



N° d'ordre :.....

**UNIVERSITE DE M'SILA**  
**FACULTE DES MATHÉMATIQUES ET DE**  
**L'INFORMATIQUE**  
**Département d'Informatique**

**Mémoire**

Présenté pour l'obtention du diplôme de Magistère

Spécialité : STIC

Option : SIC

Par ARBOUCHE NASSIMA

**SUJET**

Fouille de données biomédicales complexes  
Application à la segmentation d'images médicales

Soutenu publiquement le 24/06/2010 devant le jury composé de :

Mr. A.ZIDANI	Université de Batna(MC)	Président.
Mr. Abd elouaheb Moussaoui	UFAS(MC)	Rapporteur.
Mr. M.C.BABAHENINI	Université de Biskra(MC)	Examineur.
Mr. F.CHERIF	Université de Biskra(MC)	Examineur.
Mr. M.BOURAHLA	Université de M'sila(MC)	Examineur.

Promotion 2006 – 2007

# Dédicaces

*Je dedie ce travail à mes très chers parents,  
A toute ma famille,  
A mes amis , collègues , et tout ceux qui m'ont aidé.*

# Remerciements

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

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

Mes vifs remerciements accompagnés de toute ma gratitude vont ensuite à mon encadreur Moussaoui Abdelouaheb, maître de conférence à l'université de Ferhat abbas, pour ses conseils et ses orientations.

Mes remerciements vont aussi à toute ma famille, particulièrement, mes parents.

Je remercie ma collègue et amie hafida de répondre toujours présente quand j'avais besoin d'elle.

Je remercie aussi les responsables de l'école doctorale de Batna, pour avoir veillé à nous fournir une formation riche pendant la première année de notre formation de magistère.

Je remercie également les membres de Jury qui ont accepté de juger mon travail.

# Table des matières

<b>Introduction</b>	<b>1</b>
<b>1 Fouille de Données</b>	<b>4</b>
1.1 Introduction . . . . .	4
1.2 Définition de la fouille de donnée ou «Data Mining» . . . . .	4
1.3 Entrepôt de données « Datawarehouse » . . . . .	5
1.4 Le DM dans le processus d'ECD . . . . .	5
1.4.1 La phase de préparation . . . . .	5
1.4.1.1 L'acquisition . . . . .	6
1.4.1.2 Prétraitement des données . . . . .	6
1.4.2 Phase Fouille de données . . . . .	7
1.4.3 Phase d'interprétation et d'évaluation . . . . .	7
1.5 Méthodes et techniques du data mining . . . . .	7
1.5.1 Les méthodes du DM . . . . .	8
1.5.1.1 Les modèles prédictifs . . . . .	9
1.5.1.2 Méthodes de structuration (segmentation) . . . . .	9
1.5.1.3 La description . . . . .	9
1.5.2 Les techniques utilisées . . . . .	10
1.5.2.1 Les techniques classiques . . . . .	10
1.5.2.2 Les techniques sophistiquées . . . . .	10
1.5.2.2.1 Algorithmes de segmentation . . . . .	10
1.5.2.2.2 Règles d'association . . . . .	11
1.5.2.2.3 Algorithmes de classification, estimation et prédiction	12
1.5.2.2.4 Métaheuristiques pour l'extraction de connaissances	15
1.6 Conclusion . . . . .	16
<b>2 Fouille de Données Complexes</b>	<b>18</b>
2.1 Introduction . . . . .	18
2.2 Données complexes . . . . .	18
2.3 Fouille de données complexes . . . . .	20
2.3.1 Préparation des données complexes . . . . .	22
2.3.1.1 Prétraitement des données complexes . . . . .	22
2.3.2 Une remise en cause des outils traditionnels . . . . .	23
2.3.3 Les travaux réalisés . . . . .	23
2.3.4 La description de quelques fouilles de données . . . . .	24
2.3.4.1 La fouille de données textuelles . . . . .	24
2.3.4.2 La fouille de séquences . . . . .	24

	2.3.4.3	La fouille de données spatiales . . . . .	24
	2.3.4.4	Web mining . . . . .	25
2.4		Fouille de données biomédicales complexes : . . . . .	25
	2.4.1	Diversité des données biomédicales . . . . .	25
	2.4.2	Caractéristiques et spécificités des données biomédicales . . . . .	26
	2.4.3	Définition de la fouille de données biomédicales . . . . .	27
	2.4.4	Les travaux réalisés . . . . .	28
2.5		Conclusion . . . . .	29
<b>3</b>		<b>Méthodes de Segmentation d'Images</b>	<b>30</b>
3.1		Introduction . . . . .	30
3.2		Segmentation d'images . . . . .	31
	3.2.1	La segmentation de l'IRM cérébrale . . . . .	31
	3.2.1.1	Spécificités de la segmentation de l'IRM cérébrale . . . . .	32
3.3		Les différentes approches de segmentation . . . . .	33
	3.3.1	État de l'art du point de vue « Traitement d'images » . . . . .	34
	3.3.1.1	Approches Région . . . . .	34
	3.3.1.1.1	Le suillage . . . . .	34
	3.3.1.1.2	La croissance de région . . . . .	35
	3.3.1.1.3	Morphologie mathématique . . . . .	36
	3.3.1.2	Approches contours . . . . .	37
	3.3.1.2.1	Méthodes dérivatives . . . . .	37
	3.3.1.2.2	Modèles déformables . . . . .	38
	3.3.1.2.3	Espace-échelle . . . . .	39
	3.3.2	État de l'art du point de vue «Reconnaissance des formes» . . . . .	39
	3.3.2.1	Les méthodes de classification . . . . .	41
	3.3.2.1.1	Classification supervisée . . . . .	41
	3.3.2.1.2	Classification semi supervisée . . . . .	41
	3.3.2.1.3	Classification non supervisée (Automatique) . . . . .	42
	3.3.2.2	Méthodes supervisées . . . . .	42
	3.3.2.2.1	Approches probabilistes . . . . .	42
	3.3.2.3	Méthodes non supervisées (Automatiques) . . . . .	44
	3.3.2.3.1	La famille des C-moyennes . . . . .	45
	3.3.2.3.2	Les métaheuristiques pour l'extraction des connaissances . . . . .	50
	3.3.2.3.3	Autres méthodes . . . . .	54
3.4		Conclusion . . . . .	54
<b>4</b>		<b>Fusion de Données</b>	<b>55</b>
4.1		Introduction . . . . .	55
4.2		Éléments théoriques . . . . .	56
	4.2.1	Représentation de Connaissances ambiguës « imparfaites » . . . . .	56
	4.2.1.1	Théorie des probabilités . . . . .	56
	4.2.1.1.1	Cadre bayésien . . . . .	57
	4.2.1.2	Théorie des ensembles flous . . . . .	57
	4.2.1.3	Théorie des possibilités . . . . .	61
	4.2.1.4	Mesure de possibilité . . . . .	61
	4.2.1.4.1	Distribution de possibilité . . . . .	61

	4.2.1.4.2	Mesure de nécessité . . . . .	62
	4.2.1.5	Théorie des fonctions de croyance . . . . .	63
	4.2.1.5.1	Fonction de masse . . . . .	63
	4.2.1.5.2	Fonction de croyance . . . . .	64
	4.2.1.5.3	Fonction de plausibilité . . . . .	64
	4.2.1.5.4	Propriétés des fonctions de croyance . . . . .	64
	4.2.1.5.5	Relation entre les trois théories . . . . .	65
	4.2.1.6	Autres approches . . . . .	65
4.3		Fusion d'information . . . . .	65
	4.3.1	Caractéristiques générales des données . . . . .	66
	4.3.2	Système de fusion et type d'architecture . . . . .	67
	4.3.3	Modèles de fusion . . . . .	67
	4.3.4	Classification des types de fusion . . . . .	69
	4.3.5	Opérateurs de fusion . . . . .	69
	4.3.5.1	Classification des opérateurs de fusion . . . . .	70
	4.3.5.2	Quelques propriétés . . . . .	71
	4.3.6	Méthodes numériques de fusion . . . . .	71
	4.3.6.1	Fusion en théorie des probabilités . . . . .	71
	4.3.6.1.1	Etape de fusion . . . . .	71
	4.3.6.1.2	Etape de décision . . . . .	71
	4.3.6.2	Fusion floue et possibiliste . . . . .	72
	4.3.6.2.1	Etape de fusion . . . . .	72
	4.3.6.2.2	Etape de décision . . . . .	72
	4.3.6.3	Fusion en théorie des croyances . . . . .	73
	4.3.6.3.1	Etape de fusion . . . . .	73
	4.3.6.3.2	Etape de décision . . . . .	73
4.4		Fusion des données médicales . . . . .	74
	4.4.1	Les applications médicales de la fusion de données . . . . .	76
	4.4.2	Fusion d'image médicale . . . . .	76
	4.4.2.1	Introduction de l'information spatiale . . . . .	77
4.5		Conclusion . . . . .	78
<b>5</b>		<b>Contribution</b> . . . . .	<b>79</b>
	5.1	Introduction . . . . .	79
	5.2	Approche proposée . . . . .	80
	5.2.1	Quelle est la fusion envisagée ? . . . . .	81
	5.2.2	Les types de données à représenter ? . . . . .	82
	5.2.2.1	Modélisation de la donnée issue du dossier médical . . . . .	82
	5.2.2.1.1	Variables linguistiques . . . . .	83
	5.2.2.2	Modélisation de l'image . . . . .	84
	5.2.2.2.1	Méthode supervisée ou non ? . . . . .	85
	5.2.2.2.2	Classification floue ou non floue ? . . . . .	85
	5.2.2.2.3	l'algorithme utilisé ? . . . . .	85
	5.2.2.3	Informations supplémentaires . . . . .	93
	5.2.2.4	Comment envisage-t-on la fusion ? . . . . .	94
	5.2.3	Niveau de fusion ? . . . . .	94

5.2.4	Cadre théorique de fusion ? . . . . .	94
5.2.5	Opérateur de fusion ? . . . . .	95
5.2.5.1	Opérateur proposé . . . . .	97
5.2.5.2	Pourquoi avoir choisi des pondérations dynamiques ? . . . . .	98
5.2.6	Critère de décision . . . . .	98
5.2.7	Architecture de fusion . . . . .	100
5.2.8	Présentation de l'algorithme général . . . . .	100
5.3	Conclusion . . . . .	101
<b>6</b>	<b>Résultats et validation</b>	<b>102</b>
6.1	Introduction . . . . .	102
6.2	Choix des données biomédicales . . . . .	102
6.3	Description des volumes traités . . . . .	104
6.3.1	Fantôme du cerveau . . . . .	104
6.3.2	Images simulées . . . . .	104
6.3.3	Images réelles . . . . .	104
6.4	Critères de validation . . . . .	105
6.5	Résultats de la classification . . . . .	106
6.5.1	Résultats de classification sur des données de synthèse . . . . .	106
6.5.2	Résultats de classification sur des données réelles . . . . .	109
6.5.3	Comparaison des résultats obtenus . . . . .	111
6.6	Conclusion . . . . .	115
	<b>Conclusion et perspectives</b>	<b>116</b>
	<b>Appendices</b>	
<b>A</b>	<b>Anatomie du cerveau</b>	<b>119</b>
A.1	Introduction . . . . .	119
A.2	Système nerveux central . . . . .	119
A.2.1	L'encéphale . . . . .	120
A.2.1.1	Tronc cérébral . . . . .	121
A.2.1.2	Le cervelet . . . . .	121
A.2.1.3	Le cerveau . . . . .	122
A.2.1.4	Les méninges . . . . .	123
A.2.1.4.1	Le liquide céphalo-rachidien . . . . .	123
A.2.1.4.2	La substance grise . . . . .	124
A.2.1.4.3	La matière blanche . . . . .	124
<b>B</b>	<b>Les images IRM</b>	<b>125</b>
B.1	Introduction . . . . .	125
B.2	L'imagerie par résonance magnétique (IRM) . . . . .	126
B.2.1	Principe de fonctionnement . . . . .	126
B.2.2	Notions élémentaires sur le magnétisme . . . . .	127
B.2.2.1	Champ magnétique et moment magnétique . . . . .	127
B.2.2.2	Moment magnétique et aimantation . . . . .	127
B.2.2.3	Propriétés magnétiques de la matière . . . . .	128

B.2.2.4	Mouvement de précession et fréquence de Larmor . . . . .	129
B.2.2.5	Aimantation macroscopique d'un ensemble de spins . . . . .	130
B.2.3	Phénomène de résonance magnétique nucléaire . . . . .	131
B.2.3.1	Phase d'excitation . . . . .	131
B.2.3.2	Phase de relaxation et temps T1 et T2 . . . . .	132
B.2.4	Création d'une image IRM . . . . .	134
B.2.5	Le contraste des images . . . . .	135
B.2.6	Les agents de contrastes en IRM . . . . .	137
B.2.7	Les artefacts . . . . .	137
B.2.7.1	Artefact de mouvement . . . . .	137
B.2.7.2	Inhomogénéités RF . . . . .	138
B.2.7.3	Volume partiel . . . . .	138
B.2.7.4	Le bruit . . . . .	138

**Appendices**

<b>Bibliographie</b>	<b>139</b>
----------------------	------------

# Table des figures

1.1	Etapes du processus d'ECD. . . . .	6
1.2	Les tâches du data mining. . . . .	8
1.3	Les algorithmes du data mining. . . . .	8
2.1	Axes de complexité [35]. . . . .	20
2.2	La fusion de différents types de données. . . . .	21
3.1	L'effet de volume partiel (d'après [24]). . . . .	33
3.2	Image IRM segmentée par une approche région. . . . .	36
3.3	Segmentation par modèle déformable. . . . .	38
3.4	Exemple de scale space gaussien. [101] . . . . .	39
3.5	Principe de la classification. . . . .	39
3.6	Les méthodes de classification (d'après [83]). . . . .	40
3.7	Segmentation par méthode bayésienne. . . . .	43
3.8	Algorithme de fourmis. . . . .	53
4.1	Exemple de représentation d'une fonction d'appartenance floue [5]. . . . .	58
4.2	Exemple de sous-ensemble flou. [40] . . . . .	59
4.3	Exemple d'intersection et d'union. [40] . . . . .	60
4.4	Exemples de $\alpha$ -coupes associées au sous-ensemble flou $A$ . [40] . . . . .	60
4.5	Graphe de fusion. . . . .	68
4.6	comparaison entre les 3 théories. . . . .	75
4.7	Résultat de la classification par fusion floue de deux IRM. . . . .	77
5.1	La partie concernée par notre étude (en rouge). . . . .	80
5.2	Exemple d'une variable linguistique d'après [71]. . . . .	83
5.3	Fuzzification de la donnée âge suivant une fonction trapézoïdale. . . . .	84
5.4	la fonction de similarité . . . . .	88
5.5	Résultats de la classification des images par FCM et AntClust. . . . .	90
5.6	FCM : Sensibilité aux données aberrantes (2 classes demandées). . . . .	93
5.7	Fusion par moyenne pondérée, . . . . .	97
5.8	Fusion par moyenne pondérée [28] . . . . .	97
5.9	Architecture de la fusion proposée. . . . .	100
6.1	Segmentation des images de synthèse avec bruit. . . . .	107
6.2	Comparaison entre la méthode proposée et AntClust+FCM. . . . .	107
6.3	La sensibilité du système proposé avec changement de la valeur de l'âge . .	113
6.4	La spécificité du système proposé avec changement de la valeur de l'âge . .	113

6.5	Le recouvrement du système proposé avec changement de la valeur de l'âge	114
6.6	La similarité du système proposé avec changement de la valeur de l'âge . . .	114
6.7	Comparaison des 4 critères d'évaluation . . . . .	114
A.1	Encéphale. . . . .	120
A.2	Coupe transversale de l'encéphale. . . . .	120
A.3	Tronc cérébral. . . . .	121
A.4	Cervelet. . . . .	121
A.5	Lobes du cerveau. . . . .	122
A.6	Méninges. . . . .	123
A.7	Architecture schématique d'un neurone. . . . .	124
B.1	Acquisition par IRM. . . . .	127
B.2	Moment magnétique. . . . .	128
B.3	Mouvement de précession d'un ensemble de moments magnétiques. . . . .	129
B.4	Mouvement de précession. . . . .	129
B.5	Les deux orientations de spin d'un proton. . . . .	131
B.6	Perturbation de l'aimantation M par une impulsion RF. . . . .	132
B.7	Relaxation longitudinale. . . . .	133
B.8	Relaxation transversale. . . . .	134
B.9	Modalités IRM . . . . .	134
B.10	Signaux temporel et fréquentiel. . . . .	135
B.11	Acquisition en T1. . . . .	136
B.12	Images d'une même coupe du cerveau humain, . . . . .	136

# Liste des tableaux et Algorithmme

1.1	Méthodes du data mining nécessitant ou non un apprentissage . . . . .	15
1	Algorithme général de type ACO . . . . .	51
5.1	Risque d'atteinte par âge . . . . .	84
2	L'algorithme AntClust . . . . .	89
3	L'algorithme des C-moyennes floues . . . . .	91
4	Algorithme général . . . . .	100
6.1	Evaluations de la segmentation de la tumeur par le système développé . . .	108
6.2	Evaluations de la segmentation de la tumeur par l'approche proposée . . .	108
6.3	Evaluations de la segmentation de la tumeur par l'approche proposée . . .	108
6.4	Evaluations de la segmentation de la tumeur par AntClust+FCM . . . . .	109
6.5	Images reelles segmentées par l'approche proposée et par AntClust+FCM(1)	110
6.6	Images reelles segmentées par l'approche proposée et par AntClust+FCM(2)	111
6.7	Evaluation de la segmentation d'images réelles par l'approche proposée. . .	112
6.8	Evaluation de la segmentation d'images réelles par l'algorithme proposé . .	112
6.9	Evaluation de la segmentation d'images réelles par l'algorithme proposé. . .	112
6.10	Evaluation de la segmentation d'images réelles par AntClust+FCM. . . . .	113

# Introduction

L'ACCUMULATION des données et la croissance explosive de leur collecte génère des montagnes de données. Face à ce déluge d'informations, il fallait un outil pour les exploiter. Le domaine du Data Mining (DM) - dit aussi fouille de données - et plus généralement de l'Extraction de connaissances des données (ECD) - ou Knowledge Discovery in Databases (KDD) - sont nés de ce besoin [41] [102].

Selon KODRATOFF [174] et rappelée par ZIGHERD et al. [46], la fouille de données n'est qu'une étape d'un processus plus global appelé Extraction de Connaissances à partir de Données (ECD), ou Knowledge Discovery from Databases (KDD) en anglais.

La fouille de données a acquis une certaine maturité. En effet, Plusieurs travaux existent déjà, des méthodes et des logiciels permettent d'extraire des connaissances à partir des données généralement tabulaires, sous forme de tableaux « individus x variables ». Cependant le monde informationnel qui nous entoure est plus complexe, les données ne sont pas seulement numériques ou symboliques et peuvent se présenter de façon multimodales, elles sont souvent représentées dans des format différents (textes, images, son, vidéos, bases de données, etc.), provenant de sources différentes (données de production, scanners, satellites, enregistrements, vidéos, compte-rendus médicaux, résultats d'analyse, web, etc.), ayant une sémantique différente (langues différentes, échelles différentes, évolution de la définition d'une donnée dans le temps, etc.). De telles données sont désignées par les termes de données complexes.

Malgré les extensions méthodologiques qui ont été proposées pour exploiter ce type de donnée, La fouille de données s'est jusque là contentée de traiter un seul type de donnée à la fois : données tabulaires, données textuelles « text mining », données images « image mining », de manière indépendante des autres ce qui a conduit a un développement uni modale. La fouille de données provenant d'une seule modalité a montré ses limites dans la compréhension du contenu des informations. C'est pourquoi, une voie prometteuse pour analyser la sémantique des informations consiste en la prise en compte simultanée de la totalité des données disponibles sur une entité, ainsi l'utilisation combinée de modalités

multiples (informations visuelles, auditives et sonores, textuelles, du Web, etc.) est devenu une nécessité. L'enjeu désormais est donc de pouvoir fusionner, combiner, des informations de nature différente et les rattacher à une même unité sémantique qu'est le dossier médical par exemple. Le dossier médical est le meilleur exemple qui illustre cette problématique. En effet, les systèmes d'aide à la décision médicale nécessitent l'analyse de données hétérogènes, chaque dossier peut contenir des données tabulaires telles que des résultats d'analyse biologique, des données textuelles venant des rapports cliniques, des données images telles que des radiographies, des IRM ou des électrocardiogrammes. Dans un cadre de prise de décision, le traitement de chaque type d'information séparément a des inconvénients sérieux. Il semble donc de plus en plus nécessaire de considérer ces données différentes simultanément, entourant de ce fait toute leur complexité. Ce mémoire s'inscrit dans le cadre général du traitement de ces données complexes. Nous avons plus particulièrement utilisé les données médicales pour la mise en application de ces traitements. Notre application consiste à tenter de combiner deux sources de données différentes qui sont les images de modalité IRM et des données tabulaires pour aider le praticien dans la prise de ses décisions.

Ce manuscrit s'articule autour de ce schéma. :

**chapitre1** Le premier chapitre est consacré à la présentation de la fouille de données (DM), sa position dans le processus d'extraction des connaissances "KDD" et enfin ses différentes méthodes et techniques.

**chapitre2** Le second chapitre, se décompose en deux parties : la première abordera la problématique actuelle qui est la fouille de données complexes, traitant la particularité des ces données, tandis que la deuxième traite la fouille dans un type de données complexes bien particulier, qui est le type de données biomédicales. Nous présenterons la définition des données complexes, la fouille dans ce type de données, des exemples de prétraitement de données complexes et des outils utilisés et enfin une liste non exhaustive des travaux réalisés.

**chapitre3** Parmi les types de données complexes, l'image tient une place considérable, Le troisième chapitre portera sur les différentes méthodes de segmentation de cette dernière, on s'intéressera aux images médicales plus exactement aux images de modalité IRM.

**chapitre4** Le quatrième chapitre présente le concept de la fusion de données. Nous présentons dans un premier temps les approches permettant d'intégrer la représentation des connaissances incertaines et/ou imprécises : la théorie des probabilités, la théorie floue, 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 différents cadres théoriques précédemment présentés.

**chapitre5** Le cinquième chapitre est dédié à la présentation de notre contribution, nous présentons l'architecture de fusion proposée ,les données participant à cette fusion ainsi que leur modélisation.

**chapitre6** Le sixième chapitre présente les résultats obtenus dans le cadre de cette étude et illustre le fonctionnement du système sur des images réelles et simulées.

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

En complément, nous précisons en annexe l'anatomie du cerveau et l'imagerie IRM. En conclusion, nous dressons le bilan de ce travail et nous avançons quelques nouveaux éléments de réflexion et perspectives de recherche.

# FOUILLE DE DONNÉES

---

## 1.1 Introduction

L'avènement du Web et l'expansion des données multimédia ont généré une floraison de données disponibles dans divers domaines décisionnels (médecine, marketing, veille concurrentielle, CRM...). Face à cette masse exponentielle d'informations, la question de l'analyse de grands volumes de données s'est imposée, La fouille de données est la solution qui permet de traiter l'ensemble de ces informations afin d'en extraire la connaissance.

## 1.2 Définition de la fouille de donnée ou «Data Mining»

Il existe différentes définitions du DM dans la littérature. Les définitions données ci-dessous semblent complémentaires.

- Le DM est l'art d'extraire des connaissances à partir des données. S'il comprend l'application d'algorithmes d'apprentissage en vue de construire des modèles, il n'est qu'une étape d'un processus plus global appelé Extraction de Connaissances à partir de Données (ECD), [174] et rappelée par [46] .
- Le DM a pour but d'extraire des connaissances à partir des données par des méthodes de structuration (apprentissage non supervisé) ou par des méthodes explicatives (apprentissage supervisé), une fois les données acquises et préparées.
- D'après [58], le DM consiste, depuis un ensemble de données, à découvrir des mo-

dèles : (i) soit fonctionnels sous la forme  $f(x_1, \dots, x_n) = y$  (par exemple : une régression linéaire  $y = ax + b$ ) ; (ii) soit logique comme les règles d'association ou les arbres de décision.

On confond souvent le « data mining », au sens étroit qui désigne la phase d'extraction des connaissances, avec la découverte de connaissances dans les bases de données (KDD ou Knowledge Discovery in Databases) [63]. La fouille de données n'est en fait qu'une étape d'un processus plus global appelé Extraction de Connaissances à partir de Données (ECD) [174] [46].

### 1.3 Entrepôt de données « Datawarehouse »

Ensemble de données historisées variant dans le temps, organisé par sujets, consolidé dans une base de données unique, géré dans un environnement de stockage particulier, aidant à la prise de décision. Les entrepôts doivent accomplir trois fonctions essentielles qui sont la collecte de données de bases existantes et chargement, la gestion des données dans l'entrepôt et enfin l'analyse des données pour la prise de décision.

### 1.4 Le DM dans le processus d'ECD

L'Extraction de Connaissances à partir de Données (ECD) est un processus complexe qui se déroule suivant une série d'opérations [53]. Nous pouvons regrouper ces opérations en trois étapes majeures. Elles sont la préparation des données, la fouille de données à proprement parler qui est l'étape centrale de l'ECD et enfin la validation des modèles ainsi élaborés [47] (Voir la figure 1.1).

#### 1.4.1 La phase de préparation

La phase de préparation des données est particulièrement importante dans le processus Data Mining [135] [110], elle comporte :

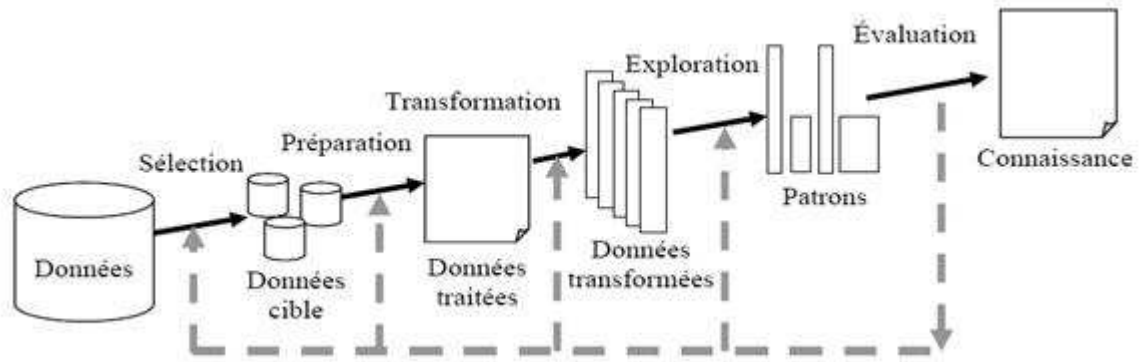


FIGURE 1.1 – Etapes du processus d'ECR.

### 1.4.1.1 L'acquisition

Les données peuvent être distribuées, stockées selon des architectures variées, elles peuvent être structurées ou non selon différents types : données tabulaires ou textuelles, images, sons ou séquences vidéo. La phase d'acquisition sert généralement à nettoyer les données qui sont rapatriées. [118]

### 1.4.1.2 Prétraitement des données

Les données issues de la phase de l'acquisition ne sont pas nécessairement toutes exploitables par des techniques de fouille de données. En effet, la plupart des techniques ne traitent que des tableaux de données numériques rangées sous forme lignes/colonnes. Certaines méthodes sont plus contraignantes que d'autres. Elles peuvent par exemple exiger des données binaires, comme c'est le cas pour la technique des règles d'association. Les données acquises peuvent être de types différents, leur préparation consiste à homogénéiser les données et à les disposer en tableau lignes/colonnes. Pour les données qui sont déjà sous forme tabulaire, une transformation telle qu'un centrage par rapport à la moyenne ou une normalisation est exigée. Les principales opérations de préparation peuvent être listées comme suit :

- **Sélection de ligne/colonne** : Elle s'effectue sur des données qui sont déjà sous forme tabulaire. Il s'agit ensuite de définir un filtre qui permet de sélectionner un sous-ensemble de lignes ou de colonnes. L'objectif est soit de réduire le nombre de données soit de sélectionner les lignes ou colonnes les plus pertinentes par rapport

aux préoccupations de l'utilisateur.

- **Le traitement des données manquantes ou aberrantes** : Certaines données peuvent être absentes et gêner ainsi l'analyse. Il faut donc définir des règles pour gérer ou pour remplacer ces données manquantes.
- **Les transformations d'attributs** : Il s'agit de transformer un attribut A en une autre variable A' qui serait, selon les objectifs de l'étude, plus approprié.
- **La construction d'agrégats** : Dans certaines situations particulières, il peut s'avérer que des agrégats d'attributs soient très importants pour la tâche d'analyse. Un agrégat d'attribut est un nouvel attribut obtenu selon une transformation précise.

## 1.4.2 Phase Fouille de données

La fouille de données concerne le data mining dans son sens restreint et est au cœur du processus d'ECD. Cette phase fait appel à de multiples méthodes issues de la statistique, de l'apprentissage automatique, de la reconnaissance de formes ou de la visualisation. Les méthodes de data mining permettent de découvrir ce que contiennent les données comme informations ou modèles utiles.

## 1.4.3 Phase d'interprétation et d'évaluation

Enfin, cette étape identifie les modèles intéressants représentant les connaissances, en se basant non seulement sur des mesures d'intérêt mais aussi sur l'avis de l'expert. Le processus présenté est itératif et plusieurs retours en arrière dans les différentes étapes sont nécessaires pour affiner les résultats.

## 1.5 Méthodes et techniques du data mining

D'après Anand et Buchner [6] le data mining offre des algorithmes et des outils pour la découverte de modèles non triviaux, implicites, non connus, potentiellement utiles et compréhensibles à partir d'une grande masse de données.

### 1.5.1 Les méthodes du DM

On y retrouve des méthodes statistiques bien établies, mais aussi des développements récents issus directement de l'informatique. Dans la littérature on parle aussi de :

- méthodes exploratoires ou descriptives (non supervisées) où il s'agit de découvrir des structures ou des comportements inattendus,
- modèles prédictifs (supervisées) où une « réponse » est à prédire.

Le data mining est en fait un ensemble de techniques complémentaires dédiées à différentes tâches. D'après Westphal et Blaxton [170], les tâches du data mining se partagent entre la classification, l'estimation, la segmentation et la description (Voir la figure 1.2).

Apté dans [8] propose une classification des algorithmes du data mining en trois catégo-

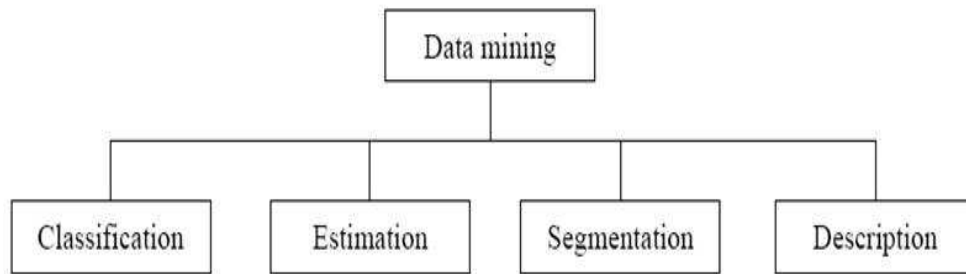


FIGURE 1.2 – Les tâches du data mining.

ries : prédiction, segmentation et recherche de patrons fréquents (Voir la figure 1.3).

Ces représentations s'accordent. Ceci permet d'arriver à la description suivante :

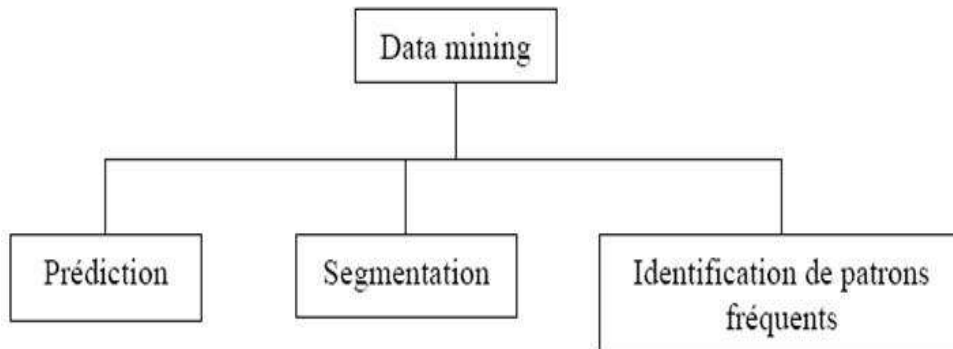


FIGURE 1.3 – Les algorithmes du data mining.

### 1.5.1.1 Les modèles prédictifs

Ont pour but de déterminer une fonction (ou un modèle) qui associe des entrées et des sorties. Les réseaux de neurone, les régressions, les arbres de décision et les règles de décision sont les méthodes les plus utilisées. Ces méthodes sont des méthodes supervisées en ce sens qu'il faut préciser quelles sont les entrées et quelle est la sortie à prédire. En fonction de la nature de la variable de sortie, deux sous catégories d'outils coexistent :

- *La classification* : Si la variable de sortie est de type discret (catégoriel), la classification aura pour rôle de construire le modèle qui permettra de classer correctement les enregistrements, c'est-à-dire d'assigner des catégories prédéfinies aux données. Parmi les techniques les plus appropriées à la classification, on trouve la classification des k moyennes et les réseaux de kohonen.
- *L'estimation* : L'estimation est similaire à la classification, excepté la variable cible est numérique au lieu d'être catégorielle. Le résultat d'une estimation est obtenu par une ou plusieurs fonctions combinant les données en entrée. Le résultat d'une estimation permet de procéder aux classifications grâce à un barème. Les réseaux de neurones sont souvent employés dans cette activité.

### 1.5.1.2 Méthodes de structuration (segmentation)

Ces méthodes regroupent toutes les techniques d'apprentissage non supervisé et de classification automatique provenant des domaines de la reconnaissance de formes, de la statistique, de l'apprentissage automatique et du connexionisme ; Les algorithmes de ces méthodes visent à identifier des ensembles d'éléments qui partagent certaines similarités, en maximisant l'homogénéité à l'intérieur de chaque ensemble et maximisant l'hétérogénéité entre les ensembles. Différentes méthodes sont utilisées pour définir ces groupes : Algorithmes hiérarchiques, réseaux de neurones, cartes topologiques de Kohonen...etc.

### 1.5.1.3 La description

Les méthodes de description uni, bi et multidimensionnelles sont numériques, pour la plupart, elles sont issues de la statistique descriptive et de l'analyse des données, ainsi que des techniques de visualisation graphique dont certaines font appel à la réalité virtuelle et à des métaphores calquées sur le modèle mental de l'humain. Elles Consistent à expliquer

les relations existantes dans les données. Les méthodes utilisées cherchent à identifier des associations entre des variables.

## 1.5.2 Les techniques utilisées

On peut dégager deux grandes catégories de méthodes d'analyse consacrées à la fouille de données. 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 :

### 1.5.2.1 Les techniques classiques

On y retrouve des outils généralistes de l'informatique ou des mathématiques : Les requêtes dans les bases de données, simples ou multicritères, 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.

### 1.5.2.2 Les techniques 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 métaheuristiques.

**1.5.2.2.1 Algorithmes de segmentation** Rappelons que l'objectif est de réaliser une partition (division en groupes disjoints) d'une population afin d'effectuer plus facilement des tâches ou des expertises. Les algorithmes tels que : K-moyennes, Méthode

d'agglomération (*méthode ascendante*) et Méthode dichotomique (*méthode descendante ou arbres de décision*) ne fournissent pas forcément le même résultat, ils correspondent à des heuristiques différentes.

– **Les avantages :**

- *Apprentissage non supervisé* : la méthode des k-moyennes et ses variantes résolvent une tâche dite non supervisée, c'est-à-dire qu'elle ne nécessite 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).
- *Facile à implanter* : la méthode ne nécessite que peu de transformations sur les données (excepté les normalisations de valeurs numériques), il n'y a pas de champ particulier à identifier, les algorithmes sont faciles à implanter et sont, en règle générale, disponibles dans les environnements de « data mining ».

– **Les 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. Ceci est général pour les méthodes de segmentation.

**1.5.2.2.2 Règles d'association** Rappelons que l'application principale est « l'analyse du panier de la ménagère », c'est à dire la recherche d'associations entre produits sur les tickets de caisse. Les domaines applicables sont tous les secteurs pour lesquels il est intéressant de rechercher des groupements potentiels de produits. Par exemple en médecine, pour rechercher des complications dues à des associations de médicaments.

– **Les 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,
  - *Méthode adaptable* : on peut faire intervenir des produits virtuels (date, jour, saison) afin de « temporaliser » les données à analyser.
- **Les 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,
  - Le coût calculatoire de la méthode est important.

**1.5.2.2.3 Algorithmes de classification, estimation et prédiction** Rappelons que ces trois tâches sont fonctionnellement liées. Elles s'appuient sur le principe de détermination d'une variable particulière d'un individu connaissant le comportement d'un échantillon donné.

**Méthode des plus proches voisins « PPV »** Cette méthode est dédiée à la classification et peut être étendue à des tâches d'estimation. Elle part de l'idée de prendre des décisions en recherchant un ou des cas similaires déjà résolus en mémoire. Cependant, il n'y a pas d'étape d'apprentissage consistant en la construction d'un modèle à partir d'un échantillon. C'est une fonction de distance et une fonction de choix de la classe qui constitue le modèle. On doit choisir a priori le nombre « k » de voisins d'un individu. On définira préalablement des distances sur chaque champ et une distance entre enregistrements.

- **Les 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 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.

– **Les 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 nécessitent 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.
- *Stocker le modèle* : le modèle est l'échantillon, il faut donc un espace de mémoire important pour le stocker ainsi que des méthodes d'accès rapides pour accélérer les calculs.
- *Définir les distances et le 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.

**Les arbres de décision** Il s'agit de générer des représentations graphiques de classifications à partir de données. Les nœuds internes de l'arbre sont des tests sur les champs, les feuilles sont les classes. Les deux algorithmes les plus connus et utilisés sont CAR (Clas-

sification And Regression Trees) et C5. Ces algorithmes sont performants et génèrent des procédures de classification exprimables sous forme de règles. Pour utiliser un arbre de décision, c'est-à-dire classer un enregistrement, il suffit de descendre dans l'arbre selon les réponses aux différents tests pour l'enregistrement considéré. Les règles des systèmes construits sont exhaustives et mutuellement exclusives. Cela signifie que pour tout enregistrement une et une seule règle s'applique.

– **Les avantages :**

- *Adaptabilité aux attributs de valeurs continus* : par exemple, dans C5, si A est un attribut continu, pour sélectionner un test, l'algorithme fait participer à la compétition tous les tests de la forme  $A > a$  où a est une valeur prise par l'attribut A dans l'ensemble d'apprentissage.
- *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.
- *Génération de règles* : C5 propose également de générer un système de règles à partir de l'arbre de décision. Le système obtenu n'est pas une simple réécriture de l'arbre car des transformations et simplifications sont effectuées.
- *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 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 des techniques qui génèrent un ensemble d'arbres votant pour attribuer la classe.

– **Les 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).

	Méthode des k-voisins (D)	Arbres de décision	Réseaux de neurones (A)	Algorithmes spécifiques
<b>Segmentation</b>	Non	Oui : méthodes dichotomique descendante (D)	Oui	-Méthode des k-moyennes (D) -Méthode d'agglomération (D)
<b>Classification</b>	Oui	Oui : CART, C5 (A)	Oui	-
<b>Estimation</b>	Oui	Oui : CART, C5 (A)	Oui	-
<b>Prédilection</b>	Oui	Oui : CART, C5 (A)	Oui	-
<b>Règles d'association</b>	Non	Non	Non	Table de co-occurrence (D)

TABLE 1.1 – Méthodes du data mining nécessitant ou non un apprentissage (D) Méthodes dynamique ou incrémentales, ne nécessitant pas d'apprentissage (A) Méthodes nécessitant une phase d'apprentissage sur un échantillon exemple

**1.5.2.2.4 Métaheuristiques pour l'extraction de connaissances** Les métaheuristiques<sup>1</sup> sont un ensemble de méthodes qui permettent de concevoir des algorithmes. Elles permettent de résoudre des problèmes d'optimisation auxquels les ingénieurs et les décideurs sont régulièrement confrontés. La majorité des problèmes d'extraction de connaissances peuvent s'exprimer comme des problèmes d'optimisation combinatoire. Or,

1. Le mot métaheuristique est dérivé de la composition de deux mots grecs : heuristique qui vient du verbe heuriskein (euriskein) et qui signifie 'trouver', et meta qui est un suffixe signifiant 'au-delà', 'dans un niveau supérieur'.

de nombreux problèmes d'optimisation combinatoire sont NP-difficiles et ne pourront donc pas être résolus de manière exacte dans un temps "raisonnable" puisque la capacité de calcul des machines évolue linéairement alors que le temps nécessaire à la résolution de ces problèmes évolue exponentiellement. Lorsqu'on s'attaque à des problèmes réels, il faut se résoudre à un compromis entre la qualité des solutions obtenues et le temps de calcul utilisé.

Les métaheuristiques sont souvent inspirées par des systèmes naturels, qu'ils soient pris en physique (cas du recuit simulé), en biologie de l'évolution (cas des algorithmes génétiques) ou encore en éthologie (cas des algorithmes de colonies de fourmis ou de l'optimisation par essais particuliers).

Selon [95] les métaheuristiques peuvent être classées en deux groupes : les méthodes à solution unique et les méthodes à population de solutions.

– **Les métaheuristiques à solution unique**

Les méthodes itératives à solution unique sont toutes basées sur un algorithme de recherche de voisinage qui commence avec une solution initiale, puis l'améliore pas à pas en choisissant une nouvelle solution dans son voisinage [163]. Les méthodes les plus utilisées extraction de connaissances sont : les méthodes de descente, le recuit simulé et la recherche tabou.

– **Les métaheuristiques à population de solutions**

Les méthodes d'optimisation à population de solutions améliorent, au fur et à mesure des itérations, une population de solutions. L'intérêt de ces méthodes est d'utiliser la population comme facteur de diversité. Les méthodes les plus utilisées sont : La recherche par dispersion (Scatter Search), Les algorithmes génétiques, Programmation génétique, Algorithmes à essaim de particules, Systèmes immunitaires artificiels, Algorithmes à Estimation de Distribution, et Les colonies de fourmis (Ants System).

## 1.6 Conclusion

Dans ce premier chapitre, nous avons présenté les principaux concepts de fouille de données ; les tâches et les méthodes les plus utilisées, ainsi que les avantages et les inconvénients de chaque méthode.

Cependant la fouille de données abordée dans ce chapitre ne traite que les types de données simples structurées sous forme de tables à deux dimensions, hors les objets du monde réel ne se présentent pas tous de façon unimodale, mais incorpore également des données textuelles, images ou graphiques, de connaissances, voire de véritables ontologies, le cha-

pitre suivant abordera la fouille dans de tels types de données c-à-d la fouille des données complexes.

---

# FOUILLE DE DONNÉES COMPLEXES

---

## 2.1 Introduction

Dans tous les domaines tels que le multimédia, la télédétection, l'imagerie médicale, les bases de données, le web sémantique, la bio informatique et bien d'autres, les données à traiter pour y extraire de la connaissance utilisable sont de plus en plus complexes et volumineuses. Face à cette complexité de données la fouille de données se trouve confronté à un nouveau challenge, Qui consiste à prendre en compte simultanément la totalité des données disponibles sur une entité. Dans le domaine médical par exemple, il faut traiter le dossier patient en prenant en compte la totalité des données qui y figurent.

## 2.2 Données complexes

Les données complexes présentent de nos jours un enjeu important pour leur structuration, leur stockage et leur analyse. Cependant, il n'existe pas de définition unique pour les données complexes. Elles dépendent directement du contexte des domaines d'applications (domaine médical, domaine de l'emploi, marketing, etc.) on peut voir le type de la données en lui-même étant complexe par exemple une image médicale ; ou bien comme un objet complexe le cas du patient, là ou il faut combiner plusieurs données hétérogènes qui se rattache à cet objet « patient » pour le décrire. [151] Plus généralement, nous considérons que les données sont complexes lorsqu'elles vérifient au moins une des conditions suivantes :

- **Nature différente** : les données relatives à un objet sont de nature différente, l'information est portée par des données de types différents (données numériques, symboliques, textes, images, sons, vidéos. . .) ;. Outre le cas classique de descripteurs numériques ou catégoriels, on a notamment le cas de données texte, image ou audio vidéo.
- **Diversité des sources** : provient de sources hétérogènes (bases de données, web, entrepôts ...) comme le montre bien l'exemple des dossiers médicaux, les données recueillies correspondent aussi bien à des questionnaires remplis par le médecin, à des mesures acquises par des appareils à des comptes rendus textuels, médicaux couplés à des ordinateurs, à des images radiologiques ou échographiques, etc.
- **Multi-structures** : les données peuvent être structurées, semi-structurées ou non structurées (bases de données relationnelles, collection de documents XML, documents textes ...)
- Multimodales : un même phénomène est décrit par plusieurs canaux ou points de vue (échelles ou langues différentes ...)
- **Multi-versions** : les données sont évolutives en termes de définition ou de valeur (bases de données temporelles, recensements périodiques ...)
- **Évolutives et distribuées** : il arrive souvent que l'on dispose de plusieurs caractérisations du même objet à des époques et/ou en des localisations différentes. Un patient est généralement suivi périodiquement par plusieurs médecins dont chacun produit une information spécifique. Ces informations s'intègrent autour d'un même sujet.
- **Liées à des connaissances externes** : la fouille intelligente des données s'appuie sur la prise en compte des connaissances externes (par des experts), dites du domaine, celle-ci pouvant se faire par le biais d'une ontologie. Dans le domaine de la cancérologie par exemple, les connaissances diagnostiques et thérapeutiques sont organisées sous forme d'arbres de décision et mises à disposition des praticiens sous la forme d'un guide des bonnes pratiques.
- **Dimensionnalité des données** : par le fait même que l'on associe différentes sources à différents moments, on multiplie les points de vue, les données, et une multitude de descripteurs potentiels. Différents problèmes d'analyse apparaissent alors, notamment, la forte dimensionnalité qui engendre à son tour des difficultés algorithmiques et méthodologiques.

Cependant, selon [35] cette première définition est insuffisante et ne couvre pas la grande variété des données complexes. Elle pourrait par contre être envisagée comme un axe de complexité parmi d'autres, comme des axes liés à la sémantique des données ou à leur traitement (*figure 2.1*). La volumétrie des données pourrait également constituer

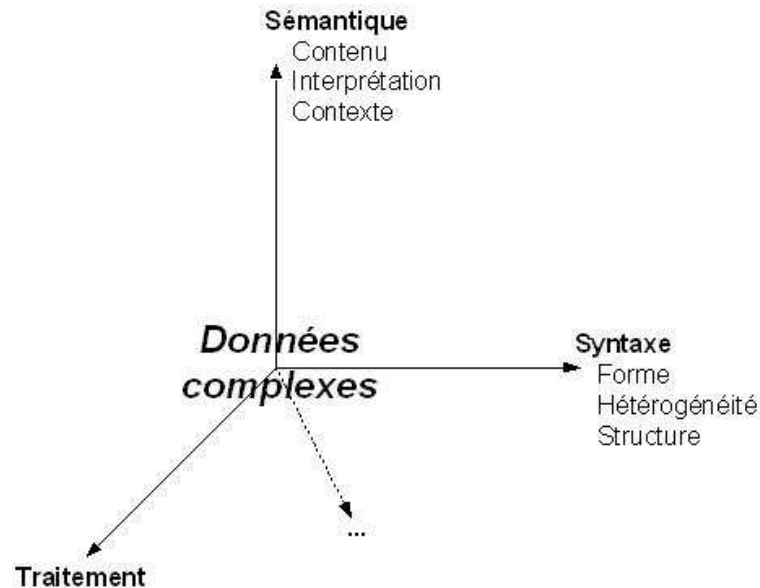


FIGURE 2.1 – Axes de complexité [35].

un tel axe. Bien que le volume des données ne soit pas une expression de complexité intrinsèque s'il est vu en termes de  $n$ -uplets d'une base de données, il devient un problème difficile en statistique ou en fouille de données lorsque c'est le nombre d'attributs qui augmente.

En conclusion, nous définissons ici un cadre permettant d'identifier ce que nous appelons des données complexes. Cependant, cette définition ne pouvant être exhaustive, nous la laissons aussi ouverte que possible à d'autres axes de complexité.

## 2.3 Fouille de données complexes

Il n'existe pas encore une définition pour la fouille de données complexes ; est ce qu'on va garder la même définition que celle de la fouille de données ? Garder la même définition en lui rajoutant quelques nouvelles notions ? Ou chercher une toute nouvelle définition ? D'après la littérature :

- Les travaux réalisés jusqu'à présent ne traitent qu'un seul type de données complexes à la fois, par exemple la fouille de données séquentielles, symboliques, textuelles et images.
- Alors que le défi reste levé, pour la prise en compte simultanée de la totalité des données disponibles sur une même entité (Voir la figure 2.2) ;

- Le problème ici se situe dans la façon de faire la combinaison entre un ensemble de données hétérogènes qui se rattache à une même entité sémantique.

Sur le plan méthodologique, il s'agit donc de définir une mesure de ressemblance ou de dissemblance entre deux objets dont la description est fournie par un complexe d'informations. Jusqu'à présent, les approches proposées ont surtout été fondées sur la juxtaposition de similarités ou de dissimilarités partielles [16]. D'après [175] les études réalisées sur la fouille de données complexes soulève souvent des problèmes similaires qui nécessitent des adaptations ou des extensions des méthodes de fouille de données traditionnelles, Ou bien le développement de nouvelles techniques avancées pour l'extraction de données complexes [86].

En général, on retrouve soit des adaptations en amont dans la phase de préparation de données, ou bien des extensions des algorithmes de fouille en y intégrant les spécificités des données complexes analysées [175].

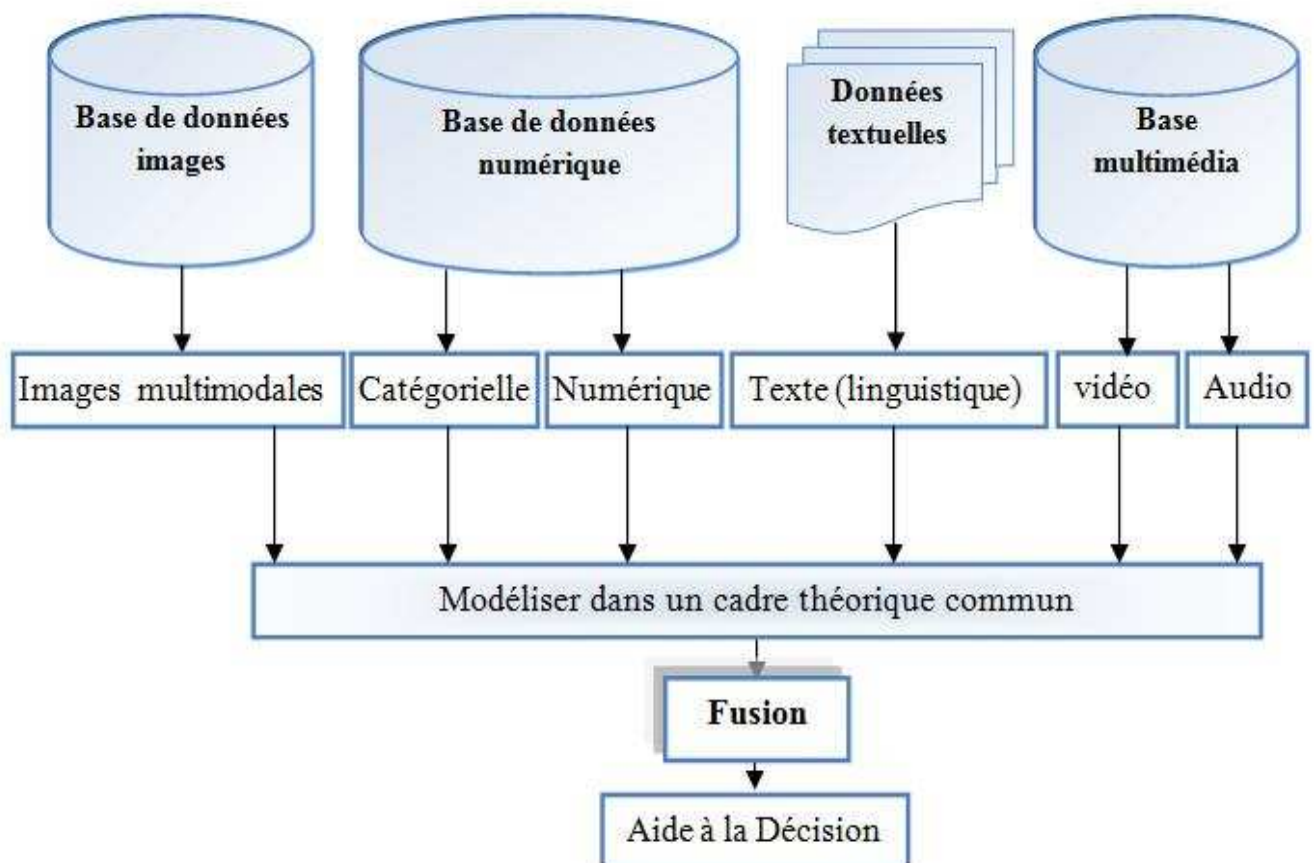


FIGURE 2.2 – La fusion de différents types de données.

### 2.3.1 Préparation des données complexes

Dans l'extraction de connaissances à partir de données complexes, plus qu'ailleurs, la préparation des données pour la fouille s'avère plus difficile. Le principal problème qui se pose est celui du choix de l'espace de représentation. En effet, à l'origine, les instances d'objets enregistrés dans la base de données sont exprimées dans un formalisme qui ne se prête pas ou peu aux traitements mathématiques qui fondent la plus part des méthodes de fouille de données. Le passage par un codage unifié, généralement sous forme vectorielle, s'avère souvent incontournable. Comment alors passer de données textuelles, images, vidéo, temporelles vers des vecteurs ? Faut-il " vectoriser " et tout aligner dans un tableau unique ? Ces choix ne sont pas neutres car ils peuvent engendrer, à leur tour, d'autres difficultés. Par exemple, le choix de coder par un vecteur numérique les textes, exige des procédés linguistiques qui sont souvent très sophistiqués et où l'intuition, les connaissances a priori du domaine et même les choix arbitraires sont couramment mis à contribution. Et cela, sans, pour autant, être certain du bon choix [34].

Comment alors réduire la dimensionnalité avec une perte minimale d'information ? Par sélection ? Par élimination ? Par projection ?

Et, surtout, comment évaluer la pertinence du nouvel espace de représentation ? De plus, les données qui arrivent sont parfois entachées de bruits et incomplètes. A défaut de redresser les anomalies, peut-on au moins prendre en compte l'incomplétude, l'incertitude et l'imprécision dans nos analyses ?

#### 2.3.1.1 Prétraitement des données complexes

Toutes les méthodes de prétraitement citées précédemment *cf. 1.4.1.1* opèrent sur des tableaux de données lignes/colonnes. Or il arrive que nous travaillions sur des données non structurées qui ne sont pas présentés sous forme de tableaux [145]. Par exemple en : *Fouille de texte* :

Nous disposons d'un ensemble de textes de longueurs variées qu'il convient de ramener à une forme tabulaire les textes de longueurs variées, l'une des technique est de :

- recenser l'ensemble des mots de tout le corpus
- calculer la fréquence de chacun de ces mots pour chaque texte représenté par une ligne ou une colonne, pour obtenir ainsi un tableau de comptage.
- codage des textes fait appel à des procédures plus élaborées qui s'appuient sur la linguistique : lemmatisation, suppression des mots vides, thesaurus ou ontologies du domaine.

*La préparation des données images, et a fortiori vidéo :*

C'est une tâche difficile pour le non-spécialiste car il faut :

- définir la liste des attributs qui décrivent l'image. (utiliser les caractéristiques de l'histogramme des niveaux de gris, les attributs de texture, etc).
- La fouille sur des images peut nécessiter d'autres transformations en amont qui relèvent plus du traitement de l'image que de l'ECD [145].

### 2.3.2 Une remise en cause des outils traditionnels

Vue l'accroissement du nombre et de la complexité des données contenues dans les entrepôts et les bases de données. La recherche dans de tels gisements est devenue une tâche très difficile et entraîne même parfois à l'extraction de faux modèles. De ce fait le développement de techniques avancées pour l'extraction de connaissances à partir des données complexes est devenu impératif. Ces algorithmes se doivent efficaces, pour pouvoir tirer profit de **cette** et exploiter les avantages de cette complexité; par exemple la réduction de la dimension, des méthodes d'approximation, échantillonnage...etc. Cependant il est essentiel de développer différentes techniques pour l'exploitation des différentes bases de données, étant donné la diversité des types de données et des objectifs.

### 2.3.3 Les travaux réalisés

Le data mining est désormais utilisé dans divers domaines, où les données ne sont pas nécessairement simples (numériques ou symboliques) ou bien structurées, par exemple comme celles qu'on peut trouver dans des bases de données relationnelles ou transactionnelles ou bien dans des entrepôts. Les données complexes sont aujourd'hui largement collectées et analysées d'une façon régulière. Une liste non exhaustive des fouilles de données complexes réalisées ou en cours, où la fouille est réalisée selon le type de données :

- Fouille de données textuelles.
- Fouille de données symboliques.
- Fouille de données spéciales.
- Fouille de données séquentielles.
- Fouille de données spatiotemporelles.
- Fouille de données dans le web (web mining).
- Fouille de données multimédia.

- Fouille de données images.
- Fouille dans les graphes

## 2.3.4 La description de quelques fouilles de données

### 2.3.4.1 La fouille de données textuelles

Est un domaine de recherche très vaste à la frontière de la recherche d'information. Elle permet d'enrichir par la sémantique (et parfois une structure) des données à l'origine non structurées. La catégorisation fait partie des méthodes les plus utilisées. Les travaux visent essentiellement à améliorer la qualité de l'apprentissage. Les algorithmes réputés les plus efficaces sont k-NN (ou les k plus proches voisins), la méthode des centroïdes et SVM [175].

### 2.3.4.2 La fouille de séquences

Contrairement aux données textuelles les données séquentielles sont bien structurées et la fouille de séquences est un problème plus restreint. Il s'agit d'étendre les méthodes de fouille de données envers la prise en compte de l'ordre séquentiel et d'optimiser les performances d'exécution.

### 2.3.4.3 La fouille de données spatiales

Permet d'extraire des propriétés spatiales cachées dans ces données et présente donc un intérêt certain pour les applications spatiales décisionnelles. La fouille de données spatiale (FDS) est aujourd'hui identifiée comme un domaine de la fouille de données à part entière. Elle résulte de la combinaison de la fouille de données et des bases de données spatiales.

#### 2.3.4.4 Web mining

Permet l'analyse de la fréquentation de sites web et du comportement des utilisateurs. Parmi ses applications on cite : la fidélisation, la mesures d'efficacité de campagnes de promotion, . . .etc

## 2.4 Fouille de données biomédicales complexes :

En santé, les experts ont accumulé ces dernières années de très grandes quantités de données (études épidémiologiques et pharmaceutiques, génomique, imagerie etc.). En effet l'accroissement des connaissances médicales est explosif et l'augmentation du nombre de paramètres nécessaires à la prise en charge des patients se fait en parallèle. Le développement ininterrompu de nouvelles méthodes d'investigations comme les techniques d'imagerie, les explorations fonctionnelles ou les techniques dérivées du génie biologique contribue à cette augmentation des connaissances. Il devient de plus en plus difficile, pour un médecin, de maîtriser les connaissances nécessaires à la pratique des soins, voire simplement d'y accéder en temps utile.

D'une façon plus générale, la gestion d'objets complexes est devenue nécessaire à la prise de décisions. Ces objets recouvrent, en plus des images, des textes en langage clair, des signaux (électriques, acoustiques, électro-physiologiques, etc.) mais aussi des ensembles de concepts reliés entre eux (connaissances physiopathologiques) ou des procédures devant être activées sous certaines conditions (protocoles de calcul d'une dose optimale d'un médicament, protocoles de soins infirmiers, etc.).

Le domaine biomédical s'inscrit parmi les domaines aux quels le data mining s'intéresse. Ainsi la communauté scientifique s'intéresse à la technologie des entrepôts de données, aux outils et techniques d'analyse en ligne connus sous l'appellation OLAP et à *la fouille de données biomédicales*.

### 2.4.1 Diversité des données biomédicales

Les informations liées aux patients peuvent être construites à partir de deux types de données :

- structurées (un champ cholestérol)

- non structurées (texte, image).

En effet le dossier médical ne se résume plus aux notes cliniques du médecin mais devient la compilation d'analyses de plus en plus sophistiquées.

On parle désormais d'objets complexes. Ces objets recouvrent, des données numériques pouvant être structurées de façon tabulaire comme les résultats d'analyses biochimiques, mais incorpore également des données textuelles comme les comptes-rendus d'observations cliniques, ou encore des graphiques tels que les tracés d'électrocardiogramme ou d'électro-encéphalogramme, ainsi que des images complexes fournies par les radiographies, échographies, IRM ou scanners ; mais aussi des ensembles de concepts reliés entre eux (connaissances physiopathologiques) ou des procédures devant être activées sous certaines conditions (protocoles de calcul d'une dose optimale d'un médicament, protocoles de soins infirmiers, etc.).

Toutes ces informations participent dans le processus de perception-décision-action en médecine, et notamment en chirurgie.

## 2.4.2 Caractéristiques et spécificités des données biomédicales

En médecine, peut-être plus que dans d'autres disciplines, les observations sur lesquelles se basent les décisions médicales sont très imparfaites. Elles peuvent être floues, ambiguës, incomplètes, incertaines. Le processus décisionnel, qu'il soit diagnostique, thérapeutique ou pronostique est donc un processus sous incertitude. Plusieurs éléments participent à cette situation [84] [130].

- Les observations sont floues parce qu'il n'y a pas de vocabulaire standardisé communément utilisé par la communauté médicale et répondant à des définitions clairement exprimées.
- Elles sont ambiguës car le malade peut exprimer une plainte et le médecin en entendre une autre. De plus, des observateurs différents ont des seuils de perception différents dans l'appréciation de la présence ou l'absence d'un signe.
- Les observations sont incomplètes car, en situation de décision, le médecin doit agir sans connaître l'ensemble des données relatives à un patient et bien entendu toute la connaissance spécifique de la situation. Ceci est particulièrement vrai dans des situations d'urgences, où un patient peut se retrouver inconscient, lorsque les résultats de laboratoire ne sont pas encore disponibles ou lorsqu'une technique particulière fournissant des informations utiles pour le diagnostic n'est pas pratiquée dans la structure de soins.
- Elles sont incertaines car les connaissances cliniques sont l'expression d'observations

statistiques sur des échantillons de patients présentant des maladies plus ou moins fréquentes, ayant des formes cliniques différentes et ne s'exprimant pas toujours par la même symptomatologie, partageant certains signes avec d'autres maladies ou présentant des réponses variables à un traitement donné.

- La reproductibilité des observations médicales est fonction des méthodes de mesures (variabilité analytique), des observateurs (variabilité intra- et inter-observateur et du sujet observé (variabilité intra et inter-individuelle)
- Données manquantes ou bruitées
- Données hétérogènes : Les données médicales peuvent se présenter sous différentes formes : descriptions textuelles, différents types d'images, valeurs numériques, signaux, multimédia...
- Données Multidimensionnelles : la multidimensionnalité signifie que de nombreux éléments "différents" de données, représentant chacun une dimension qui peut varier en valeur, et caractérisant ainsi un point d'intérêt. L'entité patient en forme le meilleur exemple, en effet il n'est pas inhabituel pour un dossier médical d'un patient de contenir 50 ou même de 100 différents types d'éléments de données. Cependant la prise en compte de l'ensemble de ces dimension n'est pas une tâche aisée et reste un défi à relever.
- Interprétations : l'interprétation des données par différents lecteurs spécialistes reste subjective.
- Questions d'éthique, légale et sociale : La difficulté rencontrée lors de la collecte des données (souvent collectées à partir de peu de patients) ce problème est dû aux problèmes d'éthique et de confidentialité des données médicales.

Ces caractéristiques typiques des données médicales ne facilite pas la tâche d'extraction des données, mais n'en forme pas pour autant un obstacle, néanmoins des adaptations des techniques de fouille ainsi que des traitements supplémentaires seront nécessaires, ce qui complique considérablement le processus de fouille, et accroît l'effort et les coûts [clinic in lab livre].

### 2.4.3 Définition de la fouille de données biomédicales

D'après K. Cios et al. [93], la fouille de données médicales peut être identifiée comme une tâche particulière d'extraction de connaissances. En effet, les données médicales sont souvent hétérogènes, soumises à des problèmes d'éthique et de confidentialité, et de ce fait elles sont souvent collectées à partir de peu de patients. Dans ce contexte, la fouille de données médicales peut s'avérer être un problème difficile et nécessitant la connaissance

d'un expert [60].

Plusieurs évolutions majeures ont marqué la pratique médicale des vingt dernières années, les données nécessaires à l'extraction de connaissance sont souvent réparties dans différents supports (dossier médical hospitalier, bases de données "locales" à un service, dossier administratif...), sous différentes formes (textuelles, multimédia...), leur exploitation à des fins de recherche clinique et évaluative nécessite alors d'une part leur intégration, d'autre part la mise au point de méthodes d'extraction de l'information adaptées aux spécificités des données médicales (exp, non déterministes, bruitées, entachées de valeurs manquantes) et la définition de processus pour mettre en œuvre ces méthodes sur des jeux de données réelles.

L'application des techniques du data mining au domaine biomédical est une aventure certes enrichissante, mais très difficile. Cette difficulté est non seulement liée aux grandes bases de données mais aussi principalement aux problèmes cités ci-dessous :

- Le premier problème concerne la spécificité des données médicales : Il s'agit de données particulières, complexes, hétérogènes, hiérarchiques souvent imprécises, subjectives, bruitées, entachées de valeurs manquantes, ou incomplètes... La phase de prétraitement apparaît alors comme une phase cruciale dans le processus de KDD. Au niveau de l'intégration des données, elle conditionne la qualité des informations contenues dans les bases de données décisionnelles et qui seront utilisées par les méthodes de fouille de données et par les outils de pilotage. Au niveau de la fouille des données proprement dite, la qualité des données et la méthode de prise en compte des données manquantes ont une influence sur les résultats obtenus par ces techniques.
- Le second problème concerne la nécessité d'une validation par les experts de l'ensemble des traitements effectués sur des données médicales, que ce soit au niveau de la phase de prétraitement ou au niveau des résultats produits par les algorithmes de fouille de données.

Ces problèmes font des algorithmes de la fouille de données classique des outils non appropriés pour traiter ce type de données. De ce fait des adaptations de ces derniers s'avèrent nécessaires avant toute application au domaine biomédical.

#### 2.4.4 Les travaux réalisés

L'utilisation des méthodes de data mining au domaine biomédical est en forte croissance. Une liste non exhaustive de quelques travaux de recherche est dressée ci-dessous :

- Fouille de règles d'association dans des images médicales [61]
- Fouille de données biomédicales basées sur une ontologie

- L’intégration des données, le data mining (fouille de données) et le support à la décision : Réalisé par IBM Research Lab –Zurich [7] ; où ils ont utilisé la techniques d’intégration (processus de combinaison de l’information). Pour intégrer ensemble les données médicales, la technique de fédération était utilisée (regroupement de différentes entités chacune gardant sa propre organisation).
- Fouille de données du génome à l’aide de modèles de Markov cachés [64]

## 2.5 Conclusion

Nous avons introduit dans ce chapitre un cadre général à la fouille de données ou plus généralement à l’extraction de connaissances à partir de données complexes.

Après avoir défini et identifié les données complexes ainsi que la fouille dans ces dernières, on a illustré la particularité de la fouille dans les données complexes et le défi qu’elle veut soulever en montrant la difficulté rencontrée lors de la phase de préparation et en présentant quelques exemples parmi ces derniers, des exemples représentatifs illustrant ce nouveau challenge peuvent être tirés du domaine biomédical. En effet, les objets manipulés dans ce domaine sont souvent des objets complexes tels : le patient, le cerveau, les images, les pathologies. . .

Dans ce document nous accordons une attention particulière aux images et aux techniques de fouille dans ces dernières, la segmentation formant l’une de ces techniques suscite un grand intérêt auprès de la communauté scientifique.

En effet, La recherche en imagerie médicale est une des disciplines les plus actives du traitement d’images. La segmentation et l’analyse d’images dans un contexte clinique reste un problème majeur de l’imagerie médicale. La multiplicité des modalités d’imagerie, ainsi que les fortes variabilités des structures et pathologies à analyser rendent cette tâche fastidieuse.

Les méthodes de segmentation d’images en général et particulier des images IRM font l’objet du chapitre suivant.

---

# MÉTHODES DE SEGMENTATION D'IMAGES

---

## 3.1 Introduction

Dès le développement des techniques d'imagerie, les images, qu'elles soient numériques ou analogiques, ont été utilisées pour aider au diagnostic et à la préparation et à la réalisation de gestes chirurgicaux. Avant celles-ci, le diagnostic s'effectuait à partir de corrélations entre la sémiologie et l'histoire.

L'imagerie médicale permet d'obtenir des images de l'intérieur du corps humain à partir de différents principes physiques (ultrasons, rayons X, rayons gamma, champ magnétique...). Elle a connu un grand développement ces dernières années, notamment grâce à l'apparition de l'imagerie par résonance magnétique (*Ref* .Annexe B). Les images IRM sont pour les médecins une source d'informations précieuse et conséquente, permettant l'analyse et le suivi des pathologies. Ces images fournissent, de façon non invasive, des représentations contrastées et de haute résolution de l'anatomie cérébrale, permettant ainsi son observation *in vivo*. Cependant, l'étude systématique et « manuelle » de clichés est longue et fastidieuse, et souffre de la variabilité d'interprétation des spécialistes. Ainsi, de nombreuses recherches ont été menées afin d'automatiser l'analyse des images ; la segmentation en forme le point de départ.

Bien que la segmentation et la reconnaissance de structures en imagerie médicale soient des thèmes de recherche anciens, ils suscitent toujours beaucoup de travaux, tant les problèmes rencontrés sont difficiles et évoluent avec les techniques d'acquisition. On Présentera dans ce chapitre une revue non exhaustive des méthodes de segmentation, en insistant particulièrement sur la segmentation d'image cérébrale en IRM.

## 3.2 Segmentation d'images

La segmentation est un traitement de bas niveau qui consiste à créer une partition de l'image observée en un certain nombre de régions selon un critère déterminé (couleur, texture, niveau de gris. . . etc) qui pourront par la suite être traitées de manières différentes [100]. La segmentation d'images est l'un des problèmes phares du traitement d'images. En effet elle constitue l'une des étapes les plus importantes pour l'analyse des données d'images [121].

La qualité de la segmentation, c'est à dire, la précision de la localisation et la non-confusion des régions, a un impact direct sur les performances de l'application. Malgré les nombreuses recherches qui ont eu lieu dans le passé sur les méthodes de segmentation, le problème de la segmentation reste toujours mal posé, car on ne sait jamais dire quelle est la segmentation idéale. L'idée est bien sûr que la région se rapproche de la notion d'objet, au sens courant du terme.

### 3.2.1 La segmentation de l'IRM cérébrale

La segmentation précise et robuste du tissu cérébral étudié par IRM est une question très importante dans beaucoup d'applications et en particulier dans l'étude de certains désordres du cerveau.

Une segmentation du tissu cérébral consiste à séparer les différents tissus tumoraux, tels que la tumeur, l'œdème et la nécrose, ainsi que les tissus normaux tels que la matière grise, la matière blanche et le liquide céphalorachidien. La segmentation est facilitée par l'utilisation de plusieurs types d'IRM. Les images pondérées en T1, T2 et en densité de protons sont notamment couramment employées en clinique. Ces trois types de séquences peuvent être complétés par d'autres types d'IRM, telles que les séquences utilisant un produit de contraste tel que le Gadolinium, FLAIR, et les images de perfusion, de diffusion et spectroscopique.

La segmentation d'images a pour objectif d'identifier, dans les images, les pixels ou voxels appartenant à une région particulière [132] [49]. Cette région peut correspondre à une structure anatomique, pathologique ou fonctionnelle. La segmentation du cerveau étudié par IRM peut comprendre :

- la segmentation de structures [166],
- la classification des tissus [89],
- la segmentation du volume d'objet (par exemple, tumeur) [62].

- La classification ou segmentation des tissus, peut, par exemple, concerner la segmentation des tumeurs [115] [120] comprenant des meningiomes et des gliomes [119].
- Elle inclut aussi la segmentation du cerveau en LCR, matière blanche (MB) et matière grise (MG) [62] [104] [158] et la quantification des lésions, par exemple liées à la sclérose en plaques [162] [176].

Les méthodes utilisées se basent sur les propriétés iconiques des structures à identifier, c'est à dire liées aux intensités des pixels ou voxels, ou sur les propriétés géométriques de ces structures. Ces propriétés correspondent à des *a priori* sur les structures à segmenter et peuvent être implicites ou explicites. Une segmentation automatique est nécessaire pour deux raisons :

- le temps du tracé manuel par un expert sur un tissu 3D spécifique est long,
- l'expertise humaine peut produire des erreurs difficilement contrôlables et reproductibles.

Par conséquent, de nombreuses études de segmentation automatique des tissus du cerveau dans les images d'IRM ont été effectuées et rapportées dans la littérature, telles que des techniques basées sur les réseaux neuronaux [149] [126], la logique floue, des méthodes statistiques [82], la fusion d'information [70]. Par la suite, nous présenterons les principales méthodes.

### 3.2.1.1 Spécificités de la segmentation de l'IRM cérébrale

La segmentation de l'IRM cérébrale présente des particularités par rapport à d'autres domaines d'applications de la segmentation [14]. Ces particularités sont liées aux spécificités des objets à segmenter (*Ref.* Annexe A) et au processus d'acquisition IRM caractérisées par :

- *Information d'intensité insuffisante* : l'intensité observée sur des images IRM n'apporte qu'une information incomplète sur les structures anatomiques sous-jacentes. Plusieurs structures peuvent avoir la même intensité, par exemple :
  - En T2 il n'y a pas de différence de contraste entre la MG et la MB.
  - En T1, certaines structures, comme le corps calleux, n'ont pas de bords précisément délimités.
  - L'intensité IRM lors de la transition entre la MG et la MB est complexe, et peut être variable.

Une solution qui permet de palier à l'insuffisance des informations présentées dans l'image, est l'utilisation simultanée de plusieurs modalités.

- *Erreurs /artefacts d'acquisition* : les principaux artefacts apparaissant en IRM ont été présentés au paragraphe (Ref .Annexe B ). Lors de la segmentation, l'artefact le plus souvent cité est l'artefact d'inhomogénéité de la radio fréquence qui se manifeste par une variation spatialement lente de l'intensité des tissus sur l'image. La plupart des méthodes de segmentation prennent en compte cet artefact, soit en l'éliminant par un prétraitement approprié, soit en le modélisant explicitement au cours de la segmentation.
- *Volume partiel* : Un problème inhérent à tout système numérique d'imagerie est l'effet de volume partiel (Voir la figure 4.7) . Si la surface entre plusieurs objets intersecte un voxel, la mesure radiométrique sur ce voxel résulte d'un mélange des contributions radiométriques des objets présents dans ce voxel. Cet effet se manifeste principalement à la frontière des principales matières (matière blanche/matière grise et liquide céphalo-rachidien/matière grise). Il se produit également lorsque des structures sont trop fines pour être visibles à la résolution de l'image (par exemple pour certains vaisseaux sanguins). Le terme « volume partiel » évoque une vision d'un pixel qui contiendrait un mélange de plusieurs tissus, le pixel à ce niveau est appelé «mixel».

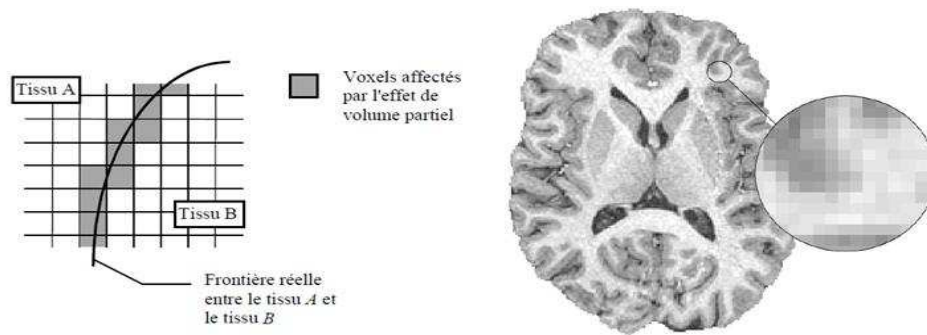


FIGURE 3.1 – L'effet de volume partiel (d'après [24]).

### 3.3 Les différentes approches de segmentation

La segmentation est une étape essentielle en traitement d'images dans la mesure où elle conditionne l'interprétation qui va être faite sur ces images. De nombreux algorithmes ont ainsi été proposés durant les dernières décennies [81] [78]. Ils sont généralement fondés sur des principes très différents comme le processus de classification, l'utilisation de modèles de contours déformables ou encore de modèles de connaissance...etc. Nous proposons ici de

parcourir, de manière non exhaustive, différentes approches présentes dans la littérature. Nous abordons différentes méthodes de segmentation des images cérébrales IRM d'abord d'un point de vue « Traitement d'images », puis du point de vue « Reconnaissance des formes ».

### 3.3.1 État de l'art du point de vue « Traitement d'images »

Dans ce premier point de vue, il s'agit d'identifier la nature particulière d'un signal bidimensionnel, si on traite individuellement des coupes, ou tridimensionnel, si on considère le volume constitué de leur empilement. La rupture du signal sera alors représentative d'un contour, et l'homogénéité des niveaux radiométriques sera caractéristique d'une région.

#### 3.3.1.1 Approches Région

Les approches de segmentation régions visent à créer une partition de l'image en un ensemble de régions homogènes au sens d'un ou plusieurs critères (Voir la figure 3.2). On recherche donc à rassembler les pixels par critères de similarité. On peut distinguer les méthodes par seuillages, croissance de régions, ainsi que d'autres qui ne seront pas détaillés dans ce manuscrit.

**3.3.1.1.1 Le seuillage** Les méthodes de segmentation par seuillages visent à discriminer les individus (pixels ou voxels) en fonction de leurs niveaux radiométriques (niveaux de gris). Elles consistent à mettre à zéro tous les pixels ayant un niveau de gris inférieur à une certaine valeur (appelée seuil, en anglais *threshold*) et à la valeur maximale les pixels ayant une valeur supérieure.

Ainsi le résultat du seuillage est une image binaire contenant des pixels noirs et blancs, c'est la raison pour laquelle le terme de binarisation est parfois employé. Cependant ces méthodes supposent donc, de manière implicite, que l'intensité des structures est suffisamment discriminante pour permettre leur bonne séparation.

Le seuillage peut être de différentes natures : globale si un seuil est utilisé pour toute l'image, locale si un seuil est utilisé pour une portion de l'image et enfin si le seuil utilisé s'ajuste selon les images/parties de l'image alors on parle de seuillage adaptatif.

De nombreuses techniques de seuillage ont été développées [9] [156] : certaines sont basées sur une analyse d'histogramme, d'autres sur des propriétés locales, comme la moyenne ou

l'écart type. Si le seuillage permet de binariser un résultat final, ce qui est toujours nécessaire, il n'est bien souvent pas suffisant pour obtenir une segmentation acceptable, en effet bien que cette technique est souvent la plus simple, rapides et simples à mettre en œuvre, mais rarement la plus efficace [56]. Elle a tendance à échouer lorsque les régions de recherche sont insuffisamment contrastées. De plus, l'absence de prise en compte de contraintes de voisinage rend ces méthodes très sensibles au bruit, les régions obtenues pouvant alors être sémantiquement peu cohérentes.

Enfin, même si théoriquement les méthodes à base de seuillages peuvent s'appliquer à des volumes multi-échos, elles n'exploitent, pour la plupart, qu'une seule pondération ou exploitent indépendamment les différentes pondérations, sans réelle exploitation de leur complémentarité [72] [161]. Ceci est alors particulièrement pénalisant dans le cadre de la recherche de tumeurs cérébrales, cette technique ne semble donc pas adaptée pour, à elles seules, segmenter les principales structures anatomiques et les éventuelles pathologies. Leur utilisation nécessite, en particulier, la prise en compte d'informations locales, liées au voisinage d'un pixel ou d'un voxel.

**3.3.1.1.2 La croissance de région** La croissance de région consiste à partir d'un pixel de l'image, appelé graine, qui appartient à la structure à segmenter, et recherche à partir de là des pixels d'intensité similaires, en suivant un critère d'homogénéité au sein de la structure à segmenter .

Ces techniques ont trouvé leur utilité pour des images cardiaques [143], ou angiographiques [172]. Dans [2], Thiran et al. proposent une méthode de croissance de régions pour la segmentation du cerveau sur des images IRM tridimensionnelles. La segmentation obtenue permet alors des rendus des surfaces cérébrales. Dans [65], le ventricule latéral et le troisième ventricule sont segmentés par croissance de régions et morphologie mathématique. La croissance de régions est guidée par des connaissances anatomiques *a priori* sur le système ventriculaire.

Tout comme les méthodes à base de seuillages, les segmentations par croissance de régions sont simples à mettre en œuvre, mais parfois coûteuses en temps de calcul. Leurs avantages résident dans la prise en compte des informations contextuelles par le mécanisme de croissance et dans la possibilité d'introduire des informations de plus haut niveau.

Néanmoins, leur inconvénient est lié à la disposition des germes initiaux. En effet, celle-ci nécessite l'intervention d'un opérateur, et que leur utilisation est restreinte au contexte monomodal [164].



FIGURE 3.2 – Image IRM segmentée par une approche région.

**3.3.1.1.3 Morphologie mathématique** La Morphologie Mathématique comprend un ensemble de techniques de traitement de l'image basées sur les formes. Son vocabulaire est celui de la théorie des ensembles. De nature fondamentalement géométrique, elle consiste à comparer les objets à analyser à un autre objet de forme connue, appelé élément structurant.

Les outils de base de la morphologie mathématique [108] (érosion, dilatation, fermeture, ouverture. . .) se retrouvent fréquemment dans des méthodes de segmentation de l'enveloppe cérébrale [116]. Ainsi dans [116], Brumer Utilise des érosions, dilatations et étiquetages en composantes connexes à la suite de seuillages afin de déconnecter les structures non cerveau de la structure cerveau et d'affiner le masque initial. Dans [52], la morphologie mathématique est couplée à un processus de seuillage adaptatif fondé sur l'histogramme de l'image.

Cette méthode est reprise dans [160] en ajoutant quelques opérations morphologiques de façon à garantir une meilleure robustesse aux variations anatomiques et des acquisitions. Toujours dans [160], Géraud propose, à la suite de la segmentation du cerveau, de coupler seuillages et morphologie mathématique pour segmenter successivement le liquide céphalo-rachidien (ventricules et liquide des sillons) et la matière grise (cortex et noyaux de gris centraux) sur des images pondérées en T1. Les résultats obtenus sont jugés très corrects par l'auteur. Il modère toutefois ses propos du fait de la dépendance successive des segmentations et de la nécessaire connaissance des niveaux radiométriques pour la détermination des seuils.

La morphologie mathématique est rarement utilisée seule pour segmenter les structures cérébrales mais fait généralement partie d'une chaîne de traitements plus complexe. À notre connaissance, il n'existe pas de méthode directe de segmentation par morphologie mathématique. Cette technique, qui s'appuie sur des informations *a priori* concernant

la forme des objets à segmenter, par nature, caractérisées par une grande variabilité de forme.

### 3.3.1.2 Approches contours

Contrairement aux approches régions, qui cherchent à former des zones homogènes, les approches contours travaillent sur les discontinuités des images afin de déterminer les contours des régions. Dans la suite, nous nous intéressons plus particulièrement aux méthodes dérivatives, aux méthodes fondées sur les modèles déformables, et aux espace-échelle. Les approches contour cherchent à extraire les contours présents dans l'image, en se basant sur l'étude des changements abrupts de la fonction de luminance.

**3.3.1.2.1 Méthodes dérivatives** Les méthodes dérivatives permettent de détecter de manière simple les ruptures dans une image ou un volume. Les contours (ou surfaces en 3D) sont assimilés à des points de forts gradients ou de dérivées secondes nulles. On retrouvera alors des opérateurs tels que les filtres de Roberts [1], de Sobel [73] ou de Prewitt [87]. Parmi la diversité des opérateurs, celui de Marr-Hildreth [32] qui fournit des contours fermés. Dans [106], Boman et al. Utilisent une approximation de cet opérateur dans un processus de segmentation. Les régions définies à partir des contours sont ensuite manuellement étiquetées. Les structures cérébrales recherchées sont alors la peau, le cerveau et le système ventriculaire. Malgré une bonne détection de la peau et du système ventriculaire, certains contours obtenus se trouvent décalés par rapport à la réalité anatomique. Dans ce cas particulier, une fermeture morphologique est recommandée.

Les méthodes dérivatives, employées pour la segmentation des images IRM cérébrales, sont relativement peu nombreuses dans la littérature. Bien que rapides et ne nécessitant pas d'apport d'information *a priori*, elles ne permettent pas, en général, d'obtenir des contours fermés [92].

Elles sont en effet très sensibles aux bruits et à la dérive du champ radiofréquence. Leur efficacité repose sur l'existence de contraste entre les structures recherchées. Si ceux-ci existent effectivement au niveau de la peau ou du système ventriculaire, ils sont beaucoup moins importants au niveau de la transition matière blanche - matière grise ou encore au niveau du liquide céphalo-rachidien. Enfin, ces méthodes fournissent fréquemment de la sous segmentation ou de la sur-segmentation [150].

**3.3.1.2.2 Modèles déformables** Les modèles déformables introduits par [91], encore appelés snakes, contours actifs ou ballons, ont connu de nombreuses applications en segmentation, mais aussi en animation, suivi de contour ou modélisation géométrique (Voir la figure 3.3).

Ils ont l'avantage, par rapport aux méthodes dérivatives, de fournir des contours ou surfaces fermés [100]. Le principe de base est de faire évoluer un objet d'une position initiale vers une position d'équilibre par la minimisation itérative d'une fonctionnelle. Ils sont en effet généralement plus robustes au bruit et aux éventuelles discontinuités dans les contours de l'image. Ils permettent en outre une interaction relativement aisée avec l'utilisateur ainsi que l'introduction de connaissances *a priori* concernant la forme de l'objet recherché.

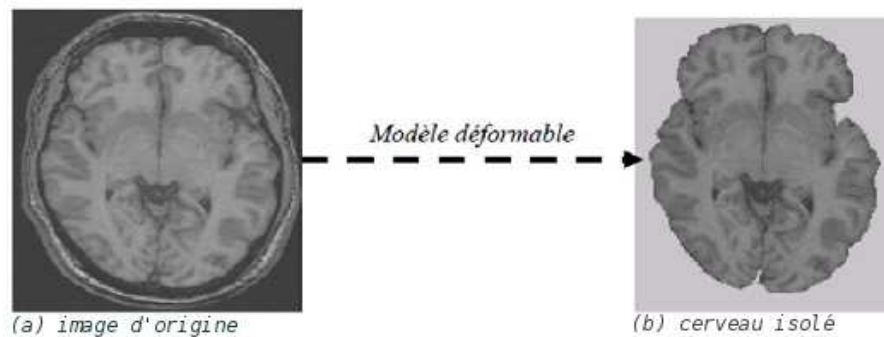


FIGURE 3.3 – Segmentation par modèle déformable.

Deux grandes familles de modèles déformables existent : les modèles paramétriques [91] [59] et les modèles géométriques [109] [26].

Les premiers sont les plus anciens et nécessitent une représentation paramétrique ou discrète. Les seconds, fondés sur la théorie d'évolution des courbes et la méthode des ensembles de niveau (level sets) introduite par Osher et Sethian [153], utilisent une représentation implicite du modèle et permettent des changements de topologie.

Cependant, si de récentes études montrent que les modèles déformables et les ensembles de niveau sont parfaitement utilisables dans le cadre de la segmentation de tissus cérébraux [18] [124] [148] [21], leur mise en œuvre reste difficile et les temps de calcul sont souvent très longs, et restent relativement sensibles à leur initialisation. Face à ce problème, les auteurs tendent à inclure des procédures d'initialisation automatique de façon à réduire l'influence des opérateurs et augmenter la reproductibilité des résultats. De plus, nous notons leur utilisation restreinte au contexte monomodal [100].

**3.3.1.2.3 Espace-échelle** La notion d'espace-échelle (en anglais scale-space) est générale et s'applique dans des dimensions arbitraires. C'est une théorie formelle pour manipuler des structures d'image à différentes échelles de telle manière que les structures puissent être successivement supprimées et un paramètre  $t$  d'échelle puisse être associé à chaque niveau dans la représentation du scale-space [101]. Lachmann [51] a exploré différentes techniques scale-space appliquées à l'imagerie médicale 3D. Une revue des différentes techniques de segmentation en scale-space a été faite par Henkel en 1995 dans [140]. Plus récemment, ces idées de segmentation multi-échelle liant les structures d'images au travers des différentes échelles a été repris par Florack and Kuijper dans [54]. Rahman et coll. proposent dans [146] d'exploiter le gradient à différentes échelles pour segmenter les images.

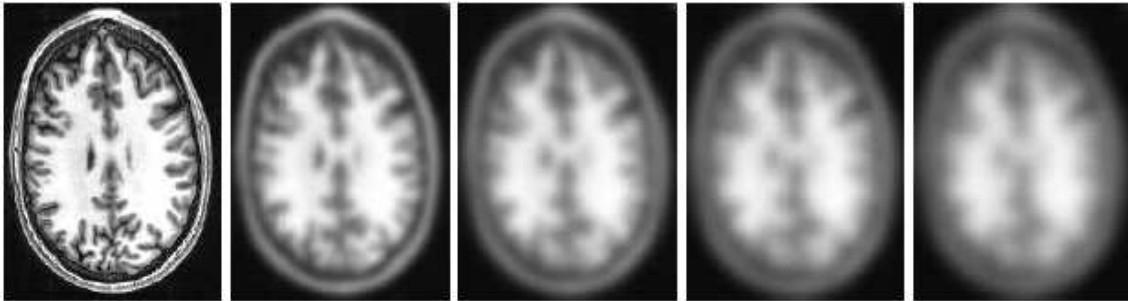


FIGURE 3.4 – Exemple de scale space gaussien. [101]

### 3.3.2 État de l'art du point de vue «Reconnaissance des formes»

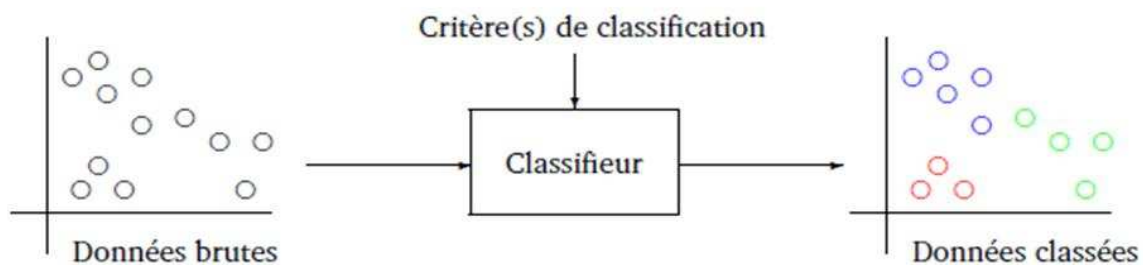


FIGURE 3.5 – Principe de la classification.

Un autre point de vue est celui de la reconnaissance des formes qui traite du problème de la prise de décision dans des problèmes de classements [17]. Le principe du regroupe-

ment se rapporte à la classification d'objets dans des groupes selon certaines propriétés de ces objets (voir figure 3.5).

La figure 3.6 présente les différentes variantes que l'on peut trouver parmi les méthodes de classification [83]. Voici une rapide description des ces méthodes :

- classification exclusive/non exclusive. Une classification exclusive est un partitionnement des objets : un objet n'appartient qu'à une classe et une seule. Au contraire, une classification non exclusive autorise qu'un objet appartienne à plusieurs classes simultanément. Les classes peuvent alors se recouvrir ;
- classification supervisée/non supervisée.
- classification hiérarchique/partitionnement. Une méthode de classification hiérarchique construit une séquence de partitions imbriquées, alors qu'un partitionnement ne construit qu'une seule partition des données.

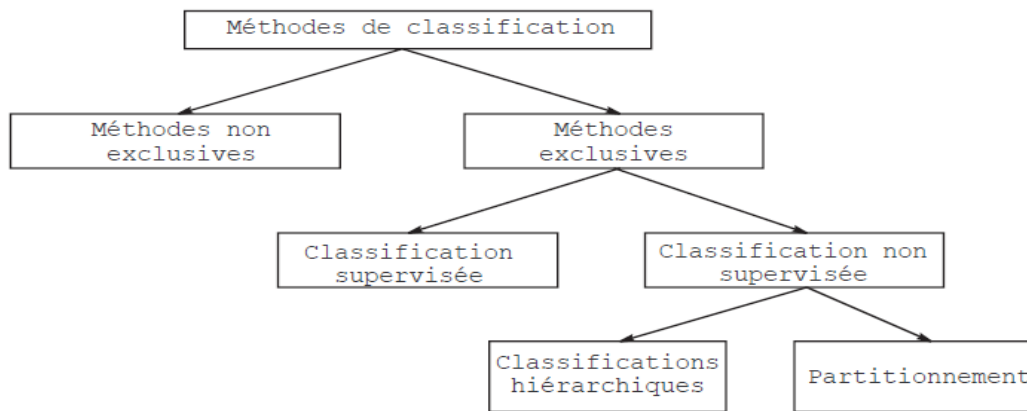


FIGURE 3.6 – Les méthodes de classification (d'après [83]).

Du point de vue de la segmentation des images, on cherche à construire des vecteurs de caractéristiques dans certaines zones typiques de l'image qui seront ensuite utilisés pour la classification des points de l'image [168]. Un problème de reconnaissance des formes nécessite alors :

- de définir les paramètres constituant le vecteur forme  $x$ , représentatif de l'état du système ; la dimension de  $x$  est celle de l'espace des caractéristiques ;
- de définir l'ensemble des états ou classes connus pour lesquels on dispose d'informations : modèle probabiliste de comportement, ensemble de vecteurs d'échantillons, etc.
- de construire une règle de décision qui, à un vecteur forme  $x$ , associe soit la décision d'affecter une classe, soit la décision de rejeter toutes les classes connues, soit une non décision.

La classification des tissus cérébraux (principalement le liquide cérébro-spinal, la matière blanche et la matière grise) est un problème abondamment abordé dans la littérature [85], il existe une grande variété d'algorithmes de regroupement de pixels, qui peuvent être classés selon trois modes d'apprentissage : supervisé, non supervisé et semi supervisé. Les méthodes supervisée et non supervisée seront détailler dans la suite, chacune selon un fondement théorique, parmi lesquels on retrouve la théorie des probabilités (*cf.4.2.1.1*), la théorie des sous ensembles flous (*cf.4.2.1.2*) la théorie de possibilité (*cf.4.2.1.3*) et la théorie des fonctions de croyance (*cf.4.2.1.5*).

### 3.3.2.1 Les méthodes de classification

**3.3.2.1.1 Classification supervisée** Les méthodes supervisées nécessitent un ensemble d'images ou de régions d'intérêt dans l'image déjà étiquetées. Les données à analyser  $X = \{x_j \in R^p, 1 \leq j \leq N\}$ ,  $X_j$  est un vecteur forme représentant le pixel  $j$ , sont donc décomposées en un ensemble d'apprentissage et un ensemble de test. 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  $C_1 \dots C_c$  par un ou plusieurs experts (radiologues). 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  $E_i$ . Tenant compte de la différence entre la structure anatomique d'un cerveau, d'un patient à un autre, la tâche d'étiquetage reste une tâche fastidieuse et longue pour les spécialistes.

**3.3.2.1.2 Classification semi supervisée** La semi-supervision intervient lorsqu'on dispose à la fois d'un ensemble de données étiquetées et non étiquetées. Généralement, les données non supervisées sont disponibles en grand nombre car peu coûteuses à produire. Au contraire, les données supervisées qui nécessitent l'expertise humaine sont plus rares mais également plus riches d'information. On distingue deux approches duales de la semi-supervision suivant le point de départ où l'on se place [154] :

- Utilisation de la donnée supervisée pour aider un algorithme de classification automatique ;
- Utilisation d'un grand volume de données non étiquetés pour améliorer un apprentissage supervisé.

**3.3.2.1.3 Classification non supervisée (Automatique)** L'intérêt des méthodes non supervisées est qu'elles ne nécessitent aucune base d'apprentissage et aucune tâche préalable d'étiquetage manuel. La seule intervention de l'expert se situe à la fin du processus pour identifier les classes calculées avec les classes biologiques [164]. Dans la suite un ensemble non exhaustif de méthodes de classification supervisées et non supervisées seront présentées.

### 3.3.2.2 Méthodes supervisées

**3.3.2.2.1 Approches probabilistes** Pour la segmentation des images IRM, les approches probabilistes se fondent sur une modélisation statistique de la radiométrie des images. Chaque région de l'image, *i.e.* chaque structure anatomique, est associée à une classe dont on recherche les caractéristiques statistiques. Dans ce cadre, chaque observation est considérée comme le résultat d'un processus stochastique et est donc associée à une variable aléatoire. La problématique est alors celle de l'estimation des densités de probabilités à partir des observations et celle de la prise de décision connaissant ces densités.

#### a. Approches paramétriques

Ce type de méthodes s'appuie sur la définition et l'estimation de densités de probabilités des vecteurs de caractéristiques associés aux données. Un label (ou classe, ou hypothèse) est alors attribué à chaque pixel/voxel des images, à partir de leur niveau radiométrique.

La difficulté principale est alors l'estimation des probabilités *a posteriori*. De façon pratique, et grâce à la théorie bayésienne, le problème sera résolu par l'estimation de la probabilité *a priori* de chaque classe et des probabilités conditionnelles aux observations.

- Dans un cadre supervisé, les densités sont obtenues grâce à l'intervention d'un opérateur (segmentation par des experts, sélection manuelle de zones d'apprentissage. . .). L'inconvénient majeur est alors la création d'un biais d'apprentissage lié à l'opérateur et le problème de la reproductibilité des résultats. La base d'apprentissage doit être suffisamment large pour estimer correctement les densités tout en étant suffisamment généraliste pour s'adapter à de nouvelles données.
- Dans un cadre non supervisé, un apprentissage automatique est nécessaire.

Cette approche convertit le problème difficile de recherche de fonction en un problème plus simple d'estimation de paramètres en imposant une forme particulière à la fonction (souvent gaussienne) [164] Cette restriction peut cependant introduire des erreurs de classification, puisque les données ne sont que rarement conformes à

des distributions paramétriques. Selon [85], les hypothèses de ce type sont mal adaptées à la segmentation d'images médicales, IRM en particulier, où l'intensité d'un pixel ne peut être dérivée d'une combinaison linéaire de probabilités distinctes attachées aux classes. Ainsi, malgré le consensus général qui affirme que chaque classe de tissu peut être modélisée par une loi gaussienne en IRM [94], nous remarquerons que les différents défauts de l'image (hétérogénéité de champ en particulier) rendent cette approximation hasardeuse lorsqu'il s'agit d'effectuer des mesures précises de volume. Afin de pallier la faible robustesse au bruit et pour régulariser les résultats, des informations contextuelles peuvent être introduites via la théorie des champs de Markov. De plus, des paramètres tels que les volumes partiels peuvent être pris en compte dans la modélisation. Cependant, un inconvénient majeur de cette théorie réside dans le fait qu'elle nécessite la connaissance parfaite des probabilités et en particulier des probabilités *a priori*. Or, dans la pratique, ces probabilités ne sont qu'estimées et portent des incertitudes et imprécisions.

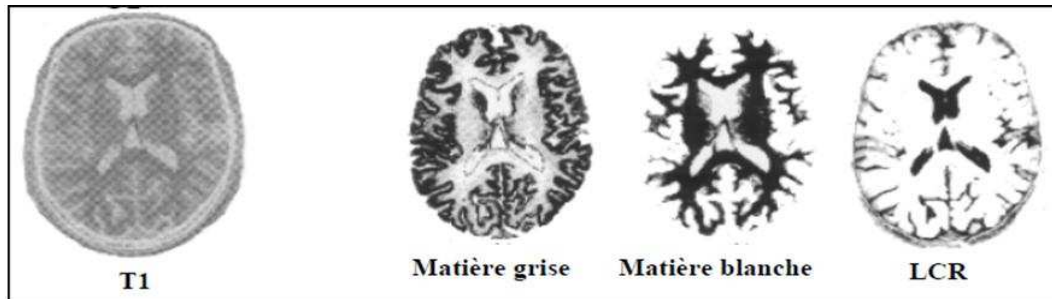


FIGURE 3.7 – Segmentation par méthode bayésienne.

### b. Approches non paramétriques

La particularité des approches probabilistes non paramétriques est leur capacité à discriminer les vecteurs formes en faisant peu d'hypothèses sur les densités de probabilités. Elles sont donc particulièrement bien adaptées lorsque les densités recherchées ne correspondent pas à des densités classiques ou paramétrables. *Les fenêtres de Parzen* et les *K-Plus Proches Voisins (K-PPV)* sont deux algorithmes supervisés classiques non paramétriques, qui se fondent sur l'estimation de la probabilité, notée  $P$ , que la réalisation  $x$  d'un vecteur aléatoire  $X$  tombe dans une région particulière. Il a été montré que l'estimateur  $\hat{P}$  de  $P$  vérifie alors  $\hat{P}(x) = \frac{r}{N}$  où  $r$  est le nombre de vecteurs formes présents dans le volume  $V$   $N$  le nombre total de vecteurs formes de l'espace des caractéristiques. Grâce à la densité  $\hat{P}$ , il est possible d'estimer les lois *a posteriori* et donc de définir la classe d'appartenance associée à chaque vecteur forme :

– Ainsi, si le volume  $V$  est choisi autour de  $x$ , alors l'estimation de  $\hat{P}(x)$  nécessite

de compter le nombre de vecteurs formes qui appartiennent au volume. C'est la technique des fenêtres de Parzen.

- Si le nombre  $r$  est fixé, il faut trouver le volume  $V$  autour de  $x$  qui contient  $r$  vecteurs formes. C'est la méthode des K-PPV.

D'après l'état de l'art présenté par Zijdenbos et Dawant [176], peu d'auteurs utilisent les fenêtres de Parzen pour segmenter les images IRM par contre les méthodes de classification fondées sur les K-PPV sont, quant à elles, plus présentes dans la littérature.

Un des principaux avantages des méthodes probabilistes non paramétriques telles que les fenêtres de Parzen ou les K-PPV est leur grande capacité d'adaptation aux données puisqu'elles ne nécessitent pas d'apport d'informations *a priori* sur les formes des distributions. Cependant, elles souffrent de la difficulté à définir un apprentissage correct des données ; la base d'apprentissage doit être suffisamment grande pour que la modélisation soit efficace, mais devient alors très fastidieuse. Elle est de plus très dépendante de l'opérateur et ne peut s'adapter aux données si ces dernières s'éloignent des données d'apprentissage ou en cas de bruit.

**Segmentation Markovienne** La théorie des champs de Markov permet d'associer le processus de segmentation à la minimisation d'une fonction d'énergie définie sur l'image à segmenter. Grâce à l'introduction de notions de voisinages entre pixels, cette approche permet d'envisager la recherche de modèles locaux de distributions selon un principe équivalent localement à celui de la classification bayésienne. Puisque les champs de Markov prennent en compte les contraintes spatiales dans la segmentation, les modalités anatomiques se prêtent tout particulièrement à ce type de méthode. Citons par exemple Held et al [90] qui segmentent des images IRM pondérées en T2 et densité de protons à l'aide d'un champ de Markov modifié permettant d'estimer les hétérogénéités dans l'image simultanément à la segmentation, ou Jaggi et al. [23] qui s'affranchissent des effets de volume partiel en IRM en estimant les coefficients d'un modèle de mélange à 5 classes (LCS, MG, MB, MB/MG et LCS/MG) à l'aide d'un champ de Markov.

### 3.3.2.3 Méthodes non supervisées (Automatiques)

Appelées aussi méthodes de regroupement (*clustering*), la classification est dite non supervisée lorsqu'aucune connaissance *a priori* n'est disponible. Cette dernière se divise en deux familles selon deux approches :

- Approches hiérarchique (ascendante, descendante)
- Approches partitionnelles.

Dans la première famille le principe est de regrouper les objets deux par deux (cas ascendant) ou à diviser l'ensemble des objets (cas descendant) récursivement jusqu'à obtenir le nombre de groupes escompté, tandis que dans la seconde famille le principe consiste construire des clusters itérativement jusqu'à stabilisation du modèle.

**3.3.2.3.1 La famille des C-moyennes** Ce sont des algorithmes de classification non supervisée basés sur différentes approches : nette comme *HCM*, floue comme le «*Fuzzy C-means*» (FCM), possibilistes comme le «*Possibilistic Cmeans*» (*PCM*) [141] [105] [142] ou évidentielle comme *ECM* récemment proposé par Masson et Denoeux dans [113]. Ces algorithmes peuvent être regroupés en deux classes d'algorithmes suivant les contraintes imposées sur les degrés d'appartenance : *les algorithmes non flous et flous (fuzzy)*. On propose d'explicitier d'abord le processus de génération de ces fonctions d'appartenances, pour détailler les deux classes d'algorithmes dans un second temps.

**Génération de fonctions d'appartenance** Notons  $X = (x_j, j = 1..N)$  l'ensemble des vecteurs forme de  $R_p$ , avec  $x_j = [x_{j1}, x_{j2}, \dots, x_{jp}]^T$ , et  $V = (v_1, \dots, v_c)$  un ensemble de vecteurs prototypes inconnus, où  $v_i$  caractérise la classe  $i$ . La génération de fonctions d'appartenance consiste à construire automatiquement des fonctions permettant d'affecter chaque  $x_j$  à une ou plusieurs classes selon certains critères. Les contraintes imposées sur les degrés d'appartenance pour chaque algorithme sont décrites comme suit :

- **HCM :**

$$\forall i : \mu_{ik} = 0 \mid 1, \sum_{k=1}^c \mu_{ik} = 1 \quad (3.1)$$

- **FCM**

$$\forall i : \mu_{ik} \in [0, 1], \sum_{k=1}^c \mu_{ik} = 1 \quad (3.2)$$

- **PCM**

$$\forall i : \mu_{ik} \in [0, 1] \quad (3.3)$$

On remarque ainsi :

$$HCM \subset FCM \subset PCM$$

Le principe de cette famille de méthodes est de minimiser itérativement une fonctionnelle en alternant une mise à jour de  $U$  et de  $V$  pour trouver un regroupement des vecteurs  $x_j$

en  $C$  classes. Ces algorithmes possèdent un paramètre contrôlant la quantité de flou noté  $m$ , on remarque que pour  $m$  grand pour les algorithmes  $FCM$  et  $PCM$ , on obtient :

$$\lim_{m \rightarrow \infty} u_{ik} = \frac{1}{C}, \forall i, k$$

En pratique, les degrés d'appartenance sont proches pour  $m$  supérieur à 20. A l'inverse, si  $m$  tend vers 1, la matrice de partition floue devient une partition stricte. Le choix de  $m$  reste ainsi un problème ouvert. Les utilisateurs de cet algorithme choisissent généralement  $m$  entre 1 et 2 pour des raisons d'efficacité. De plus,  $m = 2$  permet quelques fois de simplifier la mise à jour des prototypes en autorisant une écriture directe de ceux-ci (voir Fuzzy C-Shells).

Suivant que les vecteurs forme sont assignés à une unique classe ou à plusieurs, les algorithmes de classification sont non flous ou flous (fuzzy).

### A. Algorithmes de classification non flous

- **L'Algorithme HCM**

Dans la méthode des c-moyennes (HCM), un élément de  $X$  est attribué à une classe et une seule parmi les  $C$  proposées, respectant ainsi les contraintes 3.1 , Dans ce cas, la fonctionnelle à minimiser est :

$$J(V, U, X) = \sum_{i=1}^c \sum_{j=1}^n \mu_{ij} d^2(x_j, v_i) \quad (3.4)$$

Où  $d^2(x_j, v_i)$  représente la distance euclidienne entre l'élément de la classe et le centre de la classe.

La classification des éléments de  $X$  s'effectue de manière itérative en alternant l'étape de classification et l'étape de mise à jour des centres, jusqu'à stabilisation de la segmentation ou de la fonction objective.

Les solutions au problème s'écrivent :

– L'étape de classification :

$$\mu_A(x) = \begin{cases} 1 & \text{si } d^2(x_j, v_i) \leq d^2(x_j, v_k), k \neq i \\ 0 & \text{sinon} \end{cases}$$

– L'étape de mise à jour des centres :

$$v_i = \frac{\sum_{j=1}^n \mu_{ij} x_j}{\sum_{i=1}^n \mu_{ij}}$$

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). Les méthodes présentées dans les paragraphes suivants permettent d'obtenir une segmentation floue qui prend en compte ces aspects imprécis et incertains.

## B. Algorithmes de classification flous

### • L'algorithme FCM

L'objectif de cet algorithme est l'établissement d'une partition floue de l'ensemble des vecteurs d'attributs en C classes.

Le concept de la partition floue est défini de la manière suivante :

$$\begin{cases} \forall i, k (1 \leq i \leq c), 1 \leq k \leq n) \mu_{ik} \in [0, 1] \\ \sum_{k=1}^n \mu_{ik} < n \\ \sum_{i=1}^c \mu_{ik} = 1 \end{cases}$$

L'algorithme des c-moyennes floues (FCM) est un processus itératif qui se repose sur la recherche des valeurs prises par les centroïdes  $v_i$  des classes et par les degrés d'appartenance  $\mu_{ji}$  des voxels à ces classes, minimisent la fonction objective suivante :

$$J(V, U, X) = \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m d^2(x_j, v_i) \quad (3.5)$$

Dans la fonction 3.5,  $m > 1$  est un coefficient réglant la dureté de la partition résultante.

Les formules de mise à jour de  $U$  et  $V$  sont obtenues par introduction d'un multiplicateur de Lagrange associé à la contrainte de normalisation et en annulant les dérivées partielles du Lagrangien par rapport aux paramètres du modèle.

L'algorithme FCM consiste donc à itérer les étapes suivantes jusqu'à stabilisation de l'algorithme :

1. initialisation des valeurs des centroïdes,  $v_k$
2. calcul des fonctions d'appartenance :

$$\mu_{ik} = \left[ \sum_{j=1}^c \left( \frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{m-1}} \right]^{-1} \quad (3.6)$$

3. calcul des centroïdes :

$$v_i = \frac{\sum_{k=1}^n (\mu_{ik})^m X_k}{\sum_{k=1}^n (\mu_{ik})^m} \quad (3.7)$$

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.

#### – FCM et caractérisation des tissus cérébraux

L'algorithme FCM a été beaucoup utilisé pour la segmentation des images de cerveau, quels que soient la modalité et le type d'acquisition (mono ou multimodal). De nombreux travaux ont notamment été effectués en imagerie par résonance magnétique. *Baumgartner et al* [10] ont ainsi utilisé le FCM pour segmenter les régions activées (simulation et aires motrices) du cerveau. Les pixels 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. En IRM traditionnelle, *Clark et al.* [107] ont utilisé le FCM comme étape d'initialisation à un système expert pour segmenter des volumes de tumeur ou d'œdème sur des images pondérées en T1, T2 et en densité de protons. *Xue et coll.*, utilisent les FCM pour combiner le filtre moyen au filtre médian local afin de réaliser la segmentation locale de volumes IRM de cerveaux [173]. *Pham et coll.* ainsi que *Jiang et coll.*, utilisent les inhomogénéités d'intensité des volumes IRM dans la fonctionnelle  $L$  et obtiennent ainsi un FCM adaptatif qui permet une meilleur segmentation [48] [45] [44] [33].

#### • L'algorithme PCM

La définition d'une partition floue impose que la somme des degrés d'appartenance d'un élément à l'ensemble des classes soit égale à 1. Cette contrainte forte engendre un certain nombre de problèmes si l'on veut extraire les propriétés de chaque classe. Pour remédier à ces problèmes, *Krishnapuram et Keller* dans [141] ont proposé une méthode de classification possibiliste dont la partition floue est définie de la manière suivante :

$$\mu_A(x) = \begin{cases} \forall i, k & \mu_{ik} \in [0, 1] \\ \forall i & 0 < \sum_{k=1}^n \mu_{ik} < n \\ \forall k, \exists i & \mu_{ik} > 0 \end{cases}$$

La fonction objective à minimiser est redéfinie comme suit :

$$J(V, U, X) = \sum_{k=1}^c \sum_{i=1}^n \mu_{ij}^m d_{ik}^2 + \sum_{k=1}^c w_k \sum_{i=1}^n (1 - \mu_{ik})^m \quad (3.8)$$

La minimisation de la nouvelle fonction objective engendre alors une modification de la condition nécessaire utilisée pour la mise à jour des degrés d'appartenance qui sont maintenant calculés de la manière suivante :

$$\mu_{ik} = \frac{1}{1 + (d_{ik}^2/w_i)^{\frac{1}{m-1}}} \quad (3.9)$$

Où  $w_i$  est un réel positif déterminant le carré de la distance séparant le centre de la classe  $i$  à l'ensemble des points dont le degré d'appartenance à cette classe est égal à 0.5. Il est défini comme suit :

$$w_i = K \frac{\sum_{k=1}^c (\mu_{ij})^m d_{ik}^2}{\sum_{k=1}^c \mu_{ij}^m} \quad (3.10)$$

$k$  est le plus souvent pris égal à 1, les centres des classes sont mis à jour comme dans FCM.

Barni [105] signale le problème de coïncidence de clusters, sur différentes applications. [142] suggèrent qu'un bon choix de  $m$  pour le PCM est  $m \cong 1.5$  (plus une bonne initialisation et un bon choix de  $w_i$ ), et non pas 2 comme dans le FCM.

### – PCM et caractérisation des tissus cérébraux

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 [114] 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 [167] ont également appliqué PCM en IRM sur des vecteurs forme  $x_j$ .

### • L'algorithme ECM

Une nouvelle méthode de classification automatique, intitulée ECM (*Evidential c-means*), est proposée. Elle se fonde sur le concept de partition crédale introduit dans le cadre de la théorie des fonctions de croyance (*cf.* 4.2.1.5). Ce nouveau concept généralise ceux de partition nette, floue, probabiliste ou possibiliste. La partition crédale est obtenue grâce à un algorithme semblable à celui des c-moyennes floues. Il est fondé sur un schéma de minimisation alterné, avec dans

un premier temps, la détermination des centres des classes, et, dans un second temps, l'allocation de masses aux différents sous-ensembles de classes. Plusieurs outils d'interprétation sont proposés. Un indice de validité permet de guider le choix du nombre de classes. Les tests sur des données synthétiques montrent l'intérêt de la méthode pour l'analyse exploratoire de données [113].

Cependant d'après nos recherches il n'existe pas encore d'application de l'algorithme ECM dans le domaine de la segmentation d'imagerie et de l'IRM en particulier.

**3.3.2.3.2 Les métaheuristiques pour l'extraction des connaissances** Les métaheuristiques<sup>1</sup> sont un ensemble de méthodes qui permettent de concevoir des algorithmes. Elles permettent de résoudre des problèmes d'optimisation auxquels les ingénieurs et les décideurs sont régulièrement confrontés. La plupart des algorithmes d'optimisation sont en effet proposés dans le domaine combinatoire. Dès lors, l'adaptation de ces métaheuristiques combinatoires aux problèmes continus est profitable à un grand nombre de problèmes réels.

Les métaheuristiques sont souvent inspirées par des systèmes naturels, qu'ils soient pris en physique (cas du recuit simulé), en biologie de l'évolution (cas des algorithmes génétiques) ou encore en éthologie (cas des algorithmes de colonies de fourmis ou de l'optimisation par essaims particuliers).

Dans cette partie on s'intéresse particulièrement aux algorithmes de colonies de fourmis, qui s'inspirent respectivement de la théorie de l'évolution et du comportement de fourmis à la recherche de nourriture.

**L'optimisation par colonies de fourmis (ACO pour Ant Colony Optimisation)** Les algorithmes de colonies de fourmis sont des algorithmes inspirés du comportement des fourmis et qui constituent une famille de métaheuristiques d'optimisation.

Initialement proposé par Marco Dorigo et al. dans les années 1990, pour la recherche de chemins optimaux dans un graphe, le premier algorithme s'inspire du comportement des fourmis recherchant un chemin entre leur colonie et une source de nourriture. L'idée originale s'est depuis diversifiée pour résoudre une classe plus large de problèmes et plusieurs algorithmes ont vu le jour, s'inspirant de divers aspects du comportement des fourmis.

---

1. Le mot métaheuristique est dérivé de la composition de deux mots grecs : heuristique qui vient du verbe heuriskein (euriskein) et qui signifie 'trouver', et meta qui est un suffixe signifiant 'au-delà', 'dans un niveau supérieur'.

En anglais, le terme consacré à la principale classe d'algorithme est « Ant Colony Optimization » (abrégé ACO). Les spécialistes réservent ce terme à un type particulier d'algorithme. Il existe cependant plusieurs familles de méthodes s'inspirant du comportement des fourmis. En français, ces différentes approches sont regroupées sous les termes : algorithmes de colonies de fourmis, optimisation par colonies de fourmis, fourmis artificielles, ou diverses combinaisons de ces variantes.<sup>2</sup>

Parmi ces variantes, on cite :

- *Le système de fourmi "Ant system" (AS)* : Le premier algorithme de colonies de fourmis proposé est appelé le Ant system (système fourmi). Il vise notamment à résoudre le problème du voyageur de commerce, où le but est de trouver le plus court chemin permettant de relier un ensemble de villes [99].
- *le Max-Min Ant System (MMAS)* : Une variante efficace du Ant System est le Max-Min Ant System (MMAS), où seules les meilleures fourmis tracent des pistes et où le dépôt de phéromones est limité par une borne supérieure (empêchant une piste d'être trop renforcée) et une borne inférieure (laissant la possibilité d'être explorée à n'importe quelle solution). Cet algorithme atteint de meilleurs résultats que l'original, et évite notamment une convergence prématurée. [99]
- *le Ant Colony System (ACS)* : L'autre variante la plus connue est le Ant Colony System (ACS), où à une nouvelle règle de déplacement (appelée « règle pseudo-aléatoire proportionnelle ») s'ajoute un processus de mise à jour « locale » des éléments des pistes de phéromones, l'objectif de ce mécanisme étant d'augmenter la diversification de la recherche.

Toutes les variantes que nous venons d'exposer ont été récemment regroupées sous une description plus large : l'heuristique ACO (Ant Colony Optimization), afin de faciliter le rapprochement des méthodes entre elles et de se soustraire aux spécificités du PVC. Dans cet effort de généralisation, on peut noter l'introduction (hasardeuse) d'un processus visant à coordonner et superviser le travail des fourmis. [99] Une méthode de type ACO suit l'algorithme suivant :

---

**Table 1** Algorithme général de type ACO

---

*Initialisation* des pistes de phéromone  
*Boucler* tant que critère d'arrêt non atteint  
*Construire* les solutions composants par composants,  
 Utilisation (facultative) d'une heuristique,  
*Mise à jour* des pistes de phéromone ;

---

2. <http://fr.wikipedia.org/>

**Optimisation naturelle : pistes de phéromone** Les algorithmes de colonies de fourmis sont nés à la suite d'une constatation : les insectes sociaux en général, et les fourmis en particulier, résolvent naturellement des problèmes relativement complexes. En effet il est connu que les fourmis (appelées éclaireuses) sont capables de déterminer le chemin le plus court entre leur nid et une source de nourriture. Ceci est possible grâce à la phéromone qui est une substance que les fourmis déposent sur le sol lorsqu'elles se déplacent. Lorsqu'une fourmi doit choisir entre deux directions, elle choisit avec une plus grande probabilité celle comportant une plus forte concentration de phéromone. Ces phéromones étant attractives, les fourmis passant à proximité vont avoir tendance à suivre, de façon plus ou moins directe, cette piste. En revenant au nid, ces mêmes fourmis vont renforcer la piste. Le mécanisme de communication utilisé par les fourmis porte le nom de *stigmergie*. Les fourmis utilisent l'environnement comme support de *communication* : elles échangent indirectement de l'information en déposant des phéromones. Lorsque la fourmilière atteint une taille critique, de nouvelles fourmis (appelées reines) partent à la conquête de leur environnement pour construire de nouveaux nids. Inversement, il est possible qu'une fourmilière puisse disparaître par manque de nourriture. Les colonies de fourmis consistent un exemple caractéristique de système auto-organisé. Ce système repose sur des rétroactions positives<sup>3</sup> et négatives<sup>4</sup>.

Les algorithmes de colonies de fourmis diffèrent selon plusieurs paramètres : le comportement des éclaireuses ; les types de phéromones et leur vitesse d'évaporation ; le comportement des suiveuses en présence de phéromone ; le comportement des reines. Naturellement, l'efficacité de ces algorithmes diffère selon l'environnement : la quantité de ressources, la vitesse de renouvellement des ressources, la quantité d'obstacles, la durée de vie des fourmis, les tailles critiques de fourmilières, le nombre de colonies en compétition<sup>5</sup>

**Les fourmis artificielles** Les colonies de fourmis ont été utilisées en extraction de connaissances. On retrouve notamment leur utilisation pour effectuer des tâches de clustering [95]. Dans la nature, les fourmis offrent un modèle stimulant pour le problème du partitionnement. L'exemple du tri collectif du couvain ou de la constitution de cimetières sont les plus marquants. Certains travaux expérimentaux montrent que certaines espèces de fourmis sont capables d'organiser spatialement divers éléments du couvain : les oeufs, les larves et les nymphes [38] [55]. Le modèle de règles utilisé est relativement simple :

---

3. le dépôt de piste attire d'autres fourmis qui vont la renforcer à leur tour, création de reines pour construire de nouveaux nids

4. l'évaporation de la piste empêche le système de s'emballer, la disparition de certains nids permettent au système de se focaliser sur les zones contenant le plus de ressources

5. [www.lirmm.fr/~koriche/W-Docs/Fourmis.pdf](http://www.lirmm.fr/~koriche/W-Docs/Fourmis.pdf)

- lorsqu'une fourmi rencontre un élément du couvain, la probabilité qu'elle s'en empare est d'autant plus grande que cet élément est isolé ;
- lorsqu'une fourmi transporte un élément du couvain, elle le dépose avec une probabilité d'autant plus grande que la densité d'éléments du même type dans le voisinage est grande

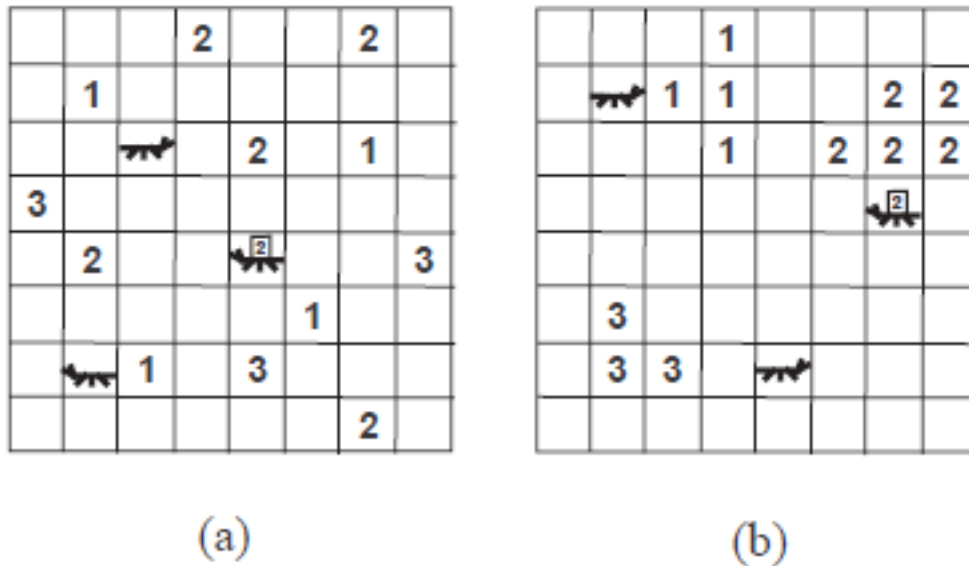


FIGURE 3.8 – Algorithme de fourmis.

Ces comportements ont inspiré les informaticiens et ont permis l'introduction de nouvelles heuristiques pour le problème du clustering. Ces algorithmes de fourmis artificielles ont pu améliorer la qualité du clustering en diversifiant la recherche de solution pour obtenir une solution plus optimale [122]

- Une des premières études relatives à ce domaine a été menée par Deneubourg et al dans [38] où une population d'agents-fourmis se déplacent aléatoirement sur une grille à deux dimensions et sont capables de déplacer des objets dans le but de les rassembler. Cette méthode a été étendue par Lumer and Faieta en 1994 sur des objets simples puis par Kuntz et al en 1997 où un problème réel est abordé dans le but de résoudre efficacement un problème d'optimisation.

- En 1999 Monmarché et al [99] proposent un algorithme de clustering hybride, où la recherche du nombre de classes est effectuée par des fourmis artificielles [123]. Ensuite, un algorithme classique en classification, les centres mobiles, est utilisé pour effacer les erreurs de classification inhérentes à la méthode stochastique.

- En 2005 Ouadfel propose dans [125] l'algorithme AntClust, ce dernier utilise les principes d'exploration stochastique et distribuée d'une population de fourmis artificielles pour four-

nir une segmentation d'une image en des classes pertinentes sans disposer d'une partition de départ et sans connaître le nombre de classes qui seront nécessaires.

**3.3.2.3.3 Autres méthodes** D'autres méthodes non supervisées sont également utilisées pour la caractérisation des tissus cérébraux. Citons rapidement celles reposant sur l'utilisation de cartes de Kohonen [29]; [36], sur la caractérisation des tissus par des paramètres fractals [37], sur une croissance de régions 3D [74], ou encore la segmentation par « fuzzy classifiers » [159] [117].

## 3.4 Conclusion

Dans ce chapitre on a passé en revue quelques méthodes de segmentation d'image en mettant l'accent sur les images médicales, particulièrement les IRM cérébrales. La particularité des IRM ainsi que la complexité du cerveau humain (*Ref* .Annexe A) font du processus de segmentation, un problème complexe difficile à résoudre, ce qui explique les innombrables recherches sur le sujet. Cependant la littérature révèle qu'aucune technique de segmentation n'est privilégiée pour l'extraction des tissus cérébraux [85], et qu'aucune méthode n'est capable à elle seule de résoudre la question de la segmentation, Ce qui a donné naissance à de nouvelles perspectives. En effet une forte tendance s'est développée en faveur de la coopération entre méthodes ou encore la segmentation par la fusion de données, cette dernière est naît de la nécessité de prendre en compte simultanément plusieurs types d'images représentant le même objet physique, en vue de la segmentation. Dans ce qui suit, nous présenterons les principes de fusion d'informations et dresserons un panorama des principales méthodes.

---

# FUSION DE DONNÉES

---

## 4.1 Introduction

Le problème de la combinaison et de l'utilisation simultanée de données et d'informations à partir de plusieurs sources se rencontre dans de nombreux champs d'application souvent liés au besoin de vouloir percevoir un environnement à partir de capteurs plus ou moins fiables, plus ou moins précis, plus ou moins efficaces.

Le terme de fusion s'étend de en plus à de plus vastes domaines et se rencontre maintenant fréquemment en représentation logique de la connaissance, ou encore en fouille de données [79].

En utilisant la redondance et la complémentarité des informations disponibles, la fusion permet d'accéder à une information plus fiable et donc d'améliorer la prise de décision. Elle permet également la prise en compte d'informations hétérogènes (numériques ou symboliques) imparfaites (imprécises, incertaines et incomplètes) modélisées sous forme de sources qu'il s'agit de combiner, agréger, fusionner.

Nous présentons dans le paragraphe suivant quatre approches permettant d'intégrer la représentation des connaissances incertaines et/ou imprécises : la théorie des probabilités, la théorie floue et des possibilités et enfin 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 précédemment présentés et enfin de justifier le cadre formel utilisé dans la suite de ce mémoire.

## 4.2 Eléments théoriques

### 4.2.1 Représentation de Connaissances ambiguës « imparfaites »

Les connaissances dont nous disposons sur une situation quelconque sont généralement ambiguës, soit parce que nous éprouvons des difficultés à les énoncer clairement (imprécision), soit parce que nous avons un doute sur leur véracité (incertitude) [15].

Les informations manipulées sont le plus souvent imparfaites. C'est d'ailleurs dans ces imperfections que se trouve une des raisons d'être de la fusion. Ces imperfections se manifestent sous de multiples formes : ambiguïté, bruit, incomplétude, imprécision, incertitude, incohérence et conflit... Il faut y ajouter le caractère variable et évolutif des informations relatives au monde dynamique.

*L'imprécision* concerne donc le contenu de l'information et porte sur un défaut quantitatif de connaissance, tandis que *l'incertitude* est relative à la vérité d'une information, caractérisant sa conformité à la réalité [39]. Par exemple, la proposition "Cet homme est grand" est imprécise, "Il pleuvra demain" est incertaine et "Il pleuvra beaucoup demain" est imprécise et incertaine. D'une manière générale, gagner en certitude fait perdre en précision.

Les approches permettant d'intégrer la représentation des connaissances incertaines et/ou imprécises et de raisonner à partir de celles-ci sont :

*Les probabilités, les ensembles flous et les possibilités et les fonctions de croyance.*

#### 4.2.1.1 Théorie des probabilités

Cette approche est la première à avoir été utilisée pour gérer des événements incertains [27]. Dans ce formalisme, la mesure entre l'information disponible et les différentes hypothèses est représentée par une distribution de probabilité conditionnelle. Cette théorie bénéficie de quatre siècles de travaux et repose donc sur des fondements mathématiques et une expérience solides, ce qui explique pourquoi c'est encore la théorie la plus utilisée pour représenter l'incertain [165].

Les axiomes régissant les lois de probabilités sont les suivants :

$$P(\textit{impossible}) = 0 \tag{4.1}$$

$$\textit{si } A \cap B = \phi, P(A \cup B) = P(A) + P(B) \tag{4.2}$$

$$P(A) + P(\bar{A}) = 1 \quad (4.3)$$

**4.2.1.1.1 Cadre bayésien** La loi de combinaison de distributions de probabilités, estimant la probabilité de l'occurrence d'un événement futur (probabilité a posteriori) en observant l'occurrence d'événements similaires passés (modélisée par une distribution *a priori*) est connue sous le nom de règle de Bayes. Elle permet après avoir fait l'estimation d'évaluer la probabilité a posteriori des hypothèses  $H_i$ . Si  $H_1, \dots, H_c$  forment un ensemble d'hypothèses mutuellement exclusives et recouvrant l'espace des hypothèses, la probabilité a posteriori d'un événement  $H_i$  parmi les  $C$  hypothèses connaissant l'information  $n_j$  à partir des lois de vraisemblance et des probabilités à priori, peut ainsi être déterminée par :

$$P(H_i | n_j) = \frac{P(H_i) \cdot P(n_j | H_i)}{\sum_{k=1}^c P(H_k) \cdot P(n_j | H_k)} \quad (4.4)$$

$P(H_i)$  est la probabilité *a priori* de l'hypothèse  $H_i$ , et  $P(n_j | H_i)$  représente la probabilité d'observer l'information  $n_j$  lorsque l'hypothèse  $H_i$  est réalisée.

Les lois  $P(n_j | H_i)$  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  $P(n_j | H_i)$  sont estimées par des lois statistiques. Ces probabilités peuvent être paramétriques ou non, dans le premier cas une forme est choisie pour  $P(n_j | H_i)$  et ses paramètres sont estimés (par maximum de vraisemblance par exemple), dans le deuxième (par exemple fenêtres de Parzen).

L'inconvénient majeur de cette théorie est qu'elle ne distingue pas l'incertitude de l'imprécision, si intrinsèquement, les probabilités expriment bien le fait qu'une réalisation est incertaine, en revanche, la notion d'imprécision est absente de cette modélisation, ce qui amène souvent à confondre les deux notions [5]. C'est ce qui a incité les chercheurs à penser au développement d'autres formes de modélisation.

#### 4.2.1.2 Théorie des ensembles flous

Pour remédier à ce problème et introduire de façon explicite la notion d'imprécision, zadeh dans [96] a introduit la théorie des ensembles flous, elle permet de prendre en compte des informations imprécises ou approximatives dans un système de décision.

- Dans la théorie des ensembles stricts « logique binaire » où un élément soit il ap-

partient soit il n'appartient pas à un ensemble. Sachant que  $S$  représente l'univers, ou espace de référence, un sous-ensemble  $A$  de  $S$  est défini par sa fonction caractéristique  $\mu(x)$ , telle que :

$$\mu_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

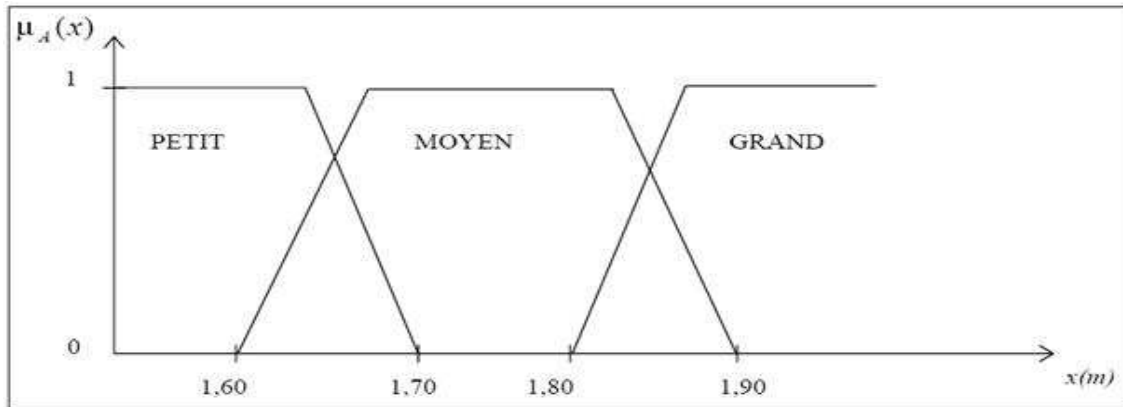


FIGURE 4.1 – Exemple de représentation d'une fonction d'appartenance floue [5].

- $\mu_A$  : est une fonction binaire, spécifiant l'appartenance (binaire) de chaque point de  $S$  à  $A$ .
- La théorie des ensembles flous traite de l'appartenance graduelle. Un sous-ensemble flou de  $S$  est défini par sa fonction d'appartenance  $\mu$  de  $S$  dans  $[0, 1]$ . Pour tout  $x$  de  $S$ ,  $\mu(x)$  est la valeur dans  $[0, 1]$  représentant le degré d'appartenance de  $x$  au sous-ensemble flou (ou ensemble flou).

Contrairement à la théorie des ensembles stricts dans cette théorie la fonction d'appartenance ne prend pas ses valeurs uniquement dans l'ensemble  $0, 1$ , mais plutôt dans tout l'intervalle  $[0, 1]$ , ainsi un élément appartiendra

- peu à un ensemble  $A$  si sa fonction d'appartenance est faible
- beaucoup si celle-ci est forte.
- pas du tout si sa fonction est nulle
- complètement si sa fonction est égale à 1

La figure 4.1) illustre un exemple où des individus sont classés selon leur taille, ainsi trois classes ont été distinguées : grands, moyens et petit, des fonctions d'appartenance peuvent être définies selon cette figure. Une caractéristique importante de la théorie des sous-ensembles flous est sa capacité à manipuler des informations de nature numérique

comme des informations de nature symbolique (proche, éloigné, au-dessus, en-dessous) [25].

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  $A$  :

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

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

$f_A$  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 4.2.1.2

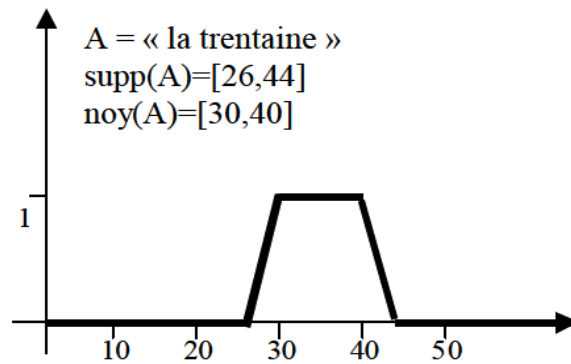


FIGURE 4.2 – Exemple de sous-ensemble flou. [40]

Les opérations pouvant être appliquées sur les ensembles flous sont : l'égalité, l'inclusion, l'union, l'intersection et enfin le complément.

### Les $\alpha$ -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

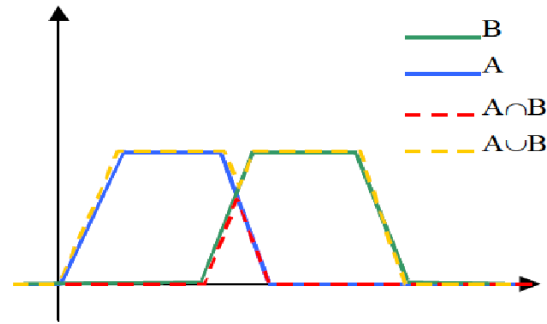
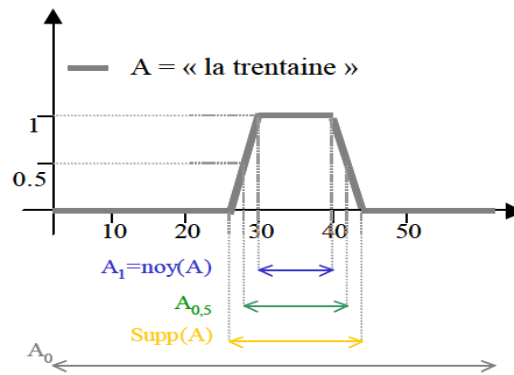


FIGURE 4.3 – Exemple d’intersection et d’union. [40]

simple de créer un sous-ensemble flou correspondant à  $A$  est de fixer une limite inférieure, notée  $\alpha$ , aux degrés d’appartenance pris en considération (voir Figure ).

Pour toute valeur  $\alpha$  de  $[0, 1]$ , on définit la  $\alpha$ -coupe  $A_\alpha$  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 :  $\chi_{A_\alpha}(x) = 1$  si et seulement si  $f_A(x) \geq \alpha$

FIGURE 4.4 – Exemples de  $\alpha$ -coupes associées au sous-ensemble flou  $A$ . [40]

L’inconvénient des ensembles flous est qu’ils représentent essentiellement le caractère imprécis des informations, l’incertitude étant représentée de manière implicite et n’étant accessible que par déduction à partir des différentes fonctions d’appartenance [67].

### 4.2.1.3 Théorie des possibilités

La théorie des possibilités dérivée de celle des ensembles flous, elle a été introduite en 1978 par Zadeh [97], puis développée par Dubois et Prade [39], constitue un cadre permettant de traiter les concepts d'imprécision et d'incertitude de nature non probabiliste. Elle peut être vue indépendamment de toute interprétation probabiliste comme une approche ordinale de l'incertain dans  $[0, 1]$ , exploitée à l'aide des mesures de possibilité et de nécessité [165].

La théorie des possibilités, dérivée des ensembles flous, permet de représenter à la fois l'imprécision et l'incertitude, par l'intermédiaire de distributions de possibilités  $\Pi$  sur un ensemble  $S$  et de deux fonctions caractérisant les événements : la possibilité  $\Pi$  et la nécessité  $N$ .

### 4.2.1.4 Mesure de possibilité

– Une mesure de possibilité est une fonction  $\Pi$  de  $C$  dans  $[0, 1]$  telle que :

$$\Pi(\phi) = 0 \quad (4.5)$$

$$\Pi(S) = 1 \quad (4.6)$$

$$\forall I \subset N, \forall A_i \subseteq S (i \in I), \Pi(\cup_{i \in I} A_i) = \sup_{i \in I} \Pi(A_i) \quad (4.7)$$

*Sup* : supremum des valeurs données (le plus grand d'entre elles dans le cas fini) :

$$\Pi(A \cup B) = \text{Max}(\Pi(A), \Pi(B)) \quad (4.8)$$

**4.2.1.4.1 Distribution de possibilité** La construction d'une distribution de possibilités permet de modéliser des informations imprécises [3,4] (par exemple, un expert fournira plus volontiers un intervalle censé contenir la valeur d'un paramètre plutôt qu'une valeur précise). La distribution de possibilités d'un paramètre  $x$  s'écrit classiquement  $\pi_x$ .  $\pi_x(u)$  est alors une valeur comprise entre 0 et 1 désigne le degré de possibilité pour que  $x = u$  :

- $\pi_x(u) = 0$  signifie que  $x = u$  est impossible,
- $\pi_x(u) = 1$  signifie que rien n'empêche  $x$  de valoir  $u$ ,
- il existe au moins une valeur  $u^*$  telle que  $\pi_x(u^*) = 1$  (condition de normalisation).

Les valeurs de  $u$  pour lesquelles  $0 < \pi_x(u) \leq 1$  forment l'ensemble flou des valeurs possibles de  $x$ . Une distribution de possibilité s'interprète comme une fonction donnant le degré de possibilité pour qu'une variable prenne la valeur  $x$  (c'est des fonctions d'appartenance à des ensembles flous),  $S$  étant l'ensemble des valeurs que peut prendre la variable. La définition d'une distribution de possibilité  $\pi$ , qui attribue à tout singleton de  $S$  un réel dans  $[0, 1]$ , et qui vérifie de plus (cas où  $S$  contient la vérité) :

$$\sup_{x \in S} \pi(x) = 1 \quad (4.9)$$

Par l'intermédiaire de distributions de possibilités  $\pi$  sur l'ensemble  $S$  et des fonctions de possibilité et de nécessité, l'imprécision et l'incertitude peuvent être représentées les deux à la fois.

$$\prod(A) = \sup\{\pi(x), x \in A\} \quad (4.10)$$

$$N(A) = \inf\{(1 - \pi(x)), x \notin A\} = 1 - \prod(A^c) \quad (4.11)$$

Mesure et distribution de possibilité peuvent être associées bijectivement.

L'ensemble non flou  $\{x \in S / \pi(x) \geq \alpha\}$  est appelé  $\alpha$ -coupe de la distribution de possibilité  $\pi$ . En particulier, la 0-coupe est le support de la distribution, et la 1-coupe est le noyau de  $\pi$ .

**4.2.1.4.2 Mesure de nécessité** Par dualité, une mesure de nécessité est définie comme une fonction  $N$  de  $C$  dans  $[0, 1]$  telle que :

$$\forall A \subseteq S; N(A) = 1 - \prod(A^c) \quad (4.12)$$

Ainsi, plus  $A$  est nécessaire, moins  $A^c$  est possible, ce qui permet d'évaluer la certitude sur la réalisation de l'événement  $A$ . Cette dualité signifie que si un événement est nécessaire, son contraire est impossible. Une mesure de nécessité vérifie les propriétés suivantes :

$$N(\emptyset) = 0 \quad (4.13)$$

$$N(S) = 1 \quad (4.14)$$

$$\forall I \subset N, \forall A_i \subseteq S (i \in I), N(\bigcap_{i \in I} A_i) = \inf_{i \in I} N(A_i) \quad (4.15)$$

*Inf* : L'infimum des valeurs concernées, c'est-à-dire la plus petite d'entre elles dans le cas fini.

Les mesures de possibilité et de nécessité ont de plus les propriétés suivantes :

- $\forall A \subseteq S, \max(\prod(A), \prod(A^c)) = 1$  , qui exprime le fait que l'un des deux ensembles  $A$

- et  $A^c$  est complètement possible ;
- $\forall A \subseteq S, \min(N(A), N(A^c)) = 0$ , qui exprime que deux événements contraires ne peuvent pas être simultanément nécessaires ;
  - $\forall A \subseteq S, \Pi(A) \geq N(A)$ , un événement doit être possible avant d'être nécessaire
  - $\forall A \subseteq S, N(A) > 0 \Rightarrow \Pi(A) = 1$
  - Puisque  $N(A) > 0 \Rightarrow \Pi(A^c) < 1$ , et  $\max(\Pi(A), \Pi(A^c)) = 1$
  - $\forall A \subseteq S, \Pi(A) < 1 \Rightarrow N(A) = 0$
  - $\forall A \subseteq S, N(A) + N(A^c) \leq 1$
  - $\forall A \subseteq S, \Pi(A) + \Pi(A^c) \geq 1$

Les deux dernières propriétés expriment la non additivité. La donnée de  $\Pi(A)$  n'est pas suffisante pour déterminer complètement  $\Pi(A^c)$ , contrairement aux mesures de probabilité.

L'incertitude attachée à un événement est exprimée par deux nombres, et non plus par un seul [71].

#### 4.2.1.5 Théorie des fonctions de croyance

La théorie de l'évidence, encore appelée la théorie des fonctions de croyance, a été introduite par Dempster [127] [128] et formalisée par Shafer [57]. Elle représente à la fois l'imprécision et l'incertitude au travers de deux fonctions : la fonction de croyance et la fonction de plausibilité [70] [3]. Ces deux fonctions sont dérivées des fonctions de masses.

**4.2.1.5.1 Fonction de masse** Le principe de la théorie des croyances repose sur la manipulation de ces fonctions de masse définies sur des sous-ensembles et non sur des singletons comme dans la théorie des probabilités. L'ensemble des hypothèses pour une source (typiquement une classe dans un problème de classification multisource) est défini sur l'espace  $X = \{H_1, H_2, \dots, H_k, \dots, H_N\}$ , appelé espace de discernement où  $H_k$  désigne une hypothèse en faveur de laquelle une décision peut être prise. Les fonctions de masse sont définies comme une fonction de  $2^x$  dans  $[0, 1]$ . En général on impose  $m(\emptyset) = 0$  et une normalisation de la forme :

$$\sum_{A_i \subseteq X} m(A_i) = 1 \quad (4.16)$$

- $m(A_i)$  est la confiance portée strictement dans  $A$  sans que celle-ci puisse être répartie sur les hypothèses qui la composent
- $A$  est un élément focal si  $m(A) \neq 0$ .

**4.2.1.5.2 Fonction de croyance** Une fonction de croyance  $Bel$  est une fonction totalement croissante de  $2^X$  dans  $[0, 1]$  définie par :

$$Bel(\phi) = 0 \quad (4.17)$$

$$Bel(X) = 1 \quad (4.18)$$

$$Bel(A) = \sum_{B_i \subset A} m(B_i) \quad (4.19)$$

$Bel(A)$  est une fonction de croyance. Inversement, à partir d'une fonction de croyance  $Bel$ , on peut définir une fonction de masse  $m$  par :

$$\forall A \in 2^X, m(A) = \sum_{B_i \subseteq A} (-1)^{|A-B|} Bel(B) \quad (4.20)$$

**4.2.1.5.3 Fonction de plausibilité** Une fonction de plausibilité  $Pl$  est également une fonction de  $2^X$  dans  $[0, 1]$  définie par :

$$\forall A \in 2^X, Pl(A) = \sum_{B_i \cap A \neq \phi} m(B) \quad (4.21)$$

La plausibilité mesure la confiance maximum que l'on peut avoir en  $A$ . A partir de la valeur de la masse d'un état, on peut définir un intervalle de probabilité. Cet intervalle contient la valeur précise de la probabilité de l'état, et est borné par les mesures croyances et plausibilité :

$$Bel(A) < P(A) < Pl(A) \quad (4.22)$$

Les mesures de croyance et de plausibilité sont liées :

$$Pl(A) = 1 - Bel(\overline{A}) \quad (4.23)$$

De ce fait, la connaissance d'une seule de ces valeurs (masse, croyance ou plausibilité) suffit à réduire les deux autres.

#### 4.2.1.5.4 Propriétés des fonctions de croyance

- $\forall A \in 2^X, Pl(A) \geq Bel(A)$
- $\forall A \in 2^X, Bel(A) + Bel(A^c) \leq 1$
- $\forall A \in 2^X, Pl(A) + Pl(A^c) \geq 1$
- $\forall A \in 2^X, Bel(A) + Bel(A^c) = 1 \Leftrightarrow Bel(A) = Pl(A)$

- $[Bel(A); Pl(A)]$  représente l'intervalle de confiance
- la longueur de cet intervalle est égal à la mesure de l'ignorance sur  $A$  et son complémentaire
- Ignorance totale quand  $m(X) = 1$
- Si  $|A| > 1 \Rightarrow m(A) = 0$  alors  $m = Bel = Pl$
- Grande souplesse de modélisation en considérant des disjonctions (hypothèses composées).

**4.2.1.5.5 Relation entre les trois théories** D'après Bernadette Bouchon-Meunier [15]; les mesures de possibilité peuvent être placées dans le cadre d'une théorie plus générale sur les fonctions de croyance, qui a l'intérêt de les situer par rapport aux probabilités. La théorie des possibilités et la théorie des probabilités apparaissent comme descendantes d'un ancêtre commun, la théorie de l'évidence. Lorsque des éléments focaux sont emboîtés, tels que  $y_1 \supseteq y_2 \supseteq \dots$ , la fonction de plausibilité a les propriétés d'une mesure de possibilité et la fonction de croyance a les propriétés d'une mesure de nécessité;

- Lorsque les éléments focaux sont des singletons, toute partie  $Y$  de  $X$  est telle que  $Cr(Y) = Pl(Y)$  et cette valeur commune est la probabilité de l'événement  $Y$ .

#### 4.2.1.6 Autres approches

Les modélisations de l'incertain se rapportant à la fusion, qu'on vient d'aborder sont les plus trouvées dans la littérature. Il existe néanmoins d'autres méthodes de modélisation des connaissances apportées par capteur sur l'état d'un système, mais elles sont utilisées de manière beaucoup plus marginale. La dessous quelques exemples de ces modélisations :

- Systèmes experts à degrés de confiance,
- filtrage de Kalman,
- réseaux de neurones.

## 4.3 Fusion d'information

La fusion d'informations est apparue afin de gérer des quantités très importantes de données multisources dans le domaine militaire. Depuis quelques années des méthodes de fusion ont été adaptées et développées pour des applications en traitement du signal.

Plusieurs sens sont donnés à la fusion d'informations, nous reprenons ici la définition proposée par [70] :

*La fusion d'informations consiste à combiner des informations hétérogène issues de plusieurs sources afin d'améliorer la prise de décision.*

La fusion de données correspond à la volonté d'utiliser simultanément plusieurs sources de données, ayant des caractéristiques éventuellement différentes, afin d'obtenir une nouvelle information de meilleure qualité.

### 4.3.1 Caractéristiques générales des données

Les données à fusionner ont des caractéristiques importantes et qui doivent donc être prises en compte avant d'entamer le processus de fusion, elles sont décrites ci-dessous :

- **Le type** : les informations à fusionner peuvent provenir de différentes sources ; observations directes, des résultats de traitements sur ces observations, de connaissances plus génériques, exprimées sous forme de règles par exemple, ou d'avis d'experts. Ces informations peuvent être exprimées sous forme numérique ou sous forme symbolique.
- **Le niveau de fusion** : On distingue généralement deux niveaux le bas niveau (typiquement les mesures originales), d'un niveau plus élevé nécessitant des étapes préliminaires de traitement, d'extraction de primitives et de structuration de l'information.
- **Hétérogénéité des données** : en présence de données hétérogènes, une modélisation ainsi qu'un traitement spécifique pour chaque type est requis. Les données fréquentes ou rares, informations factuelles ou génériques, génériques ou spécifiques, . . . etc, donnent aussi lieu à des modélisations et à des traitements différents.
- **Les informations supplémentaires** : sont les informations qui servent à guider ou aider la combinaison et qui peuvent intervenir dans le choix de la modélisation des informations à fusionner. . Il peut s'agir d'informations sur les informations à combiner, ou d'informations contextuelles, sur le domaine.
- **Imperfection** : sans l'imperfection des informations, la fusion n'a pas lieu d'être. En effet Celle-ci est toujours présente et peut prendre diverses formes : incertitude, imprécision, incomplétude, ambiguïté, conflit, redondance, ou complémentarité.

### 4.3.2 Système de fusion et type d'architecture

La fusion s'avère une tâche assez complexe et peut se décomposer en plusieurs tâches de ce fait les techniques de fusion reposent sur quatre étapes à savoir, la modélisation, l'estimation, la combinaison et la décision. Ces différentes étapes sont décrites ci-après.

- **La modélisation** : cette étape comporte le choix d'un formalisme, et des expressions des informations à fusionner dans ce formalisme. La forme de  $M_i^j$  représente une information fournie par une source  $S_j$  dépend donc du formalisme retenu.
- **L'estimation** : Une estimation du modèle retenu est souvent nécessaire, Les informations supplémentaires peuvent aider à cette estimation.
- **La combinaison** : cette étape concerne le choix d'un opérateur de combinaison, compatible avec le formalisme de modélisation retenu.
- **La décision** : c'est l'étape ultime de la fusion, qui permet de passer des informations fournies par les sources au choix d'une décision  $d_i$ .

La manière dont ces étapes sont agencées définit le système de fusion et son architecture ;

### 4.3.3 Modèles de fusion

D'après [71] cet agencement peut se faire selon quatre modèles différents, dont la description est ci-après :

- Dans le premier modèle, aucune information n'est négligée. La décision est prise à partir de toutes les  $m_i^j$ , pour toutes les sources et toutes les décisions, dans ce cas on parle d'une *décision globale (centralisée)*. La complexité de ce modèle et de sa mise en œuvre a conduit à développer des systèmes simplifiées, mais aux performances plus limitées [67].
- Dans le deuxième modèle, les décisions sont prises séparément, dans ce cas on parle de décision décentralisée où de fusion de décisions, elle consiste à prendre d'abord des décisions locales au niveau de chaque source  $s$ , pour les fusionner dans un deuxième temps en une décision globale. Ce modèle est utilisé lorsque les sources ne sont pas disponibles simultanément. Ainsi l'introduction de sources supplémentaires peut être très facilement adaptée. Ce modèle compte des inconvénients tels :
  - La mal prise en compte des relations entre capteurs et des corrélations ou dépendances possible entre sources.
  - Ce modèle conduit très facilement à des décisions locales contradictoires et la résolution de ces confits implique des arbitrages de niveau supérieur, l'indisponibilité

des informations de départ rend cette tâche difficile.

Des modèles de ce type sont souvent mis en œuvre dans des applications en temps réel.

- Le troisième modèle, Comme le modèle global mais en plus simple, est un modèle centralisé qui nécessite de disposer simultanément de toutes les sources. Il consiste à combiner par une opération de fusion  $F$  tous les  $M_i^j$  relatifs à la même décision  $d_i$ , pour obtenir une forme fusionnée  $M_i = F(M_i^1, M_i^2, \dots, M_i^m)$ . Puis une décision est prise sur le résultat de cette combinaison. Ici, aucune décision intermédiaire n'est prise, et l'information est manipulée dans le formalisme choisi jusqu'à la dernière étape, diminuant ainsi les contradictions et les conflits.

L'ajout d'éventuelles sources d'information s'avère une tâche assez difficile ce qui entraîne une perte de souplesse du modèle.

- Enfin, le dernier modèle est un modèle intermédiaire, hybride, qui consiste à choisir de manière adaptative les informations nécessaires pour un problème donné en fonction des spécificités des sources.

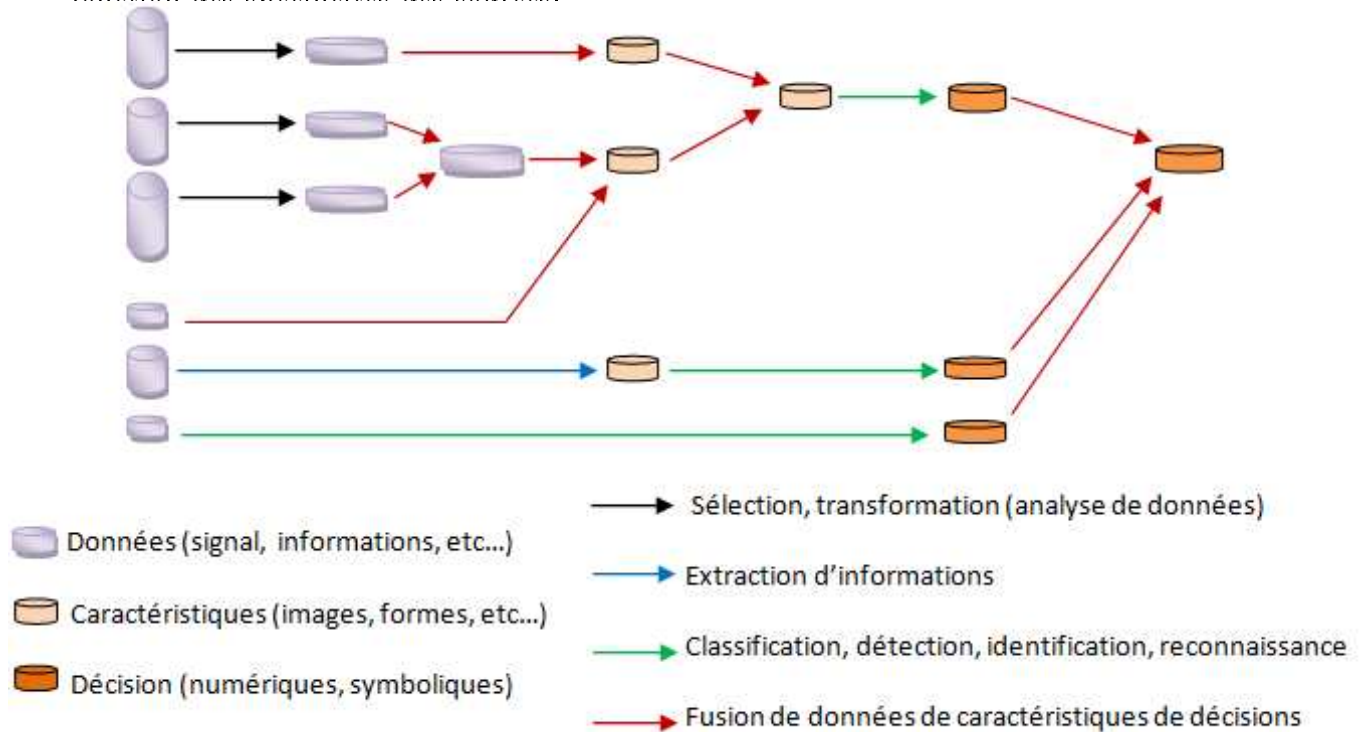


FIGURE 4.5 – Graphe de fusion.

### 4.3.4 Classification des types de fusion

Une formalisation possible de la fusion d'information introduit trois niveaux conceptuels correspondant à trois types d'information :

- **La fusion de données** c'est le niveau conceptuel le plus bas. Elle consiste essentiellement à marier des informations de bas niveau comme par exemple des primitives, dans le but de rendre l'information moins bruitée que celle obtenue avec une seule source d'information.
- **La fusion de décisions** ce type de fusion agit au niveau de l'espace de décision. Elle effectue l'association d'informations élaborées (numériques ou symboliques) qui peuvent être considérées comme des propositions de décision.
- **La fusion de modèles** ce cas est celui dans lequel les aspects complémentaires de différentes approches sont mis à partie pour combler les imperfections dont souffrent chacune d'elles indépendamment. Dans [155], le problème de la détection de contours est abordé par le biais d'un algorithme de Canny-Deriche combiné avec un réseau de neurones qui permet d'en régler les paramètres. De manière analogue, une modélisation markovienne coopère avec un réseau de neurones dans [171].

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-PET (tomographie d'émission de positron) [11] ou la fusion multi-modale [111]. Au niveau caractéristique il y a des techniques de segmentation à base de connaissances comme celles proposées par Clark [115] ou par Barra [166]. Enfin au niveau de la prise de décision la théorie de la croyance (théorie de Dempster-Shafer) propose un modèle pour la fusion [70] [69] [103].

### 4.3.5 Opérateurs de fusion

Après les travaux initiaux de L. Zadeh [96], beaucoup d'opérateurs ont été proposés dans la communauté floue, pour combiner des fonctions d'appartenance ou des distributions de possibilités. Ces opérateurs sont aussi appelés connecteurs, opérateurs de combinaison ou d'agrégation.

### 4.3.5.1 Classification des opérateurs de fusion

Il existe de nombreuses opérations de combinaison, ayant chacune un comportement différent selon le résultat désiré, parmi les principaux opérateurs on trouve en particulier les t-normes, les t-conormes, les moyennes, les sommes symétriques, et des opérateurs prenant en compte des mesures de conflit ou encore de fiabilité des sources. On les classe essentiellement en trois familles :

- *T-normes* (*T* comme *triangulaire*) : Qui a un comportement conjonctif (de type intersection des résultats).
- *T-conormes* : Qui ont un comportement disjonctif (de type de réunion des résultats) ou bien les moyennes qui ont un comportement de compromis.
- *Opérateurs adaptifs* : C'est des opérateurs dont le comportement varie en fonction du contexte et des données [67] en plusieurs catégories :
  - *Les opérateurs autonomes à comportement constant (ACC)* : c'est des opérateurs qui ont un même comportement quelque soit les valeurs à combiner et fournissent un résultat indépendant du contexte de l'agrégation.
  - *Les opérateurs autonomes à comportement variables (ACV)* : ces opérateurs ont un comportement qui varie en fonction des valeurs à fusionner, indépendamment du contexte.
  - *Opérateurs dépendants du contexte (DC)* : Le résultat de la fusion ne dépend plus seulement des données à agréger mais aussi du contexte, ce dernier est souvent défini à partir d'une connaissance *a priori*, sur le système de capteurs ou le phénomène étudié, de connaissances expertes ou d'apprentissage et est donc déterminé de manière indépendante par rapport aux données.

Une classification des opérateurs toujours suivant leur comportement et leur attitude vis-à-vis du contexte a été proposée dans [68]. Dans la suite, le comportement de  $F$  est qualifié de :

- *Sévère* : si  $F(n_1, n_2) \leq \min(n_1, n_2)$ , cet opérateur suppose que les deux capteurs sont fiables et exploite l'information commune aux deux mesures.
- *Indulgent* : si  $F(n_1, n_2) \geq \max(n_1, n_2)$ ; cet opérateur agit sur des sources *a priori* en conflit. Il augmente la certitude sur l'événement observé (et donc augmente l'imprécision...) et exprime la redondance entre les informations.
- *Prudent* : si  $\min(n_1, n_2) \leq F(n_1, n_2) \leq \max(n_1, n_2)$ ; il adopte une attitude entre ces deux extrêmes.

Un opérateur peut être de type conjonctif (sévère) ou de type disjonctif (indulgent). Un opérateur prudent correspond à une position intermédiaire entre ces deux extrêmes.

### 4.3.5.2 Quelques propriétés

Parmi les propriétés qui doivent participer à la construction des opérateurs de fusion,

- La contrainte de fermeture : la valeur retournée après la fusion doit être de même nature que les informations d'entrée.
- Un opérateur doit être si possible associatif et commutatif : l'opérateur étant alors indépendant de l'ordre de présentation des informations.
- Continu : ce qui assure la robustesse de la combinaison pour des couples de mesures voisins.
- et strictement croissant par rapport aux informations entrées [165].

## 4.3.6 Méthodes numériques de fusion

### 4.3.6.1 Fusion en théorie des probabilités

**4.3.6.1.1 Etape de fusion** Dans le modèle bayésien, la fusion peut être effectuée de manière équivalente à deux niveaux :

- Soit au niveau de la modélisation, on calcule des probabilités et ensuite on utilise la règle de Bayes dont les termes seront estimés par apprentissage :
- Soit par la règle de Bayes elle-même, où l'information issue d'un capteur vient mettre à jour l'information sur  $x$  estimée d'après les capteurs précédents (cas des informations disponibles successivement). La combinaison des informations est sous la forme d'un produit, donc fusion est conjonctive.

Une autre vision de la combinaison dans la théorie des probabilités revient à considérer la fusion comme un problème d'estimation, (Par exemple la fusion consiste à combiner les probabilités pour trouver la probabilité globale d'appartenance à la classe) et permet d'utiliser des opérateurs de combinaison différents du produit (moyenne ou moyenne pondérée, médiane).

Cependant, il est possible d'envisager une fusion purement probabiliste [164]. 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 [5].

**4.3.6.1.2 Etape de décision** La règle la plus utilisée pour la décision probabiliste et bayésienne est le maximum à posteriori, elle consiste à privilégier l'événement ayant la plus

forte probabilité a posteriori à l'issue de l'étape de fusion. Le résultat pour chaque couple de mesure est une affectation à l'événement  $H_i$  le plus probable. De nombreux autres critères ont été développés, pour une meilleure adaptation aux besoins de l'utilisateur et au contexte de sa décision, tels : maximum de vraisemblance, maximum d'entropie, marginale maximale, espérance maximale, risque minimal, ...etc [71].

#### 4.3.6.2 Fusion floue et possibiliste

**4.3.6.2.1 Etape de fusion** Dans le cadre de la fusion numérique, une application possible de la théorie floue et possibiliste consiste à définir  $\pi$  sur  $D$  (l'ensemble des décisions possibles) et la mesure  $M_i^j$  par  $M_i^j = \pi_j(d_i)$ , c'est-à-dire comme le degré de possibilité de la décision  $d_i$ , selon la source  $S_j$ . La fusion dans cette théorie consiste donc à combiner entre des distributions  $\pi_j(d_i)$  modélisant les informations issues de capteurs, pour en créer comme résultat une distribution fusionnée.

L'intérêt de fusion dans cette théorie, est qu'elle ne se limite pas à un unique mode de combinaison, et cela est possible grâce à la grande variété d'opérateurs de combinaison offerts.

Parmi les principaux opérateurs on trouve en particulier les t-normes, les t-conormes, les moyennes, les sommes symétriques, et des opérateurs prenant en compte des mesures de conflit ou encore de fiabilité des sources (cf.4.3.5). Le choix d'un opérateur se fait en fonction des ses propriétés (leur interprétation en termes de fusion de données incertaines, imprécises, incomplètes ou encore ambiguës.) et de son comportement (sévères, indulgents ou prudents), Un point toutefois aussi important a prendre en compte est le pouvoir discriminant des opérateurs.

**4.3.6.2.2 Etape de décision** La règle de décision est ici celle du maximum de possibilité (le maximum des degrés d'appartenance) :

$$x \in C_i \text{ si } \pi_i(x) = \max\{\mu_k(x), 1 \leq k \leq n\} \quad (4.24)$$

Où  $\mu_k(x)$  désigne la fonction d'appartenance à la classe  $k$  résultant de la combinaison. Chaque point de mesure se voit affecté à l'hypothèse pour laquelle il a le plus grand degré de possibilité. Des contraintes peuvent être ajoutées à cette règle pour modifier son comportement.

La qualité de décision est mesurée essentiellement selon deux critères :

- *La netteté de la décision* : le degré d'appartenance maximum correspondant à la

décision est comparé à un seuil, choisi selon les applications ou de l'opérateur de combinaison choisi.

- *Le caractère discriminant de la décision* : il est évalué par comparaison des deux valeurs les plus fortes.

Dans le cas où ces critères ne sont pas vérifiés pour un élément  $x$ , celui-ci est placé dans une classe de rejet, ou reclassifié en fonction d'autres critères.

### 4.3.6.3 Fusion en théorie des croyances

**4.3.6.3.1 Etape de fusion** La règle de Dempster est utilisée pour combiner de manière probabiliste des ensembles indépendants. Supposons que deux fonctions de masse indépendantes,  $m_1$  et  $m_2$ , existent sur un même cadre de discernement  $\Omega$ . Supposons encore que  $\Omega$  ait été divisé en deux différents sous ensembles,  $B$  et  $C$ , pour les deux fonctions de masse, respectivement. Alors, la règle orthogonale de combinaison de Dempster s'écrit :

$$m(A) = m_1(B) \oplus m_2(C) = \frac{\sum_{B \cap C = A} m_1(B)m_2(C)}{1 - K}, \forall A \neq \phi \quad (4.25)$$

où  $K = \sum_{B \cap C = \phi} m_1(B)m_2(C)$

En effet, le terme  $K$  rend compte des conflits dans les distributions de croyance provenant des sources  $m_1$  et  $m_2$  et assure que la masse combinée est normalisée en intervalle d'unité. Cette formulation de la règle de Dempster est aussi valable quand  $B$  et  $C$  sont identiques. La règle de Dempster est plutôt utilisée dans la fusion de l'information de haut niveau, c'est-à-dire des informations de nature descriptive ou symbolique. En 1996, Isabelle Bloch [69] a utilisé la théorie de l'évidence de Dempster-Shafer pour la fusion de données en imagerie médicale.

**4.3.6.3.2 Etape de décision** Dans la théorie des fonctions de croyances, plusieurs règles de décision sont possibles et sont le plus souvent appliquées au choix d'un singleton  $C_i$ .

- Le maximum de plausibilité :

$$x \in C_i \text{ si } Pl(C_i)(x) = \max\{Pl(C_k)(x), 1 \leq k \leq n\} \quad (4.26)$$

- Le maximum de crédibilité (croyance) :

$$x \in C_i \text{ si } Bel(C_i)(x) = \max\{Bel(C_k)(x), 1 \leq k \leq n\} \quad (4.27)$$

Qui est équivalent au critère du maximum de plausibilité dans le cas où le résultat de la combinaison ne porte que sur les singletons.

- Le maximum de crédibilité sans recouvrement des intervalles de confiance (sans risque d'erreur et très stricte) :

$$x \in C_i \text{ si } Bel(C_i)(x) \geq \max\{Pl(C_k)(x), 1 \leq k \leq n, k \neq i\} \quad (4.28)$$

Cette dernière condition étant particulièrement stricte et pouvant ne conduire à aucune décision.

- Le maximum de crédibilité avec rejet [112] :

$$x \in C_i \text{ si } Bel(C_i)(x) = \max\{Bel(C_k)(x), 1 \leq k \leq n\} \text{ et } Bel(C_i)(x) \geq Bel(C_i^C)(x), \quad (4.29)$$

Qui exprime que la décision doit être suffisamment non ambiguë puisque la condition sera vérifiée si la masse est très focalisée sur  $C_i$ .

- Le maximum de probabilité pignistique, celle-ci étant définie par [129] :

$$\forall C_j \in D, B \text{ et } P(C_j) = \sum_{C_j \in A} \frac{m(A)}{|A|(1 - m(\phi))} \quad (4.30)$$

Il existe d'autres règles telles :

- Des règles mixtes ont également été proposées (dans lesquelles la plausibilité est utilisée pour certaines classes et la croyance pour d'autres).
- La décision peut également être prise en faveur d'une disjonction.
- Enfin, des règles de décision avec coût.

## 4.4 Fusion des données médicales

L'utilisation de nombreuses sources d'informations complémentaires est inhérente à de nombreux processus cliniques, ce qui explique pourquoi les approches de fusion de données étaient utilisées avant même le développement d'outils de fusion de données assistée par ordinateur. Depuis, de nombreuses méthodes ont été définies et implémentées pour assister et automatiser les processus de fusion de données et pour rendre ces processus plus précis et plus objectifs.

	<b>Probabilités</b>	<b>Flou/possibilités</b>	<b>Croyance</b>
<b>Avantages</b>	<ul style="list-style-type: none"> <li>-Justification axiomatique (Cox, 1946)</li> <li>- Cadre mathématique cohérent et puissant</li> <li>- Solutions optimales au sens de certains critères</li> <li>- Grande expérience</li> <li>- Règles d'usage théoriques et heuristiques</li> <li>- Richesse des outils statistiques</li> </ul>	<ul style="list-style-type: none"> <li>- Représente l'imprécision ou l'incertitude</li> <li>-Beaucoup d'opérateurs de combinaison, avec comportements différents.</li> <li>-Grande souplesse dans les modes de combinaison</li> </ul>	<ul style="list-style-type: none"> <li>-Imprécision + incertitude.</li> <li>-Représentation de l'ignorance</li> <li>-Sous-ensembles plutôt que singletons</li> <li>-Mesure du conflit</li> <li>-Richesse de l'étape de modélisation des données.</li> <li>-Flexibilité en fonction des contraintes.</li> </ul>
<b>Limites</b>	<ul style="list-style-type: none"> <li>-Représente l'incertitude, pas l'imprécision</li> <li>- Pas de modélisation de l'ignorance</li> <li>- Difficulté d'introduire des informations qui ne se modélisent pas aisément sous forme de probabilités</li> <li>- Contraintes strictes de modélisation (additivité)</li> <li>-Contraintes strictes pour l'apprentissage</li> <li>-Hypothèses pas toujours vérifiées</li> <li>- Nécessitent beaucoup de connaissances</li> <li>-Combinaison essentiellement conjonctive</li> </ul>	<ul style="list-style-type: none"> <li>-Choix de l'opérateur adéquat.</li> </ul>	<ul style="list-style-type: none"> <li>-Combinaison se résume la plupart du temps à l'application de l'opérateur orthogonal de Dempster.</li> </ul>
<b>Modélisatio</b>	$P(C_i I_j)$	$\mu_i^j / \pi^j(C_i)$	$m_j(A), Bel_j(A), Pl_j(A), A \subseteq \{C_1, \dots, C_n\}$
<b>Fusio</b>	Bayes	$F: [0,1]^l \rightarrow [0,1]$	Règle orthogonale $\oplus$
<b>Décision</b>	MAP, etc.	Max+contraintes	Max-Bel, max-Pl, etc.
	Hypothèses apprentissage statistiques	Classification des opérateurs (comportement, propriétés, conflit)	Modélisation (éléments focaux, hypothèses composées)

Tab 4.6 – comparaison entre les 3 théories.

### 4.4.1 Les applications médicales de la fusion de données

La fusion peut être classifiée en quatre catégories principales :

1. La fusion intra sujet et intra modalité : concerne la fusion entre différents jeux de données provenant d'une même modalité pour un sujet. A cette catégorie correspondent trois principaux objectifs cliniques :
  - a) le suivi dans le temps de modifications dans l'anatomie du patient,
  - b) la comparaison de différents états
  - c) l'imagerie par soustraction.
2. La fusion intra sujet et inter modalité : concerne la fusion entre des données multimodalité d'un même sujet. Aucune modalité d'imagerie médicale ne permet d'explorer toutes les structures anatomiques ou de fournir, à la fois, des informations anatomiques et fonctionnelles à la plus haute résolution spatiale et temporelle. Cette deuxième catégorie d'applications permet de profiter de la nature complémentaire des différentes modalités d'imagerie et pour différentes zones anatomiques.
3. La fusion inter sujet et intra ou inter modalité : concerne la fusion entre des données mono ou multimodalité concernant des sujets différents. Les applications principales de cette catégorie sont liées à l'étude de la variabilité anatomique ou fonctionnelle au sein d'une population de sujets.
4. La fusion sujet et modalités : concerne la fusion entre l'espace du sujet réel et les données images. Globalement, elle concerne la mise en correspondance géométrique entre le sujet réel, les outils (comme les outils chirurgicaux, les systèmes robotiques ou les systèmes d'imagerie interventionnelle) et des images ou mesures pré ou intra opératoires. L'étape de réalisation du geste chirurgical guidée par l'image s'appuie sur ce type de fusion.

### 4.4.2 Fusion d'image médicale

L'incomplétude des informations issues de chaque source est la raison principale qui motive la technique d'analyse d'images IRM basée sur la notion de fusion des données. L'information fournie par chaque source est en général partielle, la fusion d'informations permet en effet d'obtenir des informations de plus grande qualité, et donc améliore la prise de décision médicale. Il n'existe pas de méthode universelle de combinaison des deux images  $I_1$  et  $I_2$ , l'agrégation dépend fortement de la nature des données et du contexte

applicatif.

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-PET (tomographie d'émission de positron) [11] ou la fusion multimodale [111]. Au niveau caractéristique il y a des techniques de segmentation à base de connaissances comme celles proposées par Clark [115] ou par Barra [166]. Enfin au niveau de la prise de décision la théorie de la croyance propose un modèle pour la fusion [70] [69].

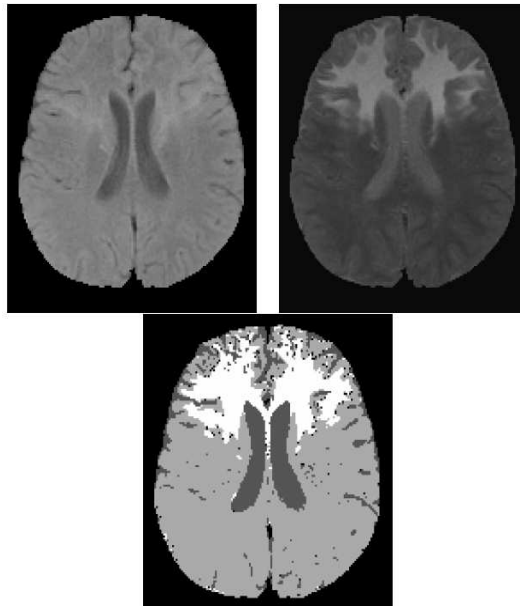


FIGURE 4.7 – Résultat de la classification par fusion floue de deux IRM. (la décision n'est prise que localement en chaque point, sans régularisation spatiale) [71]

#### 4.4.2.1 Introduction de l'information spatiale

L'information spatiale est fondamentale en traitement d'images. Son introduction dans les méthodes de fusion est cruciale, et nécessite souvent des développements spécifiques pour adapter les méthodes issues d'autres domaines. Un des objectifs les plus fréquents de ces développements est de garantir que la décision soit spatialement cohérente.

L'introduction de l'information spatiale peut se faire à différents ; niveaux modélisation, combinaison ou décision ; Elle s'avère la plus facile au niveau de la décision, plus ou moins implicite au niveau modélisation, suivant le niveau de représentation auquel on se place, et elle reste plus rare et plus délicate au niveau combinaison.

## 4.5 Conclusion

Dans ce chapitre on a exposé dans un premier temps les théories représentant les informations ambiguës et imparfaites (théorie des probabilités, floue, des possibilités et enfin des croyances), pour aborder la fusion dans chacune d'elles dans un second temps. Pour une meilleure synthèse on a dressé le tableau ci-dessous afin de faciliter et guider le choix de la théorie la plus adaptée dans le cadre de notre travail. Dans la suite de ce mémoire nous allons présenter notre contribution en justifiant le choix de la méthode de fusion de données et l'algorithme de classification adoptés, ainsi que les outils et les paramètres de chaque algorithme.

# CONTRIBUTION

---

## 5.1 Introduction

Dès les années 90 et au contact du domaine médical, la communauté informatique de l'analyse d'images a peu à peu réalisé la nécessité de ne pas se limiter à une étude de l'image elle-même mais à la prise en compte d'un contexte médical. Un médecin n'analyse pas une image isolée mais se réfère à tout un contexte clinique et social qui aide et conditionne la lecture qu'il peut faire d'une image. Cette migration d'une recherche ciblée, en analyse d'image, vers une recherche élargie, en analyse de données médicales composées d'images et d'autres informations cliniques, conduit au développement de nouvelles applications, plus spécifiques et pertinentes du point de vue clinique mais aussi souvent plus complexes à mettre en œuvre.

Notre approche se situe dans ce contexte, elle consiste en la prise en compte simultanée lors d'une segmentation d'image (notre choix s'est porté sur la modalité IRM) non seulement de l'image mais aussi des données provenant du contexte clinique ou social (âge, sexe, génétique, environnement) ou encore des données tabulaires venant du rapport clinique ou des résultats biochimiques...etc.

En présence de données hétérogènes et dans le contexte de la fouille de données complexes nous proposons un algorithme qui tente d'imiter la démarche entreprise par le médecin, ce qui revient à combiner des informations de nature hétérogènes, dans notre cas des images IRM et des données tabulaires qui contiendront les données structurées du dossier médical. Ainsi, un élément crucial dans des systèmes souhaitant aboutir à une telle prise de décision est l'existence d'un mécanisme capable de modéliser, de fusionner et d'interpréter

les informations disponibles. Les données fusionnées reflètent non seulement l'information générée par chaque source de données, mais encore l'information qui n'aurait pu être inférée par aucune des sources prises séparément.

La fusion s'opère donc sur des données hétérogènes et plutôt complémentaires qui ne s'expriment pas nécessairement dans le même cadre formel (numérique pour les images, linguistique ou numérique pour données issues du contexte clinique). L'intérêt de la fusion est ici de repousser la décision (la segmentation de structures par exemple) jusqu'à posséder suffisamment d'informations. En effet l'information de niveau de gris ou de position grossière ne suffit pas à la segmentation, mais l'agrégation de l'ensemble des connaissances permet d'inférer une décision fiable.

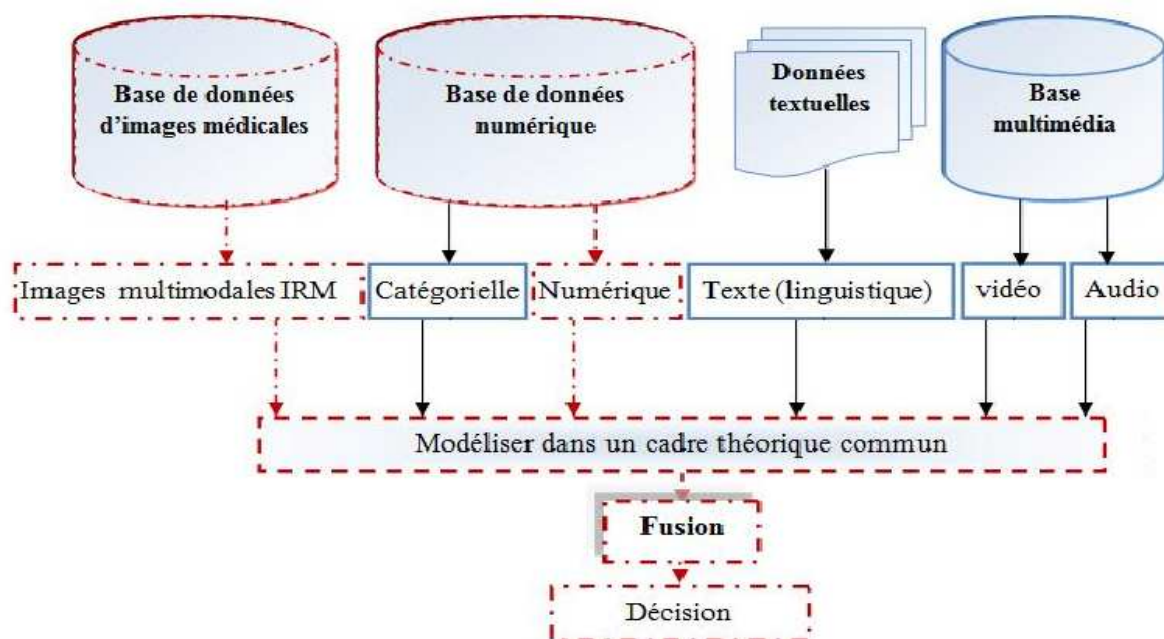


FIGURE 5.1 – La partie concernée par notre étude (en rouge).

## 5.2 Approche proposée

Afin de mettre en évidence notre approche, nous avons conçu un algorithme présenté dans 4. Cet algorithme est constitué de trois parties essentielles :

### I. *Modélisation des données*

Modélisation des informations dans un cadre théorique commun.

- *Modélisation de l'information issue du dossier médical*

Une modélisation dans la théorie des ensembles flous est appliquée à cette donnée, le choix de cette dernière est guidée par l'avis d'un expert de telle sorte qu'il y est une relation entre elle et une éventuelle pathologie pouvant apparaître sur l'image IRM.

- *Modélisation de l'image*

- Première étape : Initialisation des classes en lançant l'algorithme AntClust.
- Deuxième étape : Classification par l'algorithme FCM.

**II. Fusion de données** : Une fusion dans un cadre possibiliste est appliquée pour combiner les deux données hétérogènes modélisées précédemment.

**III. Prise de décision**

c'est l'étape ultime de la fusion, qui permet de passer des informations fournies par les sources au choix d'une décision  $d_i$ .

Notons que le but de cette étude n'est pas de construire une solution générique aux problèmes de fusion de données hétérogènes, mais seulement de vérifier si l'intégration d'une donnée dans le processus de segmentation d'une image améliorera ce dernier ou non.

Face à cette problématique, on est amené à répondre à plusieurs questions du genre : Quelle est la fusion envisagée ? Quels sont les type de données à fusionner ? Quelle modélisation utiliser pour chaque type de donnée ? Quel cadre théorique retenir pour la fusion ? Sur quel niveau s'opère la fusion ? Quel opérateur d'agrégation utiliser ? Quelle pondération donner à chacune des sources ? ...etc.

### 5.2.1 Quelle est la fusion envisagée ?

Selon [71] il existe différentes situations de fusion en traitement d'images, dans lesquelles les types d'informations ne sont pas les mêmes. Les principales situations sont les suivantes :

- *Plusieurs images du même capteur* : Les données sont relativement homogènes car elles correspondent à des mesures physiques similaires, par exemple le cas de plusieurs canaux du même satellite, d'images multi-échocs en IRM, ou encore de séquences d'images pour des scènes en mouvement.
- *Plusieurs images de capteurs différents* : L'hétérogénéité est beaucoup plus importante, les différents capteurs ne parlant pas du même aspect du phénomène. Les

images donnent chacune une vision partielle et ne sont pas informatives sur les caractéristiques auxquelles elles ne sont pas dédiées.

- *Plusieurs informations extraites d'une même image* : il s'agit de situations dans lesquelles on extrait divers types d'informations d'une image à l'aide de plusieurs détecteurs, opérateurs, classifieurs, etc., qui s'appuient sur des caractéristiques différentes des données, cherchent à extraire des objets différents, rendant ainsi les informations à fusionner souvent très hétérogènes.
- *Images et autre source d'information* : par autre source d'information, on entend par exemple un modèle, qui peut être soit particulier comme une carte, soit générique comme un atlas anatomique, des bases de connaissances, des règles, des informations issues d'experts, etc. Les informations sont à nouveau de types très différents, à la fois dans leur nature, et dans leur représentation initiale (images quand il s'agit d'une carte ou d'un atlas numérisé, mais aussi descriptions linguistiques, bases de données, etc.).

Dans le cadre de cette étude, la fusion envisagée se situe dans le dernier cas. En effet nous nous intéressons dans ce travail à une fusion de données de types différents.

## 5.2.2 Les types de données à représenter ?

Nous nous intéressons dans ce travail à une fusion de données de type hétérogène, le but est de modéliser le comportement humain via la fusion de données afin d'améliorer la segmentation. Pour ce faire, nous avons utilisé deux types d'informations différents :

- les informations extraites de l'image, de type numérique, qui concernent les tissus cérébraux.
- les informations issues du dossier médical (âge, sexe, antécédents héréditaires etc..) pouvant être numériques ou linguistiques.

De ce fait la modélisation de ces deux informations dans un cadre théorique commun est la première tâche à faire. L'étape de modélisation consiste en la représentation de l'information dans un cadre mathématique lié à une théorie particulière [67] [68].

### 5.2.2.1 Modélisation de la donnée issue du dossier médical

Des études statistiques ont permis de mettre en évidence certains facteurs de risque qui sont : l'âge, le sexe, les antécédents médicaux du patient, facteur héréditaire, . . . etc, ces informations peuvent être utilisées comme des connaissances *a priori*, et intégrées dans

le processus de segmentation. En effet ces informations participent et aident le praticien à l'interprétation des images médicales facilitant ainsi la prise de décision, de ce fait la modélisation de ces données dans un cadre théorique est la première étape pour leur futur intégration lors du processus de segmentation. Parmi ces données, on trouve celles qui sont structurées ou non, numériques (âge), catégorielles (sexe) ou textuelles (observation d'un médecin). Dans ce manuscrit on se limitera aux données numériques structurées. Il arrive souvent que les représentations numériques ne soient pas adaptées pour décrire une situation. Par exemple, si une variable a un grand domaine de variation, il peut être difficile d'affecter une valeur précise à chaque situation spécifique et on préférera alors utiliser des termes plus qualitatifs issus du langage naturel pour regrouper plus ou moins grossièrement des sous-ensembles typiques de situations intéressantes [71]. Ces types de

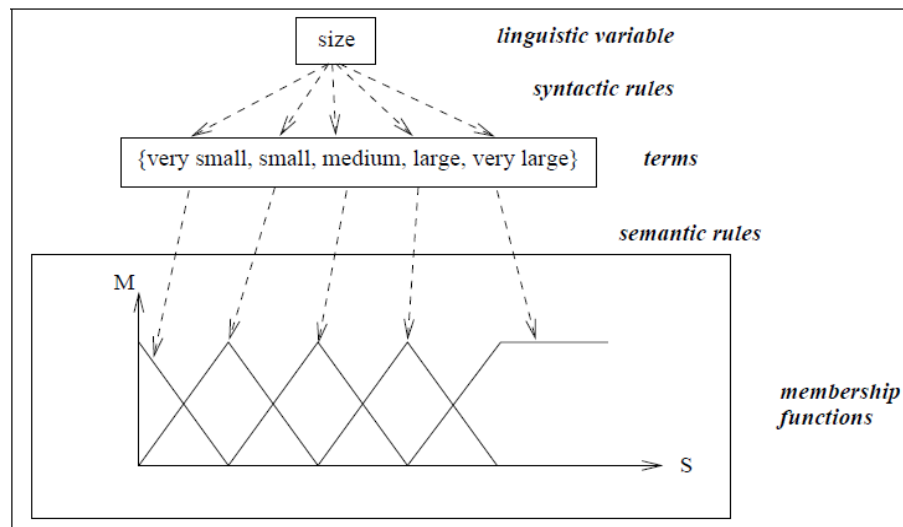


FIGURE 5.2 – Exemple d'une variable linguistique d'après [71].

représentations sont appelés variables linguistiques.

**5.2.2.1.1 Variables linguistiques** Ce sont des variables dont les valeurs sont des mots, des groupes de mots ou des phrases [98]. Leur intérêt est essentiellement que les caractérisations linguistiques peuvent être moins spécifiques que les caractérisations numériques, et nécessitent donc moins d'information pour être utilisées et manipulées dans des systèmes de raisonnement. Formellement, une variable linguistique est définie par un quintuplet  $(x; T(x); S; G; M)$  où  $x$  est le nom de la variable,  $T(x)$  l'ensemble des valeurs de  $x$  (appelées termes),  $S$  est le domaine ou univers sur lequel les valeurs de la variable sont définies,  $G$  est une règle syntaxique permettant de générer le nom  $X$  de chaque valeur de  $x$ , et  $M$  est une règle sémantique,  $M(X)$  étant l'ensemble flou défini sur  $S$  représentant

la signification de  $X$  [98] [31] [80]. Cette définition représente une conversion symbolique-numérique, et établit des liens entre le langage et les échelles numériques. Mais dans le cas de cette étude l'ensemble flou  $M(x)$  sera une simple approximation qui sera donnée par un expert.

La donnée issue du dossier médical sera donc modélisée en *théorie des ensembles flous*, une échelle de valeurs traduisant l'avis d'un expert est établie. Par exemple les variables d'entrée pour la maladie d'Alzheimer pour l'information âge seront : "risque faible", "risque moyen" et "risque haut". A chacune de ces dernières est associée une échelle de valeurs traduisant pertinemment le jugement d'un expert.

L'échelle suivante est proposée :

Risque Faible	Entre 30 et 40 ans
Risque Moyen	Entre 65 et 75
Risque Elevé	Plus de 80 ans

TABLE 5.1 – Risque d'atteinte par âge

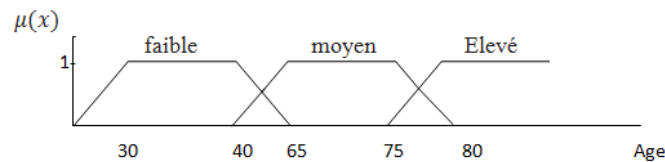


FIGURE 5.3 – Fuzzification de la donnée âge suivant une fonction trapézoïdale.

### 5.2.2.2 Modélisation de l'image

Les informations numériques extraites des images le sont ici à partir des IRM. Ces images à caractère anatomique sont utilisées pour classifier les tissus cérébraux et obtenir le degré d'appartenance de chaque voxel à ces derniers. La classification présente diverses applications, parmi lesquelles une meilleure compréhension des pathologies atteignant le fonctionnement du cerveau (démence de type Alzheimer, épilepsie, ...) ou la détection de structures cérébrales ou de régions pathologiques (ex. : tumeurs).

En reprenant la taxinomie de Bezdek et al [85]. Nous devons tout d'abord préciser si l'algorithme doit être supervisé ou non.

**5.2.2.2.1 Méthode supervisée ou non ?** Nous souhaitons obtenir un processus entièrement automatique, ce qui exclut dans un premier temps les méthodes supervisées ; d'autre part l'emploi d'un algorithme supervisé nécessite, une base d'apprentissage pour chaque classe, ce qui constitue une tâche fastidieuse pour l'expert [165]. De plus, la littérature rapporte que les méthodes supervisées en segmentation d'images médicales peuvent être très dépendantes de la base d'apprentissage. 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 des résultats, ce qui rend les méthodes supervisées inadaptées pour des études quantitatives lorsque la base d'entraînement est mal adaptée [133]. Ces raisons nous ont fait préférer une méthode non supervisée, évitant ainsi toute interaction avec un opérateur et garantissant une meilleure reproductibilité des résultats.

**5.2.2.2.2 Classification floue ou non floue ?** Les images de modalité IRM traitées dans ce mémoire présentent 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). Dans ce contexte, et d'après l'état de l'art présenté dans le chapitre 3, l'approche floue, est parfaitement adaptée à la manipulation de données incertaines et imprécises.

**5.2.2.2.3 l'algorithme utilisé ?** Avec les travaux de Bezdek, les méthodes floues de regroupement atteignent une certaine maturité. En 1973, Bezdek a soutenu une thèse de doctorat sur les mathématiques floues pour la classification [19] dans laquelle figurent les conditions nécessaires de la minimisation du critère général qui définit la famille d'algorithmes connus sous le nom des c-moyennes floues (FCM).

Beaucoup de variantes de ces algorithmes ont été par la suite mises au point afin d'en accroître les performances (Gustafson et Kessel ont proposé FCC ("Fuzzy C-Covariance", FCV (Fuzzy C-Varieties) [13], FCL (Fuzzy C-lines) [12], FCE (Fuzzy C-Elliptotypes) [76], AFCE (Adaptive Fuzzy C-Elliptotypes) [139], FCS (Fuzzy C-Shells) [138]). Cependant ces variantes sont souvent dédiées à une application particulière, certaines généralisent l'algorithme de base au cas de distances non euclidiennes, d'autres essaient d'intégrer des contraintes supplémentaires ou de tenir compte de l'environnement de chacun des éléments à grouper. Parfois les contraintes de l'algorithme de base sont assouplies, ce qui est le cas notamment dans l'approche possibiliste [141], qui présente l'une des déviations les plus originales par rapport à l'algorithme initial de Bezdek [20].

Bien qu'il existe un bon nombre de variantes du FCM, qui s'avèrent plus précises et plus spécifiques que l'algorithme des C-moyennes floues [75]. Le FCM reste comme même aujourd'hui encore utile dans un grand nombre de situations.

Le FCM faisant partie des algorithmes de partitionnement dits déterministes [75] nous paraît bien convenir dans le cadre de cette étude, et répond assez bien aux objectifs visés. En effet ces derniers sont simples à implémenter et convergent rapidement avec une solution localement optimale. Cependant leurs majeurs inconvénients est qu'ils nécessitent de fournir en entrée une partition initiale de bonne qualité ainsi que le nombre possible de classes. Ces contraintes rendent l'utilisation de ces algorithmes peu intéressante quand on veut segmenter automatiquement une image, ce qui est le cas dans le cadre de ce manuscrit.

Pour remédier à ce problème, on a donc recours à des stratégies sous-optimales ou à des heuristiques sujettes aux minima locaux. Les résultats obtenus dépendent bien évidemment de la configuration initiale choisie. Le choix des paramètres initiaux est un problème vaste de la recherche opérationnelle et de nombreuses méthodes sont possibles dans le cas de la classification [154] :

- Tirage aléatoire des centres des classes parmi les points ;
- Utilisation d'une méthode de classification hiérarchique ; On coupe la hiérarchie suivant le nombre de classes désiré et on l'estime les paramètres des classes à partir des sous-ensembles de la partition.
- Utilisation d'un algorithme de classe inférieure dans la famille des C-moyennes ; HCM -> FCM, FCM ->PCM

L'introduction d'une recherche stochastique à la place d'une recherche déterministe peut améliorer les résultats, par exemple avec un algorithme génétique (Jones and Beltrano, 1991 dans [88]; Cucchiara, 1993 dans [30] ) ou une population de fourmis. L'utilisation de fourmis artificielles pour la classification, à la place d'un AG par exemple, est pertinente dans le sens où les fourmis réelles ont ce genre de problèmes à résoudre. Dans un certain sens, le modèle de fourmis artificielles pour la classification est certainement plus proche du problème de la classification que le modèle génétique. Cependant le majeur inconvénient de ces algorithmes est la détermination des paramètres de l'algorithme et le temps d'exécution.

Pour pallier à ces problèmes, et améliorer la robustesse de notre approche nous proposons une hybridation du FCM avec l'algorithme AntClust proposé par Ouadfel [125].

Ce dernier possédant l'avantage de fournir une segmentation d'une image en des classes pertinentes sans disposer d'une partition de départ et sans connaître le nombre de classes qui seront nécessaires. De ce fait AntClust sera lancé en premier avec un nombre limité d'itération qui permettra de déterminer le nombre de classes initiales, fournissant ainsi une bonne initialisation au FCM, lequel devrait corriger des erreurs d'AntCust qui aurait mis beaucoup plus de temps à les corriger.

**L'algorithme AntClust** ANTCLUST est un algorithme de segmentation d'images proposé par Ouadfel dans [125] inspirée du système de reconnaissance chimique des fourmis. Cet algorithme est basé sur une population d'agents fourmis autonomes non intelligents capables de s'auto-organiser pour créer un comportement global complexe et faire émerger collectivement une segmentation optimale de l'image par l'intermédiaire des interactions qu'ils entretiennent avec leur environnement. AntClust utilise les principes d'exploration stochastique et distribuée d'une population de fourmis artificielles pour fournir une segmentation d'une image en des classes pertinentes sans disposer d'une partition de départ et sans connaître le nombre de classes qui seront nécessaires [125].

**Principe de AntClust** Considérons un ensemble de  $N$  pixels à regrouper en des classes aussi homogènes que possible en terme de niveau de gris. Considérons aussi une population  $A$  de  $K$   $\{a_1, a_2 \dots a_k\}$  fourmis qui coopèrent ensemble et communiquent par stigmergie pour fournir une classification optimale. Initialement les  $N$  classes sont constituées chacune d'un pixel. Au cours du processus de classification les fourmis déplacent les pixels d'une classe à une autre et tentent de regrouper dans une même classe le maximum de pixels similaires en terme de niveau de gris. Pour cela, une mesure de similarité doit être évalué entre un pixel  $p_i$  de niveau de gris  $ng_i$  et le centre de gravité  $g_k$  d'une classe  $c_k$  comme suit :

$$f(p_i, c_k) = \frac{1}{1 + \left(\frac{ng_i - g_k}{\beta}\right)^2} \quad (5.1)$$

$\beta$  est un paramètre qui contrôle la dilatation de la fonction  $f$ . Un schéma représentant l'évolution de la fonction  $f$  est donné dans la figure 5.4. La fonction de similarité  $f(\cdot)$  atteint son maximum pour  $ng_i = g_k$  et est normalisée entre 0 et 1. Quand une fourmi transporte un pixel, elle a accès à son voisinage immédiat. Elle calcule la fonction de similarité définie dans 5.1 pour chacune des cases des 8 voisins du pixel qu'elle transporte et évalue ainsi directement la possibilité de le déposer dans une des ses cases candidates. Le meilleur emplacement sera celui pour lequel la fonction de similarité est maximum. La fourmi décide alors de déposer son pixels sur cet emplacement avec une probabilité  $p_{depot}$ . Si cette décision est négative, la fourmi garde le pixel qu'elle transporte, et essaye

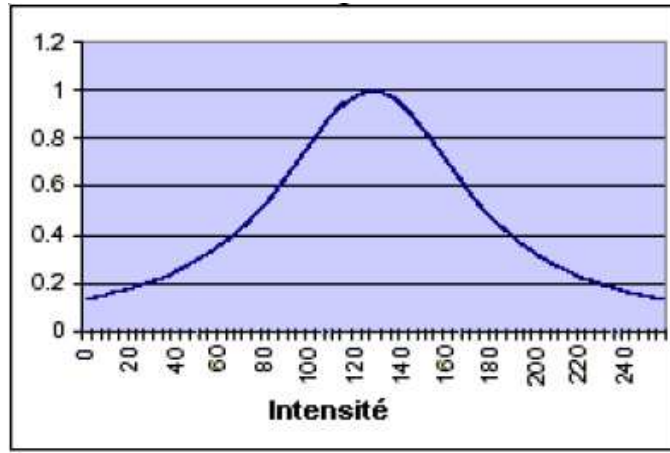


FIGURE 5.4 – la fonction de similarité .  
pour  $ng_i = 128$  et  $g_k \in [0, 255]$ ,  $\beta = 50$

d'autres cases choisies aléatoirement jusqu'à ce qu'elle arrive à le déposer.

Quand une fourmi recherche un pixel à transporter, la recherche est effectuée en utilisant un index commun contenant les pixels libres, le tri de l'index est fait par ordre croissant en fonction de la distance entre le niveau de gris du pixel et le centre de gravité de la classe où il se trouve.

#### Ramassage d'un pixel :

La probabilité de transporter un pixel  $p_i$  de sa case  $c_k$  est définie par la formule suivante :

$$p_{porter}(p_i, c_k) = \begin{cases} 1 & \text{si } |c_k| = 1 \\ q & \text{si } |c_k| = 2 \\ \frac{k_p}{k_p + f(p_i, g_k)} & \text{sinon} \end{cases} \quad (5.2)$$

#### Dépôt du pixel :

Si une fourmi transporte un pixel, elle explore son voisinage immédiat pour choisir la case vers laquelle elle se déplacera pour y déposer avec une probabilité donnée par la formule suivante :

$$p_{depot}(p_i, c_k) = \begin{cases} 1 & \text{si } f(p_i, g_k) \leq f(p_{dissim}, g_k) \\ \frac{f(p_i, g_k)}{f(p_i, g_k) + k_d} & \text{sinon} \end{cases} \quad (5.3)$$

Avec

$$f(p_{dissim}, g_k) = \min_{p_i \in k} (f(p_i, g_k))$$

Ainsi la performance de AntClust dépend d'un certain nombre de paramètres dont les valeurs peuvent dépendre ou non de l'image à segmenter :

$A$  : le nombre de fourmis

$k_d$  : contrôle la probabilité de déplacer un pixel d'une case,

$k_p$  : contrôle la probabilité de déposer un pixel dans une case,

$q$  : contrôle la probabilité de choisir un pixel dans une case contenant deux pixels.

Pour les valeurs 20, 0.3, 0.1, 0.7 correspondant respectivement aux paramètres cités la dessus, AntClust a obtenu de bons resultats sur des images de test [125].

---

**Table 2** L'algorithme AntClust

---

1. Placer les  $N$  pixels de l'image chacun dans une case du tableau,
  2. Initialiser aléatoirement les positions des  $K$  fourmis,
  3. **Pour** chaque fourmi, ramasser le pixel de la case où elle se trouve,
  4. Pendant  $T_{max}$  itérations faire
    - 4.1. **Pour** chaque fourmi  $a_i$  faire
      - 4.1.1. Revenir au nid,
      - 4.1.2. Choisir la prochaine case  $c_k$  vers laquelle elle se déplacera et y déposer le pixel  $p_i$  qu'elle transporte avec une probabilité  $p_{depot}(p_i, c_k)$
      - 4.1.3. Revenir au nid,
      - 4.1.4. Se déplacer vers la case contenant le prochain pixel  $p_i$  à transporter et le ramasser avec une probabilité  $p_{porter}(p_i, c_k)$
- 

Au cours du processus de classification, aucune nouvelle classe n'est créée mais une classe peut disparaître si la case qui correspond du tableau se vide de ses pixels. A la fin du processus de classification, le nombre de classes intéressantes de l'image correspond au nombre de cases non vides présentes sur le tableau. Le schéma général de l'algorithme AntClust est présenté dans 2.

Les résultats montrent que l'algorithme AntClust bien qu'il ne nécessite pas une connaissance du nombre probable de classes, arrive à identifier un nombre égal (ou très proche) du nombre correct de classes avec un écart type presque nul. D'un autre coté, la qualité de la classification générée par l'algorithme AntClust est très proche de la qualité optimale comme l'atteste les valeurs de l'index de Rand avec une variance intra-classe réduite ce qui traduit une bonne homogénéité à l'intérieur des classes.

Même si AntClust contient moins de paramètres par rapport aux autres algorithmes basés sur les fourmis, il reste cependant lents à l'exécution. De ce fait on propose de lancer AntClust avec un nombre limité d'itérations, pour que dans un deuxième temps intervient le FCM pour corriger les éventuelles erreurs que Antclust n'a pas corrigé.

**Choix des paramètres** On a repris les mêmes valeurs des paramètres pris dans [125] c-à-d :

	AntClust			10_FCM			100_FCM		
	$K$ ( $\sigma_K$ )	$R$ ( $\sigma_R$ )	$V$ ( $\sigma_V$ )	$K$ ( $\sigma_K$ )	$R$ ( $\sigma_R$ )	$V$ ( $\sigma_V$ )	$K$ ( $\sigma_K$ )	$R$ ( $\sigma_R$ )	$V$ ( $\sigma_V$ )
Image Synthétique 1	4 (0)	0.98 (0.004)	0.125 (0.006)	4 (0)	0.92 (0.125)	0.138 (0.006)	7 (0)	0.65 (0.125)	0.85 (0.152)
Image Synthétique 2	3 (0)	0.94 (0.125)	0.198 (0.125)	10 (0.01)	0.57 (0.07)	0.51 (0.11)	46 (0)	0.41 (0.145)	0.89 (0.125)
Image Synthétique 3	7 (0.05)	0.85 (0.02)	0.126 (0.035)	7 (0.23)	0.86 (0.008)	0.133 (0.06)	7 (0.5)	0.86 (0.062)	0.135 (0.03)

FIGURE 5.5 – Résultats de la classification des images par FCM et AntClust.

Des images synthétiques et réelles ont été utilisées,  $K$  correspond au nombre moyen de classes identifiées par les deux algorithmes,  $R$  à la valeur moyenne de l'index de Rand et  $V$  à la valeur moyenne de la variance intra classe,  $\sigma_K, \sigma_R, \sigma_V$  et représentent quant à eux aux écarts types respectifs.

- le nombre de fourmis  $A = 20$ .
- la probabilité contrôlant de déplacement d'un pixel d'une case  $K_d = 0.3$ ;
- la probabilité contrôlant de dépôt d'un pixel dans une case  $K_p = 0.1$ .
- $Q = 0.7$  pour contrôler la probabilité de choisir un pixel dans une case contenant deux pixels.
- Le paramètre  $m$  taille de la mémoire sera fixé ici à 2 pour que la mémoire de la fourmi ne stocke pas des déplacements dépassés.

**L'algorithme FCM** Fuzzy C-Means (FCM) est un algorithme de classification non-supervisée floue. Issu de l'algorithme des C-moyennes (C-means), il introduit la notion d'ensemble flou dans la définition des classes : chaque point dans l'ensemble des données appartient à chaque cluster avec un certain degré, et tous les clusters sont caractérisés par leur centre de gravité. Comme les autres algorithmes de classification non supervisée, il utilise un critère de minimisation des distances intra-classe et de maximisation des distances inter-classe, mais en donnant un certain degré d'appartenance à chaque classe pour chaque pixel. Cet algorithme nécessite la connaissance préalable du nombre de clusters et génère les classes par un processus itératif en minimisant une fonction objective. Ainsi, il permet d'obtenir une partition floue de l'image en donnant à chaque pixel un degré d'appartenance (compris entre 0 et 1) à une classe donnée. Le cluster auquel est associé un pixel est celui dont le degré d'appartenance sera le plus élevé. Les principales étapes

de l'algorithme Fuzzy C-means sont :

1. La fixation arbitraire d'une matrice d'appartenance.
2. Le calcul des centroïdes des classes.
3. Le réajustement de la matrice d'appartenance suivant la position des centroïdes.
4. Calcul du critère de minimisation et retour à l'étape 2 s'il y a non convergence de critère.

---

**Table 3** L'algorithme des C-moyennes floues

---

**Entrée :**  $\chi = \{x_1, \dots, x_n\}$  un ensemble de points,  $c$  le nombre de classes recherchées,  $v^{(0)}$  l'ensemble de prototypes initiaux,  $d$  une métrique,  $m$  le coefficient de flou,  $\varepsilon$  seuil de stabilité de la partition.

**Sortie :** Une partition floue  $U$  de  $\chi$

$t \leftarrow 0$

**repeat**

{Etape 1 : Mise à jour de la partition floue}

**for**  $i = 1$  to  $n$  **do**

**for**  $k = 1$  to  $c$  **do**

$$u_{ik}^{(t)} = \left[ \left( \sum_{l=1}^c \left( \frac{d_{il}}{d_{ik}} \right)^{\frac{2}{m-1}} \right)^{-1}$$

**end for**

**end for** {Etape 2 : Mise à jour des prototypes des classes}

**for**  $k = 1$  to  $c$  **do**

$$v_k^{(t+1)} = \frac{\sum_{i=1}^N (u_{ik}^{(t)})^m x_i}{\sum_{i=1}^N (u_{ik}^{(t)})^m}$$

**end for**

$t \leftarrow t + 1;$

**until**  $\|v^{(t+1)} - v^{(t)}\| < \varepsilon;$

---

### Choix des paramètres

- **Choix du paramètre  $m$  :** Pratiquement,  $m$  est choisi dans l'intervalle  $[1.5, 2.5]$ . Sa présence assure le réglage de la force du flou. Plus  $m$  est élevé et moins les individus ayant une forte appartenance à une classe particulière sont discriminés par rapport à ceux qui sont partagés entre plusieurs classes [24]. Une valeur comprise dans l'intervalle  $[1.5, 3]$  est généralement acceptée et Tucker suggère dans son théorème [169] de prendre  $m \geq N/(N-2)$  afin d'assurer la convergence de l'algorithme. D'après une étude effectuée par Barra dans [165], il semble que toutes les valeurs appartenant à l'intervalle  $[1.5, 3]$  conviennent, de ce fait dans le cadre de cette étude on prendra  $m = 2$ .

- **Choix de la distance** : Notre choix s'est porté sur la distance euclidienne qui forme la distance la plus usuelle et la plus rapide à calculer.
- **Le choix des vecteurs forme** : est fondamental puisque leur pertinence va permettre de discriminer les voxels entre eux. Ce choix est défini suivant le type de modalité. L'image anatomique que nous utilisons est une IRM. L'imagerie par résonance magnétique est une modalité d'imagerie multispectrale 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 T1, T2 et en densité de protons, des paramètres caractéristiques (variance sur les niveaux de gris, variance sur les gradients) peuvent aussi être calculées pour classer les textures.

Cependant, l'augmentation du nombre de caractéristiques n'améliore pas systématiquement la qualité de l'apprentissage. En effet, on est confronté au phénomène de Hughes appelé également "the curse of dimensionality" ou "peaking phenomenon" [136]. Selon [154] parmi les arguments existants en faveur de la réduction de la dimension on cite :

- La complexité des algorithmes augmente lorsque la dimension croit.
- Certains attributs sont séparément pertinents, mais le gain est faible lorsqu'ils sont combinés.
- Lorsque la dimension augmente, le nombre de paramètres libres du modèle augmente. Pour une taille des données fixée, le nombre de points impliqués dans ces estimations est insuffisant et la précision de l'estimateur décroît. Il en résulte une plus faible capacité en généralisation.

Pour toutes les raisons citées et suivant l'objectif de cette étude, on se contentera de l'utilisation du contraste d'une pondération à la fois, à savoir T1, T2, et densité de proton.

**FCM et les données aberrantes** Les données aberrantes sont un problème récurrent en classification, pour tout processus d'analyse de données. Elles constituent une difficulté majeure en classification et en particulier pour les méthodes itératives de recherche de partition telle FCM 5.6. Les poids des points aberrants peuvent être maximaux pour au moins une des classes. Ils comptent ainsi autant que les données "propres" issues de la classe dans l'estimation. Par conséquent, les centres des classes seront attirés par ces données aberrantes. Dans la conception d'un système robuste, on doit donc chercher à limiter l'influence de ces points [154]. Plusieurs approches ont été présentées dans la littérature ; par exemple RCA, une méthode de classification robuste par compétition proposée par Frigui

et al [42], qui fait diminuer explicitement l'influence des points aberrants par l'utilisation d'un M-estimateur, une autre approche FNC proposée par Davé dans [137] consiste à créer une classe de bruit et à estimer le degré d'appartenance à cette classe. Dans [43], Mac Lachlan et Peel déterminent les points aberrants par un seuillage d'une valeur  $\alpha$ , à l'inverse des deux premières méthodes, la robustesse est désormais incluse directement dans la modélisation théorique des classes.

Dans le cadre de cette étude le traitement des points aberrant se fera par la fusion des degrés d'appartenance issus de la segmentation de l'image et de la donnée issue du dossier médical, tentant ainsi de diminuer l'effet de ces données.

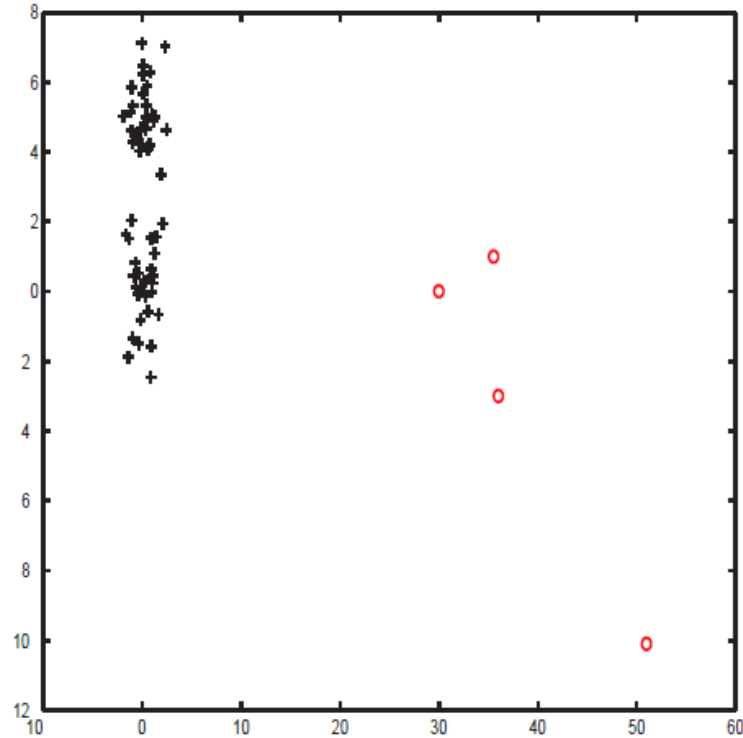


FIGURE 5.6 – FCM : Sensibilité aux données aberrantes (2 classes demandées).  
[154]

### 5.2.2.3 Informations supplémentaires

La combinaison des informations est souvent guidée ou contrainte par des informations supplémentaires, sur l'information à combiner, sur le contexte et le domaine d'application). C'est aussi une source de forte hétérogénéité. Parmi les exemples d'informations supplémentaires sur l'information on peut citer la fiabilité d'une source, soit globale, soit condi-

tionnelle aux objets observés. Ce cas est très souvent rencontré en classification d'images multi-sources, où une image peut être fiable pour une classe mais pas pour une autre. Donnons maintenant quelques exemples d'informations supplémentaires sur le domaine et le contexte [71] :

- le LCR est sombre dans les images IRM en T1,
- Les lésions apparaissent généralement hypointenses en T2, entourées d'un oedème hyperintense<sup>1</sup>.

#### 5.2.2.4 Comment envisage-t-on la fusion ?

La fusion envisagée est une fusion entre deux types de données hétérogènes, d'un coté l'image IRM et de l'autre une donnée issue de dossier médical (Âge par exemple), ces deux sources sont considérées complémentaires, car en fusionnant on cherche à apporter des informations supplémentaires sur l'existence éventuelle d'une pathologie, pour cette raison la fusion envisagée sera appliquée à l'ensemble de l'image.

#### 5.2.3 Niveau de fusion ?

La fusion envisagée porte sur le niveau données elle concerne l'agrégation entre une image de modalité IRM et une information issue du dossier médical, la fusion s'opère donc sur des données hétérogènes et plutôt complémentaires.

#### 5.2.4 Cadre théorique de fusion ?

De nombreuses théories ont été proposées pour représenter et gérer l'incertain et l'imprécis dans les informations, notamment les théories de l'incertain telles que les approches bayésiennes, la théorie des possibilités et des sous-ensembles flous et la théorie des fonctions de croyance. 4.2.

Si la théorie probabiliste repose sur une base mathématique solide, et propose un éventail d'outils très riche permettant aussi bien la modélisation, elle présente cependant un inconvénient qui la discrimine de nos choix dans le cadre de ce mémoire, car même si elle représente bien l'incertain qui entache l'information, elle ne permet pas aisément de

---

1. TUMEURS INTRACRANIENNES (147) [www.univ-lyon1.fr/servlet/](http://www.univ-lyon1.fr/servlet/)

représenter son imprécision, conduisant ainsi souvent à confondre ces deux notions [67]. La théorie des ensembles flous quant à elle s'avère plus intéressante mais reste inadapté dans le cadre de cette étude, car elle représente essentiellement le caractère imprécis des informations, l'incertitude étant représentée de manière implicite et n'étant accessible que par déduction à partir des différentes fonctions d'appartenance [67].

Il reste à choisir entre la théorie possibiliste et évidentielle, notre choix c'est porté sur la théorie possibiliste pour les raisons suivantes :

- 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, [69] ou l'incomplétude de l'information [5]). Cependant dans le cadre de cette étude on en aura pas l'opportunité d'utiliser des hypothèses composées, puisque les informations que nous extrairons des images pourront directement être modélisées par des hypothèses simples ou par des distributions de possibilité ;
- Même si la théorie des croyances laisse une place centrale à l'ignorance et permet une modélisation plus riche, elle demeure rigide quant à la combinaison des informations [67]. En effet, l'étape de combinaison en théorie des croyances se résume la plupart du temps à l'application de l'opérateur orthogonal de Dempster (*cf. 4.3.6.3.1*) 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) [165].

### 5.2.5 Opérateur de fusion ?

La notion de fusion s'applique au cas de  $n$  sources [68] [39]. 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. 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).

Deux modes de combinaison fondamentaux sont utilisables en théorie des possibilités :

le mode conjonctif (intersections généralisées) et le mode disjonctif (réunions généralisées). Le mode conjonctif utilise comme opérateurs des normes triangulaires (t-normes, par exemple : minimum, produit etc.) et s'applique dans le cas où les sources sont considérées comme toutes fiables. Le mode disjonctif est basé sur les opérateurs de t-conorme (maximum, somme probabiliste etc.) et correspond aux cas où seulement certaines sources (non identifiées) sont fiables. Le choix du mode de combinaison dépend donc du degré de cohérence entre les sources. Une règle de combinaison adaptative permet alors de passer continûment du mode conjonctif au mode disjonctif en fonction de ce dernier.

Bloch propose dans [68] la définition suivante pour le comportement de  $F$  :

- sévère si  $F(n1, n2) \leq \min(n1, n2)$ ,
- prudent si  $\min(n1, n2) \leq F(n1, n2) \leq \max(n1, n2)$ ,
- indulgent si  $F(n1, n2) \geq \max(n1, n2)$ .

En fusion d'informations multi-sources, face à des informations redondantes, deux attitudes sont possibles : soit ignorer les redondances, ce qui est approprié quand les sources peuvent ne pas être indépendantes, ou bien voir dans la redondance une raison de renforcement et de confirmation des informations.

Différents opérateurs de la littérature ont été présentés (*cf.4.3.5*). Nous avons choisi d'utiliser un opérateur à comportement constant et indépendant du contexte, notre choix s'est porté sur la moyenne, qui forme un opérateur prudent agissant comme un compromis entre les deux sources. Les deux sources (image et la donnée issue du dossier médical) présentent toutes les deux des sources fiables mais à des degrés différents. En effet, d'après les réponses données par les spécialistes, on donne plus d'importance aux informations issues d'une image IRM qu'à la donnée issue du dossier médical. Ainsi un degré de fiabilité (une pondération) est attribué à chacune des deux sources (image, donnée), soit respectivement  $t_1$  et  $t_2$ . Tel que

$$t_1 + t_2 = 1$$

et

$$t_i \in [0, 1]$$

Après l'introduction d'une pondération, on ne parle plus de moyenne, mais plutôt de moyenne pondérée appelée aussi *OWA* (Ordered Weighted Average) [144]. Ce dernier opérateur est considéré comme l'un des plus directs pour effectuer la fusion d'images [28] voir figures 5.7 et 5.8), et il est souvent sélectionné dans le cas où les contributions des différentes sources ne sont pas de même importance vis à vis de l'application ce qui est

le cas de cette étude. Dans les *OWA* les poids  $t_i$  sont définis par le rang des valeurs à

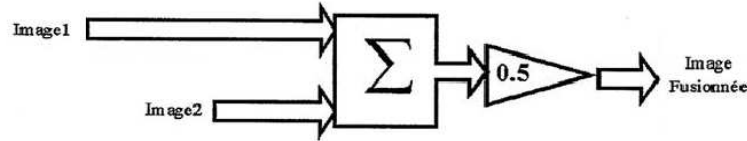


FIGURE 5.7 – Fusion par moyenne pondérée, avec des pondérations égales pour les deux images [28].

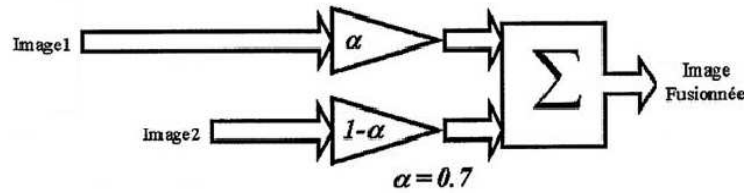


FIGURE 5.8 – Fusion par moyenne pondérée [28]

combiner.

Soient  $s_1, s_2, \dots, s_n$  ses valeurs. Elles sont ordonnées en une suite  $s_{j_1}, s_{j_2}, \dots, s_{j_n}$  telle que :

$$s_{j_1} \leq s_{j_2} \leq \dots \leq s_{j_n}$$

Alors, pour un ensemble de poids  $t_i$  vérifiant :

$$\sum_{i=1}^n t_i = 1, \forall i \leq i \leq n, t_i \in [0, 1],$$

L'opérateur *OWA* est défini par l'expression :

$$OWA(s_1, s_2, \dots, s_n) = \sum_{i=1}^n t_i s_{j_i} \quad (5.4)$$

### 5.2.5.1 Opérateur proposé

Le nouvel opérateur proposé *DOWA*, est en fait un opérateur *OWA* mais dont les pondérations sont dynamiques c'est-à-dire, le poids affecté à une source change en fonction de la valeur de l'autre source (son degré d'appartenance).

### 5.2.5.2 Pourquoi avoir choisi des pondérations dynamiques ?

Reprenant l'exemple cité dans (cf.5.2.2.1.1), le risque d'être atteint de la maladie d'Alzheimer peut être élevé, moyen, ou faible, selon le degré d'appartenance de la donnée aux 3 classes c-à-d  $\mu_{faible}$ ,  $\mu_{moyen}$ ,  $\mu_{eleve}$ .

En parallèle une classification de l'image IRM en cinq (ou six) classes est effectuée, où chaque voxel aura un degré d'appartenance  $\mu_{LCR}$ ,  $\mu_{MG}$ ,  $\mu_{MB}$ ,  $\mu_{TUMEUR}$ ,  $\mu_{FOND}$ , respectivement aux classes : *MB*, *MG*, *LCR*, *TUMEUR*, *FOND*.

**L'opérateur proposé DOWA s'écrit :**

$$DOWA(Image, Donne) = (P\mu_{pathologie} + T\mu_{donne}) \quad (5.5)$$

tel que  $P$  et  $T$  sont les pondérations attribuées respectivement à l'image et à la données. respectant la contrainte suivante :  $P = 1 - T$  leurs valeurs changent selon la valeur de  $\mu_{donnee}$  générant ainsi des pondérations dynamiques comme c'est illustré ci-dessous :

l'objectif de la fusion ici est de confirmer ou d'affirmer l'existence d'une éventuelle patho-

---

```

if  $\max(\mu_{faible}, \mu_{moyen}, \mu_{eleve}) = \mu_{faible}$  then
  attribuer à la donnée la pondération la plus faible soit  $t_1$ 
else
  if  $\max(\mu_{faible}, \mu_{moyen}, \mu_{eleve}) = \mu_{moyen}$  then
    attribuer à la donnée une pondération moyenne soit  $t_2$ 
  else
    attribuer à la donnée la pondération la plus forte soit  $t_3$ 
  end if
end if

```

---

logie qui avant la fusion n'apparaît pas clairement sur l'image IRM, mais sera plus clair après la fusion, et cela grâce l'introduction des informations issues du dossier médical.

### 5.2.6 Critère de décision

L'étape finale du processus de fusion, a pour objectif la prise de décision quant à l'appartenance d'un voxel à une classe. Les deux critères les plus couramment utilisés sont :

- le critère de maximum : la solution retenue est celle pour laquelle la distribution fusionnée atteint son maximum de degré de possibilité,
- le critère de seuil, qui consiste à choisir un seuil de degré de possibilité, et les valeurs

du paramètre pour lesquelles le degré de possibilité  $y$  est supérieur sont les solutions. Le choix du critère dépend de ce que l'on attend : une solution unique ou une plage de solutions. L'étendue de la plage solution peut par ailleurs être considéré comme un indicateur de la précision et/ou de la fiabilité du résultat obtenu [134].

La visualisation de la distribution fusionnée permet d'observer facilement la plage solution lorsqu'on applique le critère de seuil, pour cette raison notre choix s'est porté sur ce dernier critère.

$$\forall v \in I, v \in C_{pathologie} \text{ Si } \pi_{fusionne}(v) \geq \text{seuil} \quad (5.6)$$

La décision comme l'opérateur choisi est à comportement prudent, ce qui nous guidera dans le choix du seuil.

## 5.2.7 Architecture de fusion

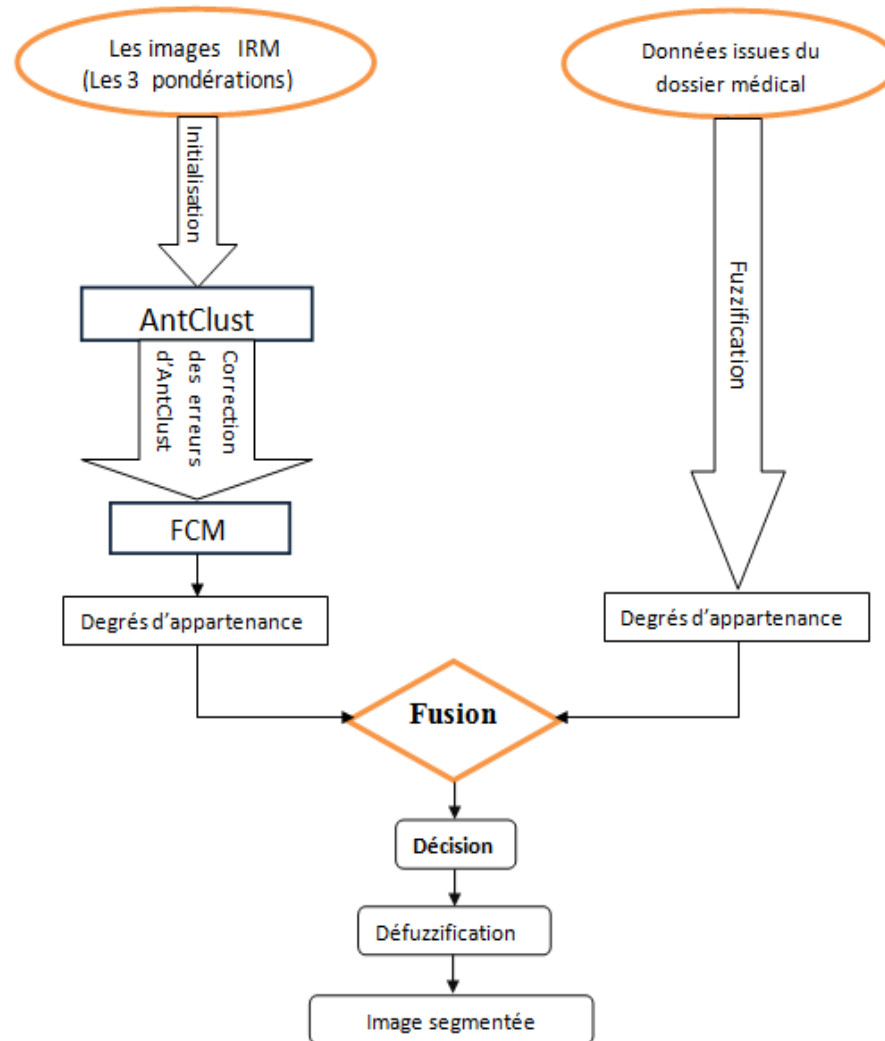


FIGURE 5.9 – Architecture de la fusion proposée.

## 5.2.8 Présentation de l'algorithme général

**Table 4** Algorithme général

<i>1. Fuzzification de la donnée issue de dossier médical</i>
<i>2. Modélisation de l'image</i>
AntClust {initialisation}
FCM {corriger les erreurs de AntClust}
<i>3. Fusion</i>
Fusion Possibiliste {entre image et donnée du dossier médical}
Prise de décision

## 5.3 Conclusion

Dans ce chapitre, nous avons illustré l'approche proposée ; il s'agit de la fusion de données dans le cadre particulier de la combinaison entre deux données de types hétérogènes (image et donnée numérique ou linguistique) . Après avoir explicité la modélisation des ces dernières, le choix de la théorie des ensembles flous a été retenue pour la représentation de la donnée numérique issue du dossier médical ; quand à la représentation de l'image, une hybridation entre l'algorithme AntClust basé sur une population d'agents fourmis et l'algorithme flou FCM a été proposée, AntClust servira de phase d'initialisation pour le FCM.

Bien qu'efficace dans de nombreux cas, le FCM possède les défauts caractéristiques des algorithmes de classification :

- choix généralement non trivial du nombre de classes,
- sensibilité aux données aberrantes.

Le premier défaut étant traité par l'introduction d'AntClust comme phase d'initialisation, le deuxième défaut quand a lui sera traité, par l'introduction d'une phase de fusion entre les deux données préalablement modélisées, palliant ainsi au problème des points aberrants et aux points mal classés dont souffre le FCM.

Après la phase de modélisation, nous avons précisé le cadre théorique de la fusion retenu, nous avons opté pour la théorie possibiliste donnant une panoplie d'opérateurs, l'opérateur prudent à comportement constant et indépendant du contexte "*moyenne pondérée OWA*" à été retenu, cependant avec une modification en fonction des données disponibles, pour enfin proposer l'opérateur *DOWA*. Pour terminer nous avons introduit la règle de décision choisie , permettant d'exploiter les résultats de la fusion.

---

# RÉSULTATS ET VALIDATION

---

## 6.1 Introduction

Dans ce chapitre, nous appliquons les méthodes proposées dans les chapitres précédents et nous les évaluons quantitativement et qualitativement à partir des différents critères d'évaluation proposés dans la littérature. Le choix des données participant à la fusion étant la première étape de ce chapitre, suivi des méthodes proposées appliquées sur ces dernières, et en fin nous présentons les résultats obtenus par le système développé, qui sont évalués sur des images réelles et simulées (mises à disposition sur le site internet brainweb<sup>1</sup>).

## 6.2 Choix des données biomédicales

La validation de notre méthode appliquée sur des données biomédicales de type hétérogène, s'avère une tâche difficile, la difficulté majeure réside en leur acquisition. En effet pour rassembler des informations se rapportant à un même patient, on se trouve souvent confronté à beaucoup de difficulté, la principale étant souvent liée à l'aspect confidentiel caractérisant ces données. La deuxième difficulté réside dans le choix des données à combiner, en effet on se trouve confronter à une panoplie de données se rapportant toutes à un même patient, théoriquement la prise en compte de toutes ces données serai la meilleure façon pour ressortir avec la meilleure décision, Paradoxalement en pratique, l'augmen-

---

1. <http://www.bic.mni.mcgill.ca/brainweb/>

tation du nombre de données n'améliore pas systématiquement la qualité de la décision (phénomène de Hughes appelé également "the curse of dimensionality").

Ainsi pour diminuer la dimension des données à combiner, en d'autres termes ne prendre que les données pertinentes, on s'est basé sur ce qui a été apporté par la littérature d'un côté, et de l'autre par les réponses données par quelque spécialistes, pour au final faire les choix ci-dessous :

- L'imagerie médicale en particulier la modalité IRM plus particulièrement l'IRM cérébrale, joue un rôle incontournable dans l'aide au diagnostic des différentes pathologies du cerveau. Pour cette raison elle a été choisie comme première donnée participant à la fusion.
- La deuxième donnée choisie est soit l'âge ou le sexe du patient, sachant que ces deux données sont dans un grand nombre de cas en lien direct avec la pathologie. En voici quelques exemples tirés de<sup>2</sup>qui illustre cette relation :
  - l'incidence des métastases cérébrales de cancers systémiques augmente avec l'âge ;
  - une macrocrânie est toujours présente dans les hydrocéphalies chroniques se révèle avant l'âge de deux ans ;
  - L'âge moyen de survenue des gliomes de bas grade (astrocytomes pilocytiques, astrocytomes diffus, oligodendrogliomes et oligoastrocytomes) est de 35 à 45 ans.
  - L'âge moyen de survenue est de 40 ans pour les astrocytomes anaplasiques et 53 ans pour les glioblastomes.
  - Les méningiomes sont plus fréquents chez la femme (sexe ratio 2/1) et leur incidence augmente avec l'âge.
  - Le lymphome d'un sujet immunocompétent, peut survenir à tout âge, mais avec un pic de fréquence autour de 60 ans, alors que l'âge moyen de survenue chez l'immunodéprimé se situe vers 30 ans.
  - les abcès du cerveau semble toucher 2 fois plus le sexe masculin que le sexe féminin et l'âge d'exposition se situe entre 30 et 45 ans.
  - 25 % environ des abcès du cerveau se rencontrent chez les enfants de moins de 15 ans (cette pathologie est exceptionnelle chez l'enfant de moins de 2 ans)<sup>3</sup> l'âge de début de la sclérose en plaques (SEP) est entre 20 à 40 ans, (moyenne 30 ans) et avec une prépondérance féminine (sex-ratio 1,7/1)<sup>4</sup>.

---

2. TUMEURS INTRACRANIENNES (147) [www.univ-lyon1.fr/servlet/](http://www.univ-lyon1.fr/servlet/)

3. <http://www.vulgaris-medical.com/encyclopedie/abces-du-cerveau-14.html>

4. <http://www.sclerose-en-plaques.apf.asso.fr/>

## 6.3 Description des volumes traités

### 6.3.1 Fantôme du cerveau

Pour valider la méthode développée, nous nous basons sur la base de données Brainweb<sup>5</sup>. 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. Ces simulations sont générées à partir d'une vérité terrain qui correspond à la classification d'une image cérébrale (de résolution millimétrique) en différentes classes correspondant aux différents tissus et structures.

La construction du fantôme est fondée sur un ensemble de 27 volumes IRM de taille  $181 \times 127 \times 181$  voxels par volume et de haute résolution ( $1\text{mm}^3/\text{voxel}$ ). Ces images ont été recalées et un volume IRM a été créé par moyennage de ces 27 volumes recalés. 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. Après corrections manuelles des résultats obtenus, les cartes floues de tissu sont finalement construites.

### 6.3.2 Images simulées

A partir des cartes floues de tissu du fantôme, des images IRM peuvent être simulées en différentes pondérations, telles que T1, T2. Pour tester nos algorithmes dans des conditions réalistes, nous avons ajouté un bruit blanc gaussien (exprimé comme un pourcentage de son écart type par rapport l'intensité moyenne du signal) aux images obtenues.<sup>3</sup>

### 6.3.3 Images réelles

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

---

5. <http://www.bic.mni.mcgill.ca/brainweb/>

image et signaux) de l'université Ferhat Abbas de Sétif avec une équipe d'un laboratoire français.

## 6.4 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 utilisés dans la littérature [147], [157] :

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

$$SE = \frac{TP}{TP + FN} \quad (6.1)$$

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 :

$$SP = \frac{TN}{TN + FP} \quad (6.2)$$

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.

Généralement, plus un test est sensible, moins il est spécifique. À l'inverse, souvent, plus il est spécifique, moins il sera sensible.

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

$$RE = \frac{TP}{TP + FN + FP} \quad (6.3)$$

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 à 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 :

$$SI = \frac{2.TP}{2.TP + FN + FP} \quad (6.4)$$

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.

Notations employées :

- TP : vrais positifs (true positive);
- FP : faux positifs (false positive);
- TN : vrais négatifs (true negative);
- FN : faux négatifs (false negative).

La complexité temporelle n'est pas considérée comme mesure de qualité de l'algorithme.

## 6.5 Résultats de la classification

Dans ce qui suit, nous présentons les expérimentations que nous avons réalisées sur 10 images, cinq de synthèses issues de BrainWeb et cinq sont des images réelles. Les résultats obtenus sont comparés avec ceux des approches classiques pour évaluer l'approche proposée. Ces résultats sont présentés dans les tableaux 6.1, 6.2, 6.3, 6.4, 6.7 et 6.8 et interprétés dans les graphes 6.3, 6.4, 6.5, 6.6 et 6.7 .

### 6.5.1 Résultats de classification sur des données de synthèse

On remarque sur les images de la figure 6.1 que la tumeur et les trois matières à savoir liquide céphalorachidien, matière blanche et la matière grise apparaissent bien pour les deux algorithmes, cependant la pathologie est mieux délimitée dans les images segmentées par la méthode proposée, particulièrement les deux premières.

Par ailleurs on a effectué une comparaison 6.2 entre les deux algorithmes mais cette fois ci en donnant pour le FCM un nombre de classes éloigné du nombre originel, dans ce cas on a constaté l'incapacité du FCM à extraire le nombre de classes contrairement

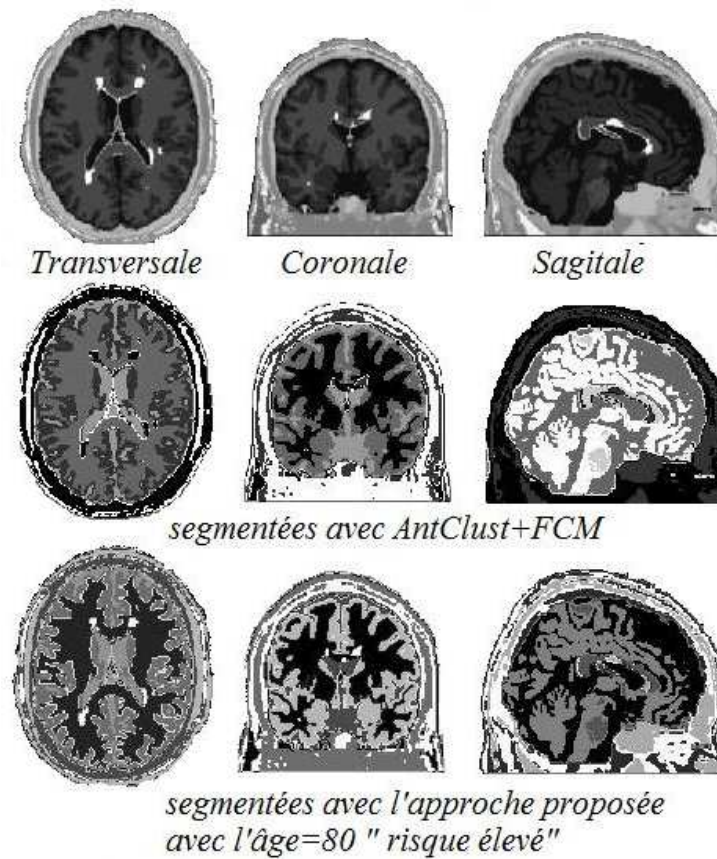


FIGURE 6.1 – Segmentation des images de synthèse avec bruit.

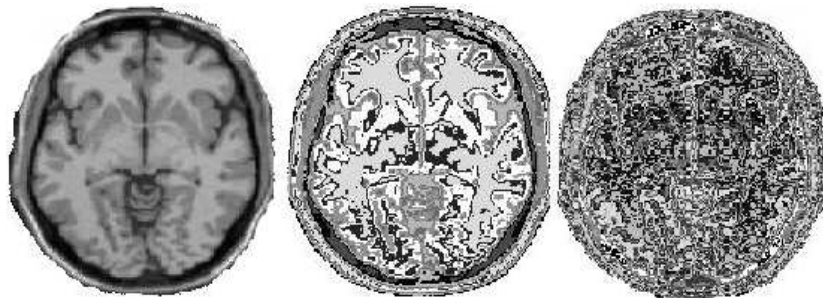


FIGURE 6.2 – Comparaison entre la méthode proposée et AntClust+FCM.

A gauche une image de synthèse  
 au milieu, la segmentation avec la méthode proposée  
 à droite, la segmentation avec FCM avec 30 classes

aux résultats donnés par notre algorithme, bien qu'il ne nécessite pas une connaissance du nombre probable de classe, arrive à identifier un nombre égal (ou très proche) du nombre correct de classes.

Tumeur				
Images	SE	SP	RE	SI
1	0,95	0,95	0,89	0,94
2	0,98	0,96	0,93	0,96
3	0,94	0,94	0,89	0,94
4	0,96	0,93	0,9	0,95
5	0,97	0,97	0,94	0,97
<b>Moyenne</b>	<b>0,96</b>	<b>0,95</b>	<b>0,91</b>	<b>0,95</b>

TABLE 6.1 – Evaluations de la segmentation de la tumeur par le système développé sur 5 coupes successives du volume BrainWeb. avec un âge à risque élevé  $\geq 80$ .

Tumeur				
Images	SE	SP	RE	SI
1	0,93	0,95	0,88	0,93
2	0,93	0,96	0,87	0,93
3	0,92	0,89	0,85	0,92
4	0,93	0,93	0,87	0,93
5	0,94	0,94	0,89	0,94
<b>Moyenne</b>	<b>0,93</b>	<b>0,93</b>	<b>0,86</b>	<b>0,92</b>

TABLE 6.2 – Evaluations de la segmentation de la tumeur par l'approche proposée sur 5 coupes successives du volume BrainWeb. avec un âge à risque moyen par exemple 65ans.

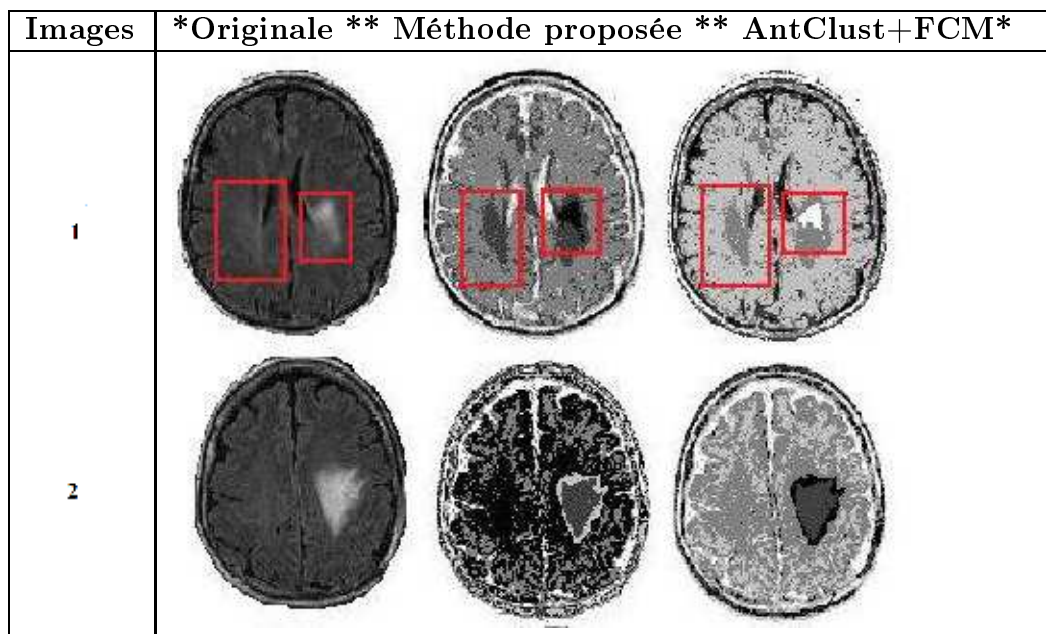
Tumeur				
Images	SE	SP	RE	SI
1	0,95	0,93	0,87	0,93
2	0,93	0,92	0,81	0,9
3	0,94	0,88	0,83	0,91
4	0,96	0,9	0,86	0,93
5	0,94	0,91	0,86	0,93
<b>Moyenne</b>	<b>0,94</b>	<b>0,91</b>	<b>0,85</b>	<b>0,92</b>

TABLE 6.3 – Evaluations de la segmentation de la tumeur par l'approche proposée sur 5 coupes successives du volume BrainWeb. avec un âge à risque faible d'atteinte par exemple 20ans.

Tumeur				
Images	SE	SP	RE	SI
1	0,91	0,88	0,82	0,90
2	0,85	0,87	0,74	0,85
3	0,92	0,89	0,85	0,92
4	0,93	0,93	0,87	0,93
5	0,95	0,91	0,87	0,93
<b>Moyenne</b>	<b>0,91</b>	<b>0,89</b>	<b>0,83</b>	<b>0,91</b>

TABLE 6.4 – Evaluations de la segmentation de la tumeur par AntClust+FCM sur 5 coupes successives du volume BrainWeb.

### 6.5.2 Résultats de classification sur des données réelles




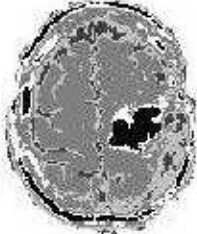

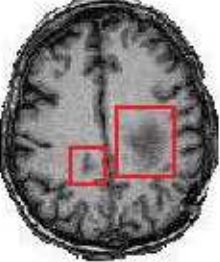

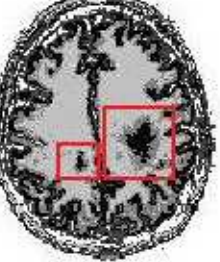
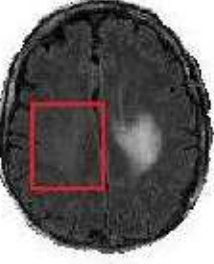
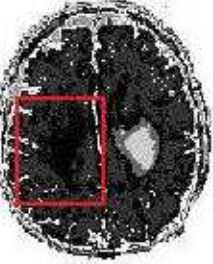
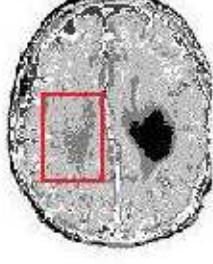
Images	*Originale **	Méthode proposée **	AntClust+FCM*
3			
4			
5			

TABLE 6.5 – Images réelles segmentées par l’approche proposée et par AntClust+FCM(1)

On remarque que la tumeur est détectée pour les deux algorithmes mais la qualité de détection est meilleure avec l’algorithme proposé comme c’est illustré dans les images du tableau 6.5 particulièrement dans les image 1,4 et 5 et les quatre images du tableau 6.6 .


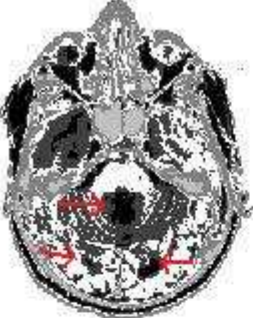
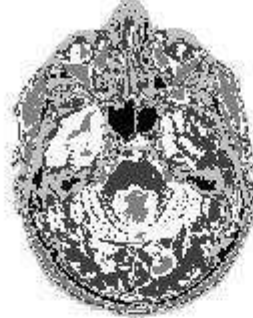
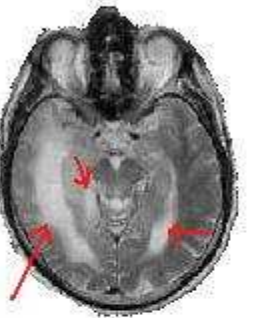


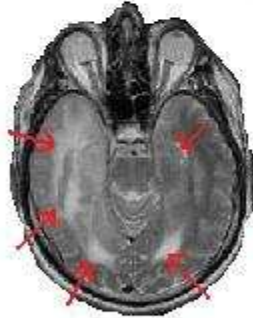
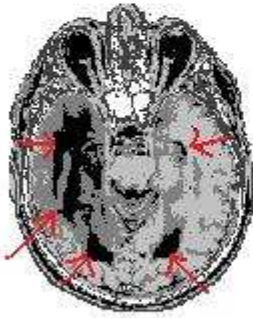

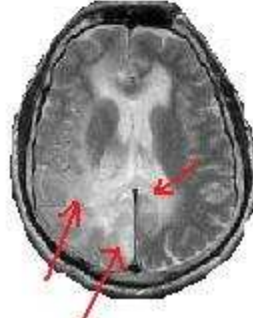


Images	*Originale ****	Méthode proposée ***	AntClust+FCM *
1			
2			
3			
4			

TABLE 6.6 – Images réelles segmentées par l’approche proposée et par AntClust+FCM(2)

### 6.5.3 Comparaison des résultats obtenus

Les tableaux de résultats et les graphes comparatifs illustrent les points suivants :

– les résultats obtenus après le calcul des différents critères d’évaluation à savoir la

Tumeur				
Images	SE	SP	RE	SI
1	0,97	0,92	0,93	0,96
2	0,93	0,94	0,84	0,92
3	0,96	0,91	0,9	0,95
4	0,94	0,93	0,89	0,94
5	0,95	0,91	0,88	0,93
<b>Moyenne</b>	<b>0,95</b>	<b>0,92</b>	<b>0,89</b>	<b>0,94</b>

TABLE 6.7 – Evaluation de la segmentation d’images réelles par l’approche proposée. Risque élevé d’atteinte, par exemple l’âge=80.

Tumeur				
Images	SE	SP	RE	SI
1	0,97	0,92	0,92	0,96
2	0,93	0,92	0,81	0,90
3	0,92	0,91	0,86	0,92
4	0,97	0,87	0,86	0,92
5	0,94	0,86	0,82	0,9
<b>Moyenne</b>	<b>0,95</b>	<b>0,87</b>	<b>0,85</b>	<b>0,92</b>

TABLE 6.8 – Evaluation de la segmentation d’images réelles par l’algorithme proposé l’âge du patient est pris égal à 65 ans "risque moyen d’être atteint".

Tumeur				
Images	SE	SP	RE	SI
1	0,97	0,86	0,89	0,94
2	0,93	0,90	0,79	0,88
3	0,92	0,87	0,83	0,91
4	0,91	0,87	0,81	0,9
5	0,94	0,86	0,82	0,9
<b>Moyenne</b>	<b>0,93</b>	<b>0,87</b>	<b>0,83</b>	<b>0,90</b>

TABLE 6.9 – Evaluation de la segmentation d’images réelles par l’algorithme proposé. le risque d’être atteint est faible par exemple l’âge=20 ans.

sensibilité, la spécificité, le recouvrement ainsi que la similarité, donne l’avantage à l’algorithme proposé par rapport à l’algorithme classique FCM hybridé avec Ant-Clust.

- Les résultats montrent l’intérêt de la prise en compte de la donnée "âge" dans le processus de segmentation. En effet le taux de recouvrement est passé de 0.83 dans

Tumeur				
Images	SE	SP	RE	SI
1	0,89	0,94	0,83	0,91
2	0,94	0,89	0,81	0,90
3	0,95	0,80	0,78	0,88
4	0,91	0,93	0,84	0,91
5	0,94	0,83	0,84	0,91
<b>Moyenne</b>	<b>0,93</b>	<b>0,88</b>	<b>0,82</b>	<b>0,90</b>

TABLE 6.10 – Evaluation de la segmentation d’images réelles par AntClust+FCM.

le cas où la donnée âge n’est pas prise en compte à 0,90 quand cette dernière est prise en compte avec risque élevé et 0,86 quand le risque est moyenne, par contre quand le risque est faible les résultats obtenus sont très proches de ceux obtenus par AntClust+FCM.

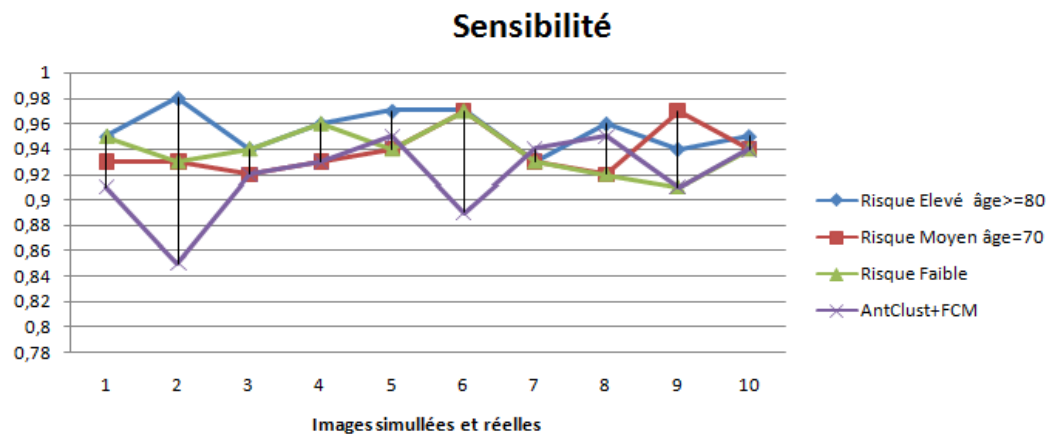


FIGURE 6.3 – La sensibilité du système proposé avec changement de la valeur de l’âge

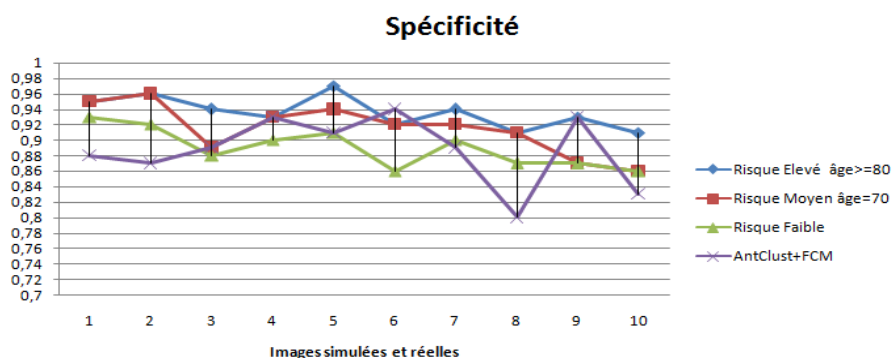


FIGURE 6.4 – La spécificité du système proposé avec changement de la valeur de l’âge

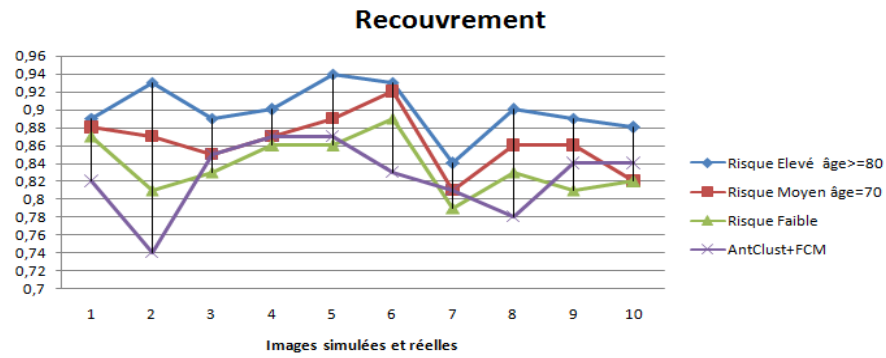


FIGURE 6.5 – Le recouvrement du système proposé avec changement de la valeur de l'âge

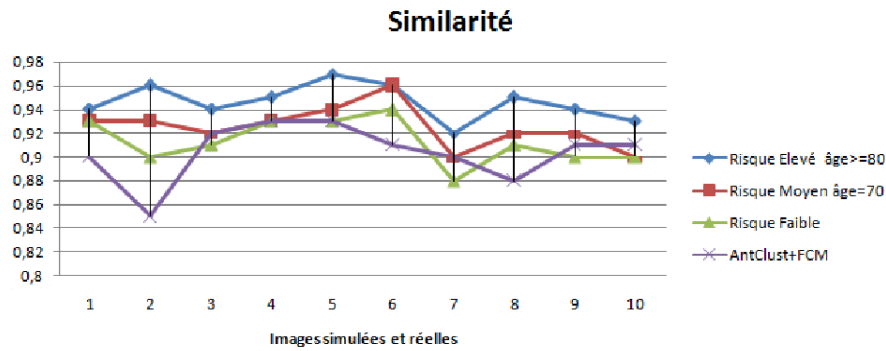


FIGURE 6.6 – La similarité du système proposé avec changement de la valeur de l'âge

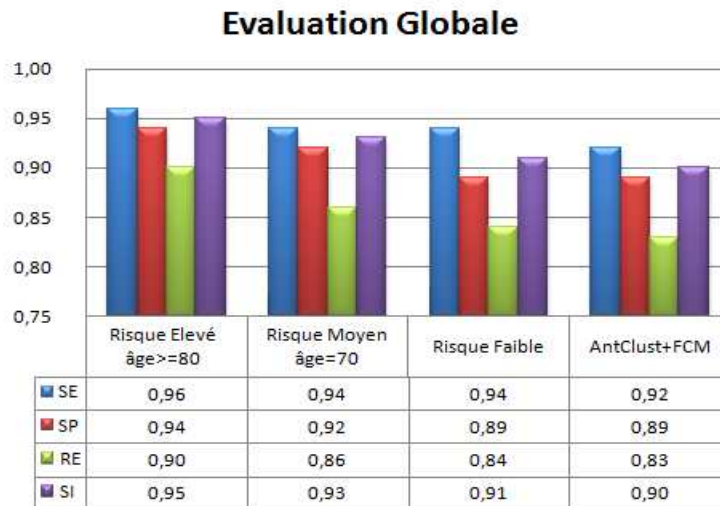


FIGURE 6.7 – Comparaison des 4 critères d'évaluation pour les 3 cas du système proposé avec le AntCust+FCM

## 6.6 Conclusion

Dans cette section, nous nous sommes intéressés à la classification des IRM par l'approche proposée. Dans un premier temps, afin de montrer et de valider l'intérêt et l'efficacité de cette approche, nous avons travaillé sur des données de synthèse puis sur des images réelles. Le traitement lors de cette partie s'est fondé sur le modèle de fusion explicite dans le chapitre précédent. Les résultats obtenus, soulignent le potentiel de la fusion de données, et laisse la porte ouverte à de nombreux autres projets cliniques, (par exemple faire intervenir plusieurs données de différents types dans le processus de fusion).

Cependant notre algorithme présente quelques inconvénients :

- La complexité temporelle ; cependant comme on l'a mentionné plus haut cette dernière n'est pas considérée comme une mesure de qualité dans le cadre de cette étude, car l'objectif principal reste la qualité de segmentation.
- Néanmoins l'inconvénient majeur de notre approche est qu'elle ne permet pas la détection de toutes les tumeurs cérébrales, car ces dernières peuvent survenir à des âges différents et chez les deux sexes confondus et possèdent des caractéristiques de forme, de radiométrie et de localisation très diverses.

# Conclusion

L'EXPLOITATION des données collectées à partir des différents supports est en passe de devenir un enjeu industriel important. En effet, les objets du monde réel ne se présentent pas tous de façon unimodale : un dossier médical, par exemple, comporte certes des données numériques pouvant être structurées de façon tabulaire comme les résultats d'analyses biochimiques, mais incorpore également des données textuelles comme les comptes-rendus d'observations cliniques, des graphiques tels que les tracés d'électrocardiogramme ou d'électro-encéphalogramme, voire des images complexes fournies par les radiographies, échographies, IRM et scanners.

L'objectif de ce travail est d'élargir le champ d'utilisation des méthodes de fouille de données pour qu'elles soient adaptés aux spécificités et aux caractéristiques des données complexes, plus particulièrement aux données biomédicales. Parmi les données médicales nous nous sommes intéressés particulièrement aux images de modalité IRM, qui forme un outil puissant permettant l'observation in vivo de l'anatomie cérébrale et qui est de plus en plus utilisé en routine clinique.

L'incomplétude des informations issues de chaque source est la raison principale qui motive la technique d'analyse d'images IRM basée sur la notion de fusion des données. L'information fournie par chaque source est en général partielle, la fusion d'informations permet en effet d'obtenir des informations de plus grande qualité, et donc améliore la prise de décision médicale. La lecture des IRM en routine clinique étant implicite, sans aucune assistance informatique. Sa rapidité ne témoigne en rien de la facilité de la démarche mais de l'expertise du médecin. En effet, l'interprétation d'un examen IRM est un processus complexe. D'une part, Le médecin compare les images à sa représentation mentale de l'anatomie cérébrale normale. D'autre part, prend en compte les variations du signal sur les différentes pondérations. Il tient également compte du contexte clinique (âge, sexe, origine ethnique, facteurs de risque, séméiologie clinique, évolution...). En présence de

toutes ces informations il parvient à réaliser ainsi une "fusion mentale" des données de plusieurs séquences d'acquisition.

En présence de données hétérogènes et dans le contexte de la fouille de données complexes nous avons proposé un algorithme qui tente d'imiter la démarche entreprise par le médecin, ce qui revient à combiner des informations de nature hétérogènes, dans notre cas des images IRM et des données tabulaire qui contiendront les données structurées du dossier médical.

Dans ce contexte, et appliqué aux données biomédicales, nous avons proposé une architecture de fusion dans un cadre possibiliste de données de type hétérogène (donnée numérique ou linguistique et une image de modalité IRM).

Pour la modélisation des deux données participant à la fusion, La théorie des ensembles flous a été retenue pour la représentation de la donnée numérique issue du dossier médical ; quand à la représentation de l'image, une hybridation entre l'algorithme AntClust basé sur une population d'agents fourmis et l'algorithme flou FCM a été proposée.

Pour pallier aux défauts du FCM à savoir, l'initialisation et la sensibilité aux données aberrantes, AntClust servira de phase d'initialisation, découvrant ainsi automatiquement le nombre de classes, quand à la sensibilité aux données aberrantes, elle est traitée lors de la phase de fusion avec la donnée issue du dossier médical.

Malgré les difficultés rencontrées lors de la validation de notre méthode sur les données biomédicales (les causes principales sont liées à l'acquisition et la multidimensionnalité des données), les premiers résultats obtenus sont cependant assez encourageants et ouvrent des porte à d'autres perspectives. En effet, les perspectives et améliorations de ce travail sont nombreuses et importantes.

- Dans le domaine de la modélisation, nous nous sommes restreints à la représentation de deux données,
- l'image de modalité IRM qui représente une donnée non structurée avec un degré de complexité considérable,
- et une donnée tabulaire structurée.

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. Il peut par exemple s'agir d'un atlas flou de structures cérébrales, de nouvelles techniques d'acquisition IRM pour la mise en évidence de phénomènes physico-chimiques reliés à la pathologie étudiée, ou encore de contraintes de formes floues de structures pa-

- thologiques (tumeurs par exemple), voire à incorporer des données non structurées telles les données textuelles comme les comptes-rendus d'observations cliniques, des graphiques tels que les tracés d'électrocardiogramme ou d'électro-encéphalogramme.
- Dans ce manuscrit notre application ne s'est pas porté sur un cas clinique précis, hors il serai plus judicieux de l'appliquer à des cas cliniques bien définis.

# ANATOMIE DU CERVEAU

---

## A.1 Introduction

Le système nerveux peut être divisé en deux parties :Le système nerveux central et le système nerveux périphérique.

- Le système nerveux central est l'ensemble constitué par l'encéphale (cerveau, cervelet et tronc cérébral) et la moelle épinière. Il est baigné par le liquide céphalorachidien contenu par les méninges. L'encéphale est protégé sur le plan osseux par la boîte crânienne, alors que la moelle épinière qui traverse le canal rachidien est protégée par la colonne vertébrale. Le cerveau est la partie la plus volumineuse de l'encéphale et comporte les deux hémisphères cérébraux (reliés par leurs commissures et creusés des ventricules latéraux), le thalamus, l'hypothalamus et l'épithalamus.
- Le système nerveux périphérique est constitué des nerfs crâniens et spinaux avec leurs ganglions. Les nerfs périphériques connectent le système nerveux central avec les organes sensoriels, les muscles et les glandes.

## A.2 Système nerveux central

Le système nerveux central est constitué de la moelle épinière logée dans le canal vertébral, et de l'encéphale. Dans ce qui suit nous décrivons l'anatomie de ce dernier, et notamment les composants qui nous intéressent dans le cadre de cette étude.

## A.2.1 L'encéphale

L'encéphale (*figure A.1 A.2*) est constitué du tronc cérébral, du cervelet et du cerveau. L'ensemble flotte dans un coussin protecteur de liquide céphalo-rachidien (LCR).

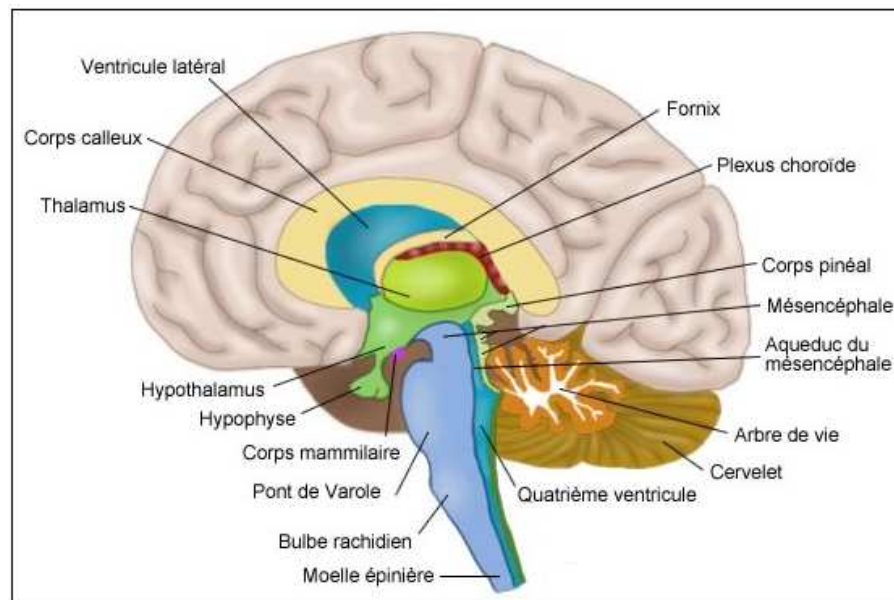


FIGURE A.1 – Encéphale.

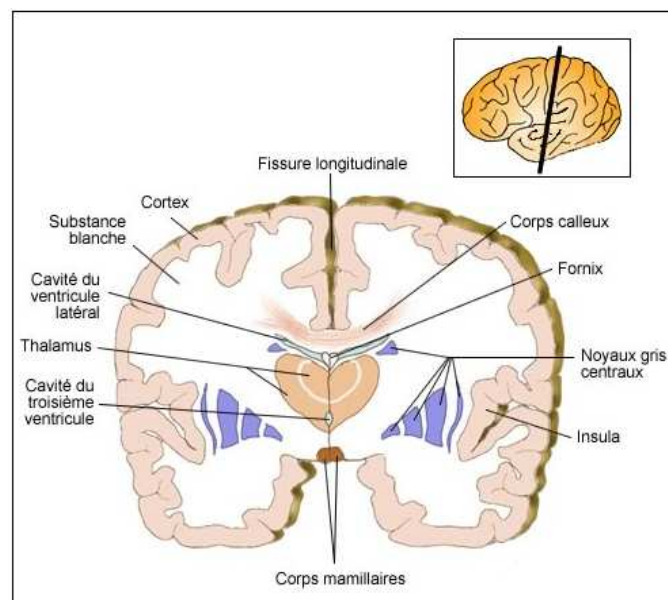


FIGURE A.2 – Coupe transversale de l'encéphale.

### A.2.1.1 Tronc cérébral

Le tronc cérébral (*figure A.3*) est situé dans la fosse postérieure du crâne, sous le cerveau. Il représente un carrefour majeur du système nerveux central entre le cerveau, en haut, le cervelet, en arrière, et la moelle épinière, en bas.

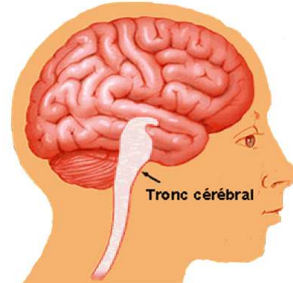


FIGURE A.3 – Tronc cérébral.

### A.2.1.2 Le cervelet

Le cervelet (*figure A.4*) correspondant à la partie de l'encéphale évoquant la forme d'un chou-fleur, située immédiatement sous le cerveau et en arrière du tronc cérébral (extrémité supérieure de la moelle épinière). Le cervelet est situé à l'arrière du crâne et est recouvert par la partie de la boîte crânienne appelée l'occiput, délimitant la fosse postérieure. Cette partie de l'encéphale est recouverte par la tente du cervelet, qui sépare le cerveau des hémisphères cérébraux situés immédiatement au-dessus de lui.

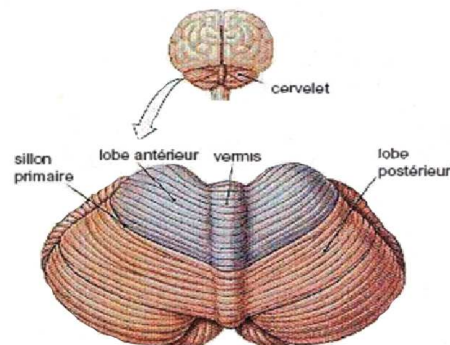


FIGURE A.4 – Cervelet.

### A.2.1.3 Le cerveau

Le cerveau est constitué de deux hémisphères qui sont réunis par le cerveau moyen appelé également diencéphale et le corps calleux. Ces hémisphères cérébraux représentent environ 83% de la masse de l'encéphale. On peut les comparer à un chapeau de champignon couronnant le pied constitué lui-même par le tronc cérébral. Le cerveau est disposé au-dessus du tronc cérébral lui-même faisant suite à la moelle épinière. Il repose également sur le dessus du cervelet et est séparé de celui-ci. Chaque hémisphère cérébral contrôle la moitié du corps se trouvant du côté opposé. C'est l'ensemble de ces trois structures qui constitue l'encéphale à l'intérieur de la boîte crânienne. Chaque hémisphère est subdivisé par des scissures formant plusieurs lobes (*figure A.5*) :

- Le lobe pariétal au-dessus.
- Le lobe temporal sur les côtés.
- Le lobe occipital en arrière.
- Le lobe frontal en avant.

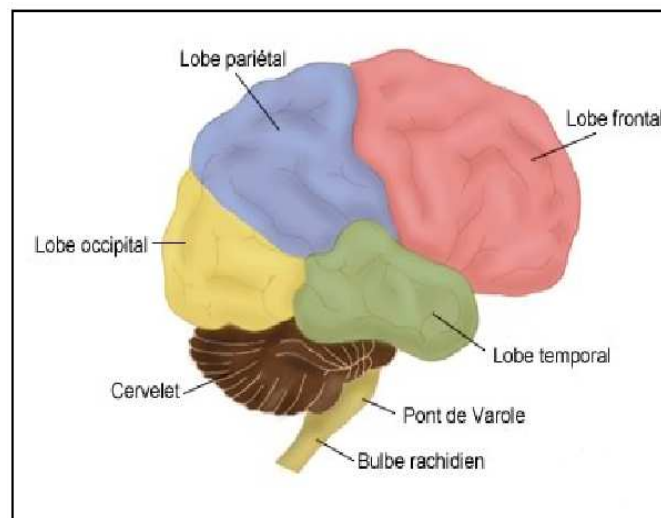


FIGURE A.5 – Lobes du cerveau.

A leur tour ces différents lobes sont creusés par des sillons qui forment des plis appelés également circonvolutions. À l'intérieur du cerveau, proprement dit, on constate la présence de cavités : les ventricules.

### A.2.1.4 Les méninges

L'encéphale est entouré de couches protectrices appelées les méninges. Elles sont constituées de trois membranes successives (*figure A.6*) :

- la plus interne de ces membranes est la pie-mère, fine, compacte, qui contient les structures veineuses qui ramènent le sang du cerveau vers le coeur ;
- la membrane intermédiaire est l'arachnoïde, ainsi appelée en raison de sa ressemblance avec une toile d'araignée (le LCR circule entre celle-ci et la pie-mère) ;
- la membrane la plus externe est la dure-mère, dure et résistante, formée de tissu conjonctif qui protège les couches sous-jacentes des os les entourant.

Les méninges, qui sont des membranes protectrices retiennent également le liquide céphalorachidien, servant lui-même de matelas protecteur. Il comprend deux types de tissu nerveux la substance grise et la substance blanche.

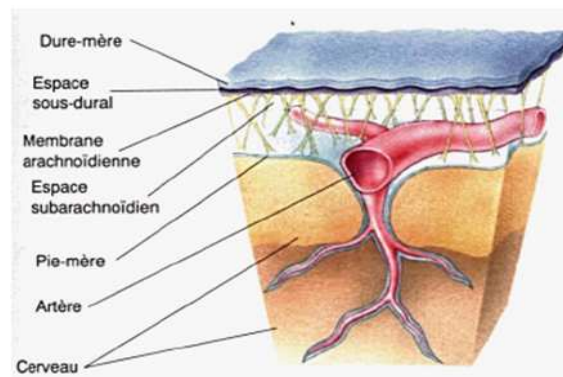


FIGURE A.6 – Méninges.

**A.2.1.4.1 Le liquide céphalo-rachidien** est une substance transparente constituée de 99% d'eau. Il forme un coussin protecteur pour le système nerveux central (SNC) et le nourrit. Le LCR est contenu dans le SNC, dans deux zones anatomiques bien distinctes :

- le système interne constitué par les ventricules cérébraux , qui sont au nombre de quatre : les ventricules latéraux, un dans chaque hémisphère cérébral, le troisième ventricule entre les deux thalamus , et le quatrième ventricule en arrière du tronc cérébral. Les ventricules latéraux communiquent avec le troisième ventricule via le trou de Monro , et le quatrième ventricule avec le troisième ventricule par l'intermédiaire de l'aqueduc de Sylvius . Ce sont les plexus choroïdes contenus dans les ventricules cérébraux qui sécrètent le liquide céphalo-rachidien ;

- le système externe appelé également système de résorption : il s'agit des cavités situées en dessous de l'arachnoïde (cette cavité est plus précisément située entre la pie-mère et l'arachnoïde).

Le compartiment ventriculaire et le compartiment de résorption communiquent entre eux au niveau de la partie inférieure du cervelet par un foramen ventriculaire appelé trou de Magendie.

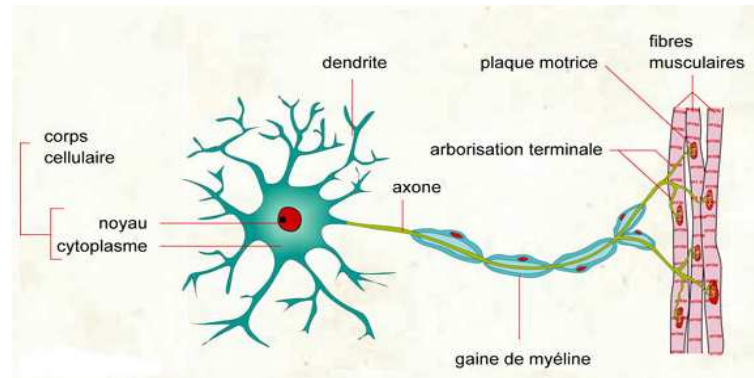


FIGURE A.7 – Architecture schématique d'un neurone.

**A.2.1.4.2 La substance grise** La matière grise (*figure A.2*) est répartie en deux types de structures : le cortex et les noyaux. D'un point de vue topologique, le cortex peut être représenté par une sphère dont l'épaisseur varie de 4 à 5 mm et recouvre une surface totale de plus de 2 m<sup>2</sup>. Afin de pouvoir loger dans la boîte crânienne le cortex est replié formant des arêtes (circonvolutions) et des vallées (fissures). Ce tissu correspond au cœur des cellules nerveuses (neurones) (*figure A.7*), que l'on retrouve également sous forme d'îlots disparates perdus dans la substance blanche et correspondant aux fibres nerveuses (axones, dendrites) recouvertes d'une gaine de myéline. Le cortex cérébral de l'homme contient environ 80% des corps cellulaires de tout le système nerveux. La forme et la position des plis corticaux sont très variables d'un individu à l'autre. Les noyaux gris sont des regroupements de substance grise, situés à l'intérieur de l'encéphale. Ils sont constitués du noyau caudé, du putamen, du globus pallidus, du corps de Luys (noyau sous-thalamique) et de la substance noire.

**A.2.1.4.3 La matière blanche** La matière blanche (*figure A.2*) est comprise entre le cortex, les noyaux gris centraux et les ventricules. Elle est composée d'axones qui établissent les connexions entre les corps cellulaires du cortex et d'autres parties du cerveau.

# LES IMAGES IRM

---

## B.1 Introduction

L'imagerie médicale est certainement l'un des domaines de la médecine qui a le plus progressé ces vingt dernières années. Ces récentes découvertes permettent non seulement un meilleur diagnostic mais offrent aussi de nouveaux espoirs de traitement pour de nombreuses maladies. Cancer, épilepsie... l'identification précise de la lésion facilite déjà le recours à la chirurgie, seule solution thérapeutique pour certains malades. De telles techniques permettent également de mieux comprendre le fonctionnement de certains organes encore mystérieux, comme le cerveau. Depuis la naissance des techniques radiologiques dans les années 1960, plusieurs méthodes ont été mises au point. Les plus fréquentes sont basées sur l'utilisation des rayons X, ultrasons et champs magnétiques. Cette partie approfondira particulièrement une méthode puissante, de plus en plus utilisée par les médecins, l'*Imagerie par Résonance Magnétique (IRM)*. L'imagerie par résonance magnétique est une modalité d'imagerie en plein développement, aussi bien technologique que dans ses applications cliniques. Cet essor tient à l'originalité de ses fondements, au croisement de la physique nucléaire, des mathématiques, de la chimie, de la biologie et de la physiologie. Ce sont ces particularités techniques qui la différencient des autres modalités d'imagerie médicale. Elles lui confèrent des capacités multiples : imagerie morphologique avec un grand contraste tissulaire, imagerie fonctionnelle, du mouvement, des flux, de la perfusion ou même de la composition chimique [66].

## B.2 L'imagerie par résonance magnétique (IRM)

Le phénomène de résonance magnétique nucléaire a été décrit en 1946 par Bloch et Purcell [50]. Lauterbur [131] 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. L'imagerie par résonance magnétique (IRM) est une technique d'imagerie médicale totalement non-invasive et indolore permettant d'obtenir une vue 2D, 3D, voire 4D (3D+t), d'une partie du corps, notamment du cerveau, cette technique est extrêmement précise spatialement, et peut permettre la détection de maladies ou de dysfonctionnements du corps humain. Elle est basée sur le principe de la résonance magnétique nucléaire (RMN). Grâce aux différentes séquences, on peut observer différents tissus avec des contrastes très élevés. Le sigle IRM regroupe en fait un ensemble de techniques d'imagerie : outre des images anatomiques et fonctionnelles, l'IRM permet d'obtenir des images de l'organisation spatiale des vaisseaux cérébraux (angiographie IRM), de l'orientation des principaux faisceaux de fibres blanches (IRM de diffusion), de la distribution de sang (IRM de perfusion) ainsi que de la concentration locale de certains métabolites énergétiques et peut-être, bientôt, de certains neurotransmetteurs (spectroscopie IRM).

### B.2.1 Principe de fonctionnement

Contrairement aux scanners qui fournissent une image basée sur les propriétés des rayons X, avec un contraste identique à celui obtenu par la radiologie conventionnelle, l'IRM (*figure B.1*) n'utilise pas de rayonnement ionisant. Elle utilise les propriétés de la matière et ne concerne que les noyaux d'hydrogène, très abondants dans le corps humain. Un aimant cylindrique creux fournit un champ magnétique intense environ 60 000 fois supérieur au champ magnétique terrestre. Ce champ magnétique fait apparaître au sein des différents tissus de l'organisme une aimantation qui est due aux noyaux d'hydrogène, comme si chaque noyau d'hydrogène se comportait comme une petite boussole. Le phénomène de résonance magnétique permet de mesurer cette aimantation. Pour cela, on perturbe les noyaux d'hydrogène en émettant vers le patient une onde électromagnétique. Sous l'effet de cette impulsion, les noyaux d'hydrogène font basculer leur aimantation par rapport à la direction du champ magnétique. A la fin de l'impulsion, l'aimantation revient à sa position d'équilibre : c'est le retour de l'aimantation à cet état d'équilibre qui est mesuré.



FIGURE B.1 – Acquisition par IRM.

## B.2.2 Notions élémentaires sur le magnétisme

L'IRM est une technique basée sur l'observation de la résonance magnétique nucléaire (RMN) des protons de l'eau. En effet, l'eau constitue environ 70% du corps humain et le proton d'hydrogène est naturellement abondant et très sensible en RMN [4]. Une compréhension des images nécessite avant tout une compréhension des bases physiques du phénomène de résonance magnétique.

### B.2.2.1 Champ magnétique et moment magnétique

Il existe un phénomène d'interaction à distance entre deux objets aimantés. Le couple moment magnétique/champ magnétique permet de décrire les interactions magnétiques entre ces deux objets. Ce sont ces interactions qui nous intéressent dans le phénomène de résonance magnétique nucléaire. L'unité permettant de quantifier un champ magnétique est la *tesla* ( $T$ ).

### B.2.2.2 Moment magnétique et aimantation

Le noyau d'hydrogène, constitué d'un proton, possède des propriétés magnétiques. On peut représenter le moment magnétique sous la forme d'un vecteur en rotation sur lui-même : ceci caractérise le spin du proton (*figure B.2*).

Bien qu'aucune unité particulière n'ait été définie pour quantifier le moment magnétique,

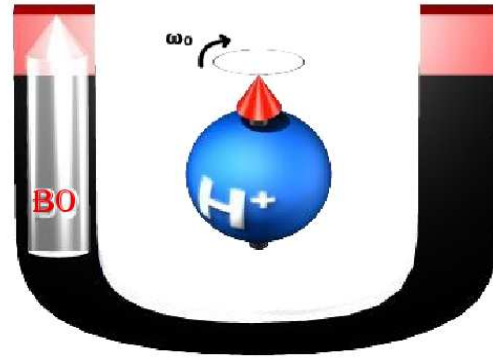


FIGURE B.2 – Moment magnétique.

la formation de toute image en IRM passe par la mesure du module et/ou de l'angle de ce vecteur moment magnétique. Cette entité est souvent remplacée par la notion *d'aimantation*. Au sens strict, l'aimantation représente le moment magnétique par unité de volume ; en réalité, le terme aimantation est souvent employé à la place de l'expression moment magnétique. En IRM, la confusion est d'autant plus légitime que la mesure se fait par unité de volume qui, en raison de la nature numérique des images, est le voxel.

### B.2.2.3 Propriétés magnétiques de la matière

Un atome est constitué d'un noyau, lui-même constitué de nucléons (neutrons et protons) et d'électrons qui gravitent autour de ce noyau . Considérons l'atome d'hydrogène abondamment présent dans le corps humain sous forme de molécules d'eau. Son noyau est composé d'un proton de charge  $+e$  et de masse  $m$  et caractérisé par une propriété quantique intrinsèque appelé *Spin* pouvant adopter deux états d'énergie possible ( $\pm 1/2$ ) (*figureB.3*). Le spin est généralement associé à un mouvement de rotation de moment angulaire  $J$  et de moment magnétique  $\mu$  tel que

$$\vec{\mu} = \gamma J$$

où  $\gamma$  représente le rapport gyromagnétique du proton. Placé dans un champs magnétique statique  $\vec{B}_0$ , les noyaux atomiques s'orientent selon l'axe de ce champ et précessent autour de celui-ci à la fréquence de Larmor B.1.

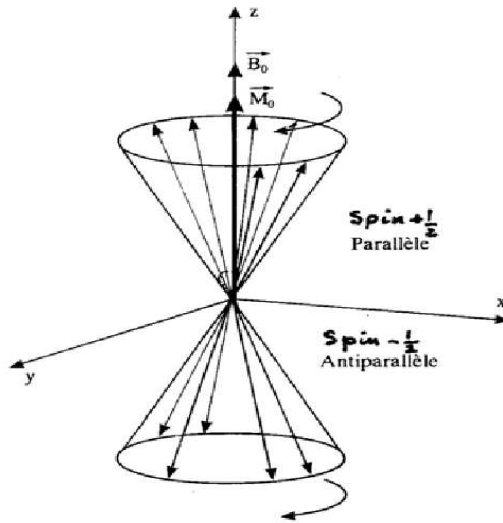
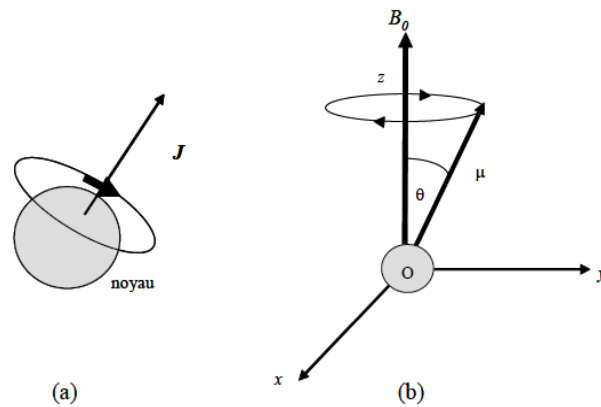


FIGURE B.3 – Mouvement de précession d'un ensemble de moments magnétiques.

#### B.2.2.4 Mouvement de précession et fréquence de Larmor

Les spins ont un mouvement de rotation et décrivent un cône autour de l'axe de  $B_0$  : c'est le mouvement de précession (*figures B.3B.4*). On peut apparenter le mouvement



(a) Mouvement de rotation d'un noyau autour du moment angulaire  $J$  appelé spin. (b) Mouvement de précession d'un noyau atomique autour de l'axe du champ  $B_0$

FIGURE B.4 – Mouvement de précession.

de précession au mouvement d'une toupie qui bascule et reste en équilibre en rotation. La vitesse de précession est proportionnelle à l'intensité du champ magnétique. On la

caractérise par la fréquence de précession  $\omega_0$  (nombre de rotation par seconde) :

$$\omega_0 = \gamma B_0 \quad (\text{B.1})$$

A l'état de repos, ces vecteurs ont une orientation aléatoire. La résultante magnétique de l'ensemble est donc nulle. Soumis à un champ magnétique intense appelé  $B_0$ , les spins s'orientent dans l'axe du champ, soit dans le même sens ("parallèle"), soit dans le sens contraire ("antiparallèle") *figure B.4*. Pour une population donnée de noyaux, la somme de toutes ces petites aimantations qui précèdent à des vitesses angulaires identiques, est susceptible de donner une aimantation non nulle  $\vec{M}_0$  dans la direction du champ magnétique  $\vec{B}_0$ .

### B.2.2.5 Aimantation macroscopique d'un ensemble de spins

Le moment magnétique des protons peut être décomposé en une composante longitudinale et une composante transversale. Le mouvement de précession se traduit ainsi par une rotation de la composante transversale autour de l'axe longitudinal<sup>1</sup>. Dans un champ magnétique  $B_0$ , la proportion de spins dans le sens de  $B_0$  (parallèles) est supérieure à celle des spins orientés dans le sens contraire (antiparallèles) : statistique de Boltzmann B.2.

$$\frac{N_p}{N_p} = \exp\left(\frac{E_p - E_p}{kT}\right) = \exp\left(\frac{\gamma h B_0}{kT}\right) \approx 1 + \frac{\gamma h}{kT} B_0 \quad (\text{B.2})$$

avec

- $\frac{N_p}{N_p}$  = proportion de spins parallèles/spins antiparallèles.
- $\gamma$  = rapport gyromagnétique.
- $h$  = constante de Planck /  $2\pi$
- $k$  = constante de Boltzmann.
- $T$  = température (en Kelvins).
- $B_0$  = intensité du champ magnétique statique (en Tesla)

L'écart de population entre spin "parallèles" et "antiparallèles" est proportionnel à l'amplitude du champ magnétique principal. Dans l'axe longitudinal, la résultante magnétique d'un groupe de spins est donc une aimantation tissulaire dans la même direction et le même sens que  $B_0$  (*figure B.5*).

Les mouvements de précession des spins ne sont pas en phase : les composantes transversales microscopiques s'annulent et il n'y a donc pas d'aimantation transversale résultante

---

1. <http://www.e-mri.org/fr/>

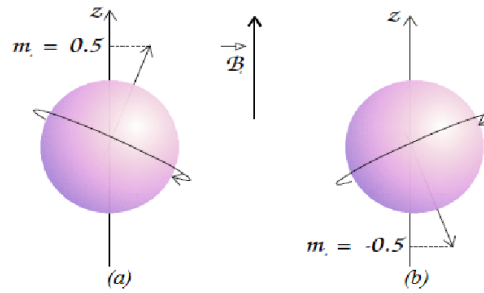


FIGURE B.5 – Les deux orientations de spin d'un proton.

(a)représente l'état de basse énergie d'une particule pour lequel la projection de son spin sur l'axe  $z$  est parallèle au champ magnétique.

(b)représente l'état de haute énergie, moins stable que le précédent. L'axe  $z$  est parallèle au vecteur champ magnétique statique.

lorsque l'on considère un groupe de spins. Alors que les spins obéissent aux lois de la physique quantique, l'aimantation macroscopique suit les principes de la physique classique. C'est cette dernière qui va être considérée par la suite. La différence de population des spins augmente proportionnellement avec l'intensité du champ magnétique statique  $B_0$ , les autres paramètres et la température du corps humain étant constants.

### B.2.3 Phénomène de résonance magnétique nucléaire

Il peut y avoir une interaction entre une onde de radiofréquence (onde RF) et les spins en précession : c'est le phénomène de résonance, qui correspond à un transfert d'énergie entre deux systèmes ayant la même fréquence. L'onde de radiofréquence va entraîner un phénomène de résonance seulement si sa fréquence est la même que celle de précession des spins. Elle va apporter de l'énergie au système de spins : c'est la *phase d'excitation*. Lorsque l'émission de radiofréquence est interrompue, le système va restituer l'énergie absorbée pour retourner à l'état d'équilibre de départ : c'est la *phase de relaxation*.

#### B.2.3.1 Phase d'excitation

L'onde RF provoque une bascule de l'aimantation tissulaire, résultante d'un groupe de spins placés dans un champ magnétique. Cette bascule est liée à un transfert d'énergie vers le système de spins. Au niveau quantique, elle se traduit par une modification des niveaux énergétiques des spins (passage de l'état de basse énergie "parallèle" vers le niveau

de haute énergie "antiparallèle") et de leurs phases.

L'angle de bascule de l'aimantation tissulaire dépend de l'intensité et de la durée pendant laquelle est appliquée l'onde RF. Pour une représentation mathématique plus simple, on étudie la bascule dans un repère  $XYZ$  tournant à la fréquence de précession (*figure B.6*). On peut décomposer l'aimantation tissulaire en une composante longitudinale (selon l'axe

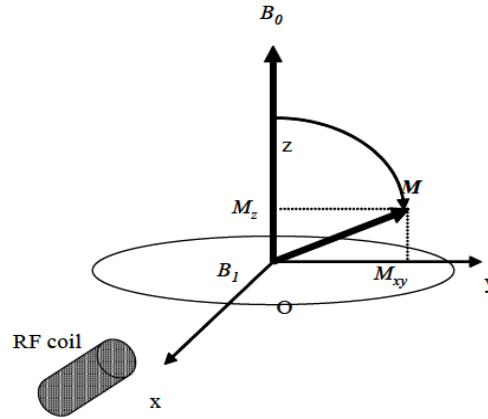


FIGURE B.6 – Perturbation de l'aimantation  $M$  par une impulsion RF.  
*l'angle de bascule est fonction de la durée de cette impulsion*

$Z$ , dans l'axe de  $B_0$ ) et une composante transversale. La phase d'excitation, et la bascule, se traduisent par une diminution de l'aimantation longitudinale (qui peut aller jusqu'à s'inverser), et l'apparition d'une aimantation transversale (sauf pour une bascule à  $180^\circ$ ).

- L'aimantation longitudinale est la traduction de la différence de population entre spins "parallèles" et "antiparallèles".
- L'aimantation transversale est la traduction de la mise en phase plus ou moins complète des spins.

Si l'on prend l'exemple d'une onde RF  $90^\circ$  : il n'y a plus de composante longitudinale car il y a égalisation des populations "parallèle" et "antiparallèle". Par contre, il apparaît une composante transversale car les spins sont mis en phase.

### B.2.3.2 Phase de relaxation et temps $T_1$ et $T_2$

La relaxation correspond au retour à l'équilibre de l'aimantation tissulaire. Elle s'accompagne d'une émission d'énergie sous la forme d'ondes RF qui constituent le signal enregistré en RMN. Elle se décompose en deux phénomènes obéissant à des mécanismes très différents :

- la *relaxation longitudinale* qui correspond à la repousse de la composante longitudinale,  $T_1$ ,
  - la *relaxation transversale*, qui correspond à la chute de l'aimantation transversale,  $T_2$ .
- La *repousse longitudinale* correspond au retour au niveau de basse énergie ("parallèle") des spins excités. C'est la *relaxation spin-réseau*. La croissance de l'aimantation longitudinale lors de la relaxation suit une courbe exponentielle (*figure B.7*). Elle est caractérisée par le temps  $T_1$ . Le  $T_1$  correspond au temps nécessaire pour que l'aimantation longitudinale retourne à 63% de sa valeur finale. Le temps  $T_1$  est de l'ordre de 200 à 3000ms pour un champ de 1,5Tesla. Il est caractéristique des tissus et dépend de la mobilité des molécules où sont engagés les noyaux d'hydrogène. Le temps  $T_1$  est plus court pour de grosses molécules.
- La *relaxation transversale* est due au déphasage des spins : le déphasage est lié aux

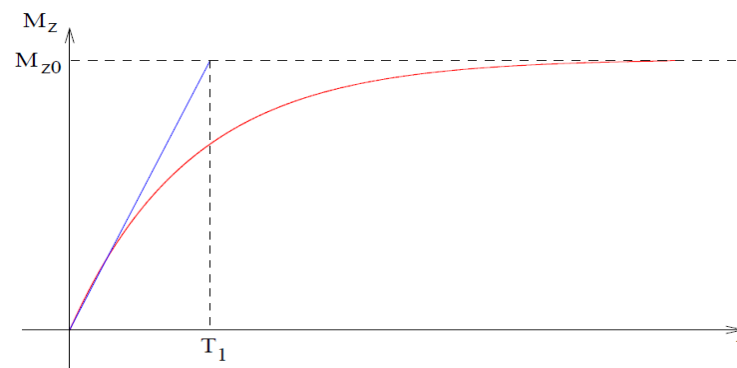


FIGURE B.7 – Relaxation longitudinale.

*Retour à l'équilibre de la composante longitudinale de l'aimantation résultante selon la constante de temps  $T_1$ .*

interactions *spins-spins* qui créent des hétérogénéités de champ, et donc de fréquence de précession. Les spins n'ayant pas exactement la même fréquence, ils se déphasent rapidement. Cette chute de l'aimantation transversale suit une courbe exponentielle décroissante caractérisée par le temps  $T_2$  (*figure B.8*). Le  $T_2$  correspond au temps mis par l'aimantation transversale pour revenir à 37% de sa valeur initiale. La relaxation transversale est beaucoup plus rapide que la relaxation longitudinale :  $T_2$  est toujours plus court ou égal à  $T_1$ .

La modalité  $T_1$  permet d'observer les tissus du cerveau : matière blanche (MB), matière grise (MG) et le liquide céphalo-rachidien. La modalité  $T_2$  ne permet pas de distinguer la matière grise de la matière blanche, mais met en évidence les lésions et le LCR (*figure B.9*).

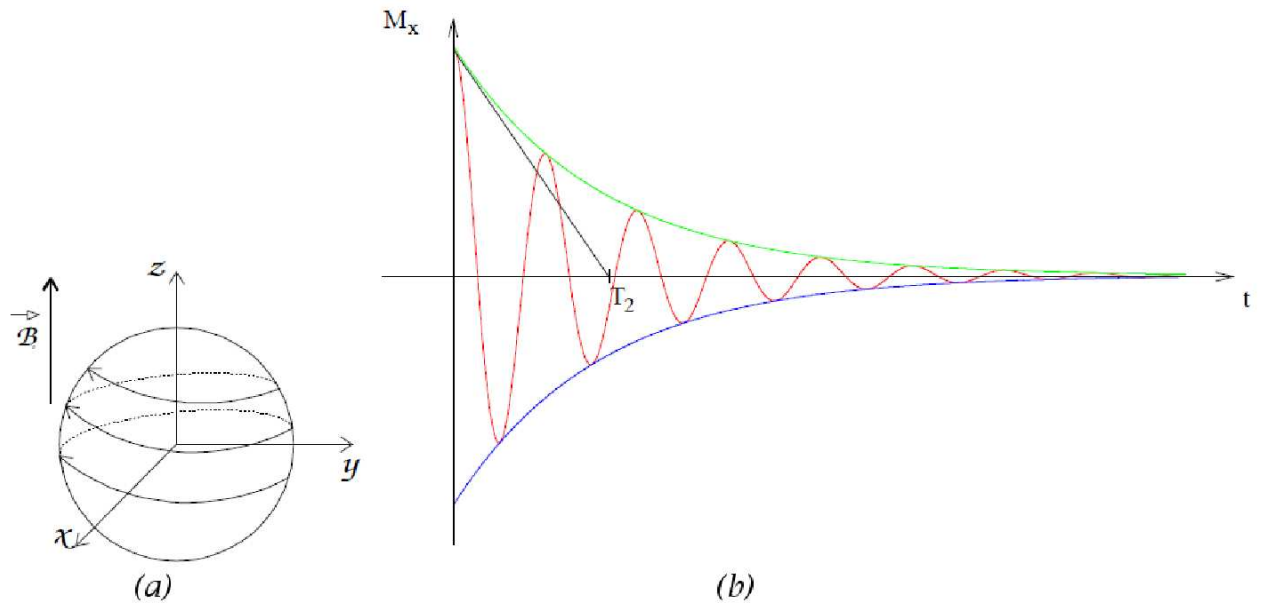
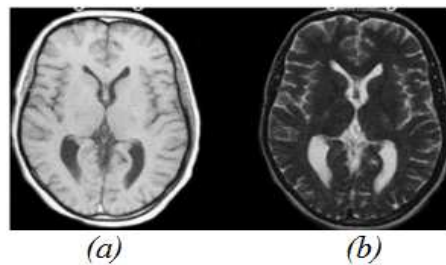


FIGURE B.8 – Relaxation transversale.

(a) Retour à l'équilibre après application de l'impulsion.

(b) Signal de précession libre sur un axe quelconque du plan transversal.



Coupes axiales.

(a) image IRM pondérée en T1.

(b) image IRM pondérée en T2.

FIGURE B.9 – Modalités IRM .

## B.2.4 Création d'une image IRM

Le patient est allongé sur un lit d'examen mobile qui doit ensuite être rentré dans le tunnel de l'aimant. L'IRM soumet le corps à un champ magnétique  $B_0$  très puissant qui va orienter les atomes d'hydrogène dans la même direction. Pour réaliser une image IRM, une antenne émet une onde de radio fréquence  $B_1$  courte de fréquence égale à la fréquence de précession des protons, modifiant l'orientation des protons. Sous l'effet de cette onde, les protons basculent d'un angle  $\alpha$ , généralement de  $90^\circ$  ou de  $180^\circ$  par rapport à  $M_0$ . À

l'arrêt de l'impulsion, le retour à l'équilibre s'établit : il y a réduction rapide du moment transversal  $M_T$  et repousse du moment longitudinal  $M_L$ . Pendant ce retour à l'équilibre, le proton continue à précesser et induit un champ magnétique *FID* (Free Induction Decay). C'est ce signal qui est recueilli par les bobines de réception puis transformé en signal électrique (*figure B.10*) [25] [77]. Ils restituent alors de l'énergie enregistrable sous forme de signal, capté par une antenne réceptrice. Puis en analysant grâce à un ordinateur les signaux de résonance fournis par ces mouvements, on peut obtenir une image tridimensionnelle des organes, présentée en coupes successives. La problématique liée au codage

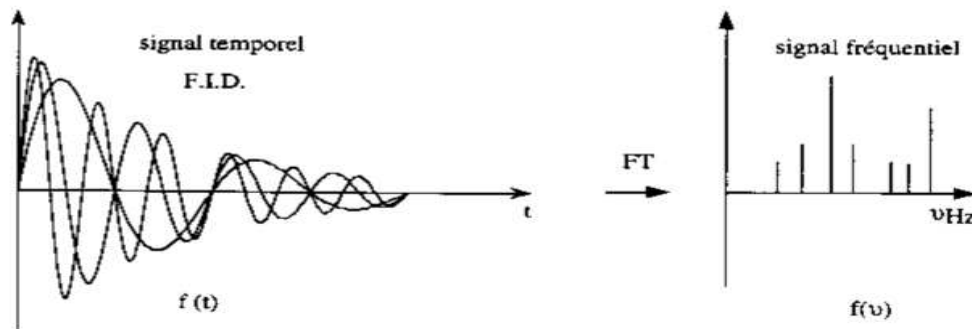


FIGURE B.10 – Signaux temporel et fréquentiel.

de l'image est alors de savoir différencier les signaux issus des différentes zones examinées. Ce codage est réalisé grâce au couplage de trois gradients différents :

- le gradient de coupe sélective sélectionne le plan de coupe, celui-ci pouvant être quelconque dans l'espace,
- le gradient de codage de phase sélectionne les lignes dans le plan de coupe sélectionné,
- le gradient de fréquence sélectionne les colonnes dans le plan de coupe sélectionné.

L'orientation des coupes, qui dépend en particulier du gradient de coupe sélective, varie en fonction de la pathologie recherchée. On retrouve cependant trois orientations principales : axiale, coronale et sagittale (*figure B.11*). Ainsi en codant l'espace en fréquence et en phase des signaux émis, il devient possible de reconstruire une image par application d'une transformée de Fourier inverse. L'IRM permet l'acquisition d'un volume complet, soit par acquisition de coupes 2D successives, soit par l'acquisition d'un volume 3D.

## B.2.5 Le contraste des images

Les paramètres qui influencent le contraste de l'image IRM peuvent se regrouper en deux grandes classes :

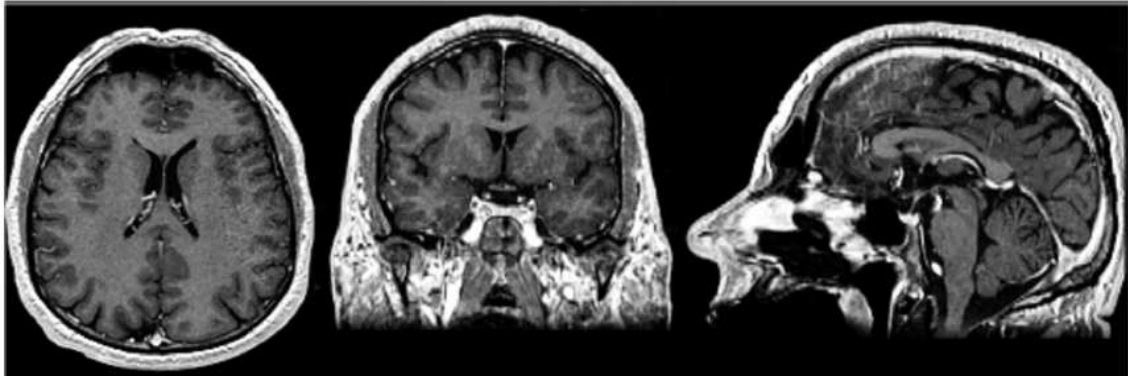


FIGURE B.11 – Acquisition en T1.  
Plans axial, coronal et sagittal (de gauche à droite)

- La première est constituée de paramètres intrinsèques liés directement aux tissus observés. Il s'agit de la densité en protons  $\rho$ , des temps de relaxation  $T1$  et  $T2$ , de la présence d'un agent de contraste ou encore de la vitesse des fluides circulant.
- La seconde est constituée de paramètres liés à l'appareil lui-même (en particulier l'intensité et la constance du champ magnétique  $B_0$ ) et à la séquence d'acquisition [22] [152]. Cette dernière dépend essentiellement de deux paramètres principaux qui sont : Le temps de répétition TR séparant deux impulsions de l'onde radio-fréquence. Le temps d'écho TE séparant l'impulsion de la lecture du signal. Parmi ces paramètres, seuls les paramètres d'acquisition sont modulables et conduisent à l'obtention d'images dites « pondérées en T1 », « pondérées en T2 » ou encore « pondérées en  $\rho$  ».

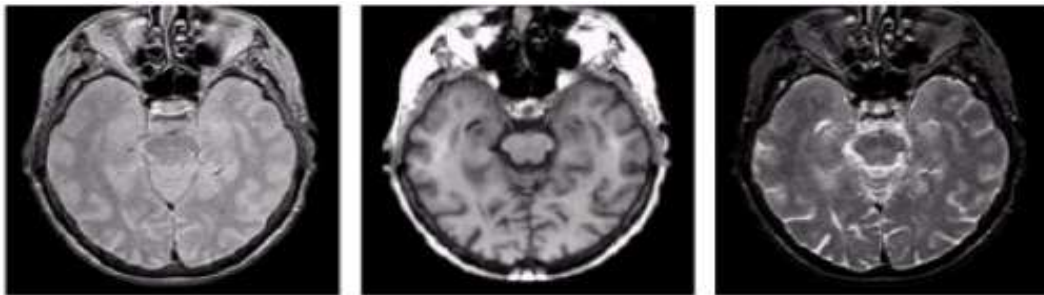


FIGURE B.12 – Images d'une même coupe du cerveau humain, pondérées respectivement en densité de protons, en T1 et en T2.

Pour une séquence donnée, les variations locales des paramètres caractéristiques des tissus créent des contrastes entre les différentes régions de l'image. Souvent les paramètres d'acquisition sont choisis de façon à ce que seul l'un de ces paramètres soit prédominant. Trois types de contraste peuvent donc être obtenus (figure B.12) :

- pour les images pondérées en densité de protons  $\rho$ , le contraste est obtenu lorsque l'intensité de l'image dépend essentiellement de la densité locale de protons et beaucoup moins des constantes de relaxation  $T1$  et  $T2$ . Il est obtenu en utilisant un  $TR$  long et un  $TE$  court.
- pour les images pondérées en  $T1$  (contraste  $T1$ ) : une image est pondérée en  $T1$  lorsque le  $TR$  et le  $TE$  sont courts. Plus le  $T1$  du tissu est court, et plus ce tissu apparaît avec une intensité élevée sur l'image.
- pour les images pondérées en  $T2$  (contraste  $T2$ ) : le contraste en  $T2$  est obtenu en utilisant un  $TE$  long et un  $TR$  long. Contrairement au contraste  $T1$ , les régions de  $T2$  court apparaissent alors plus sombres sur l'image.

## B.2.6 Les agents de contrastes en IRM

Bien que deux visions différentes d'une même structure soient possibles, il peut être encore intéressant d'utiliser des produits de contraste. Le gadolinium est le produit de contraste utilisé en IRM. Il porte le numéro atomique 64 et c'est un métal de haut poids atomique : 157. Le gadolinium se distribue dans les tissus de la même façon que les produits de contraste iodés utilisés en radiologie et pour les scanners. On peut utiliser le gadolinium dans les mêmes conditions que l'iode, en sachant qu'il entraînera une hyperdensité sur les images pondérées en « T1 » lorsque la structure anatomique fixera le produit de contraste. Le gadolinium n'est pas un produit à base d'iode et il est bien toléré. L'injection en intraveineuse n'est pas très douloureuse. Cependant il faut signaler à l'équipe si le patient n'est allergique ou si a une maladie des reins.

## B.2.7 Les artefacts

Les artefacts en IRM peuvent déformer l'image anatomique réelle et/ou simuler un processus pathologique. Il existe, entre autres, les artefacts suivants :

### B.2.7.1 Artefact de mouvement

L'artefact de mouvement en IRM est dû soit aux mouvements globaux du patient, soit aux mouvements d'un seul organe du patient (cœur, intestin, foie) pendant une séquence

d'imagerie. Ces artefacts se traduisent par des phénomènes d'aspects très variables : il s'agit souvent d'apparition d'images fantômes de la structure en mouvement.

### **B.2.7.2 Inhomogénéités RF**

Les inhomogénéités dans le champ magnétique principal et dans le champ généré par les impulsions RF peuvent provoquer une inhomogénéité dans la distribution des intensités des images obtenues en IRM, qui handicape les approches purement photométriques. Ce type d'artefacts n'est pas foncièrement gênant pour le clinicien. Par contre il peut être problématique pour un système de traitement d'image automatique.

### **B.2.7.3 Volume partiel**

La notion de volume partiel n'est pas propre à l'IRM, mais concerne toutes les techniques d'imagerie. Ces artefacts sont liés à la résolution des images. Ainsi, au niveau d'une interface entre deux tissus dans la direction de la sélection de coupe, l'utilisation de coupe fine permet de bien séparer les tissus. En revanche, une coupe plus épaisse contient à la fois les deux tissus : le signal résultant sera alors une moyenne du signal des tissus, d'où une perte d'information en termes de contraste. Il en résulte également une perte de la résolution spatiale, puisque le signal des petites structures va être moyenné avec celui des structures adjacentes, ce qui peut les rendre invisibles ou peu distinctes.

### **B.2.7.4 Le bruit**

Le bruit Outre les artefacts, la présence de bruit dans les images acquises, dû aux processus physiques d'acquisition ou aux propriétés des éléments visualisés, complexifient encore la tâche d'analyse des images.

# Bibliographie

- [1] *Machine perception of three-dimensional solids*, chapter 3. Roberts G, 1965.
- [2] Thiran J. P. ; Warscotte V. ; Macq B 1997. A queue-based region growing algorithm for accurate segmentation of multi-dimensional digital images. *Signal processing*, 1997.
- [3] Appriou A. *Discrimination multisignal par la théorie de l'évidence in Décision et Reconnaissance des formes en signal*. 2002.
- [4] Borel A. *Etude théorique et expérimentale d'agents de contraste pour l'imagerie médicale par résonance magnétique, thèse de chimie*. PhD thesis, 2001.
- [5] Dromign-Badin A. *Fusion d'images par la théorie de l'évidence en vue d'applications médicales et industrielles*. PhD thesis, institut national des sciences appliquées de Lyon, 1998.
- [6] BUCHNER A. G. ANAND S. S. Decision support using data mining. *Financial Times Pitman Publishers, London,*, 1998.
- [7] AndréElisseeff. L'intégration des données, le data mining (fouille de données)et le support à la décision. *IBM Research Lab -Zurich*, 2005.
- [8] Apté. Data mining : An industrial research perspective. *IEEE Computational Science and Engineering*, 1997.
- [9] Jahne B. Practical handbook on image processing for scientific applications. *Boca Raton*, 1997.
- [10] Moser E Baumgartner R, Windischberger C. Quantification in functional magnetic resonance imaging : Fuzzy clustering vs. correlation analysis. *Magnetic Resonance Imaging*, 1998.
- [11] Croisille P Poirier C Boudraa A Unterreiner R Mason J.C Revel D Behloul F, Janier M. Automatic assessment of myocardial viability based on pet-mri data fusion. *Proceedings of the 20th Annual International Conference of the IEEE on Engineering in Medicine and Biology Society*, 1998.

- 
- [12] Coray C Bezdek J and Watson J. Detection and characterization of cluster substructure : I. linear structure : fuzzy c-lines. *SIAM Journal of Applied Mathematics*, 40(2) :339–357, 1981.
- [13] Gunderson R Bezdek J, Coray C and Watson J. Detection and characterization of cluster substructure. ii. fuzzy c-varieties and convex combinations thereof. *SIAM Journal of Applied Mathematics*, pages 40 :358–372, 1981.
- [14] Marcel Bosc. *Contribution à la detection des changements dans les sequences IRM 3D multimodales*. PhD thesis, Université Louis Pasteur Strasbourg, 2003.
- [15] B Bouchon-Meunier, R.R. Yager, and L.A. Zadeh. Fuzzy logic and soft computing. *World Scientific*, 1995.
- [16] Masségli F Trousse B Boussaid O, Gançarski P. Fouilles de données complexes. *revue des nouvelles technologies-E-4*, 2005.
- [17] Ambroise C. Introduction à la reconnaissance statistique des formes. Technical report, École des Mines de Paris, 1997.
- [18] Baillard C, Hellier P, and Barillot C. Segmentation of brain 3d mr images using level sets and dense registration. *Medical Image Analysis*, page 5 :185–194, 2001.
- [19] BEZDEK J. C. *Fuzzy mathematics in pattern classification*. PhD thesis, Cornell Univ, 1973.
- [20] BEZDEK J. C. A convergence theorem for the fuzzy isodata clustering algorithms. *IEEETrans. Pattern Anal. Machine Intell*, (1) :1–8, 1980.
- [21] Ciofolo C and Barillot C. Brain segmentation with competitive level sets and fuzzy control. *International Conference on Information Processing in Medical Imaging*, 2005.
- [22] GUINET C and GRELLET J. Introduction à l'irm : de la théorie à la pratique. *Abrégés d'imagerie radiologique, Masson,,* 1992.
- [23] JAGGI C. Approche markovienne pour la segmentation 3d des tissus cérébraux en irm. *Colloque GRETSI*, 1997.
- [24] Jaggi C. *Segmentation par méthode markovienne de l'encéphale humain par résonance magnétique : théorie, mise en oeuvre et évaluation*. PhD thesis, Université de Caen, 1998.
- [25] Anne-Sophie CAPELLE. *Segmentation d'images IRM multi-écho tridimensionnelles pour la détection des tumeurs cérébrales par la théorie de l'évidence*. PhD thesis, l'université de Poitiers, 2003.
- [26] Kimmel R Caselles V and Sapiro v. Geodesic active contours. *Proceedings of ICCV'95, Cambridge*, 1995.

- 
- [27] Yville A .L Clark J.J. Data fusion for sensory information processing systems. *Kluwer Academic Publishers, Boston*, 1990.
- [28] comeau.D. Evaluation des technique de fusion d’images pour le projet camus. *centre de recherche pour la défense, Valcartier*, 1998.
- [29] DELLA PENNA S CONFORTO S. Tissue segmentation of mri of the head by means of a kohonen map. *Proceedings of IEEE Engineering in Medicine and Biology Society 18th Annual International Conference*, 1996.
- [30] R. Cucchiara. Analysis and comparison of different genetic models for the clustering problem in image analysis. *Albrecht, R., Reeves, C., and Steele, N., editors, International Conference on Artificial Neural Networks and Genetic Algorithms*, page 423–427, 1993.
- [31] Dubois D and Prade H. Fuzzy sets and systems : Theory and applications. *Academic Press, New-York*, 1980.
- [32] Mahr D. Theory of edge detection. *Proceedings of the Royal Society of London*, 1980.
- [33] Pham D. Robust fuzzy segmentation of magnetic resonance images. *Symposium on Computer-Based Medical Systems*, 2001.
- [34] Zighed D.A. Travaux en préparation des données. [http ://eric.univ-lyon2.fr/ zighed/cv/index.html](http://eric.univ-lyon2.fr/zighed/cv/index.html).
- [35] Pôle Bases de Données Décisionnelles (BDD)Laboratoire ERIC (EA 3083) Université Lumière Lyon 2. Données complexes. [http ://bdd.univ-lyon2.fr/](http://bdd.univ-lyon2.fr/).
- [36] REDDICK W DEATON R, SUN J. Self-organized feature detection and segmentation of mr images. *Proceedings of IEEE Engineering in Medicine and Biology Society 14th Annual International Conference*, 1994.
- [37] REDDICK W DEATON R, TANG L. Fractal analysis of mr images of the brain. *Proceedings of IEEE Engineering in Medicine and Biology Society 16th Annual International Conference*, 1994.
- [38] Goss S. Franks N. Sendova-Franks A. Detrain C. Deneubourg, J.-L. and L. Chretien. The dynamics of collective sorting : robot-like ant and ant-like robots. *Meyer and Wilson*, page 356–365, 1990.
- [39] Prade H Dubois D. Possibility theory, an approach to the computerized processing of uncertainty. *Plenum Press*, 1988.
- [40] Frenoux E. *Applications et validation de méthodologies de fusion de données en imagerie cérébrale*. PhD thesis, Université d’Auvergne, 2004.

- 
- [41] Fayyad et al. Advances in knowledge discovery and data mining. *AAAI Press / MIT Press*, 1996.
- [42] Frigui H et Krishnapuram R. A robust competitive clustering algorithm with applications in computer vision. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1999.
- [43] McLachlan G et Peel D. Finite mixture models. *Eds Wiley*, 2000.
- [44] Pham D et Prince J. An adaptive fuzzy c-means algorithm for image segmentation in the presence of intensity inhomogeneities. *Pattern Recognition Letters*, 1999.
- [45] Pham D et Prince J. Adaptive fuzzy segmentation of magnetic resonance images. *IEEE Transactions on Medical Imaging*, 1999.
- [46] Zighed D. A. et RAKOTOMALALA R. Graphes d'induction. apprentissage et data mining. *Hermes Science Publication*, 2000.
- [47] ZIGHED D. A. et RAKOTOMALALA R. Extraction de connaissances à partir de données (ecd). *Techniques de l'Ingénieur*, 2002.
- [48] Jiang L et Yang W. A modified fuzzy c-means algorithm for segmentation of magnetic resonance images. *Digital Image Computing, Techniques and Applications*, 2003.
- [49] Dawant B. M. et Zijdenbos A. P. Reliability and reproducibility in manual and semi-automatic mri segmentation,. *Proceedings of IEEE Engineering in Medicine and Biology Society 15th Annual International Conference*, 1993.
- [50] BLOCH F. Nuclear induction. *Physical Review*, 1946.
- [51] Lachmann F. *Mthodes d'analyse d'images médicales pour la reconnaissance de structures crbrales*. PhD thesis, Universit Franois Rabelais - Tours, 1992.
- [52] Mangin F. segmentation du cerveau par l'utilisation d'histogramme et morphologie mathématique. *conférence internationale sur l'imagerie médicale*, 1998.
- [53] PIATETSKY-SHAPIRO G. et SMYTH P. FAYYAD U. The kdd process for extracting useful knowledge from volumes data. *Communication of the ACM*, 1996.
- [54] Kuijper A Florack L. The topological structure of scale-space images. *Journal of Mathematical Imaging and Vision*, page 12(1) :65–79, 2000.
- [55] N. Franks and A. Sendova-Franks. Brood sorting by ants : distributing the workload over the work surface. *Behav. Ecol. Sociobiol.*, page 30 :109–123, 1992.
- [56] Dugas-Phocion G. *Segmentation d'IRM Cérébrales, Multi-Séquences, et Application à la Sclérose en Plaques*. PhD thesis, École des Mines de Paris, 2006.
- [57] Shafer G. A mathematical theory of evidence. *Princeton University Press*, 1976.

- 
- [58] Wu F. Gardarin G., Pucheral Ph. Bitmap based algorithms for mining association rules. *Proceeding of European Conference BDA. Tunis*, 1998.
- [59] Barlaud M Gastaud M and Aubert G. Combining shape prior and statistical features for active contour segmentation. *IEEE Trans. Circuits Syst. Video Techn.*, 2004.
- [60] Amaury H. *Modèles et Techniques en Inférence Grammaticale Probabiliste : de la Gestion du Bruit à l'Extraction de Connaissances*. PhD thesis, Université Jean Monnet de Saint-Etienne, 2004.
- [61] Jianzhong Li Haiwei Pan and Zhang Wei. Mining interesting association rules in medical images. *Dept. of Computer Science, Harbin Institute of Technology, Harbin, P.R. China*, 2005.
- [62] Peng L. W. Lee S. Brettschneider P. D. Shah A. Henderer J. D. Schlaepfer T. E. Pearison G. D. Harris G. J., Barta P. E. Mr volume segmentation of gray matter and white matter using manual thresholding : Dependence on image brightness. *Neuroradiol*, 1994.
- [63] Lechevallier Y. Hébrail G. Data mining et analyse des données. *Analyse des données, G.Govaert éditeur, Hermes,*, 2003.
- [64] Leblond P Mari J.F Hergalant S, Aigle B. Fouille de données du génome à l'aide de modèles de markov cachés. *Extraction et Gestion de Connaissances - EGC 2005, Paris, France*, 2005.
- [65] Schnack H.G. Automatic segmentation of the ventricular system from mr images of the human brain. *Neuroimage*, 2001.
- [66] Gahide G Le Bars E Taourel P Hoa D, Micheau A. *L'IRM pas à pas*. Nb edition, Septembre 2007.
- [67] Bloch I. Fusion de données en traitement d'images : modèles d'information et décision. *Traitement du signal*, 1994.
- [68] Bloch I. Information combination operators for data fusion : A comparative review with classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 1996.
- [69] Bloch I. Some aspects of dempster-shafer evidence theory for classification of multimodality medical images taking partial. *Volume Effect into account, Pattern Recognition Letters*, 1996.
- [70] Bloch I. Fusion d'informations en traitement du signal et des images. *Lavoisier*, 2003.
- [71] Bloch I and Maitre M. Les méthodes de raisonnement dans les images brique. *VOIR Module RASIM*, 2004.

- 
- [72] KAPOULEAS I. Automatic detection of white matter lesions in magnetic resonance brain images. *Computer Methods and Programs in Biomed*, 1990.
- [73] Sobel I. Codage des images binaires et détection des contours. *Image Processing*, 1978.
- [74] AGRIS J, DE FIGUEIREDO R, HILLMAN G, and KENT T. A Novel Method for 3D Segmentation and Volume Estimation of Brain Compartments from MRI, Magnetic Resonance Imaging. *Proceedings of the 13th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 185–186, October 1994.
- [75] BEZDEK J. Pattern recognition with fuzzy objective function algorithms. *Plenum Press*, 1981.
- [76] Bezdek J. Pattern recognition with fuzzy objective function algorithms. *Plenum Press, New York, (second edition) edition*, 1987.
- [77] BITTOUN J. *L'imagerie par résonance magnétique nucléaire (IRM)*. PhD thesis, Faculté de médecine Paris Sud.
- [78] Freixenet J, Muñoz X, Raba D, Marti J, and Cufi X. Yet another survey on image segmentation : region and boundary information integration. *Proceedings of 7th European Conference on Computer Vision-Part III (ECCV '02)*, 2002.
- [79] Gebhardt J. and Kruse R. Information source modelling for consistent data fusion. *Hamid R. Arabnia and Dongping (Daniel) Zhu, editors, Proceedings of the International Conference on Multisource-Multisensor Information Fusion - Fusion'98*, 1998.
- [80] Zimmermann H. J. Fuzzy set theory and its applications. *Kluwer Academic Publisher, Boston*, 1991.
- [81] Cocquerez J.-P and Philipp S. Analyse d'images : filtrage et segmentation. *Masson*, 1995.
- [82] Judith L. Jagath C. Rajapakse, Jay N. Giedd. Statistical approach to segmentation of single-channel cerebral mr images. *IEEE Transactions on Medical Imaging*, 1997.
- [83] A. Jain and R. Dubes. Algorithms for clustering data. *Prentice Hall Advanced Reference Series*, 1988.
- [84] Jr James H. Harrison. Introduction to the mining of clinical data. *Division of Clinical Informatics, Departments of Public Health Sciences and Pathology, University of Virginia*, 2008.
- [85] Bezdek J.C., Hall L.O., and Clarke L.P. Review of mr image segmentation techniques using pattern recognition. *Medical Physics*, 1993.

- 
- [86] Micheline Kamber Jiawei Han. *Data Mining : Concepts and Techniques*. Morgan Kaufmann, second edition edition, 2006.
- [87] Prewitt J.M.S. Picture processing and psychophysics. *Academic Press*, 1970.
- [88] D. Jones and M. Beltrano. Solving partitioning problems with genetic algorithms. *Belew and Booker, 1991*, page 442–449, 1991.
- [89] Vigario R Karp E. Unsupervised mri tissue classification by support vector machines. *Proceeding of Biomedical Engineering (BioMED 2004), Innsbruck, Austria, 2004*.
- [90] Bernd J Karsten Held, Elena Rota Kops. Markov random field segmentation of brain mr images. *IEEE Transactions on Medical Imaging*, 1997.
- [91] Witkin A Kass M and Terzopoulos D. Snakes : Active contour models. *International Journal of Computer Vision*, 1987.
- [92] Batrancourt B Colliot O Angelini E Bloch I Khotanlou H, Atif J. Segmentation de tumeurs cérébrales et intégration dans un modèle de l’anatomie. *Actes de Reconnaissance des Formes et Intelligence Artificielle, RFIA , Tours, France, 2006*.
- [93] Cios K.J. and Moore G.W. Uniqueness of medical data mining. *Artificial Intelligence in Medicine*, 2002.
- [94] Germond L. Coopération entre processus guidés par les données et les modèles pour la segmentation en irm. Technical report, RR1005-I, TIMCIMAG, 1998.
- [95] JOURDAN L. *Métaheuristiques pour l’extraction de connaissances : Application à la génomique*. PhD thesis, Université des sciences et technologies de lille U.F.R. D’I.E.E.A., 2003.
- [96] Zadeh L. Fuzzy sets. *Information and Control*, 1965.
- [97] Zadeh L. Fuzzy sets as a basis for theory of possibility. *International Journal of Fuzzy Sets and Systems*, 1978.
- [98] Zadeh L.A. The concept of a linguistic variable and its application to approximate reasoning. *Information Sciences*, 1975.
- [99] Monmarché N et Venturini G Labroche N. Modélisation de la fermeture coloniale chez les fourmis pour la classification non supervisée. *Laboratoire d’Informatique de l’Université de TOURS Ecole d’Ingénieur en Informatique*, 2002.
- [100] Germond Laurence. *Trois principes de coopération pour la segmentation en imagerie de résonance magnétique cérébrale*. PhD thesis, université Joseph Fourier, 1999.
- [101] Barillot C Lecoœur J. Segmentation d’images cérébrales : Etat de l’art. Technical Report 6306, Thème BIO, Systèmes biologiques Projet VisAGeS, INRIA, 2008.

- [102] Venturi G. Lefébure R. *Le Data Mining*. Eyrolles, 1998.
- [103] Colot O Lefevre E, Vannoorenberghe P. About the use of dempster-shafer theory for color image segmentation. *First International Conference on Color in Graphics and Image Processing, Saint-Etienne, France, 2000*.
- [104] Pfefferbaum A. Lim K. O. Segmentation of mr brain images into cerebrospinal fluid spaces, white and gray matter. *Comput. Assist. Tomogr*, 1989.
- [105] Barni M, Cappellini V, and Meccocci A. Comments on a possibilistic approach to clustering. *IEEE Trans, Fuzzy Syst*, 1996.
- [106] Bomans M, Hohne K-H, Tiede U, and Riemer M. Segmentation of mr images of the head for 3-d display. *Medical Imaging, IEEE Transactions*, 1990.
- [107] CLARK M. Mri segmentation using fuzzy clustering techniques. *IEEE Engineering in Medecine and Biology*, 1994.
- [108] Schmitt M. Morphologie mathématique. *Masson*, 1993.
- [109] Sethian J. Malladi R. and Vemuri B. Shape modelling with frontpropagation : a level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995.
- [110] R. Mamdouh. Data preparation for data mining. *Morgan Kaufmann Publishers,, 2006*.
- [111] Joshua R. Mario Aguilar. Fusion of multi-modality volumetric medical imagery. *Proceedings of the 5th. International Conference on Information Fusion,, 2002*.
- [112] Bloch I Mascle S and Vidal-Madjar D. Application of dempster-shafer evidence theory to unsupervised classification in multisource remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 1997.
- [113] Denoeux T Masson M.-H. Ecm : algorithme évidentiel des c-moyennes. *Rencontres Francophones sur la Logique Floue et ses Applications (LFA '07), pages 17-24, Nîmes, France, Novembre,, 2007*.
- [114] Schenone A Masulli F. A fuzzy clustering based segmentation system as support to diagnosis in medical imaging. *Artificial Intelligence in Medicine*, 1999.
- [115] Dmitry B. Goldgof Robert. Velthuisen F. Reed Murtagh Martin S. Silbiger Matthew C. Clark, Lawrence O. Hall. Automatic tumor segmentation using knowledge-based techniques. *IEEE Transactions on medical imaging*, 1998.
- [116] Brummer M.E, Mersereau R M, Eisner R L, and Lewine. Automatic detection of brain contours in mri data sets. *Medical Imaging, IEEE Transactions on (12)*, 1993.
- [117] KRISHNAPURAM R MEDASANI S, KIM J. An overview of membership function generation techniques for pattern recognition. *International Journal of Approximate Reasoning*, 1998.

- [118] Riadh Ben Messaoud. Data mining. Technical report, Institut Universitaire de Technologie Lumière Licence C.E.STAT, 2007.
- [119] Arya Nabavi et al Michael Kaus, Simon K. Warfield. Segmentation of meningiomas and low grade gliomas in mri. *LNCS*, 1999.
- [120] Arya Nabavi Peter M. Black Ferenc A. Jolesz Ron Kikinis Michael Kaus, Simon K. Warfield. Automated segmentation of mri of brain tumors. *Radiology*, 2001.
- [121] Roger Boyle Milan Sonka, Vaclav Hlavac. Image processing, analysis, and machine vision. *Second Edition.*, Brooks/Cole, 2002.
- [122] Labroche N. *Modelisation du systeme de reconnaissance chimique des fourmis pour le probleme de classification non supervisée : application à la mesure d’audience sur internet*. PhD thesis, Laboratoire d’Informatique, Université de Tours, 2003.
- [123] Monmarche N. On data clustering with artificial ants. *A. Freitas, editor, Data Mining with Evolutionary Algorithms : Research Directions*, page 23–26, 1999. Orlando, Florida,.
- [124] Paragios N. A variational approach for the segmentation of the left ventricle in cardiac image analysis. *International Journal of Computer Vision*, 2002.
- [125] Batouche M Ouadfel S. Des fourmis pour la segmentation des images. *Journées d’étude en Informatique graphique JIG05*, 2005.
- [126] Maciunas R. J Ozkan M, Dawant B. M. Neural-networkbased segmentation of multimodal medical images : A comparative and prospective study. *IEEE Trans. Med. Imag*, 1993.
- [127] Dempster A P. Upper and lower probability function in a context of uncertainty. *Annals of math. Statistics*, 1967.
- [128] Dempster A P. A generalization of bayesian inference. *Journal of the Royal Statistical Society*, 1968.
- [129] Smets P. Constructing the pignistic probability function in a context of uncertainty. *Uncertainty in Artificial Intelligence*, 1990.
- [130] Marius FIESCHI Patrice DEGOULET. *Traitement de l’information médicale Méthodes et applications hospitalières*, chapter 7, Données et décisions médicales. 2003.
- [131] LAUTERBUR PC. Image formation by induced local interaction : examples employing nmr. *Nature*, 1973.
- [132] Xu C. et Prince J.L Pham D.L. Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2000.
- [133] PHUPHANICH S HALL L CLARKE L SILBINGER M PHILIPPS W, VELTHUIZEN R. Application of fuzzy c-means algorithm segmentation technique for tissue

- differentiation in mr images of a hemorrhagic glioblastoma multiforme. *Magnetic Resonance Imaging*, 1995.
- [134] MOYSAN J. PLOIX M.A., GARNIER V. Théorie des possibilités appliquée à l'end du béton. *LCND, IUT Aix-en-Provence, Avenue Gaston Berger*.
- [135] D. Pyle. Data preparation for data mining. *Morgan Kaufmann Publishers, Inc. San Francisco, USA*, 1999.
- [136] Bellman R. Adaptive control processes : A guided tour. *Princeton University Press, Princeton, New Jersey, U.S.A*, 1961.
- [137] Davé R. Characterization and detection of noise in clustering. *Pattern Recognition*, pages 12(11) :657–664, 1991.
- [138] Davé R. Generalized fuzzy c-shells clustering and detection of circular and elliptical boundaries. *Pattern Recognition*, pages 25 :713–721, 1992.
- [139] Davé R and Bhaswan K. Adaptive fuzzy c-shells clustering and detection of ellipses. *IEEE Transactions on Neural Network*, (3(5)) :643–662, 1992.
- [140] Henkel R. Segmentation in scale space. *Computer Analysis of Images and Patterns*, page 41–48, 1995.
- [141] Krishnapuram R and Keller J. A possibilistic approach to clustering. *IEEE Trans. Fuzz Syst*, 1993.
- [142] Krishnapuram R and Keller J. The possibilistic c-means algorithm : insights and recommendations. *IEEE Trans. Fuzzy Syst*, 1996.
- [143] Singleton H. R. and Pohost G. M. Automatic cardiac mr image segmentation using edge detection by tissue classification in pixel neighborhoods. *Magnetic Resonance in Medicine*, 1997.
- [144] Yager R. R. On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Transactions on Systems, Man, and Cybernetics*, 1988.
- [145] Zighed D.A. Rakotomalala R. Extraction de connaissances à partir des données (lcd). *Technique de l'ingénieur*, 2003.
- [146] Abdessealm A Rahman M, Chai W. Scale-space characteristics for image segmentation. *International Conference on Electrical and Electronic Technology*, 1 :28–32, août 2001.
- [147] Dojat M Richard N and Garbay C. Automated segmentation of human brain mr images using a multi-agent approach. *Artificial Intelligence in Medicine*, 2004.
- [148] Paragios N. Rousson M. and Deriche R. Implicit active shape models for 3d segmentation in mr imaging. *MICCAI (1), volume 3216 of Lecture Notes in Computer Science*, 2004.

- 
- [149] Amartur S, Piriano D, and Takefuji Y. Optimization neural networks for the segmentation of magnetic resonance images. *IEEE Transactions on Medical Imaging*, 11 :215–221, June 1992.
- [150] Dellepiane S and Venturi G. A fuzzy model for the processing and recognition of mr pathological images. *In Lecture notes in computer science, IPMI proceedings*, 1991.
- [151] FARIZY A. S. L’entreposage et la fouille des données complexes. Technical report, 2003.
- [152] LANGLOIS S. Imagerie par résonance magnétique : du signal à l’image. *Cahier du GREYC 10, Groupe de Recherche en Informatique, Image, Instrumentation de Caen*, 1997.
- [153] Osher S and Sethian J. A. Fronts propagating with curvature-dependant speed : algorithms based on hamilton-jacobi formulations. *Journal of Computational Physics*, 1988.
- [154] Christophe SAINT-JEAN. *Classification paramétrique robuste partiellement supervisée en reconnaissance des formes*. PhD thesis, Université de la Rochelle, 2001.
- [155] Mouchot M.C. Solaiman B., Koffi R. and Hillion A. An information fusion method for multispectral image classification post-processing. *IEEE Transaction. On Geoscience and Remote Sensing*, 1998.
- [156] Hlavac V. Sonka M. and Boyle R. Image processing, analysis and machine vision. *PWS Publishing*, 1999.
- [157] Fitzpatrick J.M Sonka Mand. Handbook of medical imaging. *Medical Image Processing and Analysis*.
- [158] Jinghao Xue Jalal Fadili Daniel Bloyet. Su Ruan, Cyril Jaggi. Brain tissue classification of magnetic resonance images using partial volume modeling. *IEEE Transactions on medical imaging*, 2000.
- [159] Bilgiç T and Burhan Türksen I. Measurement of membership functions : Theoretical and empirical work. *Handbook of Fuzzy Sets And Systems, Dubois et Prade eds*, 1997.
- [160] GERAUD T. *Segmentation des Structures Internes du Cerveau en IRM*. PhD thesis, Ecole Nationale Supérieure des Télécommunications de Paris, 1998.
- [161] MANJUNATH B TSAI C and JAGADEESAN R. Automated segmentation of brain mr images. *Pattern Recognition*, 1995.
- [162] Samarasekera L.S Miki Y van Buchem M.A Grossman R.I Udupa J.K, Wei L. Multiple sclerosis lesion quantification using fuzzy-connectedness principles. *IEEE Transactions on medical imaging*, 1997.

- 
- [163] Bachelet V. *Métaheuristiques parallèles hybrides : Application au QAP*. PhD thesis, USTL LIFL France, 1999.
- [164] Barra V. *floue des tissus cérébraux en IRM 3D : une approche possibiliste versus autres méthodes*. PhD thesis, Université Blaise Pascal, 1999.
- [165] Barra V. *Fusion d'Images 3D du Cerveau : Etude de Modèles et Applications*. PhD thesis, Université d'Auvergne Clermont-Ferrand (France), 2000.
- [166] Barra V and Boire J. Automatic segmentation of subcortical brain structures in mr images using information fusion. *IEEE Transactions on Medical Imaging*, 2001.
- [167] Barra V and BOIRE JY. Tissue characterization on mr images by a possibilistic clustering on a 3d wavelet representation. *Journal of Magnetic Resonance Imaging*, 2000.
- [168] Dou W. *Segmentation d'images multispectrales basée sur la fusion d'informations : application aux images IRM*. PhD thesis, Université de caen/basse-normandie, 2006.
- [169] TUCKER W. Counterexamples to the convergence for the fuzzy isodata algorithms. *The Analysis of Fuzzy Information, Boca Raton Editions, FL,*, 1987.
- [170] BLAXTON T WESPHAL C. Data mining solutions. *John Wiley, New York,*, 1998.
- [171] Solaiman B Wilinski P and Hillion A. Towards the borders between neural and markovian paradigms. *IEEE Transaction on System, Man, and Cybernetics*, 1998.
- [172] Hu X, Alperin N, Levin D N, Tan K K, and Mengeot M. Visualization of MR angiographic data with segmentation and volume-rendering techniques. *J Magn Reson Imaging*, page 539–46, Sep-Oct 1991.
- [173] Moretti B Revenu M et Bloyet D Xue J, Ruan S. Knowledge-based segmentation and labeling of brain structures from mri images. *Pattern Recognition Letters*, 2001.
- [174] Kodratoff Y. Technical and scientific issues of kdd (or : Is kdd a science?). *Algorithmic Learning Theory, Springer Verlag*, 1994.
- [175] Karine Zeitouni. *Analyse et extraction de connaissances des bases de données spatiotemporelles*. PhD thesis, Université de Versailles Saint-Quentin-en-Yvelines, 2006.
- [176] Margolin R.A Palmer A.C Zijdenbos A.P, Dawant B.M. Morphometric analysis of white matter lesions in mr images : method and validation. *IEEE Transactions on Medical Imaging*, 1994.

## Résumé

L'objectif de ce travail est d'élargir le champ d'utilisation des méthodes de fouille de données pour qu'elles soient adaptées aux spécificités et aux caractéristiques des données complexes comme les données médicales. Dans ce contexte, et appliqué aux données biomédicales, nous proposons une architecture de fusion dans un cadre possibiliste de données de type hétérogène (donnée numérique ou linguistique et une image de modalité IRM). La théorie des ensembles flous a été retenue pour la représentation de la donnée numérique issue du dossier médical ; quant à la représentation de l'image, une hybridation entre l'algorithme AntClust basé sur une population d'agents fourmis et l'algorithme flou FCM a été proposée.

Pour pallier aux défauts du FCM à savoir, l'initialisation et la sensibilité aux données aberrantes, AntClust servira de phase d'initialisation, découvrant ainsi automatiquement le nombre de classes, quand à la sensibilité aux données aberrantes, elle est traitée lors de la phase de fusion avec la donnée issue du dossier médical.

***Mots clés : Mots-clés : Fouille de données complexes, modalité IRM, Segmentation d'images, fusion de données, FCM, fourmis artificielles, antclust***

## Abstract

The aim of this work is to broaden the scope of use of methods of data mining to be adapted to the specificities and characteristics of complex data such as medical data. In this context, and applied to biomedical data, we propose a merger in a possibilistic two heterogeneous data types (numeric or linguistic data and an image modality MRI). The theory of fuzzy sets was chosen to represent the digital data from medical records, when displaying the image, Hybrids between AntClust algorithm based on a population of ant agents and fuzzy FCM algorithm was proposed. To overcome the shortcomings of FCM namely, initialization and sensitivity to outliers, AntClust serve as initialization phase, automatically discovering the number of classes, when the sensitivity to outliers, it is treated during the phase Merge with data from the medical record.)

***Keywords : complex data mining, MRI, clustering, data fusion, FCM, artificial ants, antclust***

## ملخص

الهدف من هذا العمل هو توسيع نطاق استخدام أساليب استخراج البيانات إلى أن تتكيف مع خصوصيات وسمات البيانات المعقدة مثل البيانات الطبية. في هذا السياق، تم تطبيقها على البيانات الطبية، اقترحنا دمج نوعين من البيانات غير المتجانسة (بيانات رقمية أو لغوية، و صورة الرنين المغناطيسيIRM) في نظرية «possilistic» .

تم اختيار نظرية مجموعات الضبابية لتمثيل البيانات الرقمية الآتية من السجلات الطبية، أما بالنسبة لصورة IRM، اقترحنا التنسيق بين الخوارزمي AntClust المستنبطة من (مستعمرات النمل الاصطناعي) وخوارزمي المتوسطات الضبابية FCM. للتغلب على أوجه القصور في FCM، ألا وهي التهيئة والحساسية للقيم المتطرفة، وضع AntClust بمثابة مرحلة التهيئة، واكتشاف عدد الطبقات تلقائيا، أما بالنسبة لحساسية القيم المتطرفة، فيتم التعامل معها خلال مرحلة الدمج مع بيانات السجل الطبي).

## الكلمات الرئيسية

التنقيب في البيانات المعقدة، صور الرنين المغناطيسي، التقسيم، دمج البيانات، وخوارزمي المتوسطات الضبابية FCM، النمل الاصطناعي، Antclust.