de connaissances disponibles.

Enfin, si de nombreuses études cliniques sont possibles à partir des différentes structures d'intérêt déjà segmentées, le processus de segmentation peut être appliqué à d'autres structure cérébrales, à condition que la résolution et le contraste de l'image traitée le permettent. De même, cette démarche peut être appliquée pour la segmentation et la quantification d'autres organes.

Annexe

ELÉMENTS D'ANATOMIE CÉRÉBRALE ET IRM

A.1 Introduction

Cette annexe à pour objectif de fixer le cadre applicatif que nous avons envisagé. Dans une première partie, nous décrivons brièvement l'anatomie cérébrale afin de rendre compte du contenu des images. Différentes modalités d'observation du cerveau sont ensuite présentées. La seconde partie est onsacrée à l'imagerie par résonance magnétique (IRM). Nous y décrivons dans un premier temps les principes physiques de la résonance magnétique nucléaire et la formation des images. Enfin une description des différentes caractéristiques des images IRM est présentée.

Dans cette partie, divers cours disponibles sur Internet ont été utilisés **[Cino-C1]**, **[Sapp-C1]**, **[Bitt-C1]**, **[Bruy-C1**], **[Kori-82]**, etc., ainsi que la thèse de Vincent Barra **[Barr-00]** et la thèse de Frenoux Emmanuelle **[Fren-04]**.

A.2 Quelques éléments d'anatomie cérébrale

A.2.1 Le Cerveau

A.2.1.1 Description

Bien que représentant seulement 2% du poids total du corps humain (soit environ 1,4 kilogrammes), le cerveau gère directement ou indirectement 98 % de ses fonctions. Il est responsable des fonctions humaines les plus complexes comme la pensée, la résolution de problèmes, les émotions, la conscience et les comportements sociaux, et régit les fonctions essentielles du corps comme la respiration, le processus d'alimentation, le sommeil, les mouvements et les cinq sens.

En dépit de son extrême complexité, le cerveau n'est composé que de deux types de cellules : les neurones et les cellules gliales. Les neurones sont des cellules nerveuses capables de recevoir et de transmettre l'information. Ils sont constitués d'un corps cellulaire, de plusieurs prolongements afférents appelés dendrites et d'un prolongement efférent appelé axone. Chaque neurone peut posséder jusqu'à 10 000 connexions avec d'autres neurones, ce qui conduit à un nombre très élève de réseaux interconnectés. Les cellules gliales sont quant à elles des cellules de soutien qui contribuent à assurer le bon fonctionnement des neurones, sans participer directement au transfert de l'information. Le cerveau contient plus de 100 000 millions de neurones et encore davantage de cellules gliales **[Nobl-06]**.

Le cerveau est la partie la plus volumineuse du système nerveux central. Il est placé dans la boite crânienne. Il se situe dans une enceinte liquidienne (Liquide Céphalo-Rachiden) qui a la particularité de pénétrer également à l'intérieur du cerveau dans les cavités du système ventriculaire. Il est constitué de deux hémisphères principaux. Les hémisphères sont reliés par différentes structures cérébrales comme le corps calleux ou le thalamus.

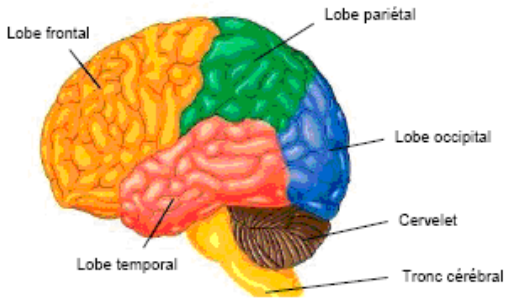


Figure A. 1 *: Structure générale de l'encéphale*.

A.2.1.2 Observation du cerveau

L'observation des coupes axiale, frontale et sagittale sont des coupes du cerveau approximativement parallèles, respectivement, au plan qui comprend nez et oreilles, au plan du visage et au plan de symétrie de la tête **[Floc-90]**. Ces coupes sont orthogonales deux à deux (*cf.* figure A .2).

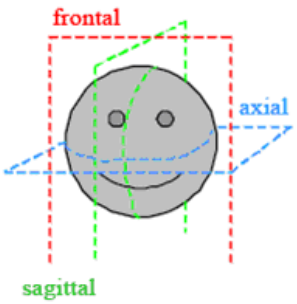


Figure A. 2 *: Plans de coupe en imagerie médicale.*

- **Coupes axiales**

Ces coupes correspondent quasiment à un plan horizontal. En imagerie par résonance magnétique, elles correspondent à un plan perpendiculaire à l'axe du champ magnétique principal.

- **Coupes sagittales**

Ces coupes sont prises dans des plans parallèles au plan inter hémisphérique. Il s'agit des vues latérales du cerveau.

- **Coupes coronales**

Se sont des coupes perpendiculaires aux coupes axiales et sagittales.

Figure A. 3 *: Les trois axes de coupe pour la visualisation du cerveau.*

A.2.1.3 Généralités sur les tissus cérébraux

Le cerveau est composé de trois tissus principaux : le liquide cérébro-spinal, la matière grise et la matière blanche **[Laro-06] [Miri-07]**.

➢ **Le liquide céphalo-rachidien**

Le liquide céphalo-rachidien ou cérébro-spinal (en anglais cerebro-spinal fluid) (LCR), baigne le cerveau et permet de le protéger. Ce fluide circule à travers une série de cavités communicantes appelé es ventricules. En plus de contribuer à absorber les coups, le LCR diminue la pression à la base du cerveau en faisant "flotter" le tissu nerveux. Il est produit par les plexus choroïdes dans les ventricules les plus hauts puis il est absorbé dans le système veineux à la base du cerveau. Le LCR circule vers le bas en évacuant les déchets toxiques et en transportant des hormones entre des régions éloignées du cerveau.

➢ **La matière grise**

La matière grise (MG), correspond au corps cellulaire des neurones avec un dense réseau de dendrites. On la trouve par exemple dans le centre de la moelle épinière et dans le cortex (écorce des hémisphères cérébraux).

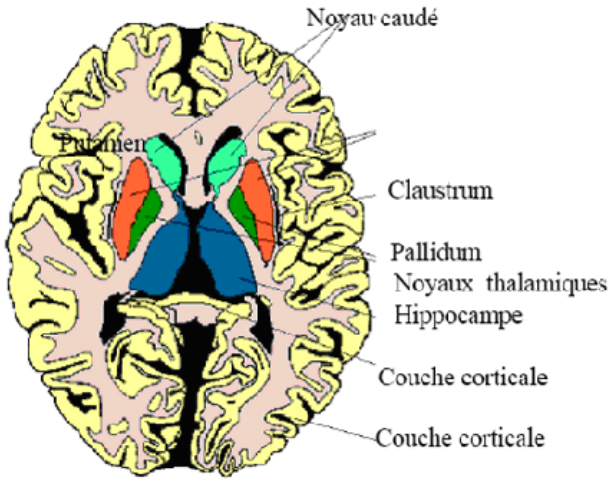


Figure A. 4*: Structures anatomiques de la matière grise*.

- **Le cortex** : Le cortex recouvre la totalité du cerveau. Sa surface est importante car il suit toutes les convolutions externes du cerveau, appelées sillons. Son épaisseur est d'environ 2 à 3 millimètres.
- **Les noyaux de base** : Les noyaux centraux, qui sont avec le cortex les seules structures de substance grise du cerveau, sont également formés de corps

cellulaires neuroniques mais avec une densité moins importante que dans le cortex. Ils sont composés des noyaux du télencéphale, ces noyaux sont appelés les noyaux de base (ou corps strié), parmi eux, on distingue les noyaux caudés et les noyaux lenticulaires.

- ✓ **Le noyau caudé :** En forme de virgule à grosse extrémité il est presque complètement enroulé autour du thalamus. Il longe à peu prés dans toute son étendue le ventricule latérale. On lui distingue trois parties : la tête, le corps et la queue.
- ✓ **Noyaux lenticulaires :** Le noyau lenticulaire se situe en dehors du noyau caudé et thalamus. Sa forme est celle d'une lentille biconvexe, triangulaire sur les coupes axiales et coronales. Le noyau lenticulaire se compose du putame (externe), et du pallidum (interne).

- **Les noyaux du diencéphale**: parmi ces noyaux :
  - ✓ **Le thalamus** : le thalamus est une masse grise qui flanque la cavité du troisième ventricule, aboutissent toutes les sensations avant qu'elles soient projetées vers la conscience. Le thalamus est connecté aux centres moteurs et coordinateurs.
  - ✓ **L'hypothalamus** : l'hypothalamus est la paroi inférieure du troisième ventricule, se prolonge jusqu'à l'hypophyse. De petite dimension, il a la charge des équilibres, physiologiques du corps. Les nouveaux atlas[8] associent dorénavant le pallidum au groupe de noyaux gris du diencéphale **[Bouc-99]**.

➢ **La matière blanche**

La matière blanche (MB) enfin, correspond aux gaines de myélines qui recouvrent les axones des neurones pour en accélérer la conduction. Ces axones myélinisés s'assemblent en faisceaux pour établir des connections avec d'autres groupes de neurones.

[8] Un atlas est une carte 3D des structures et tissus sur un cerveau particulier. Chaque voxel est classifié.

Les différents composants du cerveau sont présentés dans la figure A.5 sur des coupes IRM équivalentes.

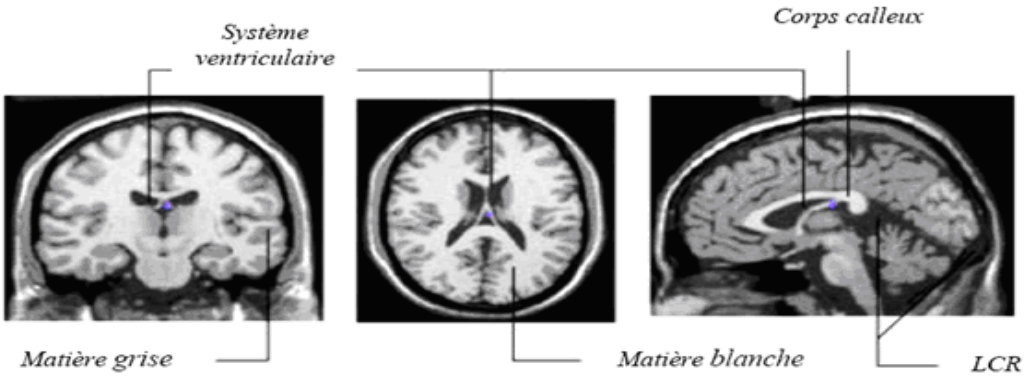


Figure A. 5 : *Coupes IRM du cerveau*.

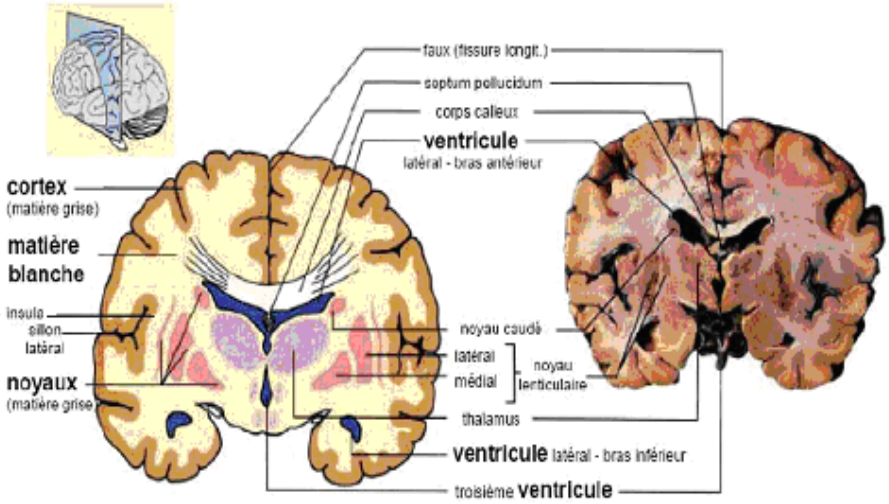


Figure A. 6 *: Différentes structures du cerveau.*

A.3 IRM (Imagerie par Résonance Magnétique) [Achm-05]

A.3.1 Introduction

La résonance magnétique nucléaire (RMN) est une technique en développement depuis une cinquantaine d'années. Le phénomène physique a été conceptualisé en 1946 par BLOCH et PURCELL **[Bloc-46]**, parallèlement. Ils ont obtenu le prix Nobel en 1952. Cette technique a été largement utilisée par les chimistes, puis les biologistes.

Lauterbur **[Laut-73]** publia les premières images obtenues par résonance magnétique nucléaire (IRM) dans la revue Nature. Ce n'est qu'en 1979 qu'un appareil permit d'obtenir des images de la tête chez l'homme. Aujourd'hui, l'IRM est devenue une technique majeure de l'imagerie médicale moderne. Potentiellement, elle est appelée encore à des développements importants; en dehors de l'image morphologique avec sa sensibilité diagnostique démontrée, l'IRM permet aujourd'hui d'autres approches : angiographiques, métaboliques (spectrométrie ^{1}H et spectrométrie phosphore), et fonctionnelles (perfusion tissulaire, volume sanguin cérébral), approches qui ne seront pas traitées dans le cadre de ce cours.

Figure A. 7 *: Appareil d'IRM* **[Laut-03].**

A.3.2 Le Phénomène Physique De La RMN [Gren-C]

- **Précession**

Placé dans un champ magnétique statique $\mathrm{B}_0 = B_0 u_z$, le moment magnétique d'un proton va tourner très rapidement autour de l'axe u_z en décrivant un cône de révolution. La fréquence avec laquelle se produit cette rotation, dite mouvement de précession, est donnée par la relation de Larmor :

$$f_0 = \gamma \frac{B_0}{2\pi}$$

où γ est le rapport gyromagnétique du proton.

Ainsi, sous l'influence de B_0, les protons produisent un moment magnétique macroscopique (ou aimantation) d'équilibre M_0 orienté dans la direction de B_0.

- **Résonance**

La résonance est un transfert d'énergie entre deux systèmes oscillant à la même fréquence. Pour faire basculer un proton d'un état d'énergie E_1 à un état E_2, il faut lui apporter une quantité d'énergie ΔE, reliée à la fréquence de résonance f_0 par la relation :

$$\Delta E = h\nu_0 = \frac{h\gamma B_0}{2\pi}$$

Lors d'une expérience RMN, l'échantillon est soumis à une onde radiofréquence (RF) créée par un champ magnétique B_1, non colinéaire à B_0, et tournant à la fréquence f_0. Les protons, qui étaient alignés selon B_0, reçoivent alors un apport d'énergie sous la forme d'une onde de pulsation égale à leur fréquence propre. Ils résonnent donc et le vecteur aimantation macroscopique est basculé de sa position d'équilibre M_0 vers une position M tant que B_1 dure.

- **Relaxation et temps de relaxation**

A l'arrêt de l'onde B_1, un signal dit de précession libre est enregistré. Il accompagne le retour à la position d'équilibre (en spirale) du vecteur M. En particulier, le retour à l'équilibre des projections de l'aimantation sur le vecteur u_z (aimantation longitudinale $M_L = (M.u_z)u_z$) et sur le plan normal à ce vecteur (aimantation transversale $M_T = M - M_L$) est mesuré. Seule M_T participe au signal en générant un signal dans l'antenne réceptrice.

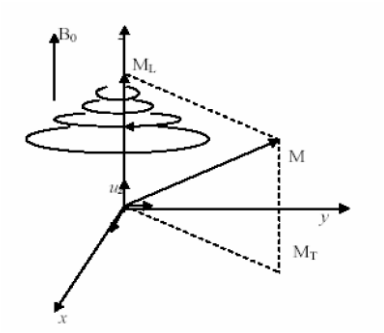


Figure A. 8 *: Retour à l'équilibre du vecteur aimantation*

Ces retours à l'équilibre suivent des cinétiques exponentielles indépendantes, qui dérivent des équations de Bloch : $M_L(t) = M_0 - (M_0 - M_L(0))e^{-t/T_1}$ et $M_T(t) = M_T(0)e^{-t/T_2}$

où T1 et T2 sont respectivement les temps de relaxation longitudinale et transversale, dépendant des propriétés physico-chimiques des tissus.

A.3.3 Paramètres de séquence

Les paramètres de séquence sont les paramètres que le manipulateur fixe sur la console pour définir la séquence IRM. Ils permettent de contrôler l'influence des différents paramètres tissulaires $\mathbf{T_1}$, $\mathbf{T_2}$ et densité de proton dans le signal "pondération" et de moduler ainsi le contraste dans l'image.

Les paramètres de séquence sont les paramètres que le manipulateur fixe sur la console pour définir la séquence IRM. Ils permettent de contrôler l'influence des différents paramètres tissulaires $\mathbf{T_1}$, $\mathbf{T_2}$ et densité de proton dans le signal et de moduler ainsi le contraste dans l'image.

- **Temps d'écho**

Le signal de précession libre ne peut être enregistré directement après l'excitation (déphasage parasite induit par les gradients). C'est pourquoi il est acquis sous la forme d'un écho de spin ou de gradient. Par définition, le délai entre le milieu de l'impulsion d'excitation et le sommet de l'écho est appelé temps d'écho, il est noté *TE*. Dans la méthode d'écho de spin, les hétérogénéités de B_0 et les différences d'aimantation des tissus sont compensées, alors quelles ne le sont pas en écho de gradient. La courbe de décroissance est donc différente pour ces deux techniques.

- **Temps de répétition**

L'image est constituée à partir de la répétition de la même séquence avec un gradient de phase *Gp* d'amplitude différente. Le temps qui sépare deux répétitions est appelé temps de répétition est noté *TR*. Le *TR*, comme le *TE,* est un facteur de contraste. S'il est suffisamment long, l'aimantation repousse tout le signal qui ne dépend pas de la vitesse d'aimantation (donc de T_1), mais essentiellement de la densité protonique. S'il est court, le système atteint après quelques répétitions un régime stationnaire et l'aimantation tend vers une valeur d'équilibre dépendant de la vitesse d'aimantation des tissus, et donc de leur T_1. L'image révèle ainsi les différences de T_1 entre les tissus. Fenêtre

- **Angle de basculement**

Si B_1 est orthogonal à B_0, ce qui est généralement le cas, le phénomène de résonance magnétique bascule l'aimantation *M* selon un axe perpendiculaire au champ principal B_0. Si *M* est basculé à 90° (excitation par une impulsion $\pi/2$), toute l'aimantation est dans le plan transversal et M_L est nulle. En cas de basculement d'un angle inférieur à 90°, seule une partie de l'aimantation est convertie en signal (M_T) et il persiste une aimantation M_L pouvant être utilisée pour une autre excitation. L'angle de basculement correspond donc à une énergie délivrée par le champ B_1. Le signal *S* sera d'autant plus faible que cet angle sera petit. En régime stationnaire, l'angle de bascule α intervient dans le contraste de l'image et gouverne la réserve en aimantation. Pour des angles petits (α<20°), la densité protonique est prépondérante. Plus α est grand, plus T_1 gouverne le contraste.

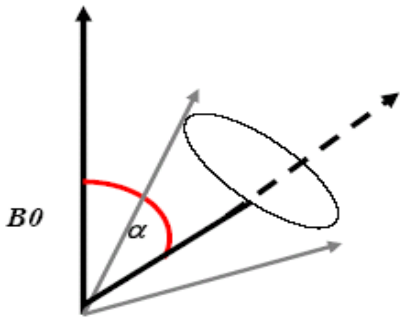


Figure A. 9 *: L'angle de basculement*

A.3.4 Séquences IRM

En modifiant les valeurs des différents paramètres de séquence, une grande diversité de volumes variant par le contraste peut être acquis **[Bosc-03]**. Le choix de l'ensemble des paramètres est alors fonction de l'étude clinique à réaliser. On peut obtenir des images pondérées en T1, en T2, en densité de protons, etc.

A.3.1.1 Image pondérée en densité de protons (ρ)

Pour un *TR* long (de l'ordre de 2s) et un *TE* court (de l'ordre de 20ms), la différence de densité protonique entre la substance grise et la substance blanche s'accentue. On obtient une

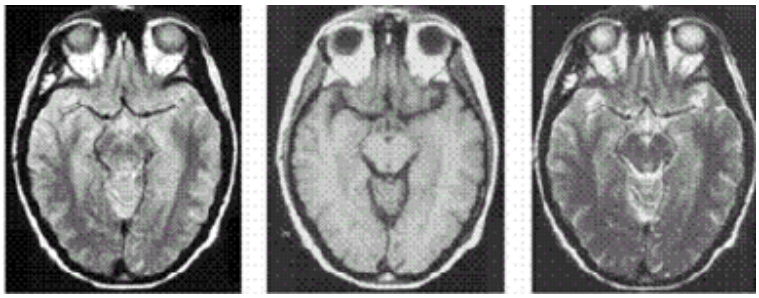
séquence pondérée en densité de protons ρ, qui reflète la localisation et la concentration des noyaux d'hydrogène des différentes structures. Les tissus sont ordonnés par niveaux de gris croissants en matière blanche (MB), matière grise (MG) et liquide cérébro-spinal (LCS)

A.3.1.2 Image pondérée en T2

Pour des TR longs (de l'ordre de 2s) et des *TE* longs (environ 90ms), la décroissance du signal domine la différence de densité protonique entre tissus, et le signal est suffisant pour réaliser une image dite pondérée en T_2, où les tissus sont ordonnés par niveaux de gris croissants en MB, MG, LCS.

A.3.1.3 Image pondérée en T1

Pour des *TR* courts (de l'ordre de 600ms), le contraste entre les tissus dépend essentiellement de leur vitesse d'aimantation, donc de T_1. Pour des *TE* courts (environ 20ms), les différences de décroissance du signal entre les tissus n'ont pas le temps de s'exprimer, rendant le contraste indépendant de T_2. Ainsi, on obtient une image pondérée en T_1, où les tissus sont ordonnés par niveaux de gris croissants en LCS, MG, MB.



Densité de protons (ρ) **T1** **T2**

Figure A. 10 *: Images IRM.*

Pour plus de détails concernant l'imagerie par résonance magnétique, le lecteur peut de référer à **[Kast-97].**

A.4 Défauts des images IRM (artefacts) [Semc-07]

Outre les trois paramètres principaux qui entrent en jeu dans la formation d'une image de résonance magnétique, un certain nombre d'autres facteurs viennent affecter la qualité des

images. Les erreurs (artefacts) d'acquisition en IRM sont de natures différentes de celles observées dans d'autres domaines du traitement d'images.
On distingue essentiellement quatre effets: le bruit, le mouvement, les variations de champ et les effets de volume partiel **[Germ-99]**.

A.4.1 Bruit

Le bruit a des origines multiples, liées en partie au bruit de l'appareillage. Dans les images par résonance magnétique, l'objectif est d'augmenter le contraste entre les tissus tout en conservant une bonne résolution et un rapport signal/bruit élevé, ces caractéristiques sont cependant contradictoires et il est nécessaire de trouver un bon compromis entre résolution et bruit. Ainsi, on peut doubler la taille des pixels pour multiplier le rapport signal/bruit d'un facteur **p**, mais la résolution est divisée par deux. Le choix d'acquisition est donc un facteur déterminant.

A.4.2 Mouvement

Le mouvement peut provenir de plusieurs sources. Il peut être lié au métabolisme comme la circulation sanguine ou la respiration (déplacement chimique). Il peut également être lie au mouvement du patient pendant l'acquisition. Dans tous les cas, le mouvement diminue la qualité de l'image et pose des problèmes d'interprétation, les mouvements de la tête, sont responsables d'artefacts dans les IRM cérébrales.

A.4.3 Variations du champ magnétique (inhomogénéité RF)

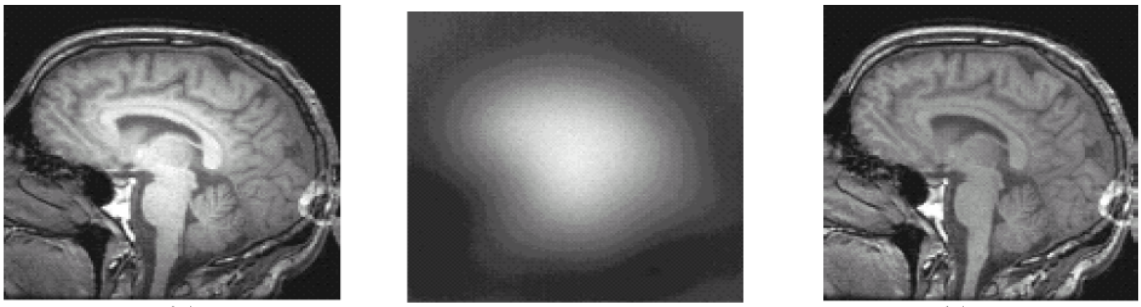
Les variations de champ ont pour conséquence une variation des intensités d'un même tissu dans une direction quelconque de l'image. Ce phénomène est dû au fait que le champ magnétique n'est pas parfaitement homogène spatialement et temporellement pendant une acquisition. Il existe de plus des non-linéarités de gradient de champ magnétique.
Des approches ont été proposées pour corriger les inhomogénéités du champ magnétique dans le cadre de pré-traitements **[Sled-98]**. Les distorsions de champ sont également analysées en détail et corrigées dans **[Lang-99].**

A.4.4 Effets de volume partiel

Les effets de volume partiel sont directement liés au processus de numérisation du signal.

Ainsi, si un pixel intersecte plusieurs objets, son niveau de gris sera une combinaison des niveaux de gris issus de chacun des objets traversés.

La prise en compte des effets de volumes partiels est nécessaire dans le cadre d'approches de segmentation dont l'objectif est d'effectuer des mesures sur les différents tissus. Cet artefact n'est pas très gênant pour le clinicien sauf dans des cas extrêmes, où le contraste entre différents tissus disparaît par exemple.



(a) (b) (c)

Figure A.11 : *L'inhomogénéité RF. (a) : Image affectée par une inhomogénéité RF, (b) : L'artefact RF isolé, (c) : Image sans artefact*

A.5 Conclusion

Nous avons présenté dans cette annexe les notions essentielles liées à l'anatomie du cerveau ainsi que les principes d'acquisition des images par résonance magnétique qui apporte une aide consistante en termes de diagnostic précoce et de caractérisation des tissus cérébraux.

Nous avons remarqué que l'imagerie à Résonance Magnétique est une technique d'imagerie médicale très utile pour l'observation du cerveau, car c'est la méthode d'imagerie qui, dans la plupart des cas, donne les informations les plus riches (la résolution spatiale habituelle est de l'ordre du millimètre).

Un volume d'IRM est divisé en deux super-classes : la super-classe dite normale et la superclasse dite anormale. La super-classe normale contient les tissus tels que MB, MG, et LCR. La super-classe anormale contient les zones pathologiques telles que tumeur, oedème et nécrose, etc.

BIBLIOGRAPHIE

[Abde-02] T-A Abdelmalik, G Laurent, R Michèle," *Architecture de fusion de données basée sur la théorie de l'évidence pour la reconstruction d'une vertèbre*", Traitement du Signal 2002 – Volume 19 – n°4

[Achm-05] A Achmirowicz J-B Beck, P-Y Delobel, C Vivarelli," *Imagerie IRM interventionnelle*", Projet Master MTS, UTC, 2004-2005.

[Agra-93] Agrawal R, Imielinski T, Swami A, "*Mining Association rules between sets of items in large database*", Proceedings of the ACM SIGMOD International Conference on Management of Data, Washington, DC, pp 207-216, May 26-28, 1993.

[Aït-06] L S Aït-Ali," *Analyse spatio-temporelle pour le suivi de structures évolutives en imagerie cérébrale multi- séquences*", thèse doctorat, devant l'Université de Rennes 1, 2006.

[Anto-90] Hendrik J. Antonisse. – Unsupervised credit assignment in knowledge-based sensor fusion systems. *IEEE transactions on systems, man, cybernetics*, vol. 20, September/October 1990, pp. 1153–1171.

[Appr-91] A Appriou, "*Probabilités et Incertitude en fusion de données Multi-Senseurs*", Revue Scientifique et Technique de la Défense, 1:27-40, 1991.

[Arif-05] M Arif , "*Fusion de Données : Ultime Etape de Reconnaissance de Formes*, *Applications à l'Identification et à l'Authentification*". Thèse de doctorat. Université de Savoie, 2005.

[Bari-94] C Barillot, D Lemoine, L Le Briquer, F Lachmann, B Gibaud, "*DataFusion in medical imaging: Merging Multimodal and Multipatient Images,Identification of Structures and 3D Display Aspects",* Yearbook of Medical Informatics,1:290-294, 1994.

[Barr-99] V Barra, JY Boire, "*Fusion de Données en Imagerie 3D du Cerveau, journée l'incertitude et l'imprécision en fusion d'informations : théories et applications*", Clermont-Ferrand, 16 Juin 1999.

[Barr-00] V Barra V, "*Fusion d'Images 3D du Cerveau : Etude de Modèles et Applications*", Ph.D. Thesis, Université d'Auvergne, Clermont-Ferrand (France), 2000.

[Barr-05] P Barralon, "*Classification et Fusion de Données Actimétriques pour la Télévigilance Médicale*", Thèse, L'université Joseph Fourier, 2005.

[Ben-06] R Ben Messaoud, "*Data mining*", Institut Universitaire de Technologie lumiere Licence C.E.STAT, 2006.

[Bens-96] A Bensaid, L Hall, J Bezdek, L Clarke, "*Partially Supervised Clustering for Image Segmentation*", Pattern Recognition, 29:859-871, 1996.

[Bezd-81] J Bezdek, "*Pattern Recognition with Fuzzy Objective Function Algorithms*", Plenum Press, New York, 1981.

[Bezd-93] J Bezdek, L Hall, L Clarke, "*Review of MR Image Segmentation Techniques using Pattern Recognition*", Medical Physics, 20:1033-1048, 1993.

[Bezd-97] R Nikhil, Pal and Kuhu Pal, J Bezdek "*AMixed C-Means Clustering Model*", 0-7803-3796-4/1997IEEE.

[Bitt-J] J Bittoun, "*L'imagerie par résonance magnétique*", cours de PCEM, UFR médicale Kremlin Bicêtre, Paris.

[Bloc-46] F Bloch, Nuclear Induction, "*Physical Review",* 70:460-474, 1946.

[Bloc-94] I Bloch," *Fusion de Données en traitement d'images : modèles d'information et décision*". Traitement du signal, 1994, vol.11, n°6, pp 435 - 446.

[Bloc-95] I Bloch, "*Fondements des probabilités et des croyances* ": une Discussion des Travaux de Cox et Smets, *15ème Colloque GRETSI, Juan les Pins,* 909-912, 1995.

[Bloc-96a] I Bloch, "*Information Combination Operators for Data Fusion: A Comparative Review with Classification*", IEEE Transactions on Systems, Man, and Cybernetics, 1:52-67, 1996.

[Bloc-96b] I Bloch, Some Aspects of Dempster-Shafer, "*Evidence Theory for Classification of multi-modality medical Images taking Partial*", Volume Effect into account, Pattern Recognition Letters, 17:905-913, 1996.

[Bloc-03] I Bloch, M Maître, J Le Cadre , V Nimier, R Reynaud, M Rombaut, F Ealet, B Collin, C Garbay, "*Fusion d'informations en traitement du signal et des images*", 2003.

[Bloc-04] I Bloch , M Maitre , "*Les méthodes de raisonnement dans les images Brique" VOIR Module RASIM*", 2004.

[Boir-00] V Barra, JY Boire, "*Tissue Characterization on MR Images by a possibilistic Clustering on a 3D Wavele Representation"*, Journal of Magnetic Resonance Imaging, 11:267-278, 2000.

[Bouc-93] B Bouchon-Meunier, "*La logique Floue*", Que saisje ? No. 2702, Edition Presses Universitaires de France, 1993.

[Bouc-95] B Bouchon-Meunier," *La Logique floue et ses Applications*", Addison-Wesley, 1995.

[Bouc-99] A Boucher, "*Une approche décentralisée et adaptative de la gestion d'information en vision*". Thèse de doctorat, Université Joseph Fourier, Grenoble. 1999.

[Bosc-03] M Bosc, "*Contribution à la détection des changements dans les séquences IRM 3D multimodales",* Thèse de doctorat. Université Louis Pasteur Strasbourg, 2003.

[Cino-L] L Cinotti, Imagerie médicale, cours de PCEM2 (CERMEP), http://www.cermep.fr/docs/cinotti/imagep2.pdf

[Clar-93] L Clarke, R Velthuizen, S Phuphanich, J Schellenberg, J Arrington, M Silbiger, "*MRI : Stability of three Supervised Segmentation Techniques*", Magnetic Resonance Imaging, 11:95-106, 1993.

[Clar-98] L Clark, C Matthew, Hall, Dmitry B. Goldgof, Robert. Velthuizen, F. Reed Murtagh, Martin S. Silbiger. "*Automatic tumor segmentation using knowledge-based techniques*". *IEEE Transactions on medical imaging*, vol. 17, n2, 1998, pp. 187–201.

[Cleu-04] G Cleuziou,"*Une méthode de classification non- supervisée pour l'apprentissage de règles et la recherche d'information*". These de doctoral, université d'Orléans, 2004.

[Coli-99] A Colin, J-Y Boire, "*MR/SPECT Fusion for the Synthetis of High-Resolution 3D functional Brain Images: a Preliminary Study*", Computer Methods and Programs in Biomedicine, 60:107-116, 1999.

[Cond-91] B Condon, "*Multi-Modality Image Combination: Five techniques for simultaneous MR-SPECT display*", Computerized Medical Imaging and Graphics, 5:311-318, 1991.

[Cout-05] O Couturier, "*contribution à la fouille de données : règles d'association et interactivité au sein d'un processus d'extraction de connaissances dans les données*", 2005.

[Dave-90] R Dave, "*Fuzzy Shell Clustering and Application to Circle Detection in Image Processing*", International Journal of General Systems, 41:343-355, 1990.

[Demp-67] A P Dempster, "*Upper and lower probability function in a context of uncertainty"*, Annals of math. Statistics, vol 38, pp. 325-339, 1967.

[Demp-68] A P Dempster, " *A generalization of bayesian inference"*, Jour. of the Royal Statistical Society, vol 30, pp. 205-247, 1968.

[Dess-05] Besse P, "*Data mining II. Modélisation Statistique & Apprentissage*", 2005.

[Dou-06] W Dou, "*Segmentation d'images multi spectrales basée sur la fusion d'informations : application aux images IRM*", Thèse doctorat de l'université de Caen, Spécialité : Traitement du Signal et des Images, 2006.

[Drom-98] A Dromigny-Badin, *"Fusion d'images par la Théorie de l'Evidence en Vue d'Applications Médicales et Industrielles"*, Thèse de doctorat, Institut National des Sciences Appliquées de Lyon, 1998.

[Dubo-80] D Dubois , H Prade, "*Fuzzy Sets and Systems: Theory and applications*", Academic press, New york, 1980.

[Dubo-85] D Dubois, H Prade, "*A Review of Fuzzy Set Aggregation Connectives*", Information Sciences, 36:85-121, 1985.

[Dubo-88] D Dubois D, H Prade , "*Possibility Theory, an approach to the omputerized processing of uncertainty*", Plenum Press, 1988.

[Dubo-91] D Dubois D, H Prade ,"*Combination of fuzzy information in the framework of possibility theory*". Data Fusion in Robotics and Machine Intelligence, pp.481–505, 1991.

[Dubo-99] D Dubois D, H Prade, R Yager, "*Merging Fuzzy Information, in Fuzzy Sets in Approximate Reasoning and Information System*", the Handbook of Fuzzy Sets Series, Kluwer Academic Publishers, 1999.

[Dubo-01] D Dubois D, H Prade, "*Possibility theory, probability theory and multiplevalued logics*",a clarification, Annals of Mathematics and Artificial Intelligence, Eds: Kluwer, Dordrecht, 32:35-66, 2001.

[Duda-73] R. Duda and P. Hart. "*Pattern Classification and Scene Analysis*". Wiley, New- York, 1973.

[Dunn-74] J Dunn. "*A fuzzy relative of the isodata process and its use in detecting* compact *well-separated clusters*". Journal of Cybernetics, 3(3) :32_57, 1974.

[Fabi-96] P Fabiani, "*Représentation dynamique de l'Incertain et Stratégie de prise en compte d'Informations pour un Système autonome en Environnement évolutif*", Thèse, Département d'Etude et de Recherche en Automatique, *ONERA*, 1996.

[Faou-04] N.-E El Faouzi,"*Fusion de données*", Rapport INRETS-LICIT, 2004.

[Fayy-96] U Fayyad , G Piatetsky-Shapiro , Smyth P, "*From Data Mining to Knowledge Discovery in Databases*", Dans aimag KDD overview pp 1-34, 1996.

[Fayy-98] Fayyad U, Piatetsky-Shapiro G, Smyth P, "*From Data Mining to Knowledge Discovery in Databases*", Advices in Knowledge Discovery and Data Mining, MIT Press, 1:pp 1-36, 1998.

[Fémé-03] J L Femenias, "*Probabilités et statistiques pour les sciences physiques",* paris : Dunod , 2003,418p.

[Fiot-06] Fiot C, "*Quelques techniques de fouille de données*", Master Pro, 2005/06.

[Floc-90] P Le Floch-Prigent, M Molho, and H Outin. 1990. "*Imagerie de l'encéphale, Anatomie et observations cliniques*." Laboratoires Sandoz.

[Fren-04] E Frenoux , "*Applications et validation de méthodologies de fusion de données en imagerie cérébrale*", Thèse, Présentée à l'Université d'Auvergne pour l'obtention du grade de Docteur d'Université, 2004.

[Gath-89] I Gath, B Geva, "*Unsupervised Optimal Fuzzy Clustering*", IEEE Transactions on Pattern Analysis and Machine Intelligence*,* 11:773-781, 1989.

[Germ-99] L Germond, "*Trois principes de coopération pour la segmentation en imagerie de résonance magnétique cérébrale"*, Thèse de doctorat L'université Joseph Fourier 1999.

[Gesu-94] V DI Gesu, L Romeo, "*An Application of Integrated Clustering to MRI Segmentation*", Pattern Recognition Letters, 45:731-738, 1994.

[Gill-00] R Gilleron and M Tommasi. "*Découverte de connaissances à partir de données*". Technical report, Grappa - Université de Lille 3, 2000.

[Gren-C] J.F. LE BAS, Chu - Grenoble ,"*Imagerie par résonance magnétique* ".

[Hadd-02] Med, H, Haddad, "*Extraction et impact des connaissances sur les performances des systèmes de recherche d'information*", these Université Joseph Fourier-Grenoble1, 2002.

[Hall-97] J Bezdek, LHall, M Clark, D Goldgof, L Clarke, "*Medical Image Analysis with Fuzzy Models*", Statistical Methods in Medical Research, 6:191- 214, 1997.

[Held-97] K Held, H Rota Kops, BJ Krause, WM Wells, R Kikinis, HW Muller-Gärtner, " *Markov random field segmentation of brain MR images*", IEEE Transactions on Medical Imaging, 16(6):878-886, 1997.

[Jagg-97] C Jaggi, S Ruan, J Fadili, D Bloyet, "*Approche markovienne pour la segmentation 3D des tissus cérébraux en IRM*", actes du 16ème colloque GRETSI, 327-330, 1997.

[Jaya-96] K,Jayarma . Udupa, Supun Samarasekera. – Fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation. *Graphical models and image processing*, vol. 58, n3, 1996, pp. 246–261.

[Jdl-91] U.S. Department of defense," *Data fusion lexicon*", Data Fusion Subpanel of the Joint Directors of Laboratories, Technical Panel for C3, 1991.

[Just-88] M Just, M Thelen, "*Tissue Characterization with T1, T2 and Proton Density Values:Results in 160 Patients with Brain Tumors*", Radiology, 169:779-785, 1988.

[Kame-91] M S Kamel , S Z Selim , "*A threshold fuzzy c-means algorithm for semi- fuzzy clustering*". Pattern Recognition 24, 1991, pp. 825-833.

[Kant-03] M Kantardzic. "*Data Mining – Concepts, Models, Methods,and Algorithms*". IEEE Press, Piscataway, NJ, USA, 2003.

[Kauf-73] A Kaufmann , "*Introduction à la Théorie des Sous_ensembles Flous. Vol. 1 Elements Théoriques de Base*",Edition MASSON, Paris, 1973.

[Kauf-90] L Kaufman , P Rousseeuw, "*Finding Groups in Data, an Introduction to Cluster Analysis*". Wiley, Canada, 1990.

[Kell-93] R Krishnapuram , J M Keller." *A Possibilistic Approach to Clustering".* IEEE Transactions on Fuzzy Systems, 1(2) :98{110, 1993.

[Kers-95] P R Kersten, "*Fuzzy Order Statistics and Its Application to Fuzzy Clustering*", Naval Air Warfare Center Weapons divion Report, August 1995.

[Khod-97] L.Khodja, "*Contribution à la Classification Floue non Supervisée*". Thèse *de* Doctorat , université de Savoie, France, 1997

[Kivi-84] K Kivinitty, "*NMR relaxation Times in NMR Imaging*", Annals of Clinical Research, 40 : 4-6, 1984.

[Kodr-98] Y Kodratoff., "*techniques et outils de l'extraction de connaissances à partir des données*", Signaux n°92 pp 38-43, Mars 1998.

[Kori-82] J.G. Koritké and H. Sick. "*Atlas de Coupes Sériées du Corps Humain*". Vol. 1 : Tête, Cou, Thorax. Urban & Schwarzenberg, Munich-Vienne- Baltimore, 1982.

[Kris-93] R Krish napuram and J.M.Keller, "*A Possibility Approach to Clustering*", IEEE Transactions of Fuzzy Systems, vol. 1, No. 2 pp. 98-110, May 1993.

[Kris-94] R Krishnapuram, "*Generation of membership functions via possibilistic clustering*", in Proceedings of the 3rd IEEE Conf. Fuzzy Syst., Orlando, FL, USA, July 1994, vol. 2, pp. 902-908.

[Kris-96] R Krishnapuram, J M Keller."*The Possibilistic CMeans Algorithm: Insights and Recommendations*", IEEE Trans. on Fuzzy Systems, 4(3):385-393, 1996.

[Kwan-96] R Kwan, A Evans, G Pike, "*An Extensible MRI Simulator for Post Processing Evaluation*", *SPIE, Visualization in Biomedical Computing,* 1131:135-140, 1996.

[Laro-05] D-T. Larouse,"*Discovering Knowledge in data An Introduction to Data mining*", Central Connecticut State University, 2005.

[Lang-99] S. Langlois, M. Desvignes, J.M. 1999. « *A simple Approach to Correcting the effects of non-linear gradient fields* ». Journal of Magnetic Resonance Imaging, 9(6), pp. 821–831.

[Laro-06] Morin Yves 2006 "*Larousse médical "* édition larousse.

[Lass-99] V Lasserre, "*Modélisation floue des Incertitudes des Mesures de Capteurs*", Thèse, Université de Savoie, 1999.

[Laut-73] PC Lauterbur, "*Image formation by induced local interaction: examples employing NMR*", *Nature*, 242:190-191, 1973.

[Laut-03] E Lefevre, "*Imagerie par résonance magnétique*", La science de l'image Département de physique, génie physique et optique, Université Lavai, 2003.

[Leco-05] G Lecomte, "*Analyse d'images Radioscopiques et Fusion d'informations multimodales pour l'amélioration du contrôle de pièces de fonderie*", Thèse Doctorat, Présentée devant L'institut Nationai des Sciences Appliquées de Lyon , 2005.

[Lefe-00] E Lefevre, P Vannoorenberghe, and O Colot, "*About the use of dempster-shafer theory for color image segmentation*", *in:* First International Conference on Color in Graphics and Image Processing, Saint-Etienne, France, October 2000.

[Levi-89] D Levin ,X Hu, K Tan, S Galhotra, C Pelizzari, G Chen, R Beck, C Chen , M Cooper, J Mullan, "*The Brain: Integrated 3D Display of MR and PET Images,Radiology*", 172:783-789, 1989.

[Lieb-07] J. Lieber (fortement mais librement inspire du cours d'Amedeo Napoli), ,"*Fouille de données : notes de cours* ",2007.

[Lure-03] C Lurette," *Développement d'une technique neuronale auto-adaptative pour la classification dynamique de données évolutives. Application à la supervision d'une presse hydrauloque*.", Thèse, Université des sciences et technologies de LILLE 2003.

[Magn-99] V Magnotta, D Heckel, N Andresen, T Cizaldo, P Corson, J Ehrhardt, W Yuh, "*Measurement of Brain Structures with Artificial Neural Networks: 2D and 3D Applications*", Radiology, 211:781-790, 1999.

[Masu-99] F Masulli, A Schenone, "*A Fuzzy Clustering based Segmentation System as Support to Diagnosis in Medical Imaging, Artificial Intelligence in Medicine*", 16:129-147, 1999.

[Meda-98] S Medasani, J Kim, R Krishnapuram, "*An Overview of Membership Function Generation Techniques for Pattern Recognition*", International Journal of Approximate Reasoning, 19:391-417, 1998.

[Mich-02] J Michel Delorme, M Teisseire M,"*L'apport de la fouille de données dans l'analyse de texte* ", 2002.

[Miya-06] S. Miyamoto, "*Classification and Clustering: A Perspective toward Risk Mining*" Sixth IEEE International Conference on Data Mining 2006.

[Miri-07] S Miri,"Segmentation des structures cérébrales en IRM: intégration de contraintes topologiques", Master 2 ISTI [R], Université Louis Pasteur Strasbourg, 2007.

[Moha-98] N Mohamed, M Ahmed, A Farag, "*Modified Fuzzy C-Means in Medical Image Segmentation*", Proceedings of IEEE Engineering in Medicine and Biology Society 20th Annual International Conference, 3:1377-1380, 1998.

[Nobl-06] V Noblet, "*Recalage non rigide d'images cérébrales 3D avec contrainte de conservation de la topologie*", Thèse, l'Université Louis Pasteur - Strasbourg I, 2006.

[Ohas-84] Y Ohashi, "*Fuzzy clustering and robust estimation ", In* 9th *Meeting of SAS Users Group International,* Floride, 1984.

[Patw-06] K P Detroja, R.D. Gudi , S.C. Patwardhan "*A possibilistic clustering approachto novel fault detection and isolation*" K.P. Detroja et al. / Journal of Process Control 16 (2006) 1055–1073 .

[Pedr-98] W Pedrycz, F Gomide. "*An introduction to fuzzy sets analysis and design*" The MIT Press, 1998.

[Pena-99] J Pena, J Lozano, P Larranaga, "*An empirical Comparison of four Initialization Methods for the k-means Algorithm*", Pattern Recognition Letters, 20:1027-1040, 1999.

[Phil-95] W Philipps, R Velthuizen, S Phuphanich, L Hall, L Clarke, M Silbiger ,"*Application of Fuzzy C-Means Algorithm Segmentation Technique for Tissue Differentiation in MR Images of a Hemorragic Glioblastoma Multiforme*", Magnetic Resonance Imaging, 13:277-290, 1995.

[Piat-96] E Piat , "*Fusion de Croyances dans le Cadre combiné de la Logique des Propositions et de la Théorie des Probabilités. Application à la Reconstructio*n *de Scèneen Robotique mobile",* Thèse, Université de Compiègne, 1996.

[Pony-00] O Pony, X Descombes, J Zerubia," *Classification d'images satellitaires hyperspectrales en zone rurale et périurbaine*", INRIA Rapport de Recherche, n° 4008, 2000.

[Prad-00] H Prade," *Possibility theory in information fusion*", 3rd international conférence on information fusion, pages PS, 5-24, 2000.

[Rich-04] N. Richard, M. Dojat, and C. Garbay. "*Automated segmentation of huma brain MR images using a multi-agent approach*". Artificial Intelligence in Medicine, 30 :153–175, 2004.

[Sand-91] S Sadeh , "*La Combinaison d'Informations incertaines et ses Aspects Algorithmiques*", Thèse, Université Paul Sabatier, Toulouse, 1991.

[Sapp-M] D Sappey-Mariner," *Résonance Magnétique Nucléaire*", cours pour l'université Claude Bernard-Lyon1.http://cism.univ-lyon1.fr/dominic/enseignement.html

[Semc-07] M Semchedine ,"*Système Coopératif Hybride de Classification dans un SMA : Application à la segmentation d'images IRM*",Magister , université Farhat Abbas ,Setif, 2007.

[Shaf-76] G Shafer, "*A Mathematical Theory of Evidence, Princeton University Press*", 1976.

[Shap-91] G. Piatetsky-Shapiro.Discovery, analysis, and presentation of strong rules. In G. Piatetsky-Shapiro and W. J. Frawley, editiors, Knowledge Discovery in databases, pages 229-238. AAAI/ MIT press,1991.

[Sled-98] J.G. Sled, A.P. Zijdenbos, and A.C. Evans 1998. " *A Nonparametric Method for Automatic Correction of Intensity Nonuniformity in MRI Data ".* IEEE Transactions on Medical Imaging, 17(1), pp. 87–97.

[Smet-90] P Smets, "*The Combination of Evidence in the Transferable Belief Model*", IEEE Transactions on Pattern Analysis and Machine Intelligence, 12:447-458, 1990.

[Stok-97] R Stokking, K Zuidereld, H Hulshoff Pol, P Van Rijk, M Viergever, "*Normal Fusion for Three-Dimensional Integrated Visualisation of SPECT and MR Brain Images*", The Journal of Nuclear Medicine, 38:624-629, 1997.

[Suck-99] J Suckling, T Sigmundsson, K Greenwood, E Bullmore, "*A modified Fuzzy Clustering Algorithm for Operator Independent Brain Tissue Classification of Dual Echo MR Images*", Magnetic Resonance Imaging, 17:1065-1076, 1999.

[Tuff-02] S Tufféry,"*data mining et scoring, Bases de données et gestion de la relation client,"* Groupe bancaire francais, universités de Rennes 1 et paris-Dauphine, 2002.

[Tuff-05] S Tufféry, "*data mining et statistique décisionnelle, l'intelligence dans les bases de données*", Groupe bancaire francais, universités de Rennes 1 et paris-Dauphine , 2005.

[Vale-01] M L Valet, " *Un système flou de fusion coopérative : application au traitement d'images naturelles*", thèse de doctorat, Université de Savoie, 2001.

[Velt-93] R Velthuizen, L Hall, L Clarke, A Bensaid, J Arrington, M Silbiger , "*Unsupervised Fuzzy Segmentation of 3D Magnetic Resonance Brain Images*", SPIE IS&T Conference, San Jose, CA, 1905:627-635, 1993.

[Velt-95] R Velthuizen , L Clarke , S Phuphanich, L Hall , A Bensaid , J Arrington , H Greenberg, M Silbiger, "*Unsupervised Measurement of Brain Tumor Volume on MR Images*". Journal of Magnetic Resonance Imaging, 5:594- 605, 1995.

[Wald-90] E Waltz and J Llinas," *Multisensor data fusion*". Artech House, 1990.

[Wald-92] L Wald, "*Data fusion, Definition and architectures, Fusion of images of different spatial resolutions*". Presses de l'Ecole des Mines Paris, 2002.

[Zade-65] L Zadeh," *Fuzzy Sets, Information and Control"*, 8:338-353, 1965.

[Zade-78] L Zadeh , "*Fuzzy Sets as a Basis for Theory of Possibility, International Journal of Fuzzy Sets and Systems*", 1:3-28, 1978.

[Zhan-94] Y.J. et al Zhang. "*Objective and quantitative segmentation evaluation and comparison*". *Signal Processing*, vol. 39, n3, 1994, pp. 43–54.

[Zigh-00] Zighed D.A., Duru G., Auray J.P., "*Sipina, méthode et logiciel*", A. Lacassagne 2000.

[Zigh-01] D A Zighed , Y Kodratoff, A Napoli A. "*Extraction de connaissance à partir d'une base de donnée*" Bulletin AFIA'01,2001