



HAL
open science

Fusion d'images multimodales pour l'aide au diagnostic du cancer du sein

Yosra Ben Salem

► **To cite this version:**

Yosra Ben Salem. Fusion d'images multimodales pour l'aide au diagnostic du cancer du sein. Traitement des images [eess.IV]. Ecole nationale supérieure Mines-Télécom Atlantique, 2017. Français. NNT : 2017IMTA0062 . tel-01803019

HAL Id: tel-01803019

<https://theses.hal.science/tel-01803019>

Submitted on 30 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

**UNIVERSITE
BRETAGNE
LOIRE**

THÈSE / IMT Atlantique

sous le sceau de l'Université Bretagne Loire

pour obtenir le grade de

DOCTEUR D'IMT Atlantique

Spécialité : Signal Image, Vision

École Doctorale Mathématiques et STIC
et en cotutelle avec l'École Nationale d'Ingénieurs de Tunis

Présentée par

Yosra Ben Salem

Préparée dans le département Image & traitement de
l'information

Thèse soutenue le 09 décembre 2017

devant le jury composé de :

Ali Khenchaf

Professeur, ENSTA Bretagne / président

Dorra Sellami Masmoudi

Professeur, Ecole Nationale d'Ingénieurs de Sfax / rapporteur

Jean-Paul Haton

Professeur, Université Henry Poincaré - Nancy 1 / rapporteur

Kamel Hamrouni

Professeur, Ecole Nationale d'Ingénieurs de Tunis / examinateur

Imed Riahd Farah

Professeur, Institut Supérieur d'Arts et Multimédias de la Manouba / examinateur

Basel Solaiman

Professeur, IMT Atlantique / directeur de thèse

Safya Belghith

Professeur, Ecole Nationale d'Ingénieurs de Tunis / invitée

**Fusion d'images multimodales
pour l'aide au diagnostic du
cancer du sein**

*« Si quelqu'un veut sérieusement rechercher la vérité,
il ne doit pas faire choix d'une science particulière;
elles sont toutes unies entre elles
et dépendent les unes des autres. »*

- Descartes -
Les Règles pour la direction de l'esprit
(1628 - 1629)

Remerciements

C'est avec un grand plaisir que je réserve cette page en signe de gratitude et de profonde reconnaissance à tous ceux qui ont contribué de près ou de loin à l'élaboration et à l'évaluation de ce travail.

Je voudrais d'abord remercier M. Ali Khenchef, professeur à l'Ecole Nationale Supérieure de Techniques Avancées (ENSTA) Bretagne, qui m'a fait l'honneur de présider le Jury.

Mes vifs remerciements vont à Mme Dorra SELLAMI MASMOUDI, professeur à l'Ecole Nationale d'Ingénieurs de Sfax (ENIS), pour m'avoir fait l'honneur de rapporter et de contribuer à l'évaluation de mon travail.

Mes remerciements vont également à M. Jean Paul Haton, professeur à l'Université Henri Poincaré Nancy1, qui m'a honoré de rapporter ma thèse, et d'évaluer mon travail.

Je remercie, aussi, M. Imed Riadh FARAH, professeur à l'Institut Supérieur d'Arts et Multimédias (ISAMM), pour avoir accepté d'examiner mon travail de thèse.

Je tiens à exprimer ma profonde reconnaissance à mon co-Directeur de thèse M. Basel SOLAIMAN, professeur et chef du département ITI à IMT Atlantique, pour sa disponibilité et son immense gentillesse et modestie qu'il m'a adressée durant les stages que j'ai passés au département ITI. La qualité de ses orientations et ses encouragements m'ont toujours motivé d'aller vers l'avant et surmonter tous les obstacles.

Je tiens à exprimer mon immense gratitude, aussi, à mon co-Directeur de thèse M. Kamel HAMROUNI, professeur à l'Ecole Nationale d'Ingénieurs de Tunis, pour sa disponibilité et la qualité de sa direction et de ses orientations tout au long de la réalisation de ce travail. Sa patience, sa bonté et sa modestie m'ont incité de puiser dans son savoir et m'ont poussé à avancer et progresser.

J'adresse mes sincères remerciements, également, à M. Karim SAHEB ETTA-BAA, Maître assistant à l'Institut Supérieur d'Informatique et des Technologies de Communication de Hammam Sousse, pour sa précieuse contribution à l'encadrement de ma thèse.

Je remercie tout particulièrement, Mme Sana BOUSSETTA, Chef du service Radiologie à l'Hopital Régional de BEN AROUS, qui a accepté de me fournir les images nécessaires à l'évaluation de ce travail de recherche. Ses explications et

ses orientations m'ont été d'une grande importance dans l'aboutissement de ce travail.

Mes vifs remerciements s'adressent aussi à tous les membres du laboratoire LR-SITI (ENIT) et du département ITI (IMT Atlantique) et à tous ceux qui m'ont aidé, encouragé et supporté tout au long de ces années de thèse.

Résumé

LE cancer du sein est le cancer le plus répandu chez les femmes de plus de 40 ans. En effet, des études ont montré qu'une détection précoce et un traitement approprié du cancer du sein augmentent de manière significative les chances de survie. La mammographie constitue le moyen d'investigation le plus utilisé dans le diagnostic des lésions mammaires. Cependant, cette technique peut être insuffisante pour montrer les structures du sein et faire apparaître les anomalies présentes et le médecin peut faire appel à d'autres modalités d'imagerie telle que l'imagerie IRM. Ces modalités sont généralement complémentaires. Par conséquent, le médecin procède à une fusion mentale des différentes informations sur les deux images dans le but d'effectuer le diagnostic adéquat. Pour assister le médecin et l'aider dans ce processus, nous proposons une solution permettant de fusionner les deux images. Bien que l'idée de la fusion paraisse simple, sa mise en oeuvre pose de nombreux problèmes liés non seulement au problème de fusion en général mais aussi à la nature des images médicales qui sont généralement des images mal contrastées et présentant des données hétérogènes, imprécises et ambiguës. Notons que les images mammographiques et les images IRM présentent des représentations très différentes des informations, étant donné qu'elles sont prises dans des conditions distinctes. Ce qui nous amène à poser la question suivante : Comment passer de la représentation hétérogène des informations dans l'espace image, à un autre espace de représentation uniforme.

Afin de traiter cette problématique, nous optons pour une approche de traitement multi-niveaux : niveau pixel, niveau primitives, niveau objet et niveau scène. Nous modélisons les objets pathologiques extraits des différentes images par des ontologies locales. La fusion est ensuite effectuée sur ces ontologies locales et résulte en une ontologie globale contenant les différentes connaissances sur les objets pathologiques du cas étudié. Cette ontologie globale sert à instancier une ontologie de référence modélisant les connaissances du diagnostic médical des lésions mammaires. Un raisonnement à base de cas est exploité pour fournir les rapports diagnostic des cas les plus similaires pouvant aider le médecin à prendre la meilleure décision.

Dans le but de modéliser l'imperfection des informations traitées, nous utilisons la théorie des possibilités avec les différentes ontologies. Le résultat fourni est présenté sous forme de rapports diagnostiques comportant les cas les plus similaires au cas étudié avec des degrés de similarité exprimés en mesures de possibilité. Un modèle virtuel 3D complète le rapport de diagnostic par un aperçu simplifié de la scène étudiée.

Mots clés : *Fusion de connaissances - Images multimodales - Représentation Conceptuelle des connaissances - Ontologies - Raisonnement à Base de cas - Représentation virtuelle 3D des connaissances.*

Abstract

THE breast cancer is the most prevalent cancer among women over 40 years old. Indeed, studies evinced that an early detection and an appropriate treatment of breast cancer increases significantly the chances of survival. The mammography is the most used tool in the diagnosis of breast lesions. However, this technique may be insufficient to evince the structures of the breast and reveal all anomalies. The doctor can use additional imaging modalities such as MRI (Magnetic Reasoning Image). Therefore, the doctor proceeds to a mental fusion of the different information on the two images in order to make the adequate diagnosis. To assist the doctor in this process, we propose a solution to merge the two images. Although the idea of the fusion seems simple, its implementation poses many problems not only related to the paradigm of fusion in general but also to the nature of medical images that are generally poorly contrasted images, and presenting heterogeneous, inaccurate and ambiguous data. Mammography images and IRM images present very different information representations, since they are taken under different conditions. Which leads us to pose the following question : How to pass from the heterogeneous representation of information in the image space, to another space of uniform representation from the two modalities ?

In order to treat this problem, we opt to a multilevel processing approach : the pixel level, the primitive level, the object level and the scene level. We model the pathological objects extracted from the different images by local ontologies. The fusion is then performed on these local ontologies and results in a global ontology containing the different knowledge on the pathological objects of the studied case. This global ontology serves to instantiate a reference ontology modeling knowledge of the medical diagnosis of breast lesions. Case-based reasoning (CBR) is used to provide the diagnostic reports of the most similar cases that can help the doctor to make the best decision.

In order to model the imperfection of the treated information, we use the possibility theory with the ontologies. The final result is a diagnostic reports containing the most similar cases to the studied case with similarity degrees expressed with possibility measures. A 3D symbolic model complete the diagnostic report with a simplified overview of the studied scene.

Keywords : *Knowledge Fusion - Multimodal Images - Conceptual knowledge representation - Ontologies - Case Base Reasoning - Symbolic Knowledge Representation.*

Table des matières

1	Traitement d'images médicales du sein	8
1.1	Images mammographiques	9
1.1.1	Examen mammaire par mammographie	9
1.1.2	Difficultés du traitement d'images mammographiques	11
1.1.3	Travaux du traitement d'images mammographiques	13
1.2	Images IRM mammaires	20
1.2.1	Examen mammaire par IRM	20
1.2.2	Apport de l'IRM	22
1.2.3	Difficulté du traitement d'images IRM	23
1.2.4	Travaux sur le traitement d'images IRM mammaires	23
1.3	Conclusion	27
2	Information et modélisation des connaissances	28
2.1	Les images médicales et les concepts de donnée, information et connaissance	29
2.2	Gestion des imperfections	32
2.2.1	Imperfections de l'information	32
2.2.2	Modélisation des imperfections de l'information	33
2.3	Modélisation de connaissances à partir des images	36
2.3.1	Modèles	37
2.3.2	Signatures	38
2.3.3	Ontologies	39
2.4	Modélisation 3D de scènes réelles	41
2.4.1	La modélisation 3D dans le domaine médical	42
2.4.2	La modélisation 3D du sein féminin	44
2.5	Conclusion	46
3	Fusion d'images et aide au diagnostic médical	47
3.1	Fusion d'information	48
3.1.1	Les niveaux de fusion	49
3.1.2	La fusion d'images mammaires	50
3.2	Aide au diagnostic par raisonnement à base de cas	53
3.2.1	Principe du raisonnement à bases des cas	54
3.2.2	Travaux sur le raisonnement à base de cas en diagnostic médical	55

3.3	Conclusion	58
4	Approche proposée pour la fusion multimodale	59
4.1	Architecture de l'approche proposée	60
4.2	Génération des ontologies locales	63
4.2.1	Couche Données : Pré-traitement et Segmentation	64
4.2.2	Couche Informations : Extraction de caractéristiques	65
4.2.3	Construction des ontologies locales	71
4.3	Fusion des ontologies locales	74
4.3.1	Mesure de similarité	74
4.3.2	Fusion ontologique	77
4.4	Instanciation	79
4.5	Raisonnement et Aide au diagnostic	79
4.5.1	Le cas requête	81
4.5.2	La base de cas	82
4.5.3	Mesure de similarité	83
4.5.4	Les cas possibles	83
4.5.5	Instanciation de nouveaux cas	83
4.6	Modèle symbolique de représentation des connaissances	84
4.6.1	Choix du modèle de représentation de connaissances	84
4.6.2	Représentation des connaissances et intégration des données	85
4.7	Conclusion	86
5	Résultats et Evaluation	89
5.1	Bases d'images	90
5.2	Description du scénario d'application	90
5.3	Résultats de l'étape du traitement des images	92
5.3.1	Résultats du prétraitement et de la segmentation	93
5.3.2	Résultats de l'extraction de caractéristiques	93
5.3.3	Validation des ontologies locales résultantes	94
5.4	Résultats et évaluation de la fusion des ontologies locales	97
5.5	Résultats et évaluation du raisonnement médical basé sur l'ontologie globale	99
5.5.1	Validation de cas similaires	100
5.5.2	Vérification des diagnostics proposés	101
5.6	Evaluation globale	102
5.7	Conclusion	103
	Bibliographie	108
A	Résultats du pré-traitement des images mammaires	121
A.1	Prétraitement et segmentation des Mammographies	121
A.2	Prétraitement et segmentation des IRM	124

Table des figures

1	Exemple de fusion de deux ontologies	5
1.1	Variation de la densité mammaire suivant l'âge de la patiente	10
1.2	Mammographie de seins normaux	10
1.3	Les différentes formes des masses	11
1.4	Les différents contours des masses	11
1.5	Les différents types de micro-calcification par la classification BiRADS	12
1.6	Les différentes formes des micro-calcifications	12
1.7	Les différentes formes de répartitions des micro-calcifications	13
1.8	Les formes des masses observées sur une IRM	21
1.9	Les contours des masses observées sur une IRM	21
1.10	Les différents types de rehaussement sur une IRM	22
1.11	Complémentarité des examens IRM et Mammographie	22
2.1	Hérarchie des concepts Donnée, Information et Connaissance suivant Ackoff [Ackoff, 1989]	30
2.2	Modèle numérique du fantôme FATMA [KW A, 2011]	45
2.3	Modèle numérique du fantôme FATMA [Bliznakova et al., 2012]	45
3.1	Les niveaux de la fusion d'images	50
3.2	Le modèle cyclique 4Rs du CBR	55
4.1	Approche proposée	61
4.2	Les différents niveaux d'interprétation de scènes	62
4.3	Vue d'ensemble des niveaux de traitement	63
4.4	Principe de l'approche proposée	64
4.5	Exemple de pré-traitement des images mammographiques et IRM	65
4.6	Exemple de segmentation des images mammographiques	66
4.7	Principe de l'extraction de caractéristiques des différentes formes pathologiques	66
4.8	Exemple de primitives extraites d'une forme pathologique	68
4.9	Processus d'extraction de caractéristiques sémantiques	70
4.10	Exemple d'une distribution de possibilité	70
4.11	Extrait de l'Ontologie MAMMO utilisée pour la génération des ontologies locales	72

4.12	La description d'un objet dans l'ontologie locale	73
4.13	La spécification des descripteurs d'une objet de l'ontologie	73
4.14	La description des relations dans l'ontologie locale	74
4.15	Principe de la fusion des ontologies locales	75
4.16	matrice de correspondance des objets	77
4.17	Exemple de fusion de deux ontologies locales	79
4.18	Extrait d'un exemple de fichier associé à un individu de l'ontologie .	80
4.19	Principe du raisonnement à base de cas d'aide au diagnostic médical	81
4.20	Approche proposée pour le raisonnement et l'aide au diagnostic . . .	82
4.21	Définition du modèle de représentation virtuelle de connaissances pour la représentation des objets pathologiques	85
4.22	Calcul de paramètres à partir de l'image IRM	86
4.23	Algorithme de modélisation virtuelle de connaissances d'un cas étudié	87
4.24	Aperçu du repère utilisé pour identifier les coordonnées du centre de l'objet pathologique	87
5.1	Procédure globale de validation de l'approche proposée	91
5.2	Procédure du traitement d'images mammaires	92
5.3	Principe de l'évaluation de la spécification des objets pathologiques extraits	94
5.4	Processus de validation d'une ontologie locale	95
5.5	Évaluation de performance de la phase du raisonnement	102
A.1	Procédure du prétraitement des images mammographiques	122
A.2	Procédure du prétraitement des images IRM	125
A.3	Objets pathologiques observées sur une image IRM	125

Liste des tableaux

1.1	Résumé des travaux effectués sur le traitement d'images mammographiques	19
1.2	Résumé des travaux effectués sur le traitement d'images IRM mammaires	26
4.1	Les coefficients de pondération des différents concepts	76
4.2	Algorithme du peuplement de l'ontologie de domaine	79
4.3	Définition des modèles virtuels des objets	85
5.1	Exemple d'évaluation de caractéristiques visuelles extraites d'une forme pathologique	93
5.2	Exemple de d'une matrice de validation d'une ontologie locale	96
5.3	Mesures de performance des ontologies locales construites à partir des images mammaires	97
5.4	Exemple de matrice d'appariement des objets des ontologies locales .	98
5.5	Matrice d'évaluation de l'appariement des objets	98
5.6	Évaluation de la fusion des ontologies locales	99
5.7	Algorithme d'évaluation des différents cas similaires	100
5.8	Extrait d'une matrice d'évaluation	100
5.9	Extrait d'une matrice de vérification de la contradiction et de l'incohérence des diagnostics potentiels	101
5.10	Évaluation de l'approche proposée par les experts du domaine médical	103
A.1	Résultat des filtres les plus utilisés sur une image mammographique	122
A.2	Résultats du filtrage par certaines techniques avec une étape de minimisation du caractère flou	123
A.3	EQM des résultats issus des filtres simples	123
A.4	EQM des résultats issus des filtres avec étape de minimisation du flou	123
A.5	Évaluation du médecin des résultats issus des filtres simples	124
A.6	Résultats de la segmentation semi-automatique avec l'algorithme "LevelSet" d'une mammographie prétraitée	124
A.7	Résultats de l'étape de filtrage appliqué sur le volume entier (coupe 54)	126
A.8	Résultats de l'étape de filtrage appliqué sur le volume entier (coupe 56)	127
A.9	EQM des résultats issus des filtres appliqués sur une IRM traité en volume entier	128

A.10 EQM des résultats issus des filtres appliqués sur une IRM traitée en volume entier avec étape de minimisation du flou	128
A.11 Évaluation du médecin des résultats issus des filtres simples appliqués sur une IRM traitée en volume entier	128
A.12 Résultats de l'étape de filtrage appliquée sur une IRM coupe par coupe (coupe 54)	129
A.13 Résultats de l'étape de filtrage appliquée sur une IRM coupe par coupe (coupe 56)	130
A.14 EQM des résultats issus des filtres appliqués sur une IRM traitée coupe par coupe	131
A.15 EQM des résultats issus des filtres appliqués sur une IRM traitée coupe par coupe avec étape de minimisation du flou	131
A.16 Évaluation du médecin des résultats d'une IRM traitée coupe par coupe	131

Liste des acronymes

[A]

- ADMM** Alternating Direction Method of Multipliers
- AGATE** Amsterdam Gated cardiac phantom
- AGATE** Amsterdam Gated Cardiac Phantom
- APR** Axis Parallel Rectangles

[B]

- BCGO** Breast Cancer Grading Ontology
- BFC** Bias Field Correction
- BGSA** Binary Gravitational Search Algorithm

[C]

- CBR** Case Base Reasoning
- CC** Cranio Caudal
- CGSA** Continuous Gravitational Search Algorithm
- CkNN** Citation kNN
- CLAHE** Contrast Limited Adaptive Histogram Equalization
- CSB-TV** Constraint Split Bergman-Total Variation
- CSB-TV** Constraint Split Bergman combinée avec Total Variation

[D]

- DCT** Discrete Cosine Transform
- DD** Diverse Density
- DeBRa** Detailed Breast model for Radiological studies
- DEWNN** Differential Evolution optimization Wavelet Neural Network
- DSS** Decision Support System
- DWT** Daubechies Wavelet

[E]

- EM Expectation-maximization
- EQM Erreur Quadratique Moyenne

[F]

- FCM Fuzzy C-Means
- FEM Finite Element Method

[G]

- GLCM Gray-Level Co-occurrence Matrix
- GLRLM Gray-Level Run, Length Matrix

[I]

- IRM Image par Résonance Magnétique

[K]

- kNN k-nearest neighbors

[L]

- LBPH Local Binary Patterns Histograms

[M]

- MGSA Mixed Gravitational Search Algorithm
- MIL Multiple Instance Learning
- MLO MedioLateral Oblique
- MLP Multi Layer Perceptron
- MRF Markov Random Fields
- MRI Magnetic Resonance Image
- MSGSA Mixed Gravitational Search Algorithm

[N]

- NN Neural Network

[O]

- OLTP On Line Transaction Processing

[P]

PCET Polar Complex Exponential Transform

PET Positron Emission Tomography

PSO Particle Swarm Optimization

[R]

RBC Raisonnement à Base de Cas

RBF Radial Basis Function

RID Regularised Isotropic Diffusion

[S]

SCSM Support-pixel Correlation Statistical Method

SIFT Scale-Invariant Feature Transform

SIFT Scale Invariant Feature Transform

SIL Single Instance Learning

SKFCM Spatial Kernelled Fuzzy C-Means

SVM Support Vector Machine

SVM Support Vector Machine

[T]

TNM Tumor - Node - Methastasis

Introduction Générale

LE domaine médical est un domaine reposant sur des théories et des règles spécifiques. Cependant, il représente aussi un domaine faisant intervenir un processus cognitif complexe amenant dans la majorité des cas à une prise de décision. Cette décision résulte d'un raisonnement effectué par le médecin sur certaines données et informations collectées auprès du patient et des différents examens qu'il réalise pour confirmer ou infirmer certaines hypothèses qu'il fait. Ce processus est appelé "*un diagnostic médical*".

Dans un premier temps, le médecin collecte le maximum d'information sur la situation d'un patient à travers des symptômes manifestés ou du ressenti du patient. A travers ces informations, il forme une première visualisation abstraite de la scène dans son mental. Afin de mieux cerner le problème, il procède à certains examens cliniques faisant appel à différentes techniques d'investigation permettant de faire imager l'intérieur du corps du patient. Ces techniques apportent un gain significatif dans le diagnostic médical permettant de montrer au médecin la structure interne des organes examinés pouvant être affectés par certaines anomalies. Par le biais d'une analyse des différents aspects de l'organe observé, il parvient à vérifier certaines suppositions qu'il fait à partir des premières données collectées. S'il intercepte une ambiguïté le limitant à prendre la bonne décision, il a recours à une autre technique d'investigation cherchant à avoir d'autres informations permettant d'enlever l'ambiguïté et de compléter les informations manquantes. Face aux différentes formes de données et d'informations, le médecin tente de reconstruire la scène en imaginant une représentation virtuelle combinant toutes les informations dont il dispose. Cette procédure de fusion est d'autant plus cruciale que les sources renferment des informations diverses et hétérogènes afin d'avoir une représentation la plus complète possible.

Par la suite, le médecin s'appuie sur son expérience et son expertise dans le but de faire le diagnostic adéquat au cas traité, tout en tenant compte des imperfections entachant les informations. Il a recours généralement à l'historique des cas diagnostiqués précédemment, cherchant les cas similaires afin de tirer profit des décisions prises en ce moment là. Ce processus est désigné par "le raisonnement à base de cas similaires".

Le développement continu des technologies de l'imagerie médicale, et l'avancée des travaux de recherche dans les domaines de la vision par ordinateur et de l'intelligence artificielle, ont pu contribuer à l'évolution de la pratique médicale en proposant aux médecins des systèmes d'aide au diagnostic. Ces systèmes apportent de l'aide aux experts de la santé, leur facilitant le processus cognitif suivi en proposant une assistance aux médecins allant de la modélisation des informations et des connaissances au raisonnement faisant appel à l'analyse et l'interprétation des images. Ces systèmes apportent un gain pronostic surtout pour les cas difficiles (organe ayant une structure difficile à analyser sur une image, images imparfaites masquant certaines anomalies, etc...)

Le cancer du sein est le cancer le plus répandu chez les femmes de plus de 40 ans présentant un risque alarmant pour leur survie. En effet, des études ont montré que plus tôt un diagnostic est effectué, plus les chances de survie sont importantes. Elles prétendent que la découverte précoce des petites formes pathologiques apporte un gain considérable impliquant une minimisation du taux de mortalité causé par ce cancer. La mammographie constitue le moyen d'investigation le plus utilisé pour diagnostiquer une anomalie du sein. Cependant, cette technique peut être insuffisante pour montrer les structures du sein et faire apparaître les lésions présentes, vue qu'elle nécessite une compression du sein et qu'elle représente une simple projection d'une face du sein suivant un certain angle. Afin de mieux évaluer la scène et de surmonter le problème de l'incomplétude et de l'ambiguïté, le médecin peut faire appel à d'autres modalités d'imagerie telle que l'échographie ou l'imagerie IRM. Ces modalités sont généralement complémentaires pour réaliser un diagnostic efficace. Le médecin a besoin, donc, de fusionner mentalement les deux représentations pour avoir une vue globale de la structure du sein.

Dans les domaines de la vision par ordinateurs et du traitement des images, la fusion représente un axe de recherche très prometteur dans plusieurs champs d'application (web, imagerie satellitaire, imagerie médicale, etc...). L'objectif étant de fournir aux experts un système se chargeant de la tâche de combiner les différentes données et informations dans une représentation unique permettant de les assister durant leurs processus de diagnostic et de prise de décision. Bien que l'idée de la fusion paraisse simple, sa mise en oeuvre pose de nombreux problèmes liés non seulement au problème de fusion en général mais aussi à la nature des images médicales qui sont généralement des images mal contrastées et présentant des données imprécises et ambiguës. Dans un processus de fusion, l'étape la plus importante est celle du recalage, qui est la procédure par laquelle on fait associer chaque point d'une image à son homologue dans l'autre image. Cette étape devient plus complexe si les structures de l'image ont des formes complexes et difficiles à identifier. C'est le cas des images mammaires. En effet, le sein est un organe malléable à représentation compliquée sur une image médicale. En effet, sur une image mammographique, le sein est comprimé entre deux plaques ce qui réduit son volume et permet sa capture d'un certain angle de projection. Pour un examen

d'échographie, la femme doit être allongée sur le dos. Ainsi, le sein est relâché et penché plutôt vers l'arrière. Ce qui produit une image très différente. Durant un examen d'IRM, la femme est allongée sur le ventre, permettant, alors, de libérer le sein qui est étiré vers le bas. Une autre représentation distincte est obtenue. Cette divergence des conditions, des positions et des angles de prises de vue procure des représentations très hétérogènes et complexes à recalculer puis à fusionner. C'est pour cette raison qu'un passage vers un autre espace de représentation des informations et des connaissances est inévitable pour le processus de fusion des images mammaires.

C'est dans ce contexte que se situe notre travail de recherche faisant l'objet de cette thèse et visant à proposer une solution permettant de dépasser les limites de l'hétérogénéité des sources d'informations et les différentes imperfections pour présenter une fusion aussi fiable que possible des données mammaires ; dans l'objectif de servir l'expert de la santé d'une représentation complète renfermant toute les informations nécessaires pour mener à bien son diagnostic. En d'autres termes notre travail se situe dans le cadre de la représentation combinée d'informations et de connaissances multiples constituant une scène observée par différents capteurs. Ce qui nous mène à poser la question : Comment passer de l'espace de l'image à un autre espace de représentation des connaissances ?

Notre approche est une approche multi-niveaux ascendante permettant de passer du bas niveau de représentation des données (image) à un haut niveau d'interprétation de scène en tenant compte des imperfections pouvant survenir tout au long du traitement. Notre approche est constituée principalement de cinq modules (Figure1).

Le premier module consiste en une modélisation conceptuelle des connaissances permettant de représenter les données des images en un modèle de représentation de connaissances approprié. Cette modélisation peut être garantie par une ontologie que nous qualifions de "locale" qui est construite à partir des connaissances extraites d'une image. En effet, l'ontologie est un nouveau paradigme permettant de formuler les connaissances suivant une spécification conceptuelle hiérarchisant des concepts et les liant par des relations. Une telle modélisation nécessite un traitement préalable de l'image afin d'extraire les connaissances pertinentes. Cette ontologie locale à l'image permet de décrire les objets pathologiques qui peuvent exister dans une image à travers certains concepts définis dans le domaine mammographique.

Le deuxième module permet la fusion des connaissances en combinant les différentes connaissances extraites des images en une représentation unique. Une mesure de similarité est effectuée afin d'apparier les objets similaires et distinguer les objets différents pour construire un modèle complet disposant d'une structure non redondante. Une ontologie globale résulte du processus de la fusion constituant une instance d'un cas pathologique du domaine mammographique.

Le troisième module est un module d'instanciation d'une ontologie de domaine mammographique permettant de l'enrichir par des instances du monde réel.

Le quatrième module consiste à utiliser un raisonnement à base de cas mettant

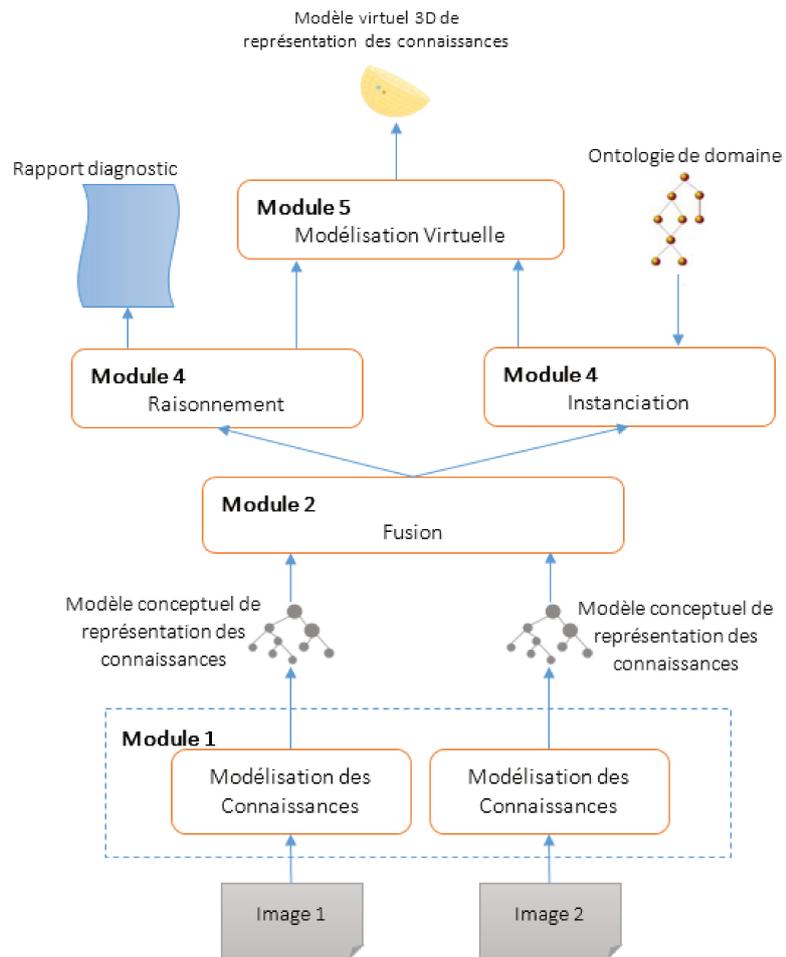


FIGURE 1 – Exemple de fusion de deux ontologies

à la disposition du médecin une proposition de certains diagnostics effectués pour des cas similaires l'aidant au processus cognitif qu'il fait afin de prendre la bonne décision. L'expert peut se servir des diagnostics potentiels ou prescrire un nouveau diagnostic s'il juge les propositions non adaptées au cas traité.

Le dernier module permet la modélisation virtuelle 3D des connaissances procurant un modèle virtuel représentant de façon simplifiée les objets pathologiques incorporés dans le sein. Il permet donc de reproduire symboliquement la scène observée.

Durant ce processus de traitement, nous devons tenir compte des différentes imperfections affectant les données, les informations et les connaissances à chaque phase. La théorie des possibilités représente un cadre puissant et prometteur pour la modélisation des informations incertaines dans un cadre ambigu. Nous tirons profit des avantages de cette théorie pour la représentation des connaissances extraites des images dans un premier temps ; et dans un deuxième temps, dans le processus de raisonnement à base de cas. Ainsi, l'expert aura un aperçu des différents diagnostics jugés possible au cas traité.

Le présent rapport présentant cette thèse s'articule autour de cinq chapitres :

Le premier chapitre rapporte une étude des différents travaux sur le traitement des images mammaires. Dans un premier temps, un aperçu général sur les images mammographiques est effectué. Cet aperçu présente la procédure de l'examen mammaire par mammographie, l'apparence des différentes lésions sur l'image mammographique et le diagnostic effectué par les médecins en se servant des images. Ensuite, la problématique du diagnostic par mammographie est posée, expliquant les problèmes liés au traitement des images. Une revue de littérature est établie exposant différents travaux récents étudiant les différents traitements effectués sur les images mammographiques. Dans une deuxième partie du chapitre, une étude des images IRM est présentée de la même manière, expliquant le déroulement de l'examen et le diagnostic des objets pathologiques tout en montrant son apport par rapport à la mammographie, la problématique du traitement des images IRM et établissant une revue de littérature des différents travaux étudiant les images IRM.

Le deuxième chapitre présente une analyse de l'état de l'art des approches utilisées pour la modélisation conceptuelle et virtuelle des informations et des connaissances. Tout d'abord, nous expliquons les spécificités des images médicales et nous exposons les types d'informations traitées par les systèmes d'interprétation de scènes et d'aide à la décision. Ensuite, nous détaillons les principaux types d'imperfections affectant les informations et les différents cadres adoptés pour leur modélisation. Puis, nous présentons les méthodes de modélisation que nous utilisons dans notre approche en rapportant les travaux récents qui les appliquent afin de modéliser les connaissances extraites des images. Finalement, nous explorons la notion de modélisation virtuelle 3D et les travaux effectués dans le cadre de la modélisation virtuelle du sein.

Le troisième chapitre présente l'état de l'art de la fusion d'information et du raisonnement à base de cas. dans un premier temps nous définissons la notion de fusion, les niveaux d'application de la fusion et nous présentons les travaux effectués dans chacun de ces niveaux. Ensuite, nous expliquons le principe du raisonnement à base de cas et nous rapportons quelques travaux exploitant ce concept dans le domaine médical.

Le quatrième chapitre détaille l'approche que nous proposons pour la fusion des images multimodales en spécifiant ses différents modules :

- Le module de la modélisation conceptuelle des connaissances qui débute par une étape d'extraction de connaissances à partir des primitives visuelles déduites des images ; ensuite une étape de modélisation de ces connaissances par des ontologies locales possibilistes est effectuée en se basant sur la structure d'une ontologie de domaine et sur le principe de la théorie des possibilités ;
- Le module de la fusion permettant de fusionner les ontologies locales définies à partir des images en utilisant un opérateur adaptatif combinant une fusion possibiliste et une fusion déterministe ;
- Le module de l'instanciation de l'ontologie de domaine en vue de son enrichissement. Afin d'instancier les concepts possibilistes nous utilisons un fichier xml contenant la description possibiliste des concepts correspondants. Ce fichier est associé à l'instance correspondante ;
- Le module du raisonnement permettant l'interprétation de la scène observée par un raisonnement à base de cas associé au principe de la théorie des possibilités pour fournir les diagnostics estimés possibles pour le cas traité.
- Le module de la modélisation virtuelle des connaissances produisant une représentation simplifiée de la scène modélisant le sein qui comporte les différents objets pathologiques.

Le dernier chapitre expose les différentes expérimentations effectuées sur les images mammaires (mammographie et IRM) et les différents tests de validation permettant d'évaluer la performance de l'approche présentée dans le chapitre précédent. Dans un premier temps nous présentons les bases de données dont nous nous servons pour l'évaluation. Dans un deuxième temps, nous détaillons les tests effectués pour chaque module de l'approche, présenté auparavant, en évaluant les différents résultats.

En conclusion, nous rappelons les objectifs atteints dans ce travail en présentant un bilan critique de notre étude tout en proposant des perspectives à court terme et à long terme permettant d'améliorer le système d'aide au diagnostic présenté et d'ouvrir de nouvelles pistes de recherche dans l'axe de la vision par ordinateur de l'intelligence artificielle.

Traitement d'images médicales du sein

Sommaire

1.1	Images mammographiques	9
1.1.1	Examen mammaire par mammographie	9
1.1.2	Difficultés du traitement d'images mammographiques	11
1.1.3	Travaux du traitement d'images mammographiques	13
1.2	Images IRM mammaires	20
1.2.1	Examen mammaire par IRM	20
1.2.2	Apport de l'IRM	22
1.2.3	Difficulté du traitement d'images IRM	23
1.2.4	Travaux sur le traitement d'images IRM mammaires	23
1.3	Conclusion	27

Une image est généralement une représentation visuelle d'une scène dont elle contient les informations visibles. Étant une des plus importantes révolution Une des plus amples révolutions de la médecine, l'imagerie médicale, a contribué au grand progrès du diagnostic des différentes pathologies attaquant l'organisme humain grâce à la visualisation des structures internes des organes. Depuis l'apparition de la radiographie en 1895, l'imagerie médicale n'a pas cessé d'évoluer : de la simple fonction photographique interne des organes informant le médecin sur la structure de leur tissu ou de l'aspect des lésions à une visualisation plus réelle renseignant sur l'activité et l'état réel des organes. Si la mammographie est la technique la plus utilisée depuis son apparition, elle est dans plusieurs situations conjointes par d'autres techniques telle que l'IRM. C'est le cas du diagnostic des maladies du sein utilisant souvent l'image mammographique couplée dans certains cas à l'IRM.

Le traitement d'images est une discipline qui définit des méthodes permettant d'améliorer la qualité des images ou d'en extraire les informations pertinentes aidant l'expert à l'interprétation de la scène étudiée. Avec l'évolution des approches de traitement d'images médicales, l'interactivité avec l'image est possible et le diagnostic est plus pertinent et plus facile.

Dans ce chapitre, nous présentons les spécificités du traitement d'images mammographiques et IRM et leurs apports dans le diagnostic médical des maladies du sein, en rapportant les travaux proposés pour leurs traitements.

1.1 Images mammographiques

Selon des études, la mammographie est le moyen le plus utilisé pour le diagnostic du cancer du sein dans un premier temps, d'une part, parce que l'examen n'est pas cher et donc accessible à tout le monde, et d'autre part elle permet dans la majorité des cas un diagnostic efficace. Dans une étude menée en Australie la mort causée par le cancer du sein diminue de 50 % chez les femmes s'exposant à un diagnostic par mammographie [Bonnie et al., 2016]. Généralement sur une mammographie les médecins cherchent à détecter les calcifications (Micro ou Macro), des masses, ou à examiner la densité mammaire. Ils peuvent aussi faire des comparaisons avec des anciennes images afin de capter les petits changements. Les statistiques des résultats par mammographie montrent un taux de spécificité pouvant atteindre 95% et un taux de sensibilité allant jusqu'à 70% pour les femmes de plus de 30 ans. Selon certains experts, la superposition des structures fibreuses majorée par la présence de la densité mammaire peut masquer certaines anomalies.

1.1.1 Examen mammaire par mammographie

La mammographie est une radiographie utilisant les rayons X appliqués sur le sein, comprimé entre deux plaques, pour détecter la présence ou non d'une tumeur. Le diagnostic par cette modalité consiste à analyser deux images pour chaque sein : une image représentant la face du sein, et une autre visualisant le sein en profil. Les différentes zones du sein absorbent les rayons envoyés en fonction de leur texture et seront par la suite représentées par des niveaux de gris différents. Généralement, sur une mammographie, les anomalies sont représentées par des régions blanches de formes arrondies ou irrégulières et de contours variés. Cependant chez une jeune femme, le tissu glandulaire du sein est représenté de la même façon, ce qui rend la détection d'anomalie dans certains cas très difficile. Avec l'âge, le tissu glandulaire diminue et le tissu graisseux devient prépondérant (Fig. 1.1) ainsi le ton de l'image mammographique devient plutôt grisâtre et par la suite la détection des régions d'anomalie devient plus facile. En observant des seins normaux, on doit trouver une symétrie de densité mammaire (Fig. 1.2) Une détection d'asymétrie doit indiquer

la présence d'une anomalie. Les spécialistes distinguent entre deux grandes classes d'anomalies de sein : les masses et les calcifications.



FIGURE 1.1 – Variation de la densité mammaire suivant l'âge de la patiente

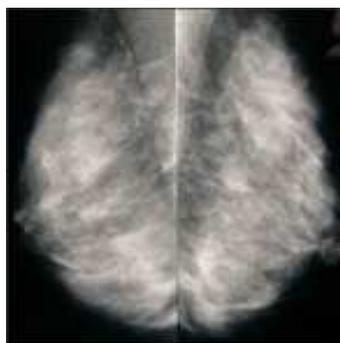


FIGURE 1.2 – Mammographie de seins normaux

a. Etude des masses

Les masses sont définies comme des lésions au niveau du tissu du sein. Elles sont généralement plus denses que le tissu et apparaissent sous la forme de régions blanches. Suivant la localisation des masses, leurs formes et leurs contours, le médecin peut décider de sa gravité (kystes, tumeur bénigne ou maline,...). On distingue cinq formes de masse qu'on peut détecter par une mammographie [Heron, 2009] (Fig. 1.3) : une forme ronde, une forme ovale, une forme lobulée, une forme irrégulière ou encore une distorsion architecturale. Les contours des masses dans l'image du sein (Fig. 1.4) peuvent être circonscrits, masqués, micro-lobulés, non distincts (ou effacés) ou encore spéculés.

Les masses régulières de forme ronde, ovale ou lobulée avec des contours définis sont généralement interprétés comme des lésions bénignes. Elles sont à l'origine des kystes, des fibroadénomes, des tumeurs phylloïdes,...[Heron, 2009]. Les masses ayant des contours irréguliers sont généralement cancéreuses. La plupart des masses observées sur une mammographie ne sont pas parfaitement régulières, ce qui implique



FIGURE 1.3 – Les différentes formes des masses

la présence de certaines irrégularités ou lobulations. Suivant le nombre de lobules observés, on peut statuer la malignité de la tumeur détectée.

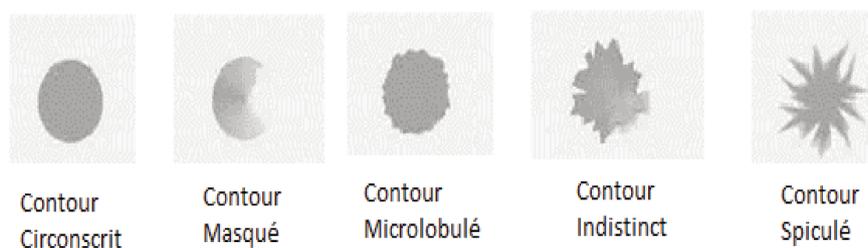


FIGURE 1.4 – Les différents contours des masses

b. Etude des micro-calcifications

Les calcifications sont des éléments souvent interprétés par des mammographies. Elles sont dues essentiellement à "des sécrétions cellulaires" ou encore à "des débris nécrotiques cellulaires". Elles sont généralement situées à l'intérieur des canalicules mammaires, dans les lobules, dans les structures vasculaires ou encore dans le tissu conjonctif intra-lobulaire. Les calcifications peuvent être des pathologies malignes (petites calcifications de moins de 0.5 mm de diamètre et de formes irrégulières) ou bénignes (grandes calcifications de forme régulière ronde ou ovale),...[Heron, 2009]. Suivant certaines classifications, on peut décider de la dangerosité des calcifications.

La classification la plus utilisée par les médecins est la classification BI-RADS [Luc, 2004]. Suivant cette classification les micro-calcifications sont diagnostiquées selon leur type (Fig. 1.5), forme (Fig. 1.6) et répartition (Fig. 1.7). Plus les micro-calcifications ont une forme irrégulière et sont plus dispersées, plus il s'agit d'un indice de malignité.

1.1.2 Difficultés du traitement d'images mammographiques

Les images mammographiques sont des images en niveaux de gris contenant des niveaux d'intensités différents et sont caractérisées par un faible contraste, ce qui



FIGURE 1.5 – Les différents types de micro-calcification par la classification BiRADS



FIGURE 1.6 – Les différentes formes des micro-calcifications

conduit à des difficultés dans l'interprétation et causer quelques problèmes au niveau du diagnostic :

- **Faux positifs** : malgré l'absence de lésions réelles, un aspect anormal de la mammographie peut conduire à de faux positifs. Ce type de diagnostic est fréquent chez les femmes jeunes où la densité mammaire est élevée augmentant l'ambiguïté de la représentation sur l'image. Le diagnostic des faux positifs requiert généralement un autre examen pour un traitement adéquat. Il représente une cause d'anxiété chez la majorité des femmes diagnostiquées.
- **Faux négatifs** : Un aspect normal de la mammographie peut conduire à un faux négatif malgré la présence de réelles lésions. Ce type de diagnostic est aussi très fréquent chez les femmes jeunes qui ont une densité mammaire élevée pouvant masquer les anomalies surtout de très petite taille sur une mammographie. Les statistiques montrent qu'entre 10 % et 30 % [Society, 2016] des cas présentent des faux négatifs. Ceci limite la détection précoce, à des premiers stades, du cancer ce qui entraîne un retard sur le diagnostic et augmente la gravité du cas et les risques de décès.

Ceci peut conduire les médecins à un sur-diagnostic et ensuite un sur-traitement du cas. En effet, le médecin ne peut pas être sûr de la malignité et du stade de la lésion. Selon les observations, il peut traiter des anomalies qui ne représentent pas un vrai problème. Selon certaines statistiques, entre 0 et 10% des cas sont sur-diagnostiqués [Society, 2016].

Les formes pathologiques apparaissent généralement sur une mammographie comme des zones blanches d'intensité élevée. De la même manière, le tissu adipeux



FIGURE 1.7 – Les différentes formes de répartitions des micro-calcifications

est représenté sur une mammographie par des zones blanches à haute intensité. Ajoutant à ceci le fait qu'une mammographie représente une projection 2D du sein comprimé, ce qui peut masquer plusieurs zones pouvant être suspectes ou cacher une zone pathologique à travers le tissu adipeux. Tous ces facteurs de la complexité de la représentation d'une image mammographique, contribuent à la difficulté que trouvent les chercheurs dans son analyse. C'est pour cela que, afin de fournir l'interprétation exacte, l'analyse d'une image mammographique reste un défi.

1.1.3 Travaux du traitement d'images mammographiques

Le traitement d'images mammographiques reste un champ de recherche très prometteur visant à apporter une aide au diagnostic médical des pathologies mammaires. Dans ce texte les travaux peuvent être classés en trois catégories : le filtrage d'images, la segmentation d'images, et la détection et la classification des formes pathologiques.

a. Prétraitement d'images mammographiques

Le pré-traitement est l'un des traitements les plus étudiés dans le domaine du traitement d'images. Il vise à améliorer la qualité de l'image (suppression du bruit, accentuation des contours, etc...) pour faciliter son interprétation. Il fait l'objet de plusieurs travaux dans le but de préparer l'image au traitement désiré.

Dans ce contexte deux chercheurs [Vedanarayanan and Nandhitha, 2016] ont proposé une méthode pour le pré-traitement d'images mammographiques consistant à utiliser le Filtre de Gabor. Dans un premier temps, une conversion de l'image en niveaux de gris est effectuée. Ensuite, un filtre de gabor de dimension 3×3 est appliqué. Le filtre de Canny est utilisé pour la détection de contours. D'après les résultats trouvés, on montre que le filtre de Gabor a permis la suppression des contours indésirables dans les images. Les images filtrées avaient une grande valeur de déviation standard et une faible entropie.

Dans leurs travaux, Makandar and Halalli [Makandar and Halalli, 2015] propose une nouvelle approche combinant certaines techniques en traitement d'images pour résoudre le problème de la détection précoce des aspects anormaux sur une image mammographique. Le traitement est effectué suivant deux grandes étapes : une

étape de pré-traitement ayant pour objectifs la suppression de l'arrière-plan de l'image grâce à une binarisation avec un seuil de 0.1 et la suppression du muscle pectoral avec la méthode de croissance des régions modifiée de façon à déterminer automatiquement le "seed point" suivant l'orientation de l'image. Finalement, un filtrage est appliqué à l'image afin d'améliorer son contraste en utilisant le filtre de Wiener et la technique d'égalisation d'histogramme (CLAHE). Une étape d'identification des masses combinant les algorithmes de Sobel pour l'identification du gradient, du "Watershead" pour la segmentation, des opérateurs morphologiques pour l'optimisation des résultats et du contour actif pour l'identification des zones suspectes.

Dans un autre travail, Makandar et Halalli [Makandar and Halalli, 2016] proposent une solution pour le pré-traitement d'images mammographique en appliquant le filtre de Wiener et CLAHE. La méthode commence par une binarisation de l'image avec un seuil 0.1, ensuite, une réduction du muscle pectoral avec un algorithme de croissance de régions modifié, enfin l'amélioration de la qualité de l'image utilisant le filtre de Wiener et la technique CLAHE. Les résultats de l'application du filtre de Wiener ont été comparés avec ceux du filtre médian, le filtre adaptatif min, max, et selon les auteurs cette méthode a donné le meilleur résultat.

Sushma et Kumar [Sushma and Kumar, 2015] ont proposé une méthode utilisant un algorithme bio-inspiré pour l'amélioration d'images mammaires dans le but de l'aide à la détection du cancer du sein. La solution consiste à trouver le meilleur facteur contraste. L'algorithme proposé commence par la détection des contours en utilisant un masque spécifié. Un seuillage de l'image est ensuite effectué. L'amélioration est effectuée suivant trois étapes : une amélioration primaire en employant deux paramètres d'optimisation (meilleure valeur locale et meilleure valeur globale) ; une amélioration secondaire consistant en une estimation probabiliste des intensités est effectuée ; et une amélioration tertiaire traitant l'image résultante en double précision et permettant de détecter la présence ou non d'une zone cancéreuse.

b. Segmentation d'images

C'est l'un des objectifs du traitement d'images qui consiste à délimiter certaines régions et les séparer du reste de l'image. La segmentation d'images médicales permet ainsi, de détecter et identifier une zone suspecte dans le tissu mammaire, ce qui reste un défi pour les chercheurs.

Dans leurs travaux, Goubalan et al., [Goubalan et al., 2016] ont proposé une solution pour la segmentation des masses malignes ayant des formes irrégulières et qui peuvent être présentes dans un tissu mammaire dense ce qui peut réduire les chances de leur détection. La solution emploie une classe spécifique de la méthode MRF (Markov Random Fields) qui s'appelle Pickard Random Field. Les auteurs

rapportent que cette méthode est plus rapide, non supervisée (requiert juste le nombre de niveau du MRF). L'image résultante est estimée d'une façon probabiliste se référant à un modèle gaussien.

Smaoui et Hlima [Smaoui and Hlima, 2016] résolvent le problème de la segmentation des zones tumorales du sein à partir d'une image mammographique en exploitant des opérateurs morphologiques (principalement l'ouverture, la fermeture et l'érosion) suivi par une extraction du contour en appliquant la méthode "Marching Squares".

Dans leurs travaux, De Cea et Yang [de Cea and Yang, 2016] ont proposé une méthode pour diminuer les faux positifs, et détecter les microcalcifications manquées sur une image mammographique. L'idée est de chercher les éléments aberrants dans un ensemble de candidats représentant des éléments suspects. Pour cela les auteurs utilisent un graphe de voisinage stochastique. La solution consiste donc à détecter les éléments ayant des caractéristiques différentes en appliquant la méthode par sélection stochastique des éléments aberrants expliquée dans [Janssens, 2013]. Cette méthode résulte en une mesure exprimant la probabilité qu'un élément soit un aberrant. Un aberrant est identifié comme un noeud qui n'a aucun bord qui pointe vers d'autres noeuds.

Dans les travaux de Angayarkanni et al. [N. Angayarkanni, 2016], une approche est proposée pour la détection des zones suspectes à partir d'une image mammographique. La solution est présentée en cinq étapes :

- La segmentation de l'image : effectuée en utilisant la méthode de seuillage de variance. Si la variance est inférieure à un certain seuil, le block est considéré comme appartenant à l'arrière-plan, sinon il est assigné à l'objet. Les blocks utilisés sont de taille 8 x 8.
- La binarisation de l'image : chaque pixel de l'image est transformé en blanc si sa valeur est supérieure à une certaine valeur v et en noir dans le cas contraire. Cette valeur v représente la valeur moyenne du block contenant les pixels considérés.
- L'amincissement de l'image : cette étape permet de réduire l'espace de l'image en supprimant les pixels non significatifs (pixels représentant l'arrière-plan : les pixels noirs) suivant un algorithme itératif. Une préservation de la connectivité des objets est prise en considération.
- La triangulation de l'image : Cette étape est effectuée en utilisant le diagramme de Voronoï pour le partitionnement de l'image suivi par la triangulation de Delaunay.
- La transformation en Distance Euclidienne : cette étape utilise la transformation en Distance Euclidienne du voisinage utilisant une fenêtre d'analyse de voisinage de taille : 3 x 3. L'image est reconvertie en niveaux de gris

c. Détection et Classification des formes pathologiques

La recherche et la localisation des objets pathologiques à partir des images médicales est une étape cruciale dans un système d'aide au diagnostic. Dans ce but, Ertosun et Rubin [Ertosun and Rubin, 2015] ont proposé récemment une solution pour la recherche et la localisation automatique des masses dans une image mammographique utilisant les réseaux de neurones à convolution. Deux modules sont envisagés : un module de classification permet de classer les mammographies en deux classes : les images contenant des masses et les images ne contenant pas des masses. Le "deep learning" est appliqué pour assurer cette classification ; et un module de localisation utilisant une approche probabiliste basée sur le "deep learning network". Afin d'analyser les anomalies présentes sur une mammographie et fournir un résultat assistant le médecin dans la prise de décision, une classification des objets pathologiques est requise. Cette classification vise, principalement à catégoriser les objets en deux grandes classes : les objets malins et les objets bénins. Une autre classification peut envisager une catégorisation par type d'anomalie.

Les travaux de Shirazi et Rashedi [Shirazi and Rashedi, 2016] concernaient la détection des tumeurs du tissu mammaire sur les mammographies. La solution proposée combine la méthode SVM (Support Vector Machine) pour la classification des caractéristiques préalablement extraites et décrites par la matrice de co-occurrence en niveaux de gris GLCM (Gray-Level Co-occurrence Matrix) , et la méthode MGSA (Mixed Gravitational Search Algorithm) pour l'optimisation des résultats et qui consiste à coupler la méthode "Binary GSA" à la méthode "Continuous GSA" .

Heriana et Soesanti [Heriana and Soesanti, 2015] ont proposé une approche utilisant l'algorithme Fuzzy C-Means pour la classification des tailles des tumeurs détectés à partir des images thermographiques du sein. La classification est effectuée selon le système TNM (Tumor - Node - Methastasis) utilisé par les médecins.

Baharuddin et al. [Baharuddin et al., 2016] ont développé un model de classification des mammographies suivant le système BIRADS. Cette classification est basée sur la logique floue utilisant trois fonctions d'appartenance : une gaussienne, une triangulaire et une trapézoïdale. La génération des règles est effectuée par l'expert. L'avantage de cette solution réside dans sa capacité à réduire l'incertitude des données. Par ailleurs, l'implication d'un expert dans le traitement d'images médicales augmente la fiabilité des résultats. Les résultats présentés indiquent que la fonction d'appartenance de type triangulaire donne les meilleures statistiques en terme de précision, spécification et sensibilité en se basant sur les règles fournies par l'expert.

Dans les travaux de Bekker et al. [Bekker et al., 2016], une classification des microcalcifications mammaires est effectuée selon le paradigme "deep-learning" utilisant des images mammographiques CC (CranioCaudal) et des images mammographiques MLO (MedioLateral Oblique). Cette méthode a été comparée à la méthode proposée par les mêmes auteurs dans des travaux antérieurs [Bekker et al., 2015] appliquant

l'algorithme EM (Expectation-maximization) utilisé avec le modèle régression logistique pour la modélisation multi-vues. L'approche proposée par les auteurs consiste à concaténer les deux représentations des mammographies et ensuite à classifier le résultat par un réseau de neurone (NN). Les vecteurs caractéristiques sont représentés par 14 caractéristiques chacun. L'apprentissage est effectué en utilisant deux couches NN avec 24 neurones (12 neurones pour chaque couche) suivant l'algorithme gradient descendant.

Swapnil et al. [Swapnil et al., 2016] ont aussi traité le problème d'analyse d'images mammographiques par classification, en proposant une méthode permettant de classifier les régions suspectes en malignes ou bénignes. D'abord, une extraction de caractéristiques est effectuée en utilisant la matrice de co-occurrences en niveaux de gris et deux autres caractéristiques définies par eux mêmes contribuant à décider de la malignité des formes pathologiques. Ensuite, la classification est effectuée en utilisant la méthode SVM (Support Vector Machine).

Wang et al. [Wang et al., 2016] ont développé un système permettant la détection et la classification des micro-calcifications. Après une étape de pré-traitement, une extraction des caractéristiques est effectuée. Deux types de caractéristiques étaient utilisés : les caractéristiques des micro-calcifications, et des caractéristiques de la texture de la zone pathologique.

Dans leurs travaux Abirami et al. [Abirami et al., 2016] ont traité aussi le problème de détection des micro-calcifications et leur classification en utilisant les réseaux de neurones artificiels. La solution commence par une étape d'Extraction des caractéristiques en utilisant l'ondelette de Daubechies DWT (Daubechies Wavelet) et l'ondelette de Haar. Ensuite, une classification est effectuée par un réseau de neurone artificiel en utilisant deux approches : le réseau Perceptron multicouche (Multi Layer Perceptron) et le réseau à Fonction de base radiale (Radial Basis Function).

Singh et al. [Singh et al., 2016] ont traité aussi la problématique de la classification des microcalcifications. Ils contribuent à l'amélioration de deux méthodes PCET (Polar Complex Exponential Transform) et DEWNN (Differential Evolution optimization Wavelet Neural Network) pour le traitement.

Saraswathi et al. [Saraswathi et al., 2016], ont proposé une nouvelle méthode pour le diagnostic du cancer du sein en appliquant une classification. Cette méthode est basée sur la transformation en courbelette couplée à un algorithme naturel. L'étape de la classification est assurée par la méthode SVM (Support Vector Machine) en utilisant un algorithme d'optimisation PSO (Particle Swarm Optimization) basé sur le principe de la volée des oiseaux.

Dans [Wang and Yang, 2016] Wang et Yang proposent une solution pour la classification des zones suspectes du tissu mammaire à partir d'une image mam-

mographique en tenant compte des caractéristiques visuelles extraites à partir de l'image et les informations pathologiques des cas traités. La méthode consiste à rechercher des cas similaires à un cas étudié afin de l'attribuer à une classe de pathologies. La procédure est décrite comme suit :

- Calculer la similarité du cas x par un classifieur $g(x)$ pré-enseigné.
- Récupérer les $2N$ cas les plus proches de $g(x)$ suivant leur degré de malignité
- Sélectionner les N premiers cas similaires à x en calculant la distance euclidienne de chaque cas à x . Ces cas sont utilisés pour la classification de x .

Des auteurs utilisent l'algorithme d'apprentissage "Adaboost" développé dans [Freund and Schapire, 1997]. L'idée proposée par les auteurs consiste à placer un poids élevé pour ces classifieurs faibles et ayant plus de précision pour la recherche des cas.

Dans leurs travaux, Aarthy et Prabu [Aarthy and Prabu, 2016] utilisent la méthode SVM pour la classification du tissu mammaire en tissu normal ou tissu anormal. La solution comporte une étape de segmentation avec la technique du seuillage d'histogramme et une étape d'extraction de caractéristiques en se basant sur la matrice GLCM et l'ondelette de Gabor.

Dans [Mina and Isa, 2015] Mina et Isa résolvent le problème de la classification des anomalies dans le tissu mammaire en appliquant les réseaux de neurones artificiels basés sur une transformation en ondelette utilisée pour l'extraction de caractéristiques.

Les travaux de Quellec, et al. [Quellec et al., 2016] traitent le problème de la classification des lésions mammaires en utilisant le paradigme Multiple Instance Learning (MIL) suivant le principe suivant :

- Décomposition du sein en des régions rectangulaires uniformes en terme de distance par rapport au contour du sein et au mamelon. La région du sein est divisée en 2 grandes parties basée sur l'axe principal du sein.
- Extraction des caractéristiques qui sont considérées suivant trois types : caractéristiques liées à la lésion (masse/Microcalcification), vecteur caractéristique contenant les caractéristiques des Masses et les caractéristiques des Microcalcifications, et caractéristiques texturales définies par la matrice GLRLM (Gray-level Run, Length Matrices); la matrice GLCM, la technique LBPH (Histograms of Local Binary Patterns), Haralick.
- Apprentissage supervisé MIL : Chaque instance est décrite par un Vecteur caractéristique et est associée à une classe. Une classe peut être associée à plusieurs instances. Les algorithmes utilisés dans cette approche sont : DD (Diverse Density), APR (Axis Parallel Rectangles), mi- SVM, MI- SVM, MIL Boost .

Cette solution est comparée à la méthode SVM utilisé seule. Les résultats exposés par les auteurs montrent la performance de l'approche traitée.

Le tableau 1.1 résume ces différents travaux.

TABLE 1.1 – Résumé des travaux effectués sur le traitement d'images mammographiques

Référence	Problématique	Solution
Vedanarayanan et Nandhitha 16	Pré-traitement d'images mammographiques	Filtre de gabor Filtre de Canny
Makandar et Halalli 16	Amélioration du contraste des	Filtre wiener Technique CLAHE (Contrast Limited Adaptive Histogram Equalization)
Baharuddin et al. 2016	Classification des objets pathologiques	Utilisation de la logique floue : Modèle de Mamdani pour le traitement de l'incertitude
Bekker et al. 2016	Classification des microcalcifications	Utilisation du Deep learning
Smaoui et Hlima, 2016	Extraction des objets tumoraux	Segmentation par les opérateurs morphologiques.Extraction des contours en utilisant \acute{e} Marching Squares \grave{a}
Swapril et al. 2016	Classification des régions suspectes	Extraction de caractéristique avec la matrice de cooccurrence en niveaux de gris Classification avec SVM
De Cea et Yang, 2016	Détection des microcalcification	Chercher des éléments aberrants parmi un ensemble d'éléments suspects en utilisant un graphe de voisinage stochastique
Wang et al, 2016	Détection et classification des microcalcifications	Extraction de caractéristiques de textures et de micro-calcification. Application du Détecteur de DOG pour la détection
Abirami et al., 2016	Détection et classification des microcalcifications	Extraction de caractéristiques avec l'ondelette de Daubechies DWTet l'ondelette de Haar Classification avec les réseaux de neurones : Multi Layer Perceptron et Fonction à base radiale
Singh et al., 2016	Classification des microcalcification	Extraction de caractéristiques en utilisant la méthode Phase Corrected PCET Classification en utilisant la méthode ADEWNN (Adaptive Differential Evolution Wavelet Neural Network)
Saraswathi et al.,2016	Classification des lésions mammaires	Extraction de caractéristiques avec la méthode Curvelet Transform Classification par SVM
Shirazi et Rashedi, 2016	Classification des lésions	Extraction des caractéristiques en utilisant la matrice de cooccurrence en niveaux de gris Classification avec SVM Optimisation avec MSGA : Mixed Gravitational Search Algorithm

Goubalan, et al., 2016	Segmentation des masses malignes	Prétraitement avec étirement d'histogramme et transformation logarithmique simple Segmentation avec Pickard Random Field
Angayarkanni, et al., 2016	Détection des zones suspectes	Segmentation en utilisant la méthode du seuillage Partitionnement de l'image avec la triangulation de Delaunay La transformation de l'image en Distance Euclidienne
Wang et Yang, 2016	Classification des zones suspectes	Extraction des caractéristiques d'un cas étudié. Recherche d'un cas similaire pour l'attribuer à une classe
Aarthy et Prabu, 2016	Classification tissu mammaire	Prétraitement avec le filtre moyen et l'égalisation d'histogramme Segmentation avec le seuillage Extraction de caractéristiques avec la GLCM et l'ondelette de Gabor Classification avec SVM
Quellec, et al., 2016	Classification des lésions mammaires	Décomposition de l'image Extraction de caractéristiques pathologiques et texturales Apprentissage supervisé MIL (Multiple Instance Learning)
Sushma et Kumar, 2015	Amélioration contraste	Trouver le meilleur facteur contraste en appliquant Le seuillage Des estimation probabiliste
Ertoşun et Rubin, 2015	Détection et localisation des objets pathologiques	Classification et localisation avec la méthode Deep learning

1.2 Images IRM mammaires

L'IRM (Imagerie par Résonance Magnétique) est une modalité très précise et est souvent utilisée pour la détection du cancer du sein. Contrairement à la mammographie, cette modalité n'est pas affectée par la densité du sein. Selon J.P. Alunni [Alunni, 2012], elle "a bien souvent été présentée comme la solution à tous les problèmes en imagerie du sein." En traitement d'images, l'IRM est une image volumique en niveaux de gris présentée comme un ensemble de coupes.

1.2.1 Examen mammaire par IRM

Le déroulement de l'examen de l'IRM mammaire nécessite l'injection d'un produit paramagnétique : "chélates de gadolinium" permettant la détection du contraste dans le tissu du sein. La présence de contraste dans l'image IRM du sein indique la présence d'une lésion. Après détection du contraste, une opération de soustraction est nécessaire pour faire apparaître l'anomalie.

Après détection de contraste, l'étape suivante de l'interprétation consiste en l'analyse des caractéristiques dynamiques et morphologiques de l'objet observé. Ensuite, la dernière étape avant de conclure, permet d'étudier l'aspect cinétique de

l'anomalie observée.

L'analyse d'une image IRM consiste à étudier trois critères décrivant les objets observés : la forme, le contour et le rehaussement. La forme d'une masse observée dans une IRM mammaire peut être : ronde, ovale, de forme lobulaire ou encore irrégulière (Fig. 1.8). L'irrégularité de la forme est souvent un signe de malignité de l'anomalie. Les spécialistes estiment que 32% des cas de formes irrégulières sont des cancers [Jalaguier-Coudray, 2009]. A l'inverse, une forme simple indique la présence d'une lésion bénigne.

Une masse, est définie aussi par la caractéristique de son contour. L'IRM mammaire permet de distinguer entre 3 types de contour de masses : lisse, irrégulier ou spiculé (Fig. 1.9). Comme pour le cas d'une mammographie, un contour irrégulier ou spiculé est dans la majorité des cas un signe de malignité de la masse observée.

Contrairement à l'image mammographique, l'IRM permet aussi de visualiser la caractéristique de rehaussement pour une masse (Fig. 1.10). Une hétérogénéité de la masse observée renseigne sur sa malignité. Inversement, la non présence de rehaussement ou l'homogénéité est signe de bénignité.

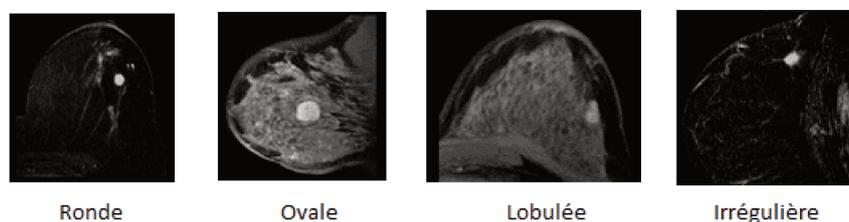


FIGURE 1.8 – Les formes des masses observées sur une IRM



FIGURE 1.9 – Les contours des masses observées sur une IRM

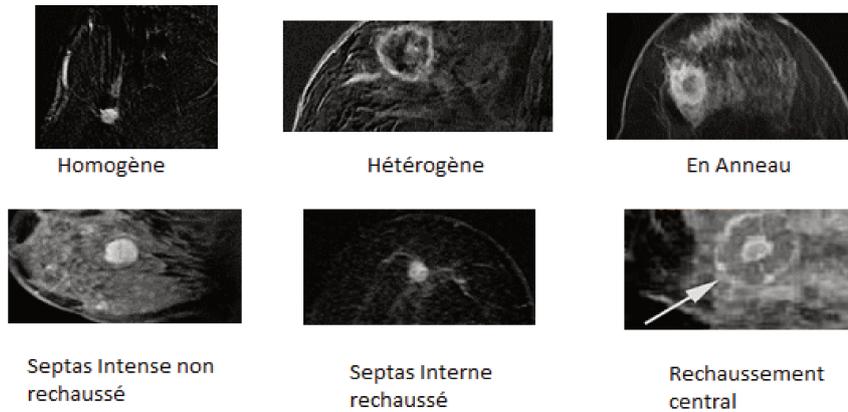


FIGURE 1.10 – Les différents types de rehaussement sur une IRM

1.2.2 Apport de l'IRM

L'examen IRM se fait généralement comme examen complémentaire afin de confirmer certains aspects douteux détectés par la mammographie. Selon les experts, l'IRM mammaire est une technique plus efficace en terme de détection et caractérisation des anomalies quand d'autres techniques échouent (Fig. 1.11). Elle est plus performante dans l'évaluation des cas à haut risque. L'avantage majeur de cette technique par rapport à d'autre réside dans sa capacité à représenter de la meilleure façon le tissu du sein dense. L'image IRM donne dans la majorité des cas une meilleure représentation permettant une prise de décision fiable. Cependant, elle peut présenter dans certains cas, des faux positifs, et dans d'autres cas l'absence de certaines lésions surtout lorsqu'il s'agit de microcalcifications (limite majeure de l'IRM). Sa qualité peut être affectée par l'état psychologique de la femme et ses mouvements lors de l'examen.

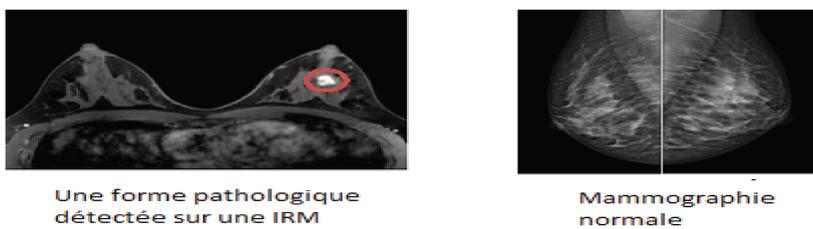


FIGURE 1.11 – Complémentarité des examens IRM et Mammographie

1.2.3 Difficulté du traitement d'images IRM

L'IRM apporte un gain important en terme de précision des résultats, et surtout par sa non sensibilité à la densité mammaire. Cependant, comme technique produisant une image en niveaux de gris, elle peut conduire aussi à des résultats erronés traduits en de faux positifs ou de faux négatifs, ce qui peut diminuer la précision des interprétations.

Un des problèmes du traitement d'une image IRM réside dans son contenu. En effet, l'image représente le sein avec toute la partie de la paroi thoracique, ce qui peut conduire à des difficultés de traitement et d'analyse de ces images. C'est pour cette raison qu'une étape de sélection de la région d'intérêt est souvent nécessaire. La représentation des limites de la peau représente également une difficulté pour l'interprétation d'images vu que ces limites sont très similaires en terme d'intensité aux zones pathologiques. Ces difficultés nécessitent un traitement approprié afin de présenter des résultats corrects apportant l'aide requise à l'expert.

1.2.4 Travaux sur le traitement d'images IRM mammaires

Similairement à la mammographie, la majorité des travaux étudiant les images IRM traitent trois problèmes : le pré-traitement, la segmentation et le problème de classification des lésions.

a. Le pré-traitement d'images IRM

Le pré-traitement est une opération de base du traitement d'images permettant de préparer l'image à l'analyse. Dans les travaux de Al-Faris, et al. [Al-Faris et al., 2015] la problématique de la réduction de la zone d'intérêt sur les images IRM mammaires, est soulevée. Vu que l'intensité du contour représentant la peau est très similaire à l'intensité d'une zone pathologique ; pouvant produire ainsi, des résultats imprécis dans la segmentation des zones tumorales, les auteurs proposent d'éliminer les contours représentant la peau des images avant la segmentation des zones pathologiques. L'approche consiste à détecter, en premier temps, les contours de la peau, en utilisant la méthode Level Set Active Contour, ensuite supprimer ces contours et garder la zone intérieure du sein en employant des opérateurs morphologiques.

Les travaux de Zhang et Peeters [Zhang et al., 2015a] résolvent le problème de l'inhomogénéité (connue comme bias field) des images IRM. Cette inhomogénéité est due principalement aux conditions de prise de l'image (interaction patient, etc...). Les auteurs proposent un algorithme permettant la correction de cette inhomogénéité en minimisant un critère d'entropie. Pour résoudre l'équation de minimisation du bias field, les auteurs utilisent la méthode ADMM (Alternating Direction Method of Multipliers). Cette méthode applique l'algorithme "Diviser pour régner".

b. La segmentation d'images IRM

La segmentation est aussi un traitement fondamental dans l'analyse et l'interprétation d'images. Plusieurs méthodes sont employées pour la segmentation d'images suivant le domaine d'application. Les travaux de Gubern-Mérida, et al. [Gubern-Merida et al., 2015] soulèvent le problème de l'estimation de la densité mammaire à partir d'une image IRM. En effet, la mesure de la densité mammaire est importante vue qu'elle représente un risque de développement d'un cancer invisible. La solution proposée consiste à estimer automatiquement la densité mammaire suivant trois étapes :

- Le Pré-traitement qui consiste à corriger l'inhomogénéité des intensités de l'image en utilisant la méthode de correction "Bias Field Correction" ;
- La segmentation du sein en utilisant une approche probabiliste qui consiste à employer un atlas probabiliste pour isoler la zone du sein ;
- la segmentation du tissu fibroglandulaire et l'estimation de la densité mammaire réalisée en appliquant l'algorithme Expectation-Maximization qui permet d'estimer la distribution des intensités du tissu du sein en distinguant les tissu graisseux du tissu fibroglandulaire.

Dans les travaux de Gnonnou et Smaoui [Gnonnou and Smaoui, 2014] la segmentation des tumeurs à partir d'une image IRM mammaire est effectuée en deux étapes : l'isolation du contour du sein, et la segmentation de la tumeur. La première étape consiste à parcourir l'image suivant 5 directions et à reproduire la valeur du premier pixel rencontré différent de 0 vers une image vide préalablement créée. D'après les auteurs, cet algorithme a l'avantage d'être rapide et applicable pour les contours fermés et ouverts. La seconde étape consiste à appliquer l'algorithme k-Means avec 6 classes sur l'image résultante de la soustraction des images "pré-post" (Images prises avant et après l'injection du produit créant le contraste) qui permet d'éliminer les superflus (tissus graisseux par exemple) sans affecter les zones tumorales.

Dans les travaux de Yang, et al. [Yang et al., 2016], une approche basée sur des calculs statistiques est introduite comme Support-pixel Correlation Statistical Method (SCSM) afin de séparer le sein de l'arrière-plan. La méthode commence par la détection des contours avec l'algorithme de Canny. Les points des contours constituent les "support-pixels". Un calcul statistique des valeurs de niveaux de gris est appliqué sur les "support-pixels". Ensuite, deux valeurs de seuil sont calculées : une pour l'objet et une pour l'arrière-plan. Ces valeurs représentent la moyenne des distributions des niveaux de gris des deux zones. Ensuite, les pixels sont groupés dans deux classes différentes suivant les deux seuils calculés. Enfin, une classe représente l'objet, l'autre classe représente l'arrière-plan.

Dans le même contexte, les travaux de Li et al. [Li et al., 2015] présentent une approche permettant de décomposer l'image en texture et structure. Cette approche se base sur la méthode CSB-TV (Constraint Split Bergman combinée avec Total Variation) pour extraire l'information dynamique de l'image. Le problème posé minimise un critère d'erreur de l'image sous la forme : $\min_u |\nabla u| + \frac{\mu}{2} \|u - f\|_2^2$ où

$f, u \in \mathbb{R}^{m*n}$ représentent l'image originale et l'image structurante de dimension $m * n$, $|\nabla u|$, $\|u - f\|_2^2$ sont des fonctions convexes. $|\nabla u|$ identifié comme un terme de régularisation pénalise les hautes variations dans u pour obtenir une image structurante lisse et $\|u - f\|_2^2$ défini comme terme de données ayant le même objectif. Le paramètre μ est fonction du poids constant modélisant une évaluation entre la fidélité des données et de la force de la régularisation. Pour une précision des résultats, la procédure de décomposition doit préserver l'image texturale intacte durant le traitement. Pour cette raison, les auteurs utilisent des informations texturales contraignant la décomposition. La méthode permet de tester différentes formes de contraintes pour trouver une meilleure valeur pour le pixel maximisant la correction des niveaux de gris.

c. Classification du contenu des images IRM

La classification du contenu des images est un problème fréquent en traitement d'images dans la plupart des domaines d'application. Dans le domaine médical, elle constitue une opération importante dans l'interprétation et l'aide à la décision. Dans leurs travaux, Song, et al. [Hong et al., 2014] traitent la problématique de la classification du tissu mammaire sur une image IRM (tissu fibroglandulaire, tissu grasseux, tissu tumoral ou tissu normal). En effet, le tissu mammaire peut comporter un tissu fibroglandulaire, un tissu grasseux ou des zones tumorales. Les auteurs proposent une approche contenant deux étapes :

- Dans un premier temps une opération de pré-traitement est effectuée en appliquant un filtre de diffusion anisotropique. Ensuite une binarisation de l'image avec le filtre d'Otsu est effectuée. Finalement une étape de sélection de la plus grande zone connectée est effectuée en appliquant une ouverture morphologique suivie d'une fermeture. Cette étape permet d'éliminer les objets non significatifs de l'image.
- Séparation du tissu fibroglandulaire du tissu grasseux en utilisant l'algorithme Kernelled Fuzzy C-Means. Cet algorithme considère les niveaux de gris des pixels sans se soucier de l'information spatiale de l'image. Or, il doit y avoir une continuité spatiale pour les pixels appartenant au même organe sur les images médicales. C'est pour cela que les auteurs ajoutent l'information spatiale à l'algorithme SKFCM

Les résultats sont comparés avec une segmentation moyennant l'algorithme FCM et KFCM et d'après les auteurs, les mesures de performance montrent que l'approche proposée présente une meilleure précision.

Les travaux de Maken and Bradley [Maken and Bradley, 2015] traitent la problématique de classification de lésions mammaires en bénignes et malignes. Le but de cette étude est d'évaluer l'approche de classification basée sur la méthode d'apprentissage semi-supervisée MIL (Multiple Instance Learning) ou la méthode SIL (Single Instance Learning) sur les images IRM mammaires. Une comparaison entre

les résultats produits par CkNN (citation kNN), kNN et la méthode traditionnelle consistant en l'extraction des caractéristiques d'une zone d'intérêt segmentée au préalable. Pour l'extraction de caractéristiques, les auteurs utilisent deux approches :

- Une approche en appliquant la technique 3D-DCT (Discrete Cosine Transform) sur l'image IRM. Cette technique donne des informations spatio-temporelles. Elle est appliquée sur chaque cube de dimension $8 \times 8 \times 4$.
- Une approche en appliquant la technique 2D-DCT sur la coupe T2 de l'image IRM. Cette technique résulte en des informations spatiales. Elle est appliquée sur chaque carré de dimension 8×8 .

Pour l'évaluation, cinq caractéristiques sont sélectionnées en utilisant la technique de la distance minimale de Hausdorff pour la méthode MIL et la distance minimale de Mahalanobis pour la méthode SIL. Les auteurs rapportent que la méthode MIL est plus performante avec les CkNN alors que la méthode SIL est plus performante avec kNN. Ils montrent à travers les résultats que la sélection des caractéristiques spatio-temporelles donnent de meilleurs résultats par rapport à la sélection des caractéristiques spatiales.

Le tableau 1.2 résume les différentes approches présentées ci-dessus.

TABLE 1.2 – Résumé des travaux effectués sur le traitement d'images IRM mammaires

Référence	Problématique	Solution
Yang, et al., 2016	Segmentation en objet et arrière-plan	Détection des contours avec l'algorithme de Canny Classification des pixels en deux classes (objet, arrière-plan) en utilisant deux valeurs de seuils
Gnonnou et Smaoui, 2014	Segmentation des Tumeurs	Isolation du contour du sein en construisant une nouvelle image Segmentation de la tumeur en appliquant une soustraction de l'image pre-injection et post injection puis en utilisant un algorithme k-means avec 6 classes
Song , et al., 2014	Classification du tissu mammaire	Prétraitement : Filtrage avec un filtre de diffusion anisotropique, une binarisation avec le filtre d'Otsu et une suppression des objets non significatifs avec des opérateurs morphologiques Séparation du tissu fibroglandulaire du tissu graisseux avec l'algorithme Kernelled Fuzzy C-Means
Zhang et Peeters, 2015	Correction de l'inhomogénéité de l'image (bias field)	Une transformation logarithmique est appliquée en premier temps. Une estimation du bias field est effectué en minimisant une fonction entropie. Minimisation du bias field avec la méthode ADMM (Alternating Direction Method of Multipliers) utilisant un algorithme diviser pour régner

Al-Faris, et al., 2015	Supprimer les contours de la peau	Détection des contours avec Level Set Suppression des contours avec des opérations Morphologiques
Gubern-Mérida, et al., 2015	Estimation de la densité mammaire	Prétraitement avec Bias Field Correction Segmentation du sein avec un atlas probabiliste Segmentation tissu Fibroglandulaire et estimation de la densité avec l'algorithme EM
(Shujuan , et al., 2015)	Segmentation d'une image en texture et structure	Approche basée sur la méthode Constraint Split Bergman couplée avec la méthode Total Variation pour l'extraction de l'information dynamique de l'image. Cette approche consiste à minimiser un critère d'erreur de l'image.

1.3 Conclusion

Le traitement d'images médicales présente souvent un apport considérable pour l'aide au diagnostic. Le sein est parmi les organes les plus étudiés dans le domaine du traitement d'images. En effet, il s'agit d'un organe délicat pouvant être le siège de certaines anomalies qui représentent dans plusieurs cas un éventuel risque de cancer. Les images traitées dans ce contexte sont majoritairement les images mammographiques et l'IRM. C'est la raison pour laquelle, nous avons réservé ce chapitre à la présentation des particularités de ces images. Pour chaque image, nous avons présenté les spécificités du diagnostic effectué, les difficultés liées à son traitement et une revue de littérature des principaux travaux effectués dans son cadre d'études.

Information et modélisation des connaissances

Sommaire

2.1	Les images médicales et les concepts de donnée, information et connaissance	29
2.2	Gestion des imperfections	32
2.2.1	Imperfections de l'information	32
2.2.2	Modélisation des imperfections de l'information	33
2.3	Modélisation de connaissances à partir des images	36
2.3.1	Modèles	37
2.3.2	Signatures	38
2.3.3	Ontologies	39
2.4	Modélisation 3D de scènes réelles	41
2.4.1	La modélisation 3D dans le domaine médical	42
2.4.2	La modélisation 3D du sein féminin	44
2.5	Conclusion	46

Les besoins des êtres humains à automatiser les tâches et les raisonnements évoluent de jour en jour, ils deviennent de plus en plus exigeants. C'est ce qui permet, d'ouvrir constamment de nouveaux champs d'études. Le domaine de la modélisation des informations et des connaissances est un champ d'études très fertile et très innovant. En effet, l'interprétation et l'analyse de scènes réelles à travers des représentations pixelliques est une tâche quasi complexe. Une focalisation sur les informations pertinentes et leur représentation plus simplifiée améliorent la qualité des connaissances dans une situation particulière. Cependant, ces connaissances sont souvent entachées par certaines imperfections et ambiguïté. Leur manque de clarté induit de l'imprécision, de l'inexactitude et de l'incertitude dans leur formulation. Le besoin de modéliser et de traiter ces imperfections a conduit les chercheurs à

proposer des outils capables de prendre en considération ces imperfections.

En effet, la théorie des probabilités fut le premier cadre modélisant l'incertitude abordée dès le XVIIème siècle avec Pascal et Fermat qui voulaient traiter certaines questions liées à un jeu de hasard [Szafarz, 1985]. Dès lors les travaux se multipliaient, donnant naissance à d'autres théories dépassant les limites des probabilités pour mieux représenter les imperfections. Tel est le cas de la théorie des possibilités, qui est une théorie pratiquement récente, ayant l'avantage de traiter plusieurs formes d'imperfection dans un seul cadre mathématique.

Par ailleurs, le développement technologique et informatique a permis d'ouvrir de nouveaux champs d'études en faisant coopérer plusieurs disciplines. Le domaine informatique est l'un des domaines qui s'est couplé avec d'autres disciplines tel que les sciences géographiques, la robotique et notamment la médecine. Ce couplage entre différentes disciplines a donné naissance à de nouveaux modes de représentation de l'information et des connaissances avec des modèles virtuels. Ces modèles constituent un apport majeur dans l'étude de certains phénomènes (évolution d'une maladie, simulation virtuelle de certains comportements, etc...), la planification pré-opératoire, la chirurgie assistée par ordinateur, etc... Ainsi, la modélisation virtuelle 3D en médecine représente un outil indispensable apportant une aide très significative au diagnostic des maladies et plus généralement à l'assistance médicale.

Dans ce chapitre nous présentons en premier temps les images médicales et nous décrivons les différents types d'informations traitées par les systèmes d'interprétation de scènes et d'aide à la décision. Ensuite, nous expliquons les principaux types d'imperfections affectant les informations et les différents cadres adoptés pour leurs modélisations. Puis, nous présentons quelques outils et les travaux qui les appliquent afin de modéliser les connaissances extraites des images. Et enfin, nous exposons la modélisation virtuelle des informations et son apport dans le domaine médical.

2.1 Les images médicales et les concepts de donnée, information et connaissance

Afin de bien mener son diagnostic, un médecin cherche à explorer l'intérieur du corps humain. De cet objectif naît l'imagerie médicale mettant en lumière cette partie du corps invisible pour les experts. Une image médicale est formée suivant des principes physiques liés à la nature des tissus explorés : La radioactivité, la réflexion des ondes ultra-sonores, la résonance magnétique, l'absorption des rayons X. Le signal émis par la source, interagit avec les tissus suivant certains phénomènes physiologiques. Cette interaction donne lieu à une représentation en niveaux de gris traduisant les caractéristiques physiques des organes prospectés et constituant l'image médicale. Cette image est utilisée dans plusieurs phases de la pratique

médicale : du simple diagnostic à l'intervention opératoire en temps réel. Selon la complexité de l'organe et le type d'anomalie examinée, l'image peut être complexe et la représentation en niveaux de gris est difficile à interpréter. C'est pour cette raison que les médecins cherchent surtout une meilleure qualité de la représentation des organes (moins de bruit, un bon contraste) ce qui requiert un traitement approprié par des outils informatiques sophistiqués.

Partant du principe que l'image est un support d'information, analyser (ou interpréter) une image revient à analyser les informations qu'elle renferme. Dans la littérature, la notion d'"information" est flanquée par deux principaux concepts : la donnée ("data"), et la connaissance ("knowledge"). Ces trois concepts (Donnée, Information, et Connaissance) interviennent dans tout processus de traitement dans la théorie de l'information. Leurs définitions et les relations qui les lient sont souvent mises en question. La réponse à cette question a nécessité plusieurs définitions suivant le domaine d'application. Souvent, on lie ces concepts suivant une hiérarchie basée sur une échelle cognitive où le bas niveau représente la perception des scènes réelles et le haut niveau s'attache au raisonnement. En effet, la donnée est vue comme l'entité de base d'une information et l'information est perçue comme l'entité de base d'une connaissance [Bossé et al., 2005]. Plusieurs chercheurs [Zeleny, 1987, Ackoff, 1989, Bierly et al., 2000, Chaffey and Wood, 2005] ont étudié cette hiérarchie nommée DIK (Data-Information-Knowledge). Elle est représentée par une pyramide ayant à la base la notion de donnée et au sommet la connaissance (Fig. 2.1). Ackoff, un théoricien des systèmes, est parmi les premiers qui ont défini cette hiérarchie.

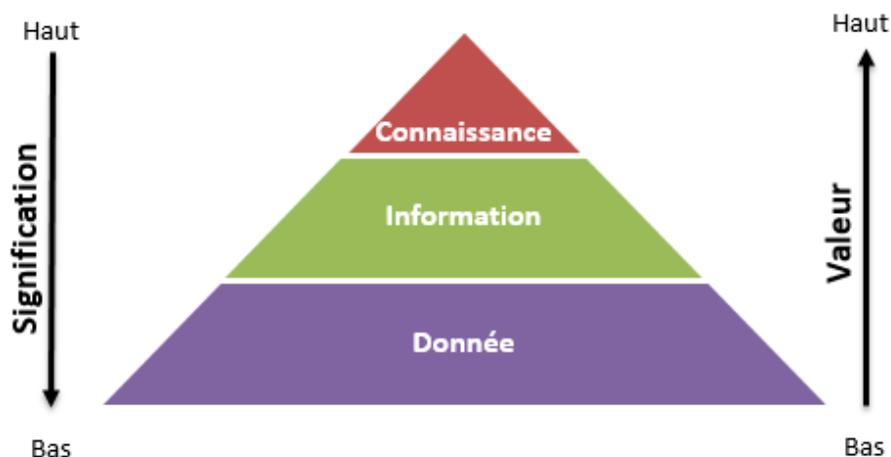


FIGURE 2.1 – Hiérarchie des concepts Donnée, Information et Connaissance suivant Ackoff [Ackoff, 1989]

Ackoff définit la donnée [Ackoff, 1989] comme le produit de l'observation représentant la matière première de tout processus cognitif. Elle existe sous n'importe quelle forme et est généralement définie par des symboles caractérisant les objets, les événements et leurs environnements. Zeleny [Zeleny, 1987] définit la donnée comme une entité n'ayant aucun sens. Elle n'a d'utilité que si elle existe dans un contexte significatif. Par exemple : le nombre 18 est une donnée qui n'a aucune signification ni aucune utilité au delà de son existence. Il peut exprimer une mesure, un cardinal, ... mais cela n'est pas spécifié.

Quant à l'information, Ackoff la présente comme une donnée augmentée par une certaine description afin de renfermer un sens. Zeleny présente l'information comme l'entité définissant le quoi. Prenons l'exemple de 18°C, l'unité de mesure °C a donné un sens au nombre 18 qui signifie désormais une mesure de température.

La connaissance est spécifiée [Davenport and Prusack, 1998] comme le fruit d'une réflexion analysant des informations collectées en se basant sur leurs contextes, certaines règles et idées et des expériences antérieures. Cette connaissance peut engendrer de nouvelles expériences et de nouvelles informations. Zeleny présente la connaissance comme l'entité exprimant le comment. Selon Ackoff, la sagesse est l'aptitude à employer les connaissances efficacement afin d'interpréter la scène observée. Zeleny décrit la sagesse par le pourquoi.

Les chercheurs, étudiant le traitement des informations, classent les informations en trois principaux types [Bloche and Maitre, 2004] : numérique, symbolique et hybride. L'information numérique caractérise l'information formulée en nombres. Elle est très utilisée dans le domaine du traitement du signal représentant des mesures des intensités des signaux, des niveaux de gris dans les images, etc. Dans une image l'information numérique constitue les valeurs des pixels (ou voxels) qui la composent. L'information symbolique est définie par Bloch comme : "toute information qui est donnée sous forme de symboles, de propositions ou de règles". En traitement d'image, manipuler des informations symboliques revient à transformer les pixels en des symboles dans le but de trouver une nouvelle représentation aidant à l'interprétation de l'image. Quant à l'information hybride, elle est vue comme une information symbolique manipulée de façon numérique.

Cette représentation, qu'elle soit une donnée, une information ou une connaissance est souvent marquée par certaines imperfections limitant sa capacité à traduire le monde réel fidèlement. C'est pour cela que plusieurs recherches étaient amenées à gérer ces imperfections.

2.2 Gestion des imperfections

Une information représente une fonction reliant un couple d'entités [Solaiman, 2001] : θ qui représente un élément définissant l'information et ω qui représente le contenu informationnel. Quatre propriétés sont définies comme indispensables [Solaiman, 2001], pour caractériser une information parfaite :

- L'exhaustivité : $\forall \theta \in \Theta, \exists \omega \in \Omega$ Telque $Inf(\theta) = \omega$
- L'exclusivité : $\forall \theta \in \Theta, \exists$ un et un seul $\omega \in \Omega$ Telque $Inf(\theta) = \omega$
- La précision de $Inf(\theta)$: θ est un singleton.
- La certitude de $Inf(\theta)$: l'identité de ω : image de θ par la fonction Inf ne comporte aucun doute.

L'absence d'une ou de plusieurs de ces propriétés implique une imperfection de l'information.

2.2.1 Imperfections de l'information

Les types d'imperfections les plus étudiés sont :

- **L'Incertitude** : Une information est dite incertaine si on n'est pas sûr de sa vérité. Elle implique une connaissance partielle de la réalité. Exemple : Considérons la phrase : "La masse détectée peut être maligne" . Cette phrase renferme une information incertaine : On n'est pas sûr si la masse est à caractère malin ou bénin. L'incertitude correspond à une ignorance totale ou partielle de la connaissance. Ce qui traduit un défaut qualitatif de la connaissance.
- **L'Imprécision** : Elle désigne le caractère non déterministe de l'information. En effet, l'information n'est pas définie de façon exacte. Cela peut être traduit par le fait que la fonction $Inf(\theta)$ n'est pas exclusive. Ce qui signifie, encore un défaut quantitatif de la connaissance. Exemple : L'affirmation "Le diamètre de la masse détectée est entre 3 et 4 mm." renferme une information imprécise puisqu'on ne peut pas savoir de façon exacte le diamètre de la masse.
- **L'Ambiguïté** : Une information ambiguë signifie qu'elle admet plusieurs interprétations. C'est à dire qu'elle manque de spécificité. Cela peut être traduit par un manque de lien binaire entre les ensembles Θ et Ω . L'ambiguïté peut être liée à d'autres types d'imperfection, principalement l'incertitude et l'imprécision. Exemple : "Le sein droit contient une zone pathologique." Dans cette situation on ne sait pas de quelle pathologie on parle.
- **L'Incomplétude** : Une information est dite incomplète si elle manque certaines parties pour caractériser la réalité. L'incomplétude représente la différence entre l'information représentée et l'information réelle. Exemple : Un sein contient deux formes pathologiques alors que sur une mammographie on n'observe qu'une seule. Avec une information incomplète, l'expert ne peut pas prendre la décision adéquate.

Les imperfections de l'information conduisent à trois solutions :

- L'omission de l'information

- La tolérance de l'imperfection
- La modélisation de l'imperfection

Dans les deux premiers cas, la solution influence la décision à prendre ce qui est très critique dans plusieurs cas. C'est pour cette raison que plusieurs études étaient établies afin de modéliser les imperfections des informations pour fournir des traitements plus fiables et efficaces.

2.2.2 Modélisation des imperfections de l'information

La modélisation des informations imparfaites désigne la transformation des informations manipulées en des expressions mathématiques exprimant l'imperfection suivant une certaine théorie. La modélisation des imperfections est une étape nécessaire pour une représentation aussi fidèle des informations traitées pour une interprétation adéquate des scènes observées et une meilleure prise de décision. Différentes théories étaient proposées afin de traiter les différentes imperfections liées aux informations. Nous distinguons quatre formalismes mathématiques permettant de modéliser les imperfections des informations.

a. Théorie des probabilités

Cette théorie représente le plus ancien cadre mathématique utilisé pour modéliser l'information imparfaite. La notion de probabilité traite toujours un événement aléatoire ce qui signifie qu'on ne peut pas conjecturer certainement sa réalisation. De ce fait, la théorie de la probabilité traite précisément l'incertitude de l'information [Kolmogorov, 1956]. D'une façon formelle, il s'agit d'une fonction positive p , appelée distribution de probabilité [Jacod and Protter, 2003], définie sur l'intervalle $[0,1]$ par :

$$\begin{aligned} p : \Omega &\longmapsto [0, 1] \\ x_n &\longmapsto p(x_n) \end{aligned} \tag{2.1}$$

où : Ω représente un univers de n éléments x .

On désigne par événement, un ensemble A , tel que $A \subseteq \Omega$. $P(A)$ s'appelle une mesure de probabilité. Cette mesure, estime le degré de l'accomplissement de l'évènement A . Elle est définie par :

$$\begin{aligned} P : 2^\Omega &\longmapsto [0, 1] \\ A &\longmapsto P(A) \end{aligned} \tag{2.2}$$

Une mesure de probabilité doit satisfaire les propriétés suivantes :

- $P(\emptyset) = 0$ et $P(\Omega) = 1$
- $P(A) = \sum_{x_n \in A} p(x_n)$
- $P(A) + P(A^c) = 1$; où A^c représente le complémentaire de A
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$; où A et B deux évènements inclus dans Ω

La théorie des probabilités représente un cadre mathématique très puissant et riche pour modéliser l'information incertaine dans plusieurs domaines. Cependant, elle se voit très pauvre pour la description de l'ignorance totale ou partielle de la connaissance ; principalement dans le cadre d'ambiguïté. Elle n'est pas capable aussi de traduire les informations subjectives émanant d'un raisonnement fait par un expert du domaine. Pour cette raison plusieurs chercheurs [Dubois and Kalfsbeek, 1990, Walley, 1996, Raufaste et al., 2003, Mouchaweh et al., 2006] ont étudié ces limitations qui peuvent affecter le processus de raisonnement ou de prise de décision.

b. Théorie de l'évidence

Appelée encore la théorie des fonctions de croyance, cette théorie a été introduite par Shafer [Shafer, 1976], en se basant sur les travaux de Dempster [Dempster, 1967]. Elle constitue le cadre mathématique permettant de modéliser l'imperfection de l'information en évaluant la crédibilité (véracité de la réalisation) d'un évènement. L'univers Ω est appelé dans ce cadre "cadre de discernement". Formellement, cette théorie se base sur l'attribution de masses m dites de croyance aux évènements par une fonction appelée "fonction de masses" en partant d'un univers Ω supposé exhaustif. Cette fonction est définie par :

$$\begin{aligned} m : 2^\Omega &\longmapsto [0, 1] \\ A &\longmapsto m(A) \quad \text{où } A \subseteq \Omega \end{aligned} \quad (2.3)$$

Cette fonction doit satisfaire les propriétés suivantes :

- $m(\emptyset) = 0$
- $\sum_{A \in \Omega} m(A) = 1$

A partir de cette fonction, deux fonctions duales sont définies : la crédibilité Cr (2.4) et la plausibilité Pl (2.5). La première estime la croyance minimale que A s'est produit. La seconde, évalue avec quel degré A est réalisable (croyance maximale).

$$Cr(A) = \sum_{B \subseteq A} m(B) \quad (2.4)$$

$$\begin{aligned} Pl(A) &= \sum_{A \cap B = \emptyset} m(B) \\ &= 1 - Cr(A^c) \end{aligned} \quad (2.5)$$

avec A^c : évènement contraire de A .

L'apport de la théorie de l'évidence réside dans l'évaluation des hypothèses jointes (une composition de contenu informationnel). Elle réussit, comme la théorie des probabilités à modéliser l'incertitude et quelques informations imprécises. Jusqu'alors l'ambiguïté reste, encore, un problème à solutionner.

c. Théorie des ensembles flous

La théorie des ensembles flous est apparue comme remède au problème de l'ambiguïté des informations en se rapprochant du raisonnement humain. En effet, la théorie des systèmes suppose que l'être humain est un système très puissant [Melese, 1968], dont le raisonnement admet les imperfections des informations traitées. Le processus cognitif humain a montré d'impressionnantes capacités à traiter des informations imparfaites dans plusieurs situations complexes. Cette performance séduisait plusieurs chercheurs dont Zadeh, fondateur de la théorie des ensembles flous [Zadeh, 1965]. Cette théorie repose sur le principe d'appartenance partielle à un ensemble dépassant, ainsi, la théorie classique admettant un principe binaire d'appartenance (un élément soit il appartient ou il n'appartient pas à un ensemble). Cette théorie a apporté un gain considérable en traitant des ensembles aux bornes mal-définies. Le formalisme de la logique floue est traduit par une fonction d'appartenance caractérisant un ensemble flou, et définie par :

$$\begin{aligned} \mu_A : \Omega &\longmapsto [0, 1] \\ x_n &\longmapsto \mu_A(x_n); \quad \text{où } A \subseteq \Omega \end{aligned} \quad (2.6)$$

La mesure μ_A estime le degré d'appartenance d'un élément (x_n) à un ensemble (A). Ainsi, trois cas se présentent :

- Si $\mu_A(x_n) = 0$ Alors x_n n'appartient pas à A
- Si $0 < \mu_A(x_n) < 1$ Alors x_n appartient partiellement à A
- Si $\mu_A(x_n) = 1$ Alors x_n appartient absolument à A

En opérant avec les ensembles flous, les opérations d'intersection et d'union sont définies de la manière suivante :

$\forall x_n \in \Omega$

- L'union de $A, B \subseteq \omega$ est un ensemble flou contenant les éléments de A ou de B tel que leur degré d'appartenance à leurs ensembles soit le plus grand.

$$\mu_{A \cup B}(x_n) = \max(\mu_A(x_n), \mu_B(x_n)) \quad (2.7)$$

- L'intersection de $A, B \subseteq \omega$ est un ensemble flou contenant les éléments de A ou de B tel que leur degré d'appartenance à leurs ensembles soit le plus petit.

$$\mu_{A \cap B}(x_n) = \min(\mu_A(x_n), \mu_B(x_n)) \quad (2.8)$$

d. Théorie des possibilités

Étendant ses travaux sur les ensembles flous, Zadeh a introduit la théorie des possibilités [Zadeh, 1978] qui a été développée ensuite par [Dubois and Prade, 1988]. Fondée dans l'intention de traiter plusieurs types d'imperfections de l'information dans un cadre mathématique unique, cette théorie a été appliquée dans plusieurs domaines et principalement dans le domaine médicale où les informations souffrent de plusieurs formes d'imperfection. Le formalisme de base de cette théorie est la distribution de possibilité (notée π) qui représente une fonction similaire à la fonction

d'appartenance de la théorie des ensembles flous. C'est une fonction de l'univers Ω à valeurs dans $[0, 1]$ définie par :

$$\begin{aligned} \pi : \Omega &\rightarrow [0, 1] \\ x_n &\mapsto \pi(x_n) \end{aligned} \quad (2.9)$$

$\pi(x_n)$ estime le degré de possibilité pour que x_n soit celui qui se produit [Dubois et al., 2007]. $\pi(x_n) = 1$ signifie que x_n est totalement possible, alors que $\pi(x_n) = 0$ signifie que x_n est totalement impossible. Ainsi, cette théorie peut envisager deux situations :

- une information complète sur l'univers Ω si $\pi(x_i) = 1$ et $\pi(x_n) = 0$; $\forall x_n \in \Omega$ tel que $x_n \neq x_i$, $x_i \in \Omega$
- une ignorance totale sur l'univers Ω si $\pi(x_n) = 1$; $\forall x_n \in \Omega$ (tous les éléments de l'univers Ω sont totalement possibles)

Dans le cadre de cette théorie, l'imperfection est formulée par deux mesures : la mesure de possibilité et la mesure de nécessité. La mesure de possibilité, associée à un évènement, estime à quel degré cet évènement est réalisable. Elle est définie par la fonction 2.10 :

$$\begin{aligned} Ps : \Omega &\mapsto [0, 1] \\ A &\mapsto \Pi(A) \end{aligned} \quad (2.10)$$

La mesure de possibilité d'un ensemble de l'univers Ω représente la plus grande distribution de possibilité attribuée à un élément de cet ensemble.

$$\Pi(A) = \max_{x \in A} \pi(x) \quad (2.11)$$

Cette fonction admet les propriétés suivantes :

- $\Pi(\emptyset) = 0$ et $\Pi(\Omega) = 1$
- $\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$; où A et B deux évènements inclus dans Ω
- $\Pi(A \cap B) = \min(\Pi(A), \Pi(B))$; où A et B deux évènements inclus dans Ω

La mesure de la nécessité, quant à elle, associée à un évènement, évalue à quel point l'évènement est certain. Elle est définie par :

$$\begin{aligned} Ps : \Omega &\mapsto [0, 1] \\ A &\mapsto N(A) \end{aligned} \quad (2.12)$$

Les propriétés liées à la dualité des deux mesures sont :

- $\Pi(A) < 1 \Rightarrow N(A) = 0$
- $N(A) > 0 \Rightarrow \Pi(A) = 1$

2.3 Modélisation de connaissances à partir des images

La modélisation de connaissances désigne le processus de spécification formelle des différentes connaissances contenues dans un support d'information. Plusieurs

modèles sont définis permettant de représenter les connaissances (Les signatures, les arbres, les réseaux de neurones, les ontologies, etc...). Dans la littérature plusieurs travaux étaient menés pour la représentation des connaissances liées aux objets d'une scène à partir des images suivant différents modèles. Dans cette section nous définissons, tout d'abord, la notion de modèle. Ensuite, nous nous limitons à la présentation des signatures et des ontologies qui sont les deux modèles que nous utilisons dans notre approche.

2.3.1 Modèles

L'histoire des modèles naissait il y a longtemps avec les beaux-arts où un modèle est défini comme une représentation réduite et simplifiée d'un objet [Chevret, 2007]. Dans la deuxième moitié du XXème siècle, le mot a été étendu vers plusieurs autres disciplines telles que les sciences humaines, la science de la terre, la médecine, etc... Il décrivait alors d'une façon précise un objet ou un phénomène réel à étudier. De ce fait, la modélisation était très utile à pronostiquer l'état de certains objets (Ex : évolution de certaines maladies) et à analyser aisément certains aspects. Dès lors, le modèle signifiait "une représentation virtuelle de phénomènes réels à partir d'algorithmes et de modèles mathématiques" [Dutang and Contamine, 2005]. Par la suite, plusieurs outils et méthodes étaient utilisés afin de concevoir les modèles adéquats. En se référant à l'encyclopédie Universalis un modèle est défini comme "une représentation schématique d'un objet ou d'un processus qui permet de substituer un système plus simple au système naturel." Ainsi un modèle permet de simuler de façon plus simple l'objet qu'il représente. La modélisation des objets et des phénomènes réels a contribué à l'évolution importante des différents domaines d'application.

La modélisation des connaissances désigne la représentation des connaissances extraites d'un support contenant les données ou les informations liées à une scène réelle suivant une représentation formelle appelée modèle de représentation des connaissances. "La modélisation des connaissances est un thème qui concerne toutes les phases d'extraction, de gestion, et de partage des connaissances. Le but premier d'un formalisme de modélisation est de permettre de représenter, modéliser, stocker et exploiter des données, des informations ou des connaissances" [H. Briand, 2008]. Dans [Bloche and Maitre, 2004], les auteurs ont présenté quelques modèles de représentation de connaissances utilisés dans les systèmes à base de connaissances en traitement d'images. Une présentation des travaux sur les systèmes d'aide à la décision clinique a été établie dans [Kong et al., 2008] avec une focalisation sur la représentation des connaissances. Dans ce contexte, Carter, affirme que le but de la représentation des connaissances est "produire des systèmes intelligents traitant des informations d'un domaine spécifique sous une forme manipulable efficacement." [Carter, 1999].

2.3.2 Signatures

La notion de signature est utilisée depuis longtemps comme un identifiant unique d'une entité. Son caractère d'exclusivité a séduit les chercheurs dans plusieurs domaines pour l'appliquer dans certains traitements. Le terme signature peut avoir diverses définitions suivant le domaine d'application où il est intégré. Généralement, une signature est une marque qui renseigne l'identité d'une personne. Elle est définie comme "un signe graphique considéré comme représentatif et même exclusif" [Denizot, 2007]. Utilisant ce terme en traitement des images numériques, une signature d'une image consiste à ajouter à l'image une composante qui permet de l'identifier de façon unique.

En effet, la signature de l'image permet de reconstituer l'information pertinente de l'image en un ensemble de propriétés [Zlatoff, 2006]. Elle est appliquée surtout en tatouage d'image afin de contrôler l'utilisation sécurisée des images ([A.Umameswari and G.R.Suresh, 2014], [Pan et al., 2010]). Cette technique trouve, aussi, toute son ampleur dans le domaine de l'indexation et de la recherche des images par le contenu où l'information contenue dans une image requête est modélisée par une signature qui sera comparée à des signatures associées à des images d'une base généralement grande ([Lee et al., 1992], [Yeh and Chang, 2006], [Veltz, 2004], [Gosselin, 2005], [Simand and Jolion, 2005]). Cette signature est définie comme un vecteur ou un fichier spécifiant des informations liées à des caractéristiques extraites d'une image afin de modéliser une identification unique de l'image ou de son contenu. Généralement, les travaux menés sur l'indexation des images et la recherche par le contenu utilisent les signatures des images de même modalité. Et par la suite la mesure de similarité entre les signatures mis en jeu est dans la majorité des cas une mesure de distance. Le processus de la mise en correspondance des images numériques par leurs signatures passe par trois étapes fondamentales :

- L'extraction des caractéristiques sur lesquelles est basée la signature.
- Le calcul de la signature de l'image qui peut être un vecteur ou un fichier de caractéristiques
- L'application d'une mesure de similarité permettant de détecter les images similaires.

a. Signatures et images mammographiques

Des travaux étaient aussi menés, dans le domaine médical, sur l'utilisation des signatures pour représenter le contenu des images médicales ([Andaloussi et al., 2009], [Krishnamurthy and Kesavan, 2015], [Pamart et al., 2015])

Dans le domaine mammographique, Lamine et al. [Lamine et al., 2014] proposent de représenter les images mammographiques par un graphe orienté d'objets où chaque objet est décrit par une signature. Cette signature est composée de 10 caractéristiques numériques. Elle est utilisée pour comparer deux graphes à travers leurs objets. Jouirou et al. [Jouirou et al., 2015] proposent de représenter des régions d'intérêt d'une mammographie par une signature composée de 16 caractéristiques

extraites en utilisant la méthode curvelette. Cette méthode décompose l'image en 4 couches, à partir de chacune des 4 premiers moments sont calculés. D'autres travaux, [Paramkusham et al., 2015, Dhanalekshmi and Phadke, 2013], proposent de symboliser les contours des masses par une signature 1D. Cette signature définit les points du contour par les distances qui les séparent du centre de la masse. Elle est ensuite utilisée pour classer les masses en bénignes ou malignes. Zyout et Togneri [Zyout and Togneri, 2016] proposent une méthode pour la détection des distorsions architecturales dans une mammographie utilisant la signature du tissu environnant l'anomalie. En effet, une signature statistique du tissu est calculée à travers les 2DIMF (2D Intrinsic Mode Functions) extraites pour chaque région d'intérêt. Cette signature est utilisée pour distinguer les tissus pathologiques des tissus normaux.

b. Signatures sémantiques

Dans plusieurs cas les informations visuelles extraites des images ne sont pas en corrélation avec la sémantique de l'image créant ce qu'on appelle le "fossé sémantique" entre la représentation numérique de l'image et son contenu sémantique. L'intérêt de l'interprétation et de l'analyse des images motive les chercheurs à adopter des stratégies tenant compte de la sémantique des images. Pour cette raison, la modélisation des images par signature passe d'une représentation d'information à une représentation de connaissances à travers des signatures sémantiques.

Dans [Gong et al., 2013] les auteurs proposent deux signatures sémantiques en se basant sur le processus cognitif humain permettant de décrire les objets à travers certains attributs ayant un sens sémantique (Ex : Un objet est rigide ou flexible). La première est une signature spécifiant les attributs liés aux objets. La deuxième signature permet de décrire un objet à travers une comparaison faite d'autres objets de référence. Dans [Wang et al., 2014], les auteurs proposent une approche pour classer des images par leurs sémantiques par rapport à une image référence. En effet, les caractéristiques visuelles extraites des images sont projetées dans des espaces sémantiques qui leur sont reliés. Les auteurs rapportent que cette solution augmente de manière significative la précision des résultats. En se basant sur ces travaux, [Patil et al., 2016] ont utilisé cette approche pour rechercher des images à travers leurs contenus en les représentant par des signatures sémantiques.

2.3.3 Ontologies

Un autre modèle très utilisé pour la représentation des connaissances est l'ontologie. Dans [Idoudi, 2017], Idoudi a collecté les principales définitions attribuées au concept d'ontologie. Une ontologie est définie comme une spécification hiérarchique conceptualisant une réalité [T.R.Gruber, 1993]. Le rôle d'une ontologie est de formaliser les connaissances d'un domaine avec une manière générique. Une ontologie est formée essentiellement par quatre composantes :

- Les concepts définis comme des entités utilisées pour la spécification des connaissances.
- Les relations représentent les liens définissant les interactions entre les concepts.
- Les axiomes décrivent les connaissances élémentaires ne pouvant pas être définies par les concepts ou les relations. Ce sont des propriétés toujours vraies des concepts ou des relations
- Les individus (appelés aussi des instances) représentent des cas particuliers des concepts génériques.

Plusieurs classifications ont été proposées pour les ontologies. [Guarino, 1997] classe les ontologies en quatre types :

- Les ontologies de haut niveau représentent des ontologies abstraites composées par des concepts généraux qui incluent des concepts plus spécifiques attachés aux sous-domaines.
- Les ontologies de domaine modélisent des connaissances d'un domaine spécifique. Elles spécifient les concepts et les relations de ce domaine
- Les ontologies de tâche (Domain-Task) représentent une spécification des ontologies de domaine. Elles servent à modéliser des connaissances liées à des tâches spécifiques ou des solutions particulières de façon indépendante du domaine.
- Les ontologies d'application représentent le niveau le plus spécifique de l'organisation des types. C'est une spécialisation des deux types précédents.

Les travaux de Gowsikhaa et al. [Gowsikhaa et al., 2012] s'intègrent dans le domaine de la recherche d'image par le contenu. Ils commencent par construire une ontologie à partir d'une image en utilisant trois types de caractéristiques de bas niveau extraites au préalable (Caractéristiques de forme, caractéristiques de couleur et caractéristiques de texture). Après extraction de caractéristiques, une étape de classification est effectuée dans le but de grouper des entités suivant certains critères de domaine. A partir de cette classification et en se référant à la terminologie du domaine, les auteurs construisent l'ontologie manuellement en définissant les concepts et les relations entre les concepts puis en instanciant cette ontologie par les valeurs des caractéristiques extraites. Cette ontologie est utilisée, ensuite, pour la recherche des images par leur contenu à partir d'une base de données.

Dans le même but, ils utilisent cette ontologie pour la recherche d'images par le contenu dans le web. Wang et al. [Wang et al., 2008] ont proposé une construction automatique d'une ontologie multi-modale à partir de Wikipedia et des caractéristiques extraites des images. En considérant une page de Wikipedia, les auteurs utilisent les mots ayant des définitions sur Wikipedia et liés vers leurs pages correspondantes pour construire une partie de l'ontologie. En ce qui concerne l'autre partie de l'ontologie, elle est définie à partir des images. Chaque image est représentée par un vecteur caractéristique. Une classification des caractéristiques est ensuite effectuée, définissant 16 concepts qui décrivent la couleur et la texture de l'image.

S'intégrant aussi dans ce cadre, les travaux de [Iskandar et al., 2007b] proposent une approche permettant de construire une ontologie à partir des images. Chaque image est annotée en appliquant la technique "equal-weight linear combination" [Iskandar et al., 2007a] par certains concepts. Ces concepts sont utilisés, par la suite, pour spécifier l'ontologie. Si une image contient un nouveau concept, l'ontologie est augmentée en lui ajoutant ce nouveau concept. Dans [Liu et al., 2010], les auteurs proposent de construire une bibliothèque d'ontologies à partir des images. Ils se servent de cette bibliothèque pour la recherche d'images par leur sémantique. Afin de transformer une image vers une ontologie, ils utilisent les caractéristiques SIFT (Scale-Invariant Feature Transform) et les caractéristiques de bas niveau liées principalement aux couleurs, aux textures et aux formes des entités présentes dans les images ; ces entités sont préalablement extraites. Ensuite, une annotation sémantique des entités est effectuée manuellement pour donner lieu à l'ontologie définissant l'image.

Dans [Wang and Yang, 2011], les auteurs présentent une approche pour la représentation d'une image par une ontologie. Opérant de la même manière, ils procèdent tout d'abord à une extraction des caractéristiques de l'image. Ces caractéristiques sont groupées en classes caractérisées par des relations, permettant de construire une base de connaissances relative à l'image. Ces connaissances servent pour son annotation. A travers cette annotation, les classes spécifiées et les relations définies entre les classes, l'ontologie décrivant l'image est construite. Dans [Liaqat, 2013], Liaqat propose de créer une ontologie floue à partir d'une image dans le but de réduire le fossé sémantique entre le traitement du système et le processus cognitif humain dans un processus de récupération du contenu et de classification des images. Dans une première étape, l'auteur propose de construire une ontologie simple à partir d'une image annotée. Dans une deuxième étape, il utilise le principe de la logique floue, pour modéliser l'imperfection de l'information représentée par l'ontologie. Il compare les deux résultats et affirme que l'ontologie floue représente le meilleur outil pour la représentation du contenu des images.

Dans le domaine mammographique, dans la majorité des travaux les auteurs utilisent les ontologies de domaine pour comprendre la sémantique des images. Les ontologies du domaine mammographique sont principalement utilisées dans l'annotation des images [Tutac et al., 2008, Bulu et al., 2013, Ilyass et al., 2015].

2.4 Modélisation 3D de scènes réelles

L'évolution des techniques de développement informatique et technologique a permis d'offrir de nouveaux outils permettant une représentation simplifiée de systèmes complexes dans plusieurs domaines. Le besoin d'interagir efficacement avec le monde réel, exige de modéliser la réalité. Dans cette intention, reproduire une scène réelle

de façon synthétique plus simplifiée, rend possible toute simulation pour en évaluer les résultats. Tel est le principal intérêt de la modélisation virtuelle des scènes réelles opérée avec les technologies de la "réalité virtuelle" ou encore la "réalité augmentée". Ces technologies sont très utilisées dans le domaine de la robotique où plusieurs simulations sont effectuées pour évaluer les environnements surtout industriels. Le domaine médical est, aussi, un domaine, où la modélisation de scènes réelles est d'une extrême utilité dans les diagnostics médicaux surtout des cas critiques.

2.4.1 La modélisation 3D dans le domaine médical

Depuis des années, l'exploration du corps humain a permis l'obtention d'une multitude d'informations à travers plusieurs moyens d'investigation. Ces informations sont utilisées, ensuite, par le clinicien pour effectuer une évaluation de la scène observée afin de diagnostiquer le cas. Certes, la multitude des images exploitées a permis de faciliter significativement le diagnostic, cependant, elles manquent toujours de précision et de fidélité au monde réel. La modélisation virtuelle d'un organe, a permis, dans ce contexte, de fournir une nouvelle représentation bénéficiant d'un certain nombre de caractéristiques renfermant de façon simplifiée l'essentiel de la scène observée. Dans le domaine médical, plusieurs chercheurs se sont intéressés à développer différents modèles virtuels de représentation des connaissances constituant un outil indispensable à l'aide au diagnostic.

Selon Chevret [Chevret, 2007], la modélisation passe par trois étapes principales :

- une phase de formulation des variables d'entrée :
- une phase d'identification : où les paramètres du modèle doivent être définis d'une façon précise à partir des variables d'entrées.
- une phase de validation : durant cette phase on obtient les variables de sortie souhaitées (le modèle virtuel 3D de représentation de connaissances).

Auparavant, les expériences sur des animaux étaient très utilisées dans les études médicales et certaines simulations. L'expérience animale peut être insuffisante vu les différences des anatomies entre les organes animaux et les organes humains. Dès l'apparition du concept de modélisation virtuelle 3D en médecine, plusieurs expériences sont devenues plus accessibles et plus faciles à étudier. Les modèles ont contribué à l'évolution rapide du domaine médical permettant d'une part d'aider les cliniciens au diagnostic des différents cas de pathologies et d'autre part leur offrir des moyens de simulations et d'étude plus sophistiqués. Les modèles peuvent ainsi être classés en deux catégories suivant leurs types. Les modèles physiques et les modèles numériques.

a. Modélisation physique

Depuis qu'on s'est aperçu de l'utilité de la modélisation des organes humains, plusieurs modèles physiques (fantômes) apparaissaient représentant différents organes

du corps. L'évolution du domaine de l'informatique et la puissance des outils de traitement de l'information a permis de produire des modèles de plus en plus réaliste. Les différents fantômes réalisés permettent surtout l'évaluation et la validation des nouvelles modalités d'acquisition d'images cliniques, l'assistance à la chirurgie ainsi que l'étude de certains phénomènes pathologiques. Les fantômes créés diffèrent en termes de complexité de la représentation et de méthodes de mise en oeuvre.

Parmi les fantômes apparus dans le domaine médical, on peut citer le fantôme AGATE (Amsterdam GATEd cardiac phantom) [Comas, 2005] et le fantôme DeBRa (Detailed Breast model for Radiological studies) [Ma et al., 2009]. AGATE est un fantôme du coeur d'une structure simplifiée, construit dans le but d'évaluer certaines mesures en "tomographie myocardique synchronisée" [Comas, 2005]. Quant au fantôme DeBRa, c'est un des premiers modèles du sein construit d'une façon détaillée afin de réaliser certaines études d'images mammaires [Ma et al., 2009]. Des formes pathologiques sont également modélisées et introduites dans le fantôme.

b. Modélisation numérique

Les modèles numériques sont des modèles basés sur des équations mathématiques permettant ainsi de créer des modèles virtuels d'organes plus ou moins complexes suivant les complexités des équations utilisées. Comas [Comas, 2005], dans son rapport décrivant ses travaux de recherches en modélisation, a classé les modèles numériques en quatre catégories :

- **Les atlas numériques** : représentent les différents organes d'un être humain à partir d'une banque de données. Les atlas sont basés généralement sur le projet de l'homme visible ("Visible Human Project") qui repose sur des images d'organes pris de cadavres de quelques hommes et femmes déclarant la permission de l'exploitation scientifique de leurs corps.
- **Les modèles basés sur les pixels** : sont très liés aux données cliniques d'un patient particulier. Les données sont extraites d'images radiologiques du patient. D'où il existe trois étapes nécessaires à la mise en oeuvre de ces modèles : une étape d'extraction de formes d'intérêt, une étape de traitement visant à personnaliser la représentation de ces formes dans le nouveau modèle et finalement l'implémentation de la représentation.
- **Les modèles mathématiques** : sont basés sur différentes primitives géométriques et des calculs analytiques afin de produire un modèle générique pouvant être utilisé pour effectuer des études d'ordre général. La complexité du modèle conçu est proportionnelle aux équations utilisées. Plus le modèle s'approche de la représentation réelle, plus il est complexe à mettre en oeuvre.
- **Les modèles hybrides** : sont des modèles basés à la fois sur des primitives géométriques et des données spécifiques aux patients. Dans certains travaux, les chercheurs utilisent un modèle générique auquel ils ajoutent des données extraites d'images du patient.

2.4.2 La modélisation 3D du sein féminin

Depuis quelques années, plusieurs chercheurs se sont intéressés à modéliser le sein de la femme pour différentes raisons. La plupart des modèles étaient réalisés pour construire des fantômes afin d'offrir aux cliniciens une représentation virtuelle pouvant leur servir à faire quelques simulations requises.

Parmi les premiers travaux en modélisation de fantôme d'un sein féminin est celui de Yu-Neifert en 1995 [Rajagopal et al., 2007]. Yu-Neifert a implémenté un modèle contenant principalement deux composantes : la peau et le tissu adipeux. Ce modèle a été reconstruit en utilisant des formes tri-linéaires dans le but de prédire l'applicabilité de l'interférométrie holographique pour la détection du cancer du sein.

S'alignant dans cette voie de recherche, en 2000, un groupe de chercheurs [Azar et al., 2000] se sont intéressés à développer un modèle du sein afin de prévoir la position de la tumeur dans un sein compressé pendant une opération de biopsie. Le travail consiste à segmenter l'image IRM d'une façon semi-automatique. Un modèle 3D est ensuite généré à partir du contour et des formes extraites de l'image à l'aide d'un programme C.

En 2001, Wu et Sullivan [Wu and Sullivan, 2001], dans le but de créer un modèle numérique pour l'aide à la détection et au diagnostic du cancer du sein, utilisent l'algorithme *marching cubes* et la technique du seuillage d'histogramme d'intensité pour la modélisation du sein à partir d'une image IRM segmentée auparavant. Dans cette même année, un autre groupe de chercheurs [Samani et al., 2001] a proposé une technique de maillage appelée "transfinite interpolation mesh generation". Cette méthode consiste en premier temps à segmenter l'image en différents types de tissus en utilisant la technique de seuillage. Des éléments sont ensuite créés pour chaque partie de l'image pour produire une maille 3D.

Entre 2003 et 2006, un groupe de chercheurs a introduit le fantôme représentant la structure la plus détaillée du sein [Ma et al., 2009] qu'ils appellent DeBRa. Il est construit par un ensemble de primitives géométriques (ellipsoïdes) et personnalisé afin d'avoir une représentation réaliste. Les tumeurs et les micro-calcifications ont été également modélisées et introduites dans le fantôme.

En 2011, et dans l'intention d'améliorer le fantôme DeBRa, KW et Alghamdi [KW A, 2011] ont travaillé sur le développement d'un modèle pour la construction d'un fantôme du sein qu'ils appellent FATMA comportant une structure détaillée pour des études dosimétriques. Leur méthode consiste à modéliser la forme extérieure du sein par un semi-ellipsoïde et une hyperbole (Fig. 2.2). La structure interne est représentée par des petits ellipsoïdes représentant les ligaments de Cooper et un ensemble de cylindres divergeant à la mamelle représentant les canaux galactophoriques. Le modèle contient aussi une représentation de la peau consistant en une

couche d'épaisseur uniforme entre la couche extérieure et la couche intérieure du tissu mammaire.

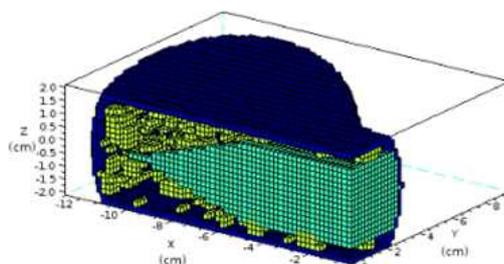


FIGURE 2.2 – Modèle numérique du fantôme FATMA [KW A, 2011]

Une approche similaire a été adoptée en 2012 par Bliznakova et al. [Bliznakova et al., 2012] dans leurs travaux de recherche consistant à modéliser un sein cancéreux (Fig. 2.3). L'apport de ces travaux réside dans la représentation de formes pathologique dans le modèle du sein proposé. En effet, les anomalies avaient des formes rondes, ovales ou irrégulières.

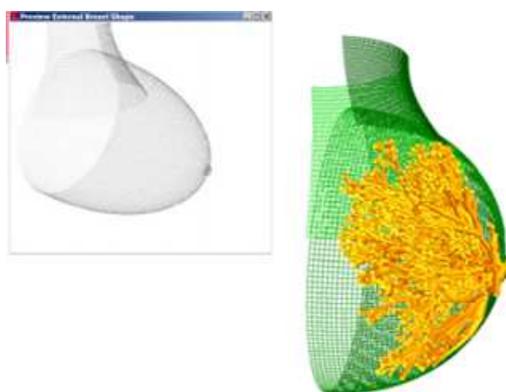


FIGURE 2.3 – Modèle numérique du fantôme FATMA [Bliznakova et al., 2012]

En 2013, un groupe de chercheurs [Wessel et al., 2013] ont développé un modèle du sein cancéreux dans le but de recalibrer deux images représentant le sein de deux façons différentes : une image IRM prise en position faciale et une image US prise dans une position dorsale. Le modèle présente l'avantage de la représentation des formes spéculées des tumeurs avec la présence du tissu dense à leur voisinage. Afin de parvenir au résultat souhaité, les formes requises ont été segmentées manuellement par Seg3D (un logiciel gratuit développé par une équipe d'étudiants dans l'université

d'UTAHA). Le logiciel SciRun a été ensuite utilisé pour la création de la géométrie du sein. Enfin, la plateforme Abaqus a servi pour l'implémentation du modèle final.

Récemment, un autre groupe de chercheurs [Han et al., 2014] ont abordé également le problème de recalage de deux images IRM de deux positions différentes : faciale et dorsale. Le modèle a été construit à partir de l'image IRM en la segmentant en quatre classes principales (Tissu adipeux, tissu fibroglandulaire, peau et muscle pectoral). L'algorithme *marching cubes* a été utilisé afin d'extraire les surfaces à modéliser et finalement le modèle est généré par un maillage tétraédrique en employant la méthode des éléments finis.

Les travaux de recherche concernant la modélisation du sein féminin pour l'aide au diagnostic des tumeurs sont multiples. Suivant les objectifs du traitement, certains chercheurs ont opté pour des modèles génériques, d'autres pour des modèles spécifiques. La plupart des travaux sont réalisés dans le but de recalibrer deux représentations différentes d'images médicales de modalités différentes. La majorité des travaux effectuent un recalage pixelique des images ou des formes étudiées.

2.5 Conclusion

L'image, en tant que support d'information, a contribué incontestablement à l'amélioration de l'examen médical. Elle permet de visualiser la structure interne des organes étudiés, offrant au médecin un outil aidant à révéler les causes pathologiques d'un certain diagnostic. Certaines conditions peuvent affecter la qualité de l'information et par la suite la précision du diagnostic. Une modélisation adéquate de l'information dans son cadre imparfait peut assister le raisonnement du clinicien.

Par ailleurs, la virtualisation est une technique qui permet de fournir un modèle virtuel 3D présentant une reproduction simplifiée d'une scène réelle observée par un capteur. Employée dans le domaine médical, elle constitue un outil captivant permettant de retrouver une représentation symbolique des différents éléments significatifs d'une scène étudiée. Ceci, permet de surmonter certaines imperfections et par la suite améliorer le diagnostic.

Dans ce chapitre, nous avons présenté les différents types d'information ainsi que les principaux types d'imperfections pouvant l'affecter. Nous avons exposé les différents cadres mathématiques traitant ces imperfections (la théorie des probabilités, la théorie de l'évidence, la théorie des ensembles flous et la théorie des possibilités). La modélisation d'informations et de connaissances a été abordée à travers les signatures et les ontologies, dont nous avons présenté quelques travaux en littérature. Finalement, nous avons abordé le principe de la modélisation virtuelle 3D d'informations.

Fusion d'images et aide au diagnostic médical

Sommaire

3.1 Fusion d'information	48
3.1.1 Les niveaux de fusion	49
3.1.2 La fusion d'images mammaires	50
3.2 Aide au diagnostic par raisonnement à base de cas	53
3.2.1 Principe du raisonnement à bases des cas	54
3.2.2 Travaux sur le raisonnement à base de cas en diagnostic médical	55
3.3 Conclusion	58

L'incomplétude de l'information est une autre imperfection limitant les experts des domaines à mieux analyser et interpréter de manière fiable les scènes observées. Ce qui conduit à un besoin de combiner les informations et un avancement des méthodes de fusion. En effet, l'utilisation de plusieurs informations issues de différents capteurs afin de mieux appréhender une situation remonte vers les premières traces de la vie sur la Terre. Depuis fort longtemps les êtres vivants ont exprimé un besoin de combiner plusieurs informations afin d'améliorer leur perception d'une situation. Et ceci dans le but de prendre les décisions adéquates en fonction des connaissances dont ils disposent. La fusion de données à proprement dit a vu le jour avec le domaine militaire pour des besoins de surveillance des champs de batailles. L'idée s'est vite transmise vers les autres domaines afin de bénéficier de ses avantages pour la qualité de la représentation des informations et de l'aisance de la prise des décisions.

Dans le domaine médicale, la prise de décision est une tâche cruciale surtout lorsqu'il s'agit de manipuler des informations complexes entachées par plusieurs imperfections. Un clinicien, s'appuie souvent sur l'historique des diagnostics ayant déjà en tête afin de fournir le diagnostic adéquat au cas présent. Cette action est

connue sous le nom de Raisonnement à Base de Cas (ou Case Base Reasoning : CBR en Anglais) dans le domaine de l'intelligence artificielle. Il est principalement utilisé dans la résolution des problèmes et la prise de décision en se basant sur des expériences antérieures. Le principe de ce paradigme consiste à reproduire le raisonnement humain pour formuler un problème et chercher la solution adéquate.

Ce chapitre est réservé à la présentation de ces deux concepts : La fusion d'information et le raisonnement à base de cas. Pour le premier cas, nous présentons les différents niveaux de fusion définis en littérature, puis nous rapportons certains travaux effectués sur la fusion d'images mammographiques dans chacun de ces niveaux. Le raisonnement à base de cas est présenté dans un deuxième temps en expliquant son principe et les travaux effectués dans son contexte.

3.1 Fusion d'information

Afin d'analyser et d'interpréter une scène, l'expert du domaine cherche souvent le maximum d'informations utiles. Une des difficultés de cette analyse est l'incomplétude des informations disponibles. Cette limite peut affecter le processus de prise de décision. Une solution très adéquate est la fusion d'information. Selon LAROUSSE, la fusion est la "réunion en un seul groupe de divers éléments distincts". Partant de cette définition la fusion d'informations se voit comme l'association d'informations issues de diverses sources. Cependant, cette information peut être vue suivant plusieurs types. Cette diversité peut affecter la définition du concept de fusion d'informations pour l'adapter au type concerné. Le type de l'information n'est pas le seul critère à être pris en considération afin de définir ce concept, l'objectif de la fusion à appliquer est aussi un aspect important à cette définition. De cette intention, il existe différentes définitions bien que les similitudes sont aussi maintes.

En 2001, Isabelle Bloch affirme que la fusion consiste à combiner des informations venant de multiples sources et d'exploiter cette combinaison pour répondre à des questions ou prendre des décisions [Boström et al., 2007]. Dans cette définition, Bloch essaie d'exprimer le concept de fusion d'informations à partir des actions faites afin d'atteindre son objectif : l'action "combiner" qui explique le principe du concept et l'action "exploiter" qui spécifie la finalité de son existence.

Dans son oeuvre "Organon", Aristote précise que "La définition fait connaître ce qu'est la chose". Ainsi, en introduisant le principe de la théorie de la définition, il explique que partant de l'idée que "la science est la recherche des causes", une définition doit exprimer "l'essence" d'un phénomène en mettant en évidence sa "cause" [Valois, 1994]. Tenant compte de cette philosophie de la définition, nous désignons par la fusion d'informations, le mécanisme permettant l'association des informations caractérisant un même problème mais de différentes provenances pour

retenir un maximum d'informations en diminuant les redondances. Cette définition a un aspect général pouvant être appliqué à n'importe quel type d'informations malgré la diversité des domaines d'applications.

3.1.1 Les niveaux de fusion

La question de hiérarchiser la fusion suivant différents niveaux est une question très abordée dans les travaux de fusion d'information et surtout des images. En effet, le choix du niveau est crucial pour faire la fusion. [Dasarathy, 1997] présentent trois niveaux pour la fusion d'information : le niveau *données*, le niveau *caractéristiques* et le niveau *décision*. Le premier niveau correspond à la fusion de données issues directement de leurs sources. Le second niveau se réfère à la fusion d'information extraites des données issues des sources. Le dernier niveau correspond à la fusion d'hypothèses issues d'un raisonnement fait sur l'information et les données sources. Cette organisation est souvent utilisée avec cette terminologie [Verlinde, 1999], où en adoptant la terminologie "*bas niveau*", "*niveau intermédiaire*" et "*haut niveau*" [Solaiman, 2001] respectivement pour chaque niveau présenté précédemment. [Wald, 2002] a proposé aussi trois niveaux de fusion avec une autre désignation : niveau *attribut*, niveau *analyse* et niveau *représentation*. Il reprend pour les premiers niveaux respectivement la terminologie de caractéristiques et de décision et définit un nouveau niveau : la fusion de représentation signifiant la fusion de modèles d'information. Dans les travaux de [Desodt-Lebrun, 1996] quatre niveaux de fusion sont définis selon le type de l'information traitée : niveau *signal*, niveau *pixels*, niveau *caractéristiques* et niveau *symbolique*. Utilisant, également le type de l'information, [Fabre, 1999] a défini trois niveaux : le *bas niveau* (qui correspond à la fusion de pixels), le *niveau numérique* (l'information numérique correspond à un attribut) et le *niveau symbolique*.

En traitant la problématique de la fusion d'images, [Mangolini, 1994] a proposée une architecture de trois niveaux niveau *pixels*, niveau *caractéristiques* et niveau *décision*. Ces différents niveaux sont présentés dans la figure 3.1 basée sur l'image présentée dans [Mangolini, 1994]. Un autre niveau a été défini pour la fusion d'images : le niveau *Score* [Ross et al., 2006]. A partir des images, un vecteur score est calculé (sa dimension est égale au nombre de source de données). Les différents scores constituant le vecteur sont combinés en un seul score permettant de fusionner les différentes données et information.

Malgré les différences dans les terminologies proposées, la fusion d'information est organisée dans la plupart des travaux suivant trois niveaux :

- Le niveau Données : Les images sont considérées comme des sources de données directement exploitables, et la fusion de fait ainsi avec les pixels constituant les images.
- Le niveau caractéristiques : A partir des images étudiées, certaines caractéristiques sont extraites suivant plusieurs critères. Une mesure de similarité

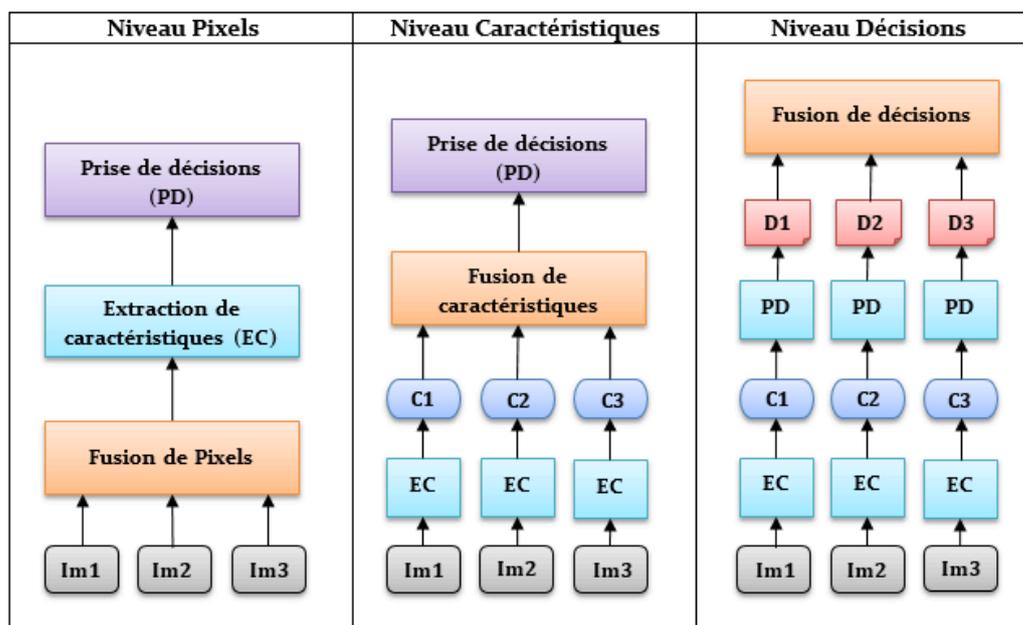


FIGURE 3.1 – Les niveaux de la fusion d'images

doit être effectuée sur ces caractéristiques pour pouvoir les combiner.

- Le niveau décision : Dans ce niveau, une décision automatique est générée à partir de l'images traitée (généralement à partir des caractéristiques extraites). La fusion est effectuée, ainsi sur les décisions générées.

3.1.2 La fusion d'images mammaires

La fusion d'images médicales est un axe de recherche étudié par plusieurs chercheurs depuis longtemps. C'est un axe qui représente toujours un challenge pour les chercheurs afin de produire un système d'aide au diagnostic aussi fiable que possible afin d'aider le médecin à bien explorer le cas et prendre la bonne décision. Le traitement des images mammaires est l'un des axes médicaux les plus étudiés. Dans ce qui suit, nous rapportons quelques travaux récents sur la fusion d'images mammaires en opérant dans un des niveaux présentés.

Dans les travaux de [Hopp et al., 2012a], une approche de fusion de bas niveau est proposée permettant de fusionner une image mammographique avec une image IRM. Cette fusion est effectuée dans le but de localiser les objets pathologiques en utilisant les deux modalités. Cette approche est basée sur trois étapes :

- Prétraitement de deux images : Cette étape consiste à préparer les deux images pour l'étape suivante. Pour les deux images une segmentation est effectuée permettant de séparer l'arrière-plan de l'objet sur l'image. Une rotation est appliquée afin d'avoir un alignement global. L'image mammographique subit

une remise à l'échelle afin de la recaler à l'image IRM.

- Recalage des deux images : Pour simuler la compression du sein produisant l'image mammographique, un modèle biomécanique est généré à partir de l'IRM. La méthode des éléments finis (FEM : Finite Element Method) est appliquée pour simuler le comportement du sein durant la déformation par la compression. Certains paramètres sont lus des métadonnées de la mammographie pour trouver la bonne déformation du modèle. Une projection du volume comprimé est générée.
- Fusion : Les deux images 2D (mammographique et la projection créée à partir de l'IRM), recalées, sont combinées.

L'approche est évaluée par mesure de précision de l'étape de recalage. Elle est calculée en divisant la distance euclidienne séparant les deux centres des lésions sur les deux images par le diamètre maximal de la lésion marquée par les experts sur la mammographie. Les auteurs montrent que sur 11 couples d'images 10 présentent une bonne précision.

Dans les travaux de [Hopp et al., 2012b] l'approche utilisant un modèle biomécanique du sein est également utilisée pour la fusion d'une image mammographique avec une image échographique. La procédure contient trois étapes :

- Le prétraitement des deux images, permettant de les préparer pour le recalage. L'image mammographique est initialement segmentée pour séparer le sein de l'arrière-plan en utilisant un algorithme de seuillage et des opérateurs morphologiques. La résolution de l'image est modifiée et une rotation est aussi effectuée afin d'obtenir un alignement exact. Pour l'image échographique, une segmentation est également effectuée en utilisant un masque elliptique pour l'approximation du contour. L'erreur de la segmentation est corrigée manuellement.
- La construction du modèle biomécanique est effectuée à partir de l'image échographique en utilisant l'algorithme de la triangulation de Delauny. Ensuite, une simulation de la compression du sein est effectuée en utilisant la méthode des éléments finis. Des paramètres utilisés (épaisseur du sein comprimé) sont extraits des métadonnées de l'image mammographique. Une projection est prise à partir de cette simulation.
- Le recalage est effectué entre l'image mammographique et la projection effectuée à partir du modèle biomécanique. La fusion est effectuée suivant trois méthodes : la projection sound speed est combinée à l'image mammographique, la projection atténuation est combinée avec l'image mammographique, les deux projections sont combinées avec l'image mammographique.

Afin d'analyser les images une classification est effectuée avec la méthode SVM.

Dans les travaux de [Minavathi et al., 2013] une approche de fusion multimodale au niveau caractéristiques a été proposée en vue de classifier les masses mammaires. La méthode procède en trois étapes :

- Extraction de caractéristiques à partir des deux images.

- Fusion : Afin de concaténer les vecteurs caractéristiques issus des deux images ; une étape de normalisation de caractéristiques est indispensable en utilisant la norme Z-score.
- Classification : en utilisant la méthode SVM (Support Vector Machine)

L'extraction de caractéristiques de l'image mammographique est effectuée en évaluant la texture du sein à partir de l'image compte tenu que la distribution de la glande mammaire est linéaire, et que sur une image elle est orientée vers le mamelon (cas normal) ou vers la zone suspecte (cas anormal). Des filtres de Gabor de différentes orientations sont utilisés, le filtre gaussien est utilisé pour filtrer les basses fréquences, et finalement le gradient est utilisé pour extraire l'orientation de chaque pixel. L'entropie, la densité, l'homogénéité et la déviation standard sont aussi considérées. Pour l'image échographique, un prétraitement est effectué permettant de supprimer le bruit additif et multiplicatif en utilisant respectivement le lissage gaussien et le filtre de diffusion anisotropique. Ensuite l'image est segmentée par la technique du contour actif. Les caractéristiques extraites sont : la spiculation du contour en mesurant l'angle de courbure à chaque point ; la forme elliptique ; l'ombre acoustique en mesurant l'intensité moyenne ; l'entropie ; l'homogénéité et la déviation standard.

Opérant dans le niveau décision, Les travaux de [Zhang et al., 2015b] présentent une approche de fusion d'images basée sur la recherche par contenu (CBIR) suivant trois principales étapes :

- Une étape de construction de graphe à partir de chaque image en se basant sur les résultats de la recherche par contenu. Cette image représente le centre du graphe. La technique du kNN est utilisée pour construire le graphe où une première couche est construite avec les kNN de l'image et la deuxième couche est construite par les kNN des kNN précédents. Les arcs reliant les noeuds du graphe sont qualifiés par le coefficient de similarité de Jaccard.
- Une étape de fusion de graphes puisque plusieurs graphes sont construits des résultats récupérés. La fusion est effectuée par l'ajout de nouveaux noeuds au graphe final ou par la consolidation des coefficients des arcs.
- Une étape de rangement des résultats et de sélection de cas similaires. Les images sont rangées suivant leurs distributions de probabilité. En effet, les auteurs définissent un noeud important comme un noeud ayant une grande probabilité de visite. Ils utilisent le modèle de navigation intelligent proposé dans les travaux de [Richardson and Domingos, 2002] pour la navigation entre les noeuds en se servant de la matrice de transition.

Quelques travaux étaient aussi proposés pour la fusion d'images mammaires au niveau score. Les travaux de [Lesniak et al., 2012] abordent la problématique de la classification des masses mammaires. Une estimation du degré de malignité des objets pathologiques est effectuée en fusionnant les informations issues des deux images : image mammographique et IRM. La procédure commence par l'extraction de caractéristiques des deux images :

- Pour l'image mammographique, 53 caractéristiques sont extraites de chaque

vue, dont la similarité, la spiculation, le gradient, la linéarité de texture, la morphologie, la localisation, la densité. Les caractéristiques des deux vues sont regroupées en un seul vecteur caractéristique contenant 106 caractéristiques.

- Pour l'image IRM, 46 descripteurs sont utilisés. Les caractéristiques extraites sont des caractéristiques cinétiques, pharmacocinétiques et morphologiques. Les images de haute résolution sont utilisées pour les caractéristiques cinétiques et morphologiques dont la "baseline", la taille de lésion, la compacité, l'élongation,... Les images de basse résolution sont utilisées pour les caractéristiques pharmacocinétiques dont le volume extracellulaire V_e , le coefficient du transfert de volume K_{Trans} , et la constante du modèle cinétique K_{ep} .

La méthode SVM linéaire est utilisée pour la sélection des caractéristiques pertinentes de chaque modalité. Une méthode SVM non linéaire est utilisée ensuite pour la classification des lésions. Après, une étape de standardisation des estimations en utilisant la valeur moyenne et la déviation standard est utilisée afin de pouvoir effectuer la fusion. La fusion est effectuée en prenant la valeur moyenne des estimations de malignité unimodales.

Dans les travaux de [Baum et al., 2011], une approche de fusion des deux images PET (Positron Emission Tomography) et IRM mammaire est proposée. Cette approche est basée sur la fusion d'images en utilisant les tables de couleurs générées par un algorithme génétique. En effet, une table de deux dimensions est définie prenant comme indices des colonnes les intensités d'une première image source, et comme indices des lignes les intensités d'une deuxième image source. Chaque pixel de l'image résultante contient la valeur correspondante de la table des couleurs.

Chaque table de couleur contient 12 nombres représentant les coefficients R Or B Ob G Og pour chaque image. L'algorithme commence par générer une population de tables de couleur aléatoirement, ensuite chaque membre de la population est évalué et classé en calculant un score d'ajustement d'une table à certains critères. Une nouvelle population est générée et le processus est itéré. L'algorithme est arrêté quand le score ne change pas à travers les dernières générations. Deux algorithmes utilisant deux tables différentes sont testés.

3.2 Aide au diagnostic par raisonnement à base de cas

Le raisonnement à base de cas (CBR), un concept de l'intelligence artificielle, a commencé en 1977 dans l'université Yale avec les travaux de Roger Schank et ses collègues [Schank, 1983]. Ils cherchaient à concevoir un système doté d'une mémoire dynamique dans le but de résoudre des problèmes en considérant les solutions adoptées dans des situations [Castro et al., 2011]. Dès lors, la notion de CBR a pris de l'ampleur et a évolué de manière significative en étendant ses capacités de raisonnement à des situations de plus en plus complexes dans divers domaines.

3.2.1 Principe du raisonnement à bases des cas

Le raisonnement à base de cas (CBR - Case Based Reasoning) est une méthodologie utilisée dans plusieurs domaines dans la prise de décision. Il est basé sur l'exploitation des solutions précédentes adoptées pour des cas similaires. L'idée fondamentale d'un CBR est "*Les problèmes similaires ont des solutions similaires*" [El-Sappagh et al., 2015]. Dans une base de cas, un cas consiste en une représentation conceptuelle des connaissances d'une certaine situation. Formellement, un cas est représenté par un couple de termes : $C_i = (P_i, S_i)$ où P_i est la description d'un problème, et S_i est la spécification de la solution. Plusieurs définitions ont été attribuées au raisonnement à base de cas. Généralement, il est défini comme une "*approche récente de résolution de problèmes*" [Aamodt and Plaza, 1994].

Dans [D.B., 1996], Leake définit le CBR comme "*raisonnement par souvenance*". Riesbeck et Schank [Riesbeck and Schank, 2013] précise que le CBR consiste à résoudre de nouveaux problèmes en adaptant des solutions utilisées pour des anciens problèmes. Le CBR est défini comme un modèle cyclique, nommé "*le modèle 4 Rs*" ("*The 4 Rs model*") [Aamodt and Plaza, 1994] (Fig. 3.2).

- La phase "Retrieve" est la phase initiale du processus du raisonnement. Elle consiste à explorer une base constituée de cas antérieurs afin d'extraire les cas les plus similaires au cas étudié.
- La phase "Reuse" est la deuxième étape du processus du raisonnement. Elle consiste à réutiliser une solution adoptée pour un cas antérieur pour résoudre le nouveau problème. Elle permet de proposer une solution à partir de solutions disponibles des cas similaires sélectionnés.
- La phase "Revise" est une phase consistant à réviser la solution proposée. Cette révision est effectuée par l'expert du domaine qui essaie d'adapter la solution au problème posé.
- La phase "Retain" est la dernière phase du processus du raisonnement. Elle consiste à retenir le nouveau cas examiné dans la base des cas afin de s'en servir dans la résolution de futurs cas.

A la fin de ce cycle, la base de cas est mise à jour et enrichie en y ajoutant la nouvelle solution retenue.

Afin d'extraire, potentiellement, les cas similaires, une mesure de similarité est calculée. Cette mesure dépend principalement du type de données utilisées (numériques, textuelles, etc...). Généralement, la mesure de similarité est déterminée par la distance de la description de chaque caractéristique définie dans les représentations relatives au nouveau cas étudié et au cas de la base.

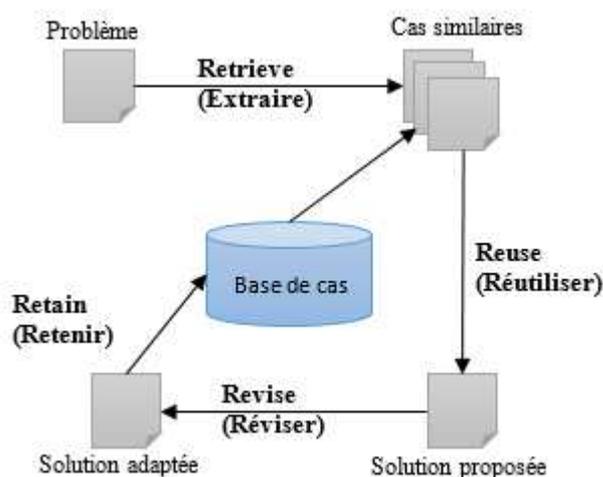


FIGURE 3.2 – Le modèle cyclique 4Rs du CBR

3.2.2 Travaux sur le raisonnement à base de cas en diagnostic médical

Vue son apport dans l'aide au diagnostic médical, le raisonnement à base de cas a été utilisé dans plusieurs travaux de recherche. La diversité des modes de représentation de l'information et des connaissances a conduit les chercheurs à adapter le concept du raisonnement à base de cas au type de l'information traitée. Ainsi, certains travaux ont combiné le concept d'ontologie au CBR. D'autres travaux ont traité l'imperfection de l'information dans les systèmes CBR.

Dans [Darzi et al., 2011], Darzi and al proposent une solution utilisant le concept de CBR pour la sélection de caractéristiques décisives dans la classification pathologique pour l'aide au diagnostic du cancer du sein. La méthode exploite tout d'abord un algorithme génétique assurant la recherche d'ensembles de caractéristiques qui donne la plus haute précision dans la classification. Ensuite, le CBR est utilisé pour évaluer chaque ensemble sélectionné.

Phase de recherche : chaque ensemble contient des caractéristiques représentant la spécification pathologique d'un patient. Chaque ensemble est représenté par un code binaire, où 1 est attribué à une caractéristique sélectionnée et 0 sinon.

Phase d'évaluation : Cette phase consiste à détecter le meilleur ensemble de caractéristiques aidant le médecin dans son diagnostic. Pour cela, chaque ensemble est évalué à travers un CBR en extrayant le cas le plus similaire en calculons la distance euclidienne. La solution du cas extrait est gardée comme solution pour le cas étudié.

Les auteurs rapportent que leur approche présente une précision de 94.74%

avec toutes les caractéristiques d'un cas et de 97.37% en utilisant une sélection de caractéristiques pertinentes. Soit alors, une amélioration de la classification au niveau de la précision.

Garrido et al. [Garrido et al., 2008] proposent une approche tirant profit des deux concepts ontologie et CBR dans un système d'aide à la décision. Cette approche est constituée principalement de trois étapes : Une étape de spécification de cas (alimentation d'une base) en exploitant une ontologie, une étape de recherche et d'extraction de cas similaires à travers le CBR, et finalement une étape d'adaptation de la solution au cas étudié.

La première étape consiste à construire une base de cas en utilisant une ontologie de domaine. Ainsi, les différents cas représentent des instances de cette ontologie. Un cas est décrit par un ensemble d'attributs auxquels sont affectés des poids pour guider la mesure de similarité dans la procédure du raisonnement.

La deuxième étape consiste à adopter un raisonnement à base de cas pour chercher une solution à un problème posé dans un nouveau cas. Ainsi, une mesure de similarité entre deux cas réfère à une mesure de distance entre deux individus de l'ontologie. La dernière étape consiste à adapter la solution du cas extrait au cas courant. Deux types d'adaptation sont utilisés :

- Adaptation des règles du raisonnement où certaines règles du raisonnement nécessitent d'être adaptées selon les valeurs des attributs du nouveau cas.
- Adaptation taxonomique de l'ontologie en ajoutant des informations supplémentaires à l'ontologie pouvant être utilisées dans de futures études.

Cette approche a été appliquée afin de fournir aux praticiens médicaux des recommandations pour le contrôle de situations de saignement.

Dans le même contexte, Samwald et al. [Samwald et al., 2015] proposent principalement la même démarche afin de mettre en oeuvre une solution permettant de formaliser les connaissances pharmaco-génomiques et présentant une assistance dans la prise de décision. Les auteurs commencent par la création d'une ontologie modélisant les connaissances pharmaco-génomiques. Les instances de l'ontologie sont utilisées comme une base de cas dans le raisonnement afin de fournir une aide à la prise de décision concernant les nouveaux cas qui se présentent.

Abdrabou et al. [Abdrabou and Salem, 2010] ont proposé aussi une approche bénéficiant de la puissance des ontologies dans la formalisation des connaissances et des avantages du raisonnement à base de cas dans un système de classification de pathologies mammaires. L'approche exploite deux frameworks : jCOLIBRI et myCBR pour atteindre cet objectif. L'approche passe par les mêmes étapes décrites précédemment :

- Utilisation d'une ontologie de domaine
- Instanciation de l'ontologie constituant une base de cas.
- Utilisation du principe du raisonnement à base de cas pour déterminer la classe (bénigne ou maligne) d'une lésion mammaire.

Abdul et al. [Abdul et al., 2014] proposent une approche combinant le CBIR et le principe de la logique floue dans le but de gérer les tâches dans un système de gestion de bases de données. La solution proposée est composée de trois modules :

- La caractérisation d'une tâche en utilisant un CBR ; une tâche est définie par un certain nombre de paramètres et un type. Une mesure de similarité est effectuée entre une nouvelle tâche et les tâches stockées dans une base de cas ; ensuite, le type de la tâche la plus similaire est attribué à la nouvelle tâche, sinon un nouveau type est lui est affecté et la base est mise à jour en y insérant cette nouvelle tâche.
- La planification en appliquant la théorie de la logique floue visant à attribuer une mesure d'appartenance aux ensembles flous : On Line Transaction Processing (OLTP) et Decision Support System (DSS).
- La détection des temps oisifs en se basant sur les deux modules précédents et en exploitant le CBR.

Ekong et al. [Ekong et al., 2012] proposent une approche pour le diagnostic des cas de dépression en combinant un CBR avec le concept de réseaux de neurones et de la théorie de la logique floue. La solution proposée commence par une étape de fuzzification des symptômes de la dépression chez un patient. Ces symptômes servent comme entrée à un réseau de neurone constitué de trois couches : la couche symptômes contenant 25 noeuds représentant les symptômes flous, la couche cachée contenant 5 noeuds correspondants aux catégories des attributs de la dépression et la couche de sortie contenant 4 noeuds représentant le degré de sévérité de la dépression. Le principe du raisonnement à base de cas est utilisé afin de chercher un diagnostic adéquat à un nouveau cas présent en s'appuyant sur une mesure de similarité floue et un système d'inférence adaptatif (adaptive neuro-fuzzy inference system).

Hullermeier et al. [Hullermeier et al., 2002] proposent une méthode adaptant un modèle linguistique à des données observées automatiquement en tirant profit du concept de raisonnement à base d'instance et de la théorie des possibilités. En effet, considérant un cas décrit par le couple (situation, résultat), les auteurs cherchent à prédire le résultat d'un nouveau cas à travers sa similarité avec des cas existants aux mêmes situations. Afin d'aboutir à cette fin, un CBR est nécessaire pour l'extraction de cas similaires. Les auteurs proposent d'exprimer la prédiction du résultat en terme de distribution de possibilités qui exprime le degré de possibilité d'un résultat obtenu. Ils utilisent pour cette intention deux types de règles d'inférence : des règles possibilistes et des règles de certitude.

Alsun et al. [Alsun et al., 2010] proposent aussi une approche tirant profit de la théorie des possibilités dans le raisonnement à base de cas pour l'aide au diagnostic médical. En effet, les connaissances sont tout d'abord modélisées dans un espace possibiliste. Ensuite, elle sont utilisées pour former une base de connaissances

possibiliste. Lorsqu'un nouveau cas se présente, un raisonnement à base de cas est appliqué en procédant à une mesure de similarité possibiliste entre la description du nouveau cas et la description de chaque cas de la base. Un certain nombre de diagnostics potentiels exprimés en terme de possibilité et de nécessité est proposés à l'expert du domaine afin d'estimer le diagnostic le plus adéquat au cas étudié.

El-Sabbagh et al. [El-Sappagh et al., 2015] proposent une approche combinant trois concepts : l'ontologie, la théorie des ensembles flous et le CBR dans le profit du diagnostic du diabète. Ils procèdent par construire une ontologie avec des attributs flous. La méthode proposée est constituée principalement par six modules :

- Module de prétraitement servant à préparer les données aux différents traitements. Dans ce module, les données sont formatées suivant la théorie des ensembles flous.
- Module de construction d'une ontologie de domaine spécifiant les connaissances du diabète.
- Module de transformation de l'ontologie rigide en une ontologie floue. Cette ontologie sert pour base de cas.
- Module d'instanciation de l'ontologie avec des cas flous.
- Module de recherche de cas similaires à un nouveau cas présent.
- Module d'extraction du cas similaire et d'adaptation du diagnostic adéquat.

3.3 Conclusion

Plusieurs imperfections peuvent affecter une information. La fusion, en tant que concept, présente un cadre très enrichissant pouvant palier au problème de l'incomplétude des informations. La modélisation des informations avec leurs imperfections dans un cadre mathématique adéquat, doit apporter un gain considérable au processus de fusion d'information et par la suite l'analyse et l'interprétation de scènes.

Approche proposée pour la fusion multimodale

Sommaire

4.1	Architecture de l'approche proposée	60
4.2	Génération des ontologies locales	63
4.2.1	Couche Données : Pré-traitement et Segmentation	64
4.2.2	Couche Informations : Extraction de caractéristiques	65
4.2.3	Construction des ontologies locales	71
4.3	Fusion des ontologies locales	74
4.3.1	Mesure de similarité	74
4.3.2	Fusion ontologique	77
4.4	Instanciation	79
4.5	Raisonnement et Aide au diagnostic	79
4.5.1	Le cas requête	81
4.5.2	La base de cas	82
4.5.3	Mesure de similarité	83
4.5.4	Les cas possibles	83
4.5.5	Instanciation de nouveaux cas	83
4.6	Modèle symbolique de représentation des connaissances	84
4.6.1	Choix du modèle de représentation de connaissances	84
4.6.2	Représentation des connaissances et intégration des données	85
4.7	Conclusion	86

LE terme diagnostic est d'origine grecque signifiant "*Connaissances à travers des signes*". Jean-Charles Sournia [Sournia, 1995] définit le diagnostic comme une "*démarche intellectuelle*" effectuée par une personne examinatrice, appelée médecin, et permettant d'identifier une maladie chez une personne examinée, appelée patient. Cette démarche sollicite d'une part, une "*collecte de données et d'informations*"

sur le patient, à travers des "signes" (symptômes) et des représentations par des "investigations complémentaires"; et d'autre part un "travail de réflexion". Le résultat de cette démarche est une spécification devant être exprimée d'une manière brève avec une terminologie appropriée dérivant du savoir du médecin. L'efficacité d'un diagnostic médical est traduite par l'aptitude du médecin à tirer la bonne conclusion en étudiant les données et les informations relatives au patient (qui sont généralement hétérogènes) tout en considérant les différentes imperfections qui peuvent les affecter.

Dans ce cadre, "Les systèmes d'aide à la décision médicale ont pour but de fournir aux cliniciens les connaissances appropriées à une situation clinique d'un patient, correctement filtrées et présentées afin d'améliorer la qualité des soins et la santé des patients" [Salis et al., 2010]. Ainsi, d'après la définition précédente, un système proposant l'aide au diagnostic doit établir les deux composantes du diagnostic : la collecte et la modélisation des données et informations, et le raisonnement approprié tenant compte l'hétérogénéité et les imperfections des informations. Dans ce contexte, nous proposons une approche permettant d'exploiter deux images multimodales, afin de fournir une aide au diagnostic médical adéquate assistant le médecin dans sa prise de décision. Ce chapitre, détaille cette approche en spécifiant ses cinq principaux modules : (1) l'extraction de connaissances à partir des images et leur modélisation par des ontologies locales en se basant sur la structure d'une ontologie de domaine ; (2) la fusion des ontologies locales ; (3) l'instanciation de l'ontologie de domaine en vue de son enrichissement ; (4) l'interprétation par un raisonnement à base de cas ; et finalement, (5) la modélisation virtuelle 3D des connaissances.

4.1 Architecture de l'approche proposée

L'objectif principal de nos travaux, est de présenter l'aide au diagnostic nécessaire pour le médecin afin de prendre la bonne décision. Par conséquent, notre approche se situe dans le cadre d'interprétation de scènes observées à travers les capteurs. Elle est basée sur trois principes : La modélisation de connaissances sous deux formes : la représentation conceptuelle des connaissances en utilisant le concept ontologie et en se servant d'une ontologie de domaine existante et la représentation virtuelle 3D des connaissances ; la fusion de connaissances ; et le raisonnement à base des cas. Notre approche est ainsi constituée de cinq étapes différentes (Fig. 4.1) :

- Une génération d'ontologies locales à partir des images mammaires.
- La fusion des différentes ontologies locales.
- L'instanciation de l'ontologie globale (ontologie de domaine).
- Le raisonnement et l'aide au diagnostic.
- La modélisation virtuelle 3D des connaissances.

Le passage de l'espace image (produit du capteur) à l'interprétation de scène (produit d'un raisonnement) nécessite un certain nombre d'étapes suivants différents niveaux de traitement. Partant d'une image, un traitement multi-niveau (Fig. 4.2)

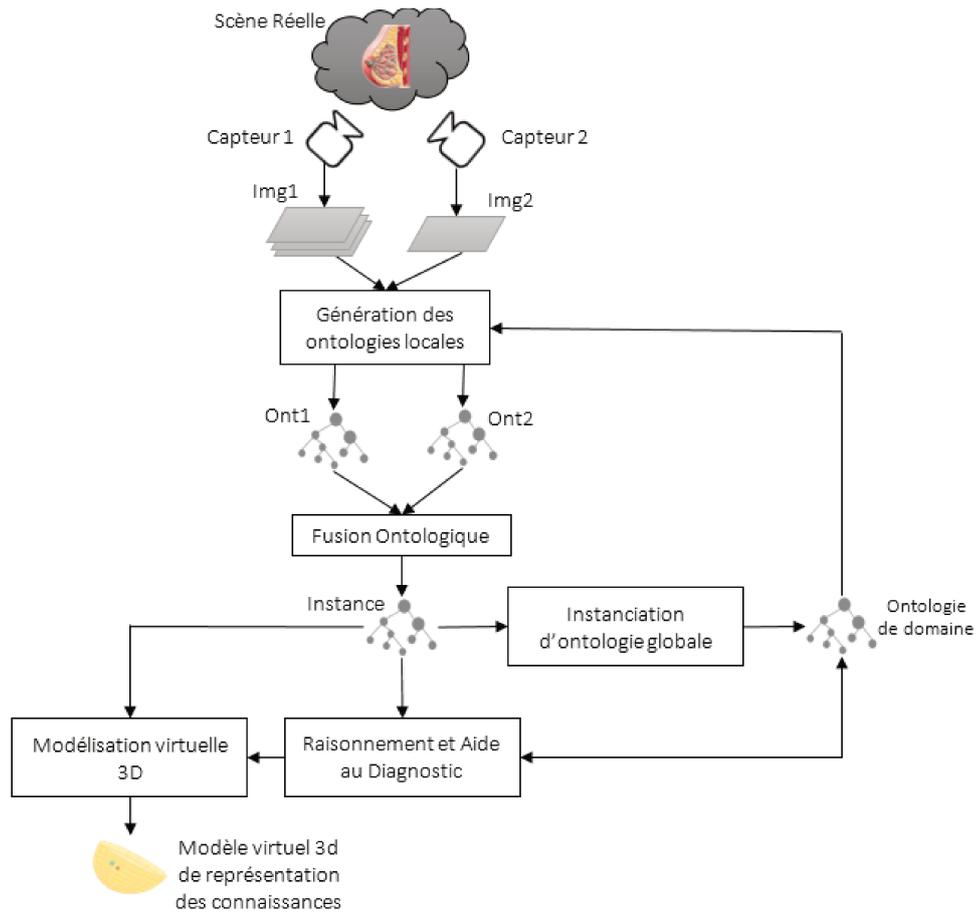


FIGURE 4.1 – Approche proposée

est effectué afin d'aboutir au résultat final.

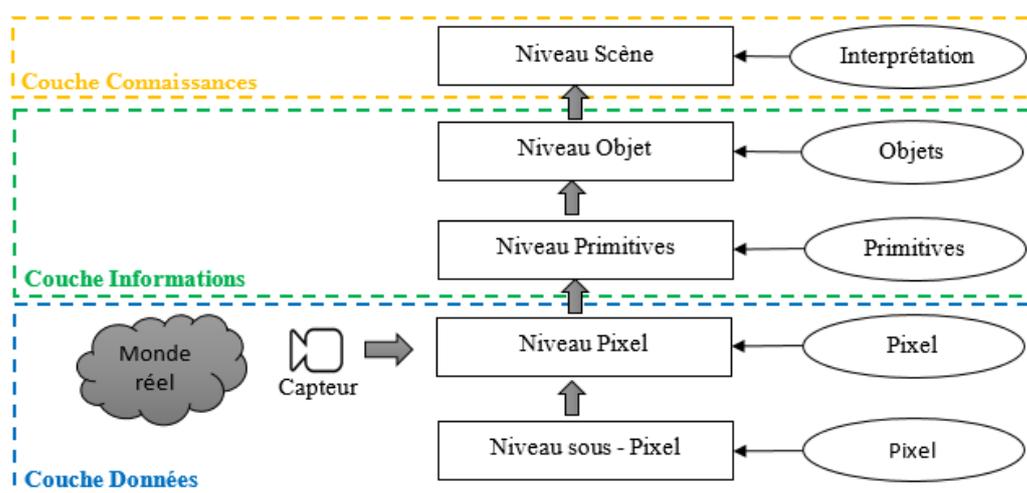


FIGURE 4.2 – Les différents niveaux d'interprétation de scènes

Nous définissons ainsi cinq niveaux de traitement :

- **Niveau "sous-pixel"** : (Granularité de l'information : information sur les pixels.) C'est un niveau qui concerne l'identification d'informations liées aux pixels. Ces informations sont liées principalement à la résolution du capteur utilisé pour l'acquisition de l'image.
- **Niveau pixel** : (Granularité de l'information : les pixels) C'est le niveau de base par lequel commence l'interprétation d'une scène imagée. Dans ce niveau, chaque pixel est analysé individuellement pour en extraire certaines informations.
- **Niveau caractéristiques** : (Granularité de l'information : caractéristiques) Les caractéristiques peuvent être aussi appelées des primitives visuelles. Ces caractéristiques représentent un ensemble d'informations déduites à partir de l'analyse des pixels.
- **Niveau objet** : (Granularité de l'information : objet) A partir des primitives visuelles extraites, des entités d'un niveau sémantique plus haut sont construites. Ces entités sont identifiées comme des objets en tenant compte de leur contexte dans la scène.
- **Niveau scène** : (Granularité de l'information : interprétation) C'est le niveau sémantique le plus haut. Dans ce niveau, les objets sont regroupés suivant les relations qui existent entre eux. A partir de toutes les informations collectées à travers les niveaux précédents, ce niveau présente une interprétation de la scène observée.

Nous organisons ces différents niveaux dans une architecture multi-couche (Fig. 4.3) tenant compte du type d'information utilisée et s'alignant avec les trois niveaux de fusion définis en littérature :

- La **"couche Données"** englobe les deux premiers niveaux, et représente le bas niveau du traitement.
- La **"couche Information"** intègre le 3^{ème} niveau, et représente le niveau intermédiaire du traitement.
- La **"couche Connaissances"** réunit les deux derniers niveaux, et représente le plus haut niveau de traitement. Elle opère sur les connaissances extraites à partir des informations.

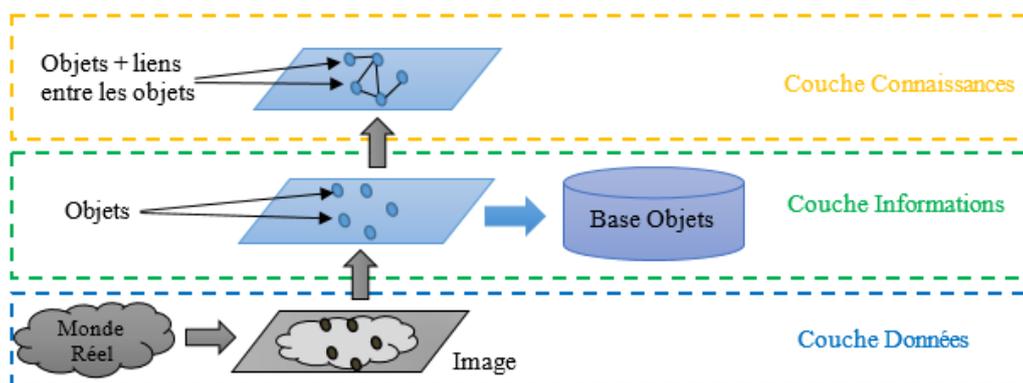


FIGURE 4.3 – Vue d'ensemble des niveaux de traitement

Dans ce qui suit, nous détaillons les différentes étapes de notre approche suivant le niveau et la couche du traitement correspondant.

4.2 Génération des ontologies locales

Chaque image est modélisée par une ontologie locale qui décrit le contenu pertinent de l'image. Durant le traitement, un passage de l'espace de l'image (bas niveau) vers l'espace de description des objets contenus dans l'image (haut niveau) est effectué. Chaque objet est modélisé par une ontologie Description Objet. Les différentes relations entre les objets extraits sont modélisées par une ontologie Relations. Ces ontologies constituent l'ontologie locale relative à une image.

Le traitement des images passe par cinq principales étapes suivant le niveau du traitement :

- Une étape de pré-traitement, dans le but de sélectionner la région d'intérêt dans l'image tout en améliorant sa qualité (contraste, suppression du bruit,...)
- Une étape de segmentation permettant de sélectionner la lésion dans l'image.
- Une étape d'extraction de caractéristiques afin de caractériser la lésion.
- La formation des objets pathologiques qui consiste en la définition de l'objet dans son contexte. Cette étape est assurée par une formation d'une ontologie Objet.
- La mise des objets dans leurs contexte en identifiant les relations entre eux.

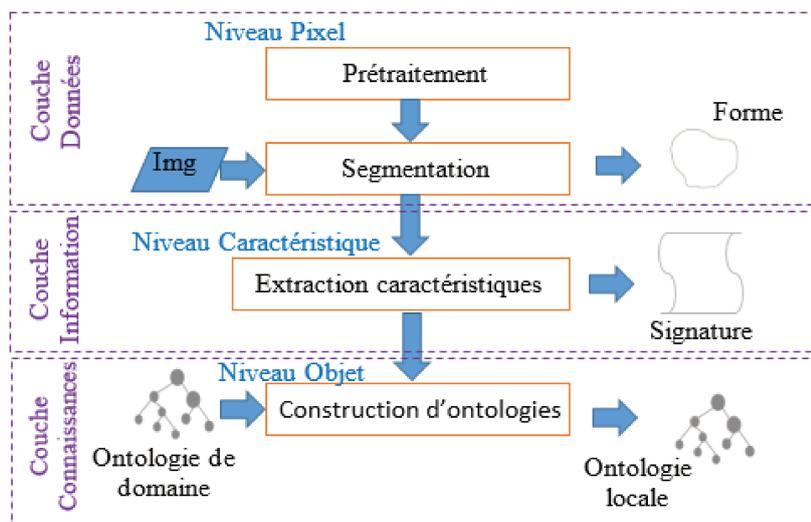


FIGURE 4.4 – Principe de l'approche proposée

4.2.1 Couche Données : Pré-traitement et Segmentation

Les images médicales, sont généralement caractérisées par un faible contraste, ce qui fait augmenter l'ambiguïté entre le tissu normal et le tissu pathologique. La netteté et la clarté des données dans une image représente un point clé pour un traitement efficace et une analyse exacte des informations.

a. Prétraitement

Le prétraitement est une étape qui vise à préparer l'image pour le traitement. Dans notre cas, elle est constituée de trois sous étapes :

- La sélection de la zone d'intérêt (sein) : Dans l'image mammographique le sein est séparée de son arrière-plan. Dans l'image IRM, cette étape consiste à séparer le sein du reste du corps imagé.
- Le rehaussement de l'image : Il consiste à améliorer son contraste sans affecter l'information pertinente à lire. Pour ce faire nous appliquons la technique de l'étirement de l'histogramme. Un exemple de cette étape est donné par la figure 4.5.
- Le filtrage : Il a pour but de supprimer le bruit présent dans l'image. Cependant, il peut causer une perte d'information considérable ; ce qui dépend du type du filtre utilisé et de ses paramètres. Pour filtrer les images mammographiques, nous avons mener une étude comparative sur les résultats obtenus en appliquant les filtres les plus utilisés dans la littérature. Nous avons calculé l'EQM (Erreur Quadratique Moyenne) et en tenant compte de l'évaluation d'un expert du domaine. Le filtre RID (Regularised Isotropic

Diffusion) est adopté. La figure 4.5-(A-c) (respectivement (B-c)) présente le résultat obtenu de l'application de ce filtre sur les images mammographiques (respectivement pour les images IRM). (L'annexe A détaille les résultats du filtrage des images.)

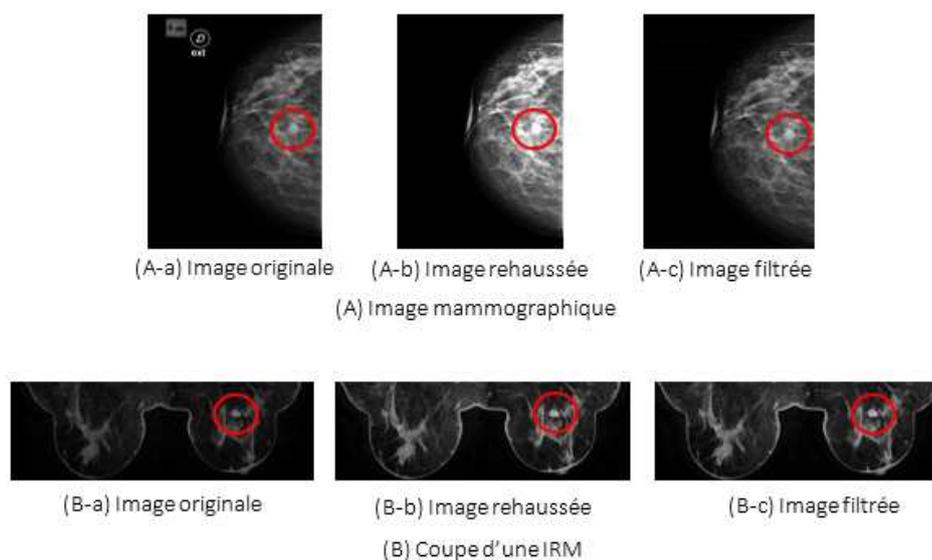


FIGURE 4.5 – Exemple de pré-traitement des images mammographiques et IRM

b. Segmentation

L'étape de segmentation vise à sélectionner les formes pathologiques à partir des images. Une zone pathologique est caractérisée généralement par une forte luminosité des pixels qui la constituent. Pour segmenter les formes pathologiques dans les deux images nous utilisons la méthode "Level-Set". Cette méthode est introduite par Osher et Sethian [Osher and Sethian, 1988]. L'idée de base de la méthode Level Set consiste à suivre une interface à travers une fonction. Cette fonction est une variété d'iso contours constitués d'un ensemble de points à une certaine distance de l'interface. Lorsque cette distance est nulle, le contour est dit de niveau 0 et il définit l'interface concernée.

Pour l'image IRM la segmentation est effectuée coupe par coupe. La figure 4.6 montre les résultats de la segmentation des images mammographiques.

4.2.2 Couche Informations : Extraction de caractéristiques

La couche Informations représente une couche de traitement intermédiaire entre la représentation bas niveau des données et la représentation haut niveau des connaissances qu'on peut déduire de l'image.

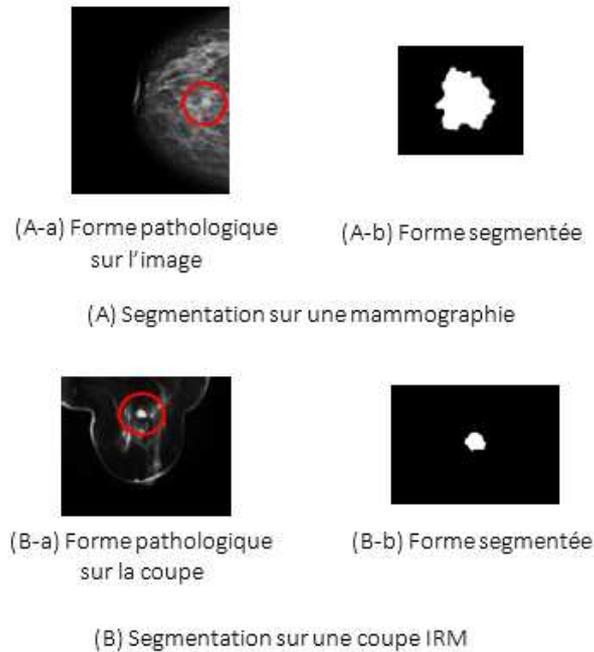


FIGURE 4.6 – Exemple de segmentation des images mammographiques

L'extraction de caractéristiques consiste en deux grandes étapes :

- L'extraction de primitives visuelles qui sont des propriétés calculées à partir des données sur l'image liées à la taille, la forme, et le contour de la forme segmentée. Cette étape résulte en une signature numérique contenant la description des différentes formes pathologiques détectées.
- L'extraction des caractéristiques sémantiques, que nous pouvons déduire à partir des primitives visuelles. Cette étape nous informant sur la sémantique liée à la taille, la forme et le contour des formes pathologiques extraites. Elle résulte en une signature sémantique contenant la description des différentes formes pathologiques étudiées.

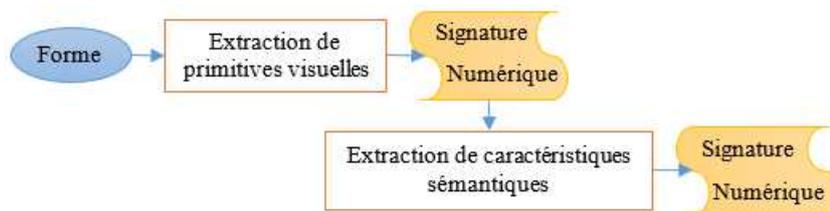


FIGURE 4.7 – Principe de l'extraction de caractéristiques des différentes formes pathologiques

Pour l'extraction de primitives visuelles, nous utilisons quatre formes essentielles pour effectuer le calcul des propriétés liées à la forme pathologique détectée. Ces quatre formes sont :

- L'enveloppe convexe (EC) : C'est la plus petite forme géométrique contenant la forme pathologique (Fig. 4.8-(d)).
- L'Ellipse minimale circonscrite (EMC) : C'est la plus petite ellipse englobant la forme pathologique (Fig. 4.8-(e)).
- Le Cercle minimal circonscrit (CMC) : C'est le plus petit cercle contenant la forme pathologique (Fig. 4.8-(e)).
- Le Cercle maximal inscrit (CMxI) : C'est le plus grand cercle inclus dans la forme pathologique (Fig. 4.8-(f)).

S'alignant avec les travaux sur les images mammaires, nous avons calculé les propriétés les plus étudiées en littérature :

- L'aire de la forme (ArF) représente la surface de la forme pathologique en terme de nombre de pixels constituant cette forme. L'aire informe sur la taille de la forme pathologique qui est un critère décisif de sa malignité.
- L'aire de l'ellipse minimale circonscrite (ArEMC).
- L'aire du cercle minimal circonscrit (ArCMC).
- L'aire du cercle maximal inscrit (ArCMxI).
- Le périmètre de la forme (PrF) représente le nombre de pixels constituant le contour de la forme pathologique (Fig. 4.8-(c)).
- Le coefficient de compacité (Cmp) est une mesure informant sur la régularité ou l'irrégularité de la forme pathologique. C'est le rapport du périmètre de la forme sur celui d'un cercle ayant la même surface. La compacité est calculée par l'expression suivante :

$$Cmp = \frac{(PrF)}{2\sqrt{\pi ArF}}$$

- L'excentricité de la forme (ExF) mesure le degré de correspondance de la forme pathologique et de l'ellipse minimale circonscrite. Elle est calculé par l'expression :

$$ExF = \frac{ArF}{ArEMC}$$

- La convexité de la forme.
- La circularité (Cr) renseigne sur la correspondance de la forme à un cercle. Elle est mesurée par le rapport de la surface de la forme sur la surface du cercle de même périmètre. $Cr = 1$ réfère à une forme circulaire, $0 < Cr < 1$ indique une forme elliptique. Elle est calculé par l'expression :

$$Cr = 4\pi \frac{ArF}{(PrF)^2}$$

- Le diamètre maximal de la forme pathologique (DmxF)
- Le diamètre minimal de la forme pathologique (DmnF)
- Diamètre maximal de l'EMC (DmxEMC)
- Diamètre minimal de l'EMC (DmnEMC)
- L'orientation de l'axe maximal de l'EMC par rapport à l'horizontale.
- La longueur radiale normalisée (NRL). Elle représente le rapport entre la somme des distances euclidiennes (DE) entre le centre de la forme pathologique

(C) et les différents points du contour (Pt_i) et le périmètre de la forme

$$NRL = \frac{\sum_{i=1}^{Pr} DE(C, Pt_i)}{Pr F}$$

- Ecart type de la longueur radiale normalisée.
- Coordonnées du centre de la forme (Fig. 4.8-(b)).

La figure 4.8 expose quelques caractéristiques extraites d'une forme pathologique.

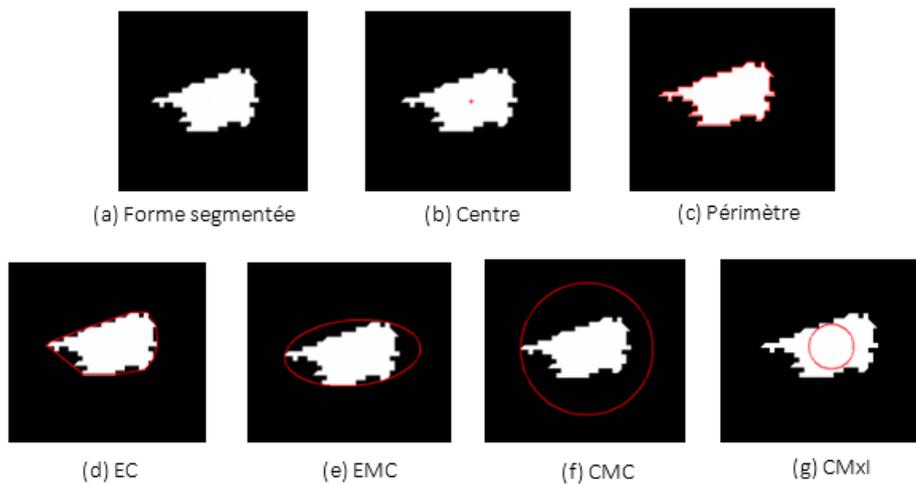


FIGURE 4.8 – Exemple de primitives extraites d'une forme pathologique

Les différentes caractéristiques sont regroupées dans un fichier signature définissant l'identité visuelle des formes pathologiques. A partir des primitives visuelles nous optons pour une identification sémantique des caractéristiques de chaque forme. Pour cela, deux principes sont appliqués : Principe de déduction des caractéristiques à travers des règles d'inférences, associé au principe de la théorie des possibilités pour se rapprocher au maximum du raisonnement humain effectué par l'expert et modéliser l'imperfection de la connaissance déduite. La figure 4.9 montre le processus d'extraction de caractéristiques suivant ces deux principes. Les experts affirment que la forme, les propriétés du contour et la taille de la lésion sont les critères prépondérants pour la caractérisation pathologique des lésions. Les caractéristiques visuelles calculées sont alors utilisées pour déduire les propriétés pathologiques des formes détectées. Des règles d'inférence basées sur la logique défaisable sont appliquées dans cette intention. La logique défaisable est une logique non-monotone proposée par Donald Nute pour formaliser le raisonnement défaisable [Raddaoui, 2013]. Elle permet de modéliser principalement la contradiction. Suivant le principe de la logique défaisable, nous définissons alors, les règles strictes et les règles défaisables. Ci-après quelques exemples de règles strictes.

- R1 : Si $cmp > s$ Alors la forme est irrégulière

- R2 : Si $Circ = 1$ Alors la forme est ronde.
- R3 : Si $ExF = 1$ Alors la forme est ovale

Les règles défaisables :

- R1 : Si $cmp < s$ Alors la forme est irrégulière. D1 : La forme peut être spiculée.
- R2 : Si $ExF < 1$ Alors la forme a l'allure elliptique D2 : La forme peut être irrégulière.
- R3 : Si $cmp < s$ Alors le contour est irrégulier.

Les caractéristiques sémantiques sont déterminées suivant la spécification clinique des descripteurs des lésions mammaires observées par les moyens d'investigation par les experts :

- **La forme** : c'est le premier descripteur vérifié sur une image mammaire. Les différentes formes évaluées, comme définies par les experts (Chapitre 1) sont :
 - La forme ronde : la masse est vue comme un cercle.
 - La forme ovale : La masse est vue comme une ellipse.
 - La forme lobulaire : La masse a une forme légèrement irrégulière.
 - La forme irrégulière : C'est une forme non-identifiable aux limites irrégulières.
 - La distorsion architecturale : Il s'agit d'une forme très irrégulière avec une très faible convexité.
- **Le contour** : De la même façon le contour représente un critère décisif dans le diagnostic des lésions mammaires. La spécification clinique du contour révèle les types suivants :
 - Contour régulier : la masse est vue comme un cercle.
 - contour indistinct : Le contour ne peut pas être identifié à partir de l'image. Il s'agit d'un contour flou.
 - Contour irrégulier : le contour présente des irrégularités moyennes.
 - Contour spéculé : C'est un contour très irrégulier, souvent associé avec des formes de type distorsion architecturale.
- **La taille** : La taille de la lésion est spécifiée de manière symbolique suivant la valeur de son diamètre et son aire selon des valeurs seuils. Les experts identifient trois tailles décisives en terme de bénignité et de malignité :
 - Petite : une masse de petite taille est signe de bénignité,
 - Moyenne : une masse de taille moyenne indique un risque de malignité,
 - Large : une masse de grande taille réfère à un risque plus grand de malignité.
- **La distribution** : La distribution concerne surtout les micro-calcifications. Nous identifions trois formes de distribution en se basant sur les observations cliniques des experts :
 - Singulière : Spécifiée généralement pour les masses. Les masses se voient comme un seul objet pathologique,
 - Dispersée : Spécifiée pour les micro-calcifications présentes de manière dispersée dans le tissu mammaire,

- Concentrée : spécifiée pour les foyers de calcification. Il s'agit de micro-calcifications condensées dans une zone du tissu mammaire.
- **La localisation** : La localisation d'une lésion est définie par les experts du domaine d'une manière symbolique ; en effet, le sein est divisé en quatre quadrants en vue de face et en 3 régions en vue de profil. Les experts localisent les lésions à travers cette décomposition.

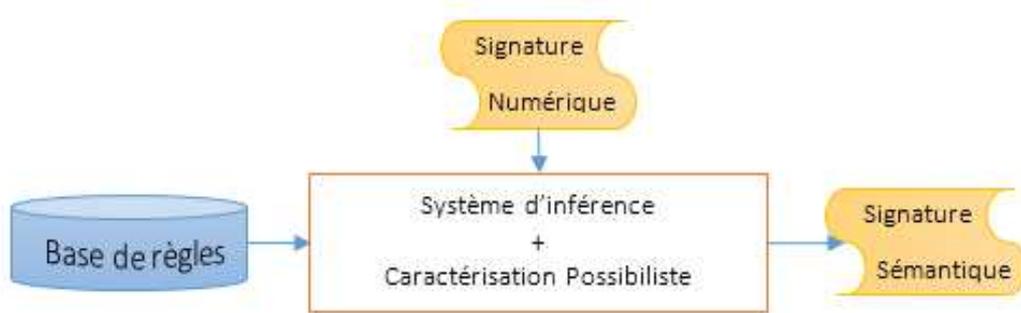


FIGURE 4.9 – Processus d'extraction de caractéristiques sémantiques

Pour modéliser l'imperfection de l'information, surtout en terme d'ambiguïté et d'incertitude nous nous servons de la théorie des possibilités pour la caractérisation pathologiques des lésions détectées. Plusieurs méthodes sont utilisées pour la définition des distributions de possibilité. Dans notre contexte, nous utilisons la définition donnée par l'expert. Les caractéristiques spécifiées par des mesures de possibilité sont : La forme, la taille, le contour, la distribution, et la localisation. Suivant les valeurs calculées pour les différentes caractéristiques sémantiques, des distributions de possibilités sont définies. La mesure de possibilité associée à un concept est calculée à travers les distributions relatives aux différents paramètres utilisés (Exemples : la circularité, l'excentricité, la convexité,...).

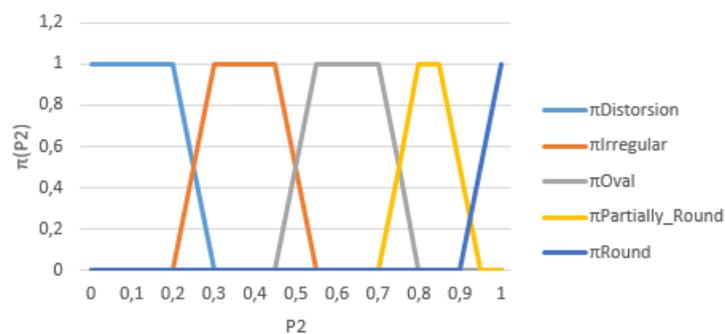


FIGURE 4.10 – Exemple d'une distribution de possibilité

La caractérisation sémantique des formes pathologiques est sauvegardée dans un

fichier représentant la signature sémantique de l'image. A partir de cette signature une ontologie locale est générée.

4.2.3 Construction des ontologies locales

Une ontologie locale est construite à partir de chaque image utilisée suivant la structure d'une ontologie de domaine. Quelques ontologies ont été proposées pour le domaine mammographique. En effet, une ontologie de domaine est conçue pour gérer toute information en relation avec ce domaine [Idoudi et al., 2014]. Par la suite, une ontologie mammographique gère les informations liées au domaine mammaire pouvant décrire les formes des lésions, les facteurs de risque, les diagnostics possibles, etc. Parmi les ontologies proposées, nous citons :

- L'ontologie MammOnto : C'est une ontologie développée par [Hu et al., 2003] dans le but de formaliser la terminologie utilisée dans la description des images mammaires, le diagnostic médical, et la spécification des examens. Elle est basée sur la spécification BI-RADS. Elle utilise un ancien langage de description et elle n'est pas accessible [Idoudi et al., 2014].
- BCGO (Breast Cancer Grading Ontology) est une ontologie proposée par le projet European Virtual Physiological Human [A. Branici, 2010] et étendue par [Idoudi et al., 2014]. Elle est développée dans l'intention d'apporter un gain de performance en traitement sur la classification des mammographies.
- "Mammographic Ontology" est une autre ontologie de domaine mammographique définie pour l'annotation d'images mammographiques et la recherche d'information par le contenu. Elle permet ainsi de passer de données de bas niveau extraites des images et de connaissances de haut niveau définies par l'ontologie.
- "Mammography Annotation Ontology" [Bulu et al., 2012] est une autre ontologie de domaine mammographique utilisée dans l'annotation d'images mammaires. Elle contribue surtout à l'annotation des lésions observées dans une image. Cette Ontologie n'est pas accessible.
- GIMY Mammographic Ontology est une ontologie du domaine mammographique développée par [Taylor and Toujilov, 2011] utilisant le langage de description le plus récent. Elle est utilisée dans une technique d'apprentissage afin d'établir une correspondance entre des annotations des stagiaires avec des annotations des experts. Elle se compose de deux ontologies : "Core Mammographic Ontology" et "Mammography Learning Ontology". La première contient la description liée à la mammographie : description anatomique et pathologique, le diagnostic, etc. La seconde est utilisée pour la correspondance entre deux ensembles d'annotation.

L'ontologie "Core Mammographic Ontology" (MAMMO.owl) est réutilisée dans les travaux de [Idoudi et al., 2016] en vue de son enrichissement sur trois niveaux : le niveau des concepts, le niveau des clusters et le niveau des relations. Nous utilisons cette ontologie pour guider la structuration des connaissances par ontologie locale.

La figure 4.11 présente un extrait de cette ontologie.

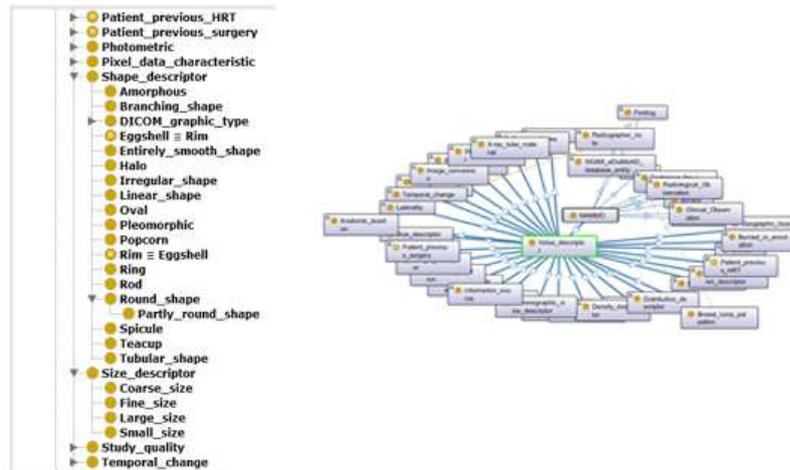


FIGURE 4.11 – Extrait de l'Ontologie MAMMO utilisée pour la génération des ontologies locales

Les différentes caractéristiques sémantiques sont mappées avec les concepts de l'ontologie de domaine. Certains concepts sont ajoutés afin de mieux détailler la description du contenu de l'image. Les figures 4.12 et 4.13 présentent les différents concepts de l'ontologie locale relative à une image. En effet, l'ontologie contient d'une part les descripteurs d'un objet (forme, contour, taille, distribution, et la localisation). Chaque descripteur est spécifié par un certain nombre de concepts. À chaque concept est associé une mesure de possibilité ω exprimant à quel degré ce concept décrit l'objet détecté. D'autre part, l'ontologie contient les paramètres d'une lésion : diamètre et aire (et qui représente le poids d'une lésion en terme clinique utilisé par les experts)

L'existence de plusieurs objets dans une seule représentation nous conduit à définir des relations entre ces objets. La figure 4.14 montre la spécification des relations entre les objets dans l'ontologie locale. Les relations sont catégorisées en deux groupes :

- **Les relations de direction** : Définissant la position d'un objet par rapport à un autre objet dans l'espace de la scène observée. Six relations sont définies :
 - "Under" : Un objet est sous un autre objet ;
 - "Behind" : Un objet est en arrière d'un autre objet ;
 - "Above" : Un objet est au dessus d'un autre objet ;
 - "In_front" : Un objet est devant un autre objet ;
 - "Right" : Un objet est à la droite d'un autre objet ;
 - "Left" : Un objet est à la gauche d'un autre objet.
- **Les relations topologiques** : Décrivent les relations de connexion entre les

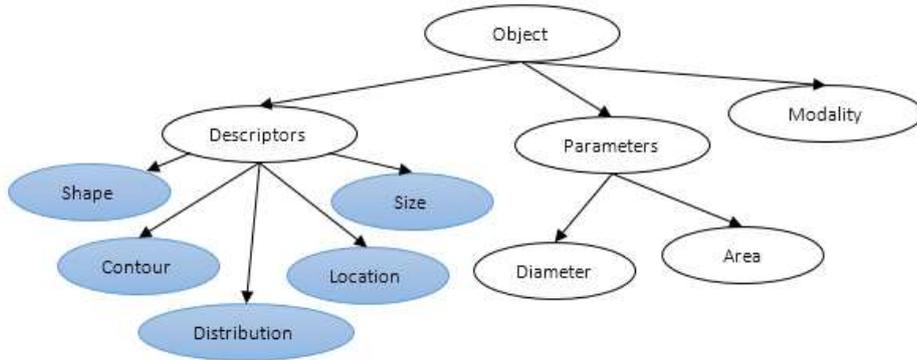


FIGURE 4.12 – La description d’un objet dans l’ontologie locale

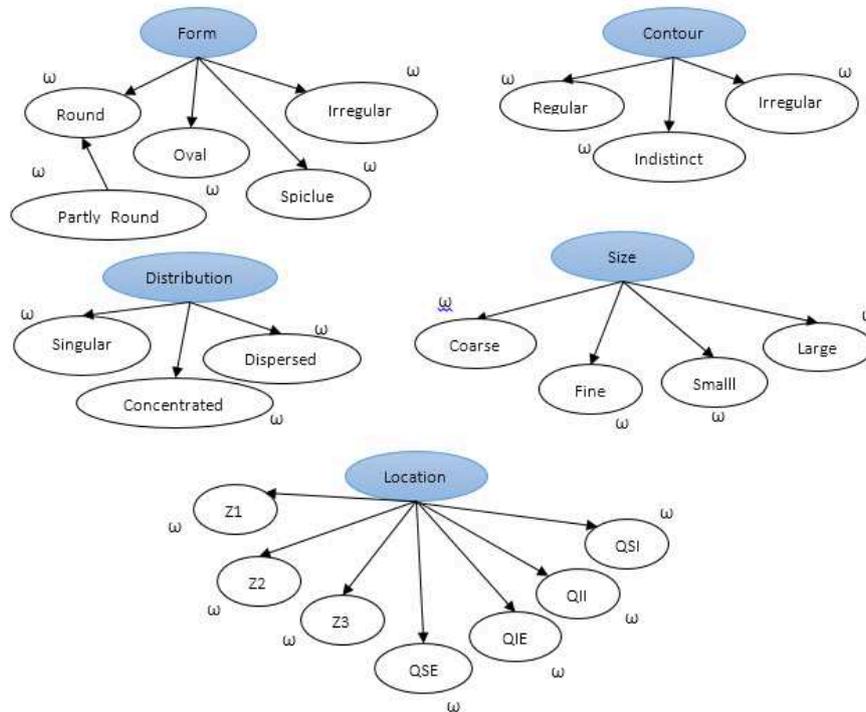


FIGURE 4.13 – La spécification des descripteurs d’une objet de l’ontologie

- différents objets. Trois relations sont définies relativement à ce groupe :
- "Connected" : Les deux objets sont connectés par leurs bords (tangents) ;
 - "Disconnected" : Aucun lien d'intersection entre les objets. Ils sont spatialement distants ;
 - "Overlapped" : Les deux objets se chevauchent.

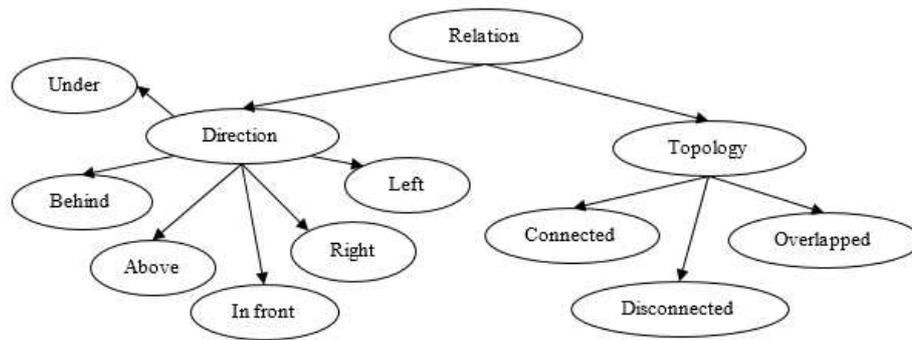


FIGURE 4.14 – La description des relations dans l'ontologie locale

4.3 Fusion des ontologies locales

Dans tout système de fusion, on cherche souvent à reconnaître les ressemblances et distinguer les différences afin de garder le maximum d'informations pertinentes sans avoir de redondances. Cette étape de comparaison est connue comme une mesure de similarité. Cette mesure de similarité est une étape primordiale pour assurer une fusion correcte des informations. Elle dépend fortement des informations manipulées. Deux mesures de similarité sont appliquées suivant les concepts constituant les deux ontologies : une mesure de similarité syntaxique appliquée sur certains concepts des ontologies et une mesure de similarité possibiliste appliquée sur les concepts avec des poids possibilistes. La fusion est ensuite effectuée en combinant les différentes connaissances suivant deux modes : le mode déterministe pour les connaissances déterministes et le mode possibiliste pour les connaissances possibilistes (Fig. 4.15)

4.3.1 Mesure de similarité

Pour comparer les deux ontologies locales et aboutir à leur fusion, une estimation de leur degré de similitude est effectuée à travers deux mesures de similarité :

- une mesure de similarité syntaxique appliquée aux concepts définis de manière déterministe ;
- une mesure de similarité possibiliste appliquée aux concepts associés à des poids possibilistes.

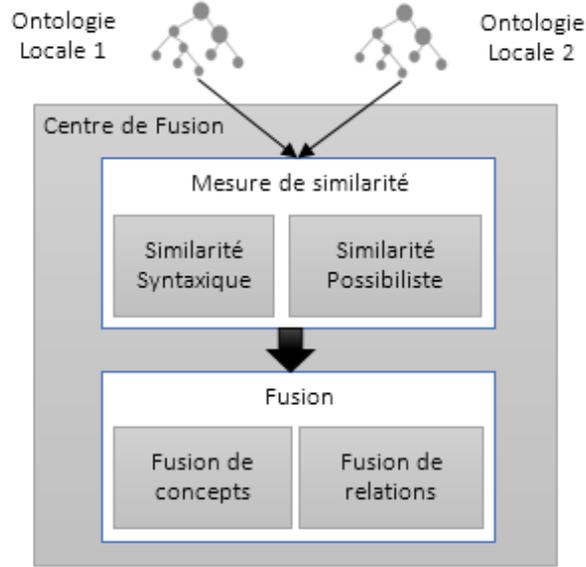


FIGURE 4.15 – Principe de la fusion des ontologies locales

a. Mesure de similarité syntaxique

La mesure de similarité syntaxique est appliquée aux attributs définis syntaxiquement selon l'équation suivante :

$$S_{sa} = \begin{cases} 1 & \text{Si } val(a_{O1}) = val(a_{O2}) \\ 0 & \text{Si } val(a_{O1}) \neq val(a_{O2}) \end{cases} \quad (4.1)$$

où : S_{sa} est la mesure de similarité des objets $O1$ et $O2$ respectivement des ontologies locales 1 et 2 suivant l'attribut a d'un concept. $val(a_O)$ est la valeur de l'attribut a d'un concept dans l'ontologie locale.

b. Mesure de similarité possibiliste

La mesure de similarité possibiliste entre deux informations est une mesure du degré de similitude entre les distributions de possibilité des informations. Dans la littérature plusieurs mesures de similarité possibilistes sont définies. Dans notre étude, nous utilisons la mesure "Information Affinity" proposée par [Jenhani et al., 2007] et décrite par l'équation :

$$S_{IA}(\pi_i, \pi_j) = 1 - \frac{D_M(\pi_i, \pi_j) + Inc(\pi_i, \pi_j)}{2} \quad (4.2)$$

où :

— D_M : une des distances de Minkowski [Hertz, 2006]. Ici nous utilisons la distance de Manhattan définie par :

$$L_1(\pi_i, \pi_j) = \frac{1}{N} \sum_{k=1}^N |\pi_i(x_k) - \pi_j(x_k)| \quad (4.3)$$

- $Inc(\pi_i, \pi_j)$: La mesure d'inconsistance entre π_i et π_j . Cette mesure estime l'inconsistance de l'intersection des deux distributions de possibilités. Elle est mesurée par :

$$Inc(\pi) = 1 - \max_{\omega \in \Omega} \{\pi(\omega)\} \quad (4.4)$$

Finalement, pour mesurer la similarité possibiliste globale entre les attributs d'un concept nous utilisons l'équation :

$$Sim_p(C_{O1}, C_{O2}) = \frac{1}{L} \sum_{i=1}^L S_i(a_{O1}, a_{O2}) \quad (4.5)$$

où L représente le nombre d'attributs d'un concept.

c. Combinaison des mesures de similarité

Les deux types de similarité calculées précédemment sont combinés donnant ainsi, une mesure globale de la similarité entre deux objets appartenant chacun à une ontologie. Cette combinaison est effectuée suivant l'équation :

$$Sim(O_1, O_2) = \sum_{k=1}^N \omega_k S_k(C_{O1}, C_{O2}) \quad (4.6)$$

où :

- N est le nombre de concepts d'un objet dans l'ontologie
- $S_k(C_{O1}, C_{O2})$ est la mesure de similarité entre les deux objets suivant le concept k, où : $S_k(C_{O1}, C_{O2}) = S_{IA}(C_{O1}, C_{O2})$ Si C_{O1} et C_{O2} sont des concepts associés avec des poids possibilistes
 $S_k(C_{O1}, C_{O2}) = S_{sa}(C_{O1}, C_{O2})$ Si A_{O1} et A_{O2} sont des concepts définis avec des valeurs déterministes.
- ω_k est un coefficient de pondération défini par l'expert comme le montre le tableau 4.1.

TABLE 4.1 – Les coefficients de pondération des différents concepts

Concepts	coefficient de pondération ω
Forme	0.03
Contour	0.04
Distance	0.1
Centre	0.12
Taille	0.18
Location	0.3
distribution	0.2
Poids	0.01
Diamètre	0.02

Finalement, une matrice de similarité, M (Fig.4.16), est obtenue contenant les valeurs de possibilité estimant le degré de similarité entre deux objets de deux

ontologies locales comparées. Cette matrice est créée au préalable et est initialisée à 0. Le nombre de lignes correspond au nombre d'objets de la première ontologie locale et le nombre de colonnes correspond au nombre d'objets de la deuxième ontologie locale. Si l'objet i de la première ontologie locale est comparé à l'objet j de la deuxième ontologie, $M(i,j)$ prend la valeur S_k qui représente le degré de similarité des deux objets. La valeur 0 indique que les deux objets sont distincts et la valeur 1 représente la similarité totale des deux objets.

	1	2	..j..	M
1	0	0	0	0
2	0	0.2	0	0.6
..i..	0	0.5	1	0
N	0	0	0	0

FIGURE 4.16 – matrice de correspondance des objets

4.3.2 Fusion ontologique

La fusion entre les deux ontologies locales est effectuée selon le degré de similarité évalué par la mesure de similarité. Deux objets sont estimés similaires si $\text{sim} \geq 0.75$. Dans ce cas, les différents concepts et les relations correspondantes sont combinés et un seul objet résulte. Autrement, l'objet garde sa définition dans l'ontologie globale.

a. Fusion de concepts

Comme notre ontologie est définie avec deux types de concepts : des concepts avec des valeurs déterministes et des concepts avec des valeurs possibilistes, la fusion de concepts des deux objets est effectuée, respectivement, de manière déterministe et possibiliste.

Pour la fusion déterministe, les valeurs d'attributs sont combinées dans un intervalle. En effet, le traitement d'images issues des différentes sources donne généralement des valeurs légèrement différentes pour le même objet. Pour modéliser cette imprécision, nous proposons de combiner les valeurs dans un intervalle. Par conséquent, le médecin aura un intervalle approximatif de valeurs possibles.

Pour la fusion possibiliste, trois opérateurs sont définis :

- Opérateurs convergents utilisés lorsque les sources sont supposées fiables et non contradictoires ;

- Opérateurs divergents utilisés lorsqu'on a besoin de collecter le maximum d'informations ;
- Opérateurs hybrides (appelés encore adaptatifs) utilisés lorsqu'on désire avoir un maximum d'informations fiables quand la fiabilité des sources est mise en question.

Dans nos travaux, nous utilisons l'opérateur adaptatif proposé par Dubois et Prade dans [Dubois and Prade, 1994]. Cet opérateur permet ainsi de passer du mode conjonctif au mode disjonctif en fonction de l'accroissement du conflit. Il est défini par l'équation suivante :

$$\max \left(\frac{\pi_{\wedge}}{h(\pi_1, \pi_2)}, \min(\pi_{\vee}, 1 - h(\pi_1, \pi_2)) \right) \quad (4.7)$$

où :

- π_{\wedge} représente l'opérateur conjonctif de la fusion.
- π_{\vee} représente l'opérateur disjonctif de la fusion.
- $h(\pi_1, \pi_2)$ est une mesure de conflit donnée par :

$$h(\pi_1, \pi_2) = 1 - \max_{x \in \Omega} \min(\pi_1(x), \pi_2(x)) \quad (4.8)$$

L'opérateur conjonctif et disjonctif utilisés ici sont respectivement la T-Norme (4.9) et la T-Conorme (4.10) définis par Zadeh. Ce sont les opérateurs les plus utilisés dans ce contexte.

$$\pi_{\wedge}(x) = \pi_1(x) \wedge \pi_2(x) = \min(\pi_1(x), \pi_2(x)) \quad (4.9)$$

$$\pi_{\vee}(x) = \pi_1(x) \vee \pi_2(x) = \max(\pi_1(x), \pi_2(x)) \quad (4.10)$$

b. Fusion de relations

Les objets en relation avec deux objets similaires sont comparés puis combinés selon le principe suivant :

- Si les deux objets similaires sont en relation avec deux objets similaires, les deux objets sont fusionnés en un seul objet identifié en relation avec l'objet initial résultant.
- Si les deux objets similaires sont en relation avec deux objets différents, ces derniers sont ajoutés (tous les deux) en relation avec l'objet initial résultant.

Pour les objets différents, des calculs de leurs localisations les uns par rapport aux autres sont effectués en utilisant les coordonnées de leurs centres ainsi que leur distances par rapport au mamelon et au muscle pectoral. A partir de ces calculs, les relations correspondantes entre eux sont établies et définies dans l'ontologie globale.

La figure 4.17 présente un exemple de processus de fusion de deux ontologies locales. Chacune des deux ontologies locales contient deux objets. Les différents objets sont comparés deux à deux et une matrice de similarité est obtenue. D'après la matrice de similarité, les objets Obj12 et Obj21 sont similaires. Ces deux objets sont combinés et constituent l'objet Obj3 de l'ontologie globale. Les objets Obj11 et Obj22 sont reproduits respectivement en Obj1 et Obj2.

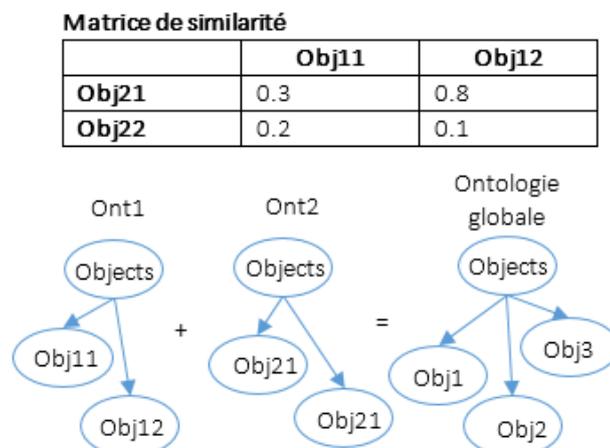


FIGURE 4.17 – Exemple de fusion de deux ontologies locales

Pour chaque cas i faire :

1. L'instance i est mappée aux concepts et relations de l'ontologie du domaine,
2. Mise à jour du fichier XML en ajoutant un nouvel individu et en lui attribuant les degrés de possibilité correspondants,
3. Validation de la consistance de l'ontologie.

TABLE 4.2 – Algorithme du peuplement de l'ontologie de domaine

4.4 Instanciation

L'instanciation (ou le peuplement) d'une ontologie consiste à associer des instances (appelées encore des individus) à des concepts ou / et des relations de l'ontologie de domaine. En effet, comme nous utilisons des poids possibilistes à certains concepts de l'ontologie, nous proposons d'utiliser un fichier XML, contenant les différentes instances. Ce fichier est associé à l'ontologie de domaine utilisée. La figure 4.18 montre un extrait d'un exemple du fichier xml instanciant l'ontologie. La procédure de l'instanciation effectue une mise à jour de ce fichier par une nouvelle instance à chaque fois.

Le tableau 4.2 présente l'algorithme du peuplement de l'ontologie à travers des instances extraites d'images ayant des poids possibilistes :

4.5 Raisonnement et Aide au diagnostic

Face à un problème donné, l'expert doit examiner les informations et les connaissances disponibles qu'il utilise dans un processus de raisonnement afin d'en tirer des conclusions sur le cas étudié. Durant ce processus, le médecin s'appuie, généralement, sur son expérience vécue dans le diagnostic médical. Cela revient à dire qu'il essaie

```
▼<Individuals>
  ▼<individual>
    <NamedIndividual IRI="Cas1"/>
    ▼<Concept IRI="Round_shape">
      <degree value="0"/>
    </Concept>
    ▼<Concept IRI="Oval">
      <degree value="0.8"/>
    </Concept>
    ▼<Concept IRI="Irregular_shape">
      <degree value="0.3"/>
    </Concept>
    ▼<Concept IRI="Large_size">
      <degree value="0.1"/>
    </Concept>
    ▼<Concept IRI="Small_size">
      <degree value="0.8"/>
    </Concept>
  </individual>
  ▼<individual>
    <NamedIndividual IRI="Cas2"/>
    ▼<Concept IRI="Partly_round_shape">
      <degree value="0.4"/>
    </Concept>
    ▼<Concept IRI="Irregular_shape">
      <degree value="0.8"/>
    </Concept>
    ▼<Concept IRI="Large_size">
      <degree value="0.1"/>
    </Concept>
    ▼<Concept IRI="Small_size">
      <degree value="0.75"/>
    </Concept>
  </individual>
</Individuals>
```

FIGURE 4.18 – Extrait d'un exemple de fichier associé à un individu de l'ontologie

de se souvenir des cas qu'il a diagnostiqués précédemment et de chercher ceux les plus similaires au cas présent afin de faire correspondre un certain diagnostic. Ce processus représente le raisonnement à base de cas, qui est une méthodologie très utilisée dans la discipline de l'intelligence artificielle et surtout dans le domaine médical, dans l'objectif d'aider le médecin au diagnostic et l'assister à la prise de décision. Le principe du raisonnement à base de cas adopté consiste à confronter le cas requête avec des cas déjà diagnostiqués et formant un ensemble appelé base des cas. Une mesure de similarité entre les descriptions des cas est calculée dans un contexte possibiliste. Les cas les plus similaires sont retenus et le médecin a, à sa disposition, les diagnostics possibles extraits des cas retenus avec des mesures de possibilité exprimant le degré de correspondance du cas sélectionné au cas requête. Le médecin examine les diagnostics possibles et décide quel diagnostic est adéquat au cas étudié. Si le médecin évalue les diagnostics possibles comme inadéquats, il peut ajouter un nouveau diagnostic qui enrichira, par la suite, la base de cas. Les figures 4.19 et 4.20 présentent ce principe.

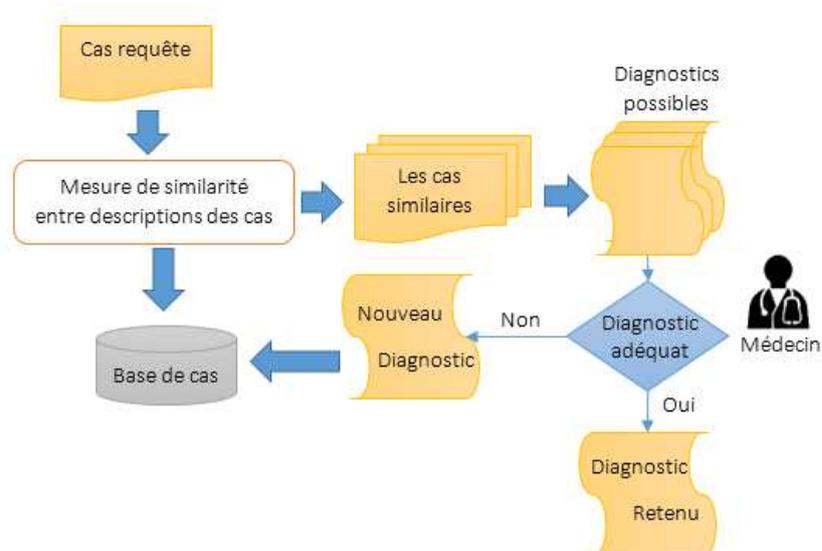


FIGURE 4.19 – Principe du raisonnement à base de cas d'aide au diagnostic médical

4.5.1 Le cas requête

Le cas requête représente le cas examiné par le médecin à travers les deux modalités d'images utilisées. En traitant les images relatives à ce cas, les informations et les connaissances concernant le cas sont représentées par la description ontologique expliquée précédemment. A cette description, le médecin peut ajouter certaines informations cliniques qu'il collecte auprès du patient durant un questionnaire d'examen clinique. Ces informations sont ajoutées à l'instance correspondante dans l'ontologie

de domaine. Ainsi, le cas requête contient seulement une description clinique et requiert un diagnostic décisionnel adéquat.

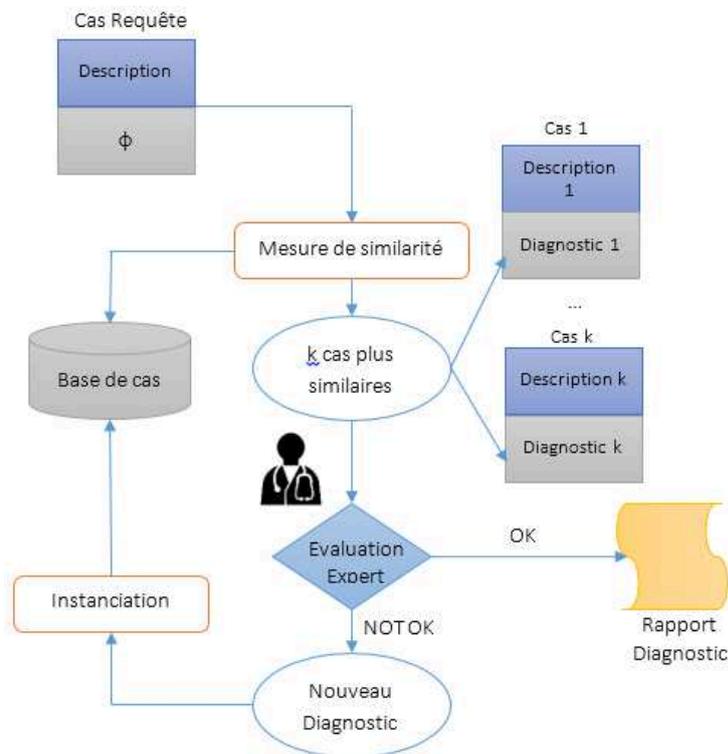


FIGURE 4.20 – Approche proposée pour le raisonnement et l'aide au diagnostic

4.5.2 La base de cas

La base de cas représente l'historique des diagnostics déjà effectués pour des cas étudiés. Elle modélise, alors, l'expérience passée dans les diagnostics dans le domaine d'application. Cette base contient tous les cas diagnostiqués. Chaque cas diagnostiqué représente une instance de l'ontologie du domaine. Il est représenté par deux parties essentielles 4.20 :

- Une partie description qui contient les connaissances disponibles extraites des images étudiées et décrivant l'état pathologique du patient. En terme formel, cette partie représente la description du problème ;
- Une partie diagnostic qui contient la décision prise par l'expert pour diagnostiquer le cas étudié. Elle modélise l'interprétation décisionnelle du cas examiné. En terme formel, cette partie constitue la solution possible du problème évoqué.

Par exemple, pour un cas donné, la partie descriptive représente les connaissances extraites des images et modélisées par une ontologie locale contenant les différentes

propriétés des lésions décelées (forme, contour, taille,...) ; la partie diagnostic contient l'interprétation faite par le médecin selon le type de la lésion(kyste, fibroadénome,...) et la décision prise (biopsie, ablation,...). Cette base de cas est utilisée pour évaluer le degré de similitude entre un cas examiné et les cas y présents en vue de trouver un éventuel diagnostic.

4.5.3 Mesure de similarité

La mesure de similarité a pour objectif de confronter le cas requête aux différents cas déjà diagnostiqués afin d'obtenir un diagnostic adéquat. Pour aboutir à cette fin, le cas examiné (cas requête) est comparé à chaque cas présent dans la base de cas. En effet, il s'agit ici de comparer les cas par leurs descriptions. Notons ici que la description des cas est effectuée suivant la même structure définie par l'ontologie de domaine. Par conséquent leur comparaison est possible. Le résultat de cette opération est un ensemble de cas similaires au cas étudié. Une valeur possibiliste est attribuée à chaque cas similaire détecté estimant son degré de similitude à la requête.

4.5.4 Les cas possibles

Les cas possibles présentés au médecin sont les cas retenus du processus de mesure de similarité suivant une valeur seuil de possibilité définie par l'expert. Seuls les cas ayant une mesure de possibilité supérieure à cette valeur seuil sont maintenus. Ces cas sont présentés au médecin classés par ordre décroissant des mesures de possibilité ; de manière à ce que le cas le plus possible est présenté en premier. Le médecin aura alors à sa disposition un rapport contenant les cas les plus similaires par description et les diagnostics potentiels qui leurs sont associés. Le médecin évalue alors les différents diagnostics possibles et sélectionne le plus adéquat au cas requête. Si le système ne présente aucun cas possible, ou le médecin estime qu'aucun cas des cas similaires sélectionnés ne représente le diagnostic efficace pour le cas requête, il spécifie un nouveau diagnostic.

4.5.5 Instanciation de nouveaux cas

Dans le cas où le médecin estime le rapport diagnostic fourni manquant ou discordant, il procède à la spécification d'un nouveau diagnostic pour le cas examiné. Cette nouvelle solution est affectée à ce dernier et le cas est instancié dans la base de cas, permettant son enrichissement pour d'éventuelles recherches à venir. L'instance correspondante dans l'ontologie de domaine est par conséquent mise à jour.

4.6 Modèle symbolique de représentation des connaissances

Le modèle symbolique de représentation des connaissances est un modèle simplifié permettant de visualiser de façon symbolique la scène étudiée. En effet, dans notre contexte d'aide au diagnostic du cancer du sein, ce modèle sert à fournir un aperçu sommaire des différents objets pathologiques dans le sein.

Selon le domaine d'application, la validation d'un modèle impose un certain nombre de critères. D'une façon générale, et afin d'avoir un bon modèle permettant une manipulation aisée et efficace des objets d'intérêt, certaines propriétés doivent être confirmées [Moron, 1996] [Brown, 1981] :

- **La capacité** : Tout objets d'intérêt pourra être représenté dans le modèle.
- **La concision** : Tout objet doit être décrit d'une façon concise en termes de taille, verbosité, redondance...
- **La stabilité** : réfère aux méthodes de calcul numérique des paramètres du modèle.
- **La précision** : Tout objet ne doit pas avoir des ambiguïtés dans sa représentation dans le modèle.
- **La sensibilité** : le modèle doit différencier les objets représentés de façon similaires.
- **La fiabilité** : le modèle doit être représenté d'une façon claire, correcte et robuste.

4.6.1 Choix du modèle de représentation de connaissances

La définition du modèle de représentation de connaissances de la scène observée sert à coder les différents objets à modéliser par des symboles distincts. Ainsi, chaque objet est représenté par un symbole afin de fournir une reproduction simplifiée de la scène étudiée.

La première étape à considérer est la détermination des différents objets à modéliser. Pour le cas des connaissances mammaires nous retenons les objets suivants :

- **Le sein** : C'est l'organe étudié. Il représente l'objet principal modélisant une scène 3D. C'est l'espace contenant les objets pathologiques à détecter et étudier.
- **Les objets pathologiques** : Ce sont les objets devant être diagnostiqués pour traiter le cas de la patiente présente. Les médecins distinguent plusieurs types d'objets pathologiques dans le contexte des lésions mammaires suivant leurs caractéristiques. Dans notre représentation, nous considérons les masses, les microcalcifications et les carcinomes.

Le tableau 4.3 récapitule les différents objets pathologiques ainsi que les symboles associés. La figure 4.21 présente un aperçu général du modèle virtuel choisi pour la

représentation d'objets pathologiques.

TABLE 4.3 – Définition des modèles virtuels des objets

Objet	spécification	symbole
Organe principal :		
Sein	Objet principal représentant la scène étudiée	demi-ellipse
Objets pathologiques :		
Masses :		
Masses Bénignes (MB)	objets pathologiques représentant habituellement des formes et des contours réguliers	sphère bleue
Masses Malignes (MM)	objets pathologiques représentant habituellement des formes et des contours irréguliers	sphère orange
Microcalcifications (MC) :		
Microcalcifications Régulières (MCR)	dépôt de petits objets pathologiques représentant habituellement des formes et des contours réguliers	sphère verte
Microcalcification Irrégulières (MCI)	dépôt de petits objets pathologiques représentant habituellement des formes et des contours irréguliers	sphère jaune
Désorganisation architecturale	Risque éventuel d'un carcinome (un cancer)	cube rouge

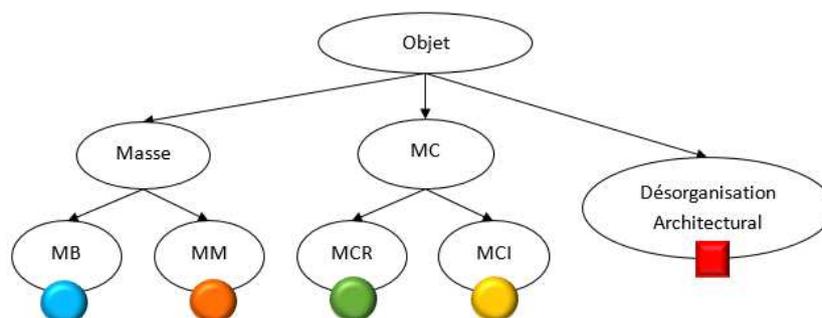


FIGURE 4.21 – Définition du modèle de représentation virtuelle de connaissances pour la représentation des objets pathologiques

4.6.2 Représentation des connaissances et intégration des données

La représentation virtuelle de connaissances passe par trois étapes principales :
 — La modélisation du sein

- La modélisation des objets pathologiques
- L'intégration d'objets pathologiques dans le modèle du sein.

a. Le modèle virtuel du sein

Le modèle virtuel du sein est représenté par un demi-ellipsoïde en calculant les diamètres nécessaires à partir de l'image IRM comme le montre la figure ??.

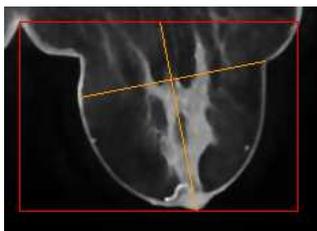


FIGURE 4.22 – Calcul de paramètres à partir de l'image IRM

b. Modélisation virtuelle des objets pathologiques

La modélisation virtuelle des différents objets pathologiques est effectuée suivant les caractéristiques spécifiant chaque objet. La figure 4.23 schématise l'algorithme de modélisation d'objets relatifs à un cas étudié. En effet, chaque cas représente une instance de l'ontologie globale. Il dispose d'une description et d'un diagnostic issu de l'étape du raisonnement adopté. Par conséquent, une sélection de paramètres nécessaires : principalement les diamètres et la décision sur l'identité pathologique de l'objet (masse, microcalcifications, ou carcinome), permet d'identifier le symbole approprié (comme défini dans la sous-section précédente) pour l'objet en question. L'objet, étant modélisé par le symbole adéquat, est ensuite intégré dans le modèle 3D du sein préalablement préparé. L'algorithme est itéré sur le nombre d'objets détectés dans l'examen du cas.

c. Intégration des objets pathologiques dans le modèle du sein

L'intégration des différents objets modélisés dans le modèle virtuel du sein est effectué suivant la position de l'objet dans l'espace du sein. Cette position est déterminée par les coordonnées du centre de l'objet dans le repère ayant comme origine le mamelon (Fig. 4.24).

4.7 Conclusion

La conception d'un système d'aide au diagnostic médical nécessite un certain nombre d'étapes, afin de garantir un système fiable pouvant fournir l'aide requise pour le médecin. L'approche que nous proposons dans le cadre de cette thèse, combine les connaissances extraites de deux sources différentes représentant la même

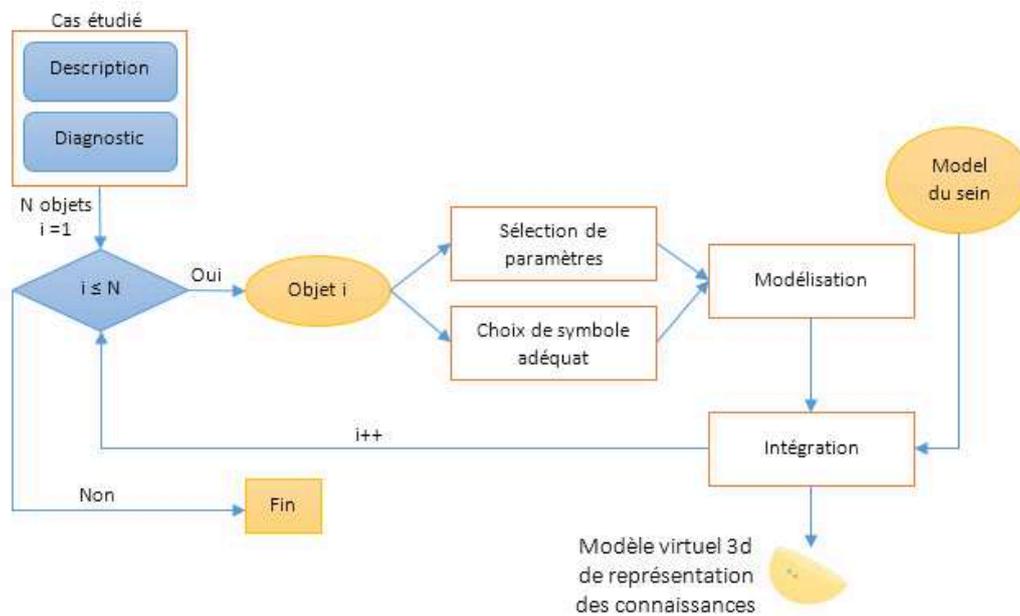


FIGURE 4.23 – Algorithme de modélisation virtuelle de connaissances d’un cas étudié

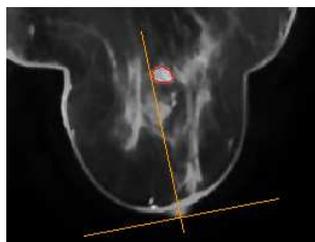


FIGURE 4.24 – Aperçu du repère utilisé pour identifier les coordonnées du centre de l’objet pathologique

scène et présente au médecin une interprétation à travers un raisonnement à base de cas similaires.

Le présent chapitre présente en premier temps le principe de l'approche dans sa globalité puis détaille les différentes étapes. Notre approche est constituée principalement de quatre étapes :

- L'extraction de connaissances et leur modélisation en ontologies locales en se basant sur une ontologie de domaine.
- La fusion des ontologies locales en une seule ontologie globale en combinant les différents concepts et leurs relations
- L'instantiation de l'ontologie de domaine avec l'ontologie globale obtenue.
- Le raisonnement possibiliste à base de cas en vue de l'obtention d'une interprétation du cas étudié à travers un cas similaire.

Résultats et Evaluation

Sommaire

5.1 Bases d'images	90
5.2 Description du scénario d'application	90
5.3 Résultats de l'étape du traitement des images	92
5.3.1 Résultats du prétraitement et de la segmentation	93
5.3.2 Résultats de l'extraction de caractéristiques	93
5.3.3 Validation des ontologies locales résultantes	94
5.4 Résultats et évaluation de la fusion des ontologies locales .	97
5.5 Résultats et évaluation du raisonnement médical basé sur l'ontologie globale	99
5.5.1 Validation de cas similaires	100
5.5.2 Vérification des diagnostics proposés	101
5.6 Evaluation globale	102
5.7 Conclusion	103

Dans ce chapitre, nous présentons les expériences menées afin de tester notre approche. Nous exposons les différents résultats de son évaluation et de sa vérification. Dans un premier temps, nous présentons les détails des bases de données utilisées pour l'évaluation de l'approche. Ensuite, nous détaillons les tests effectués pour chaque phase de l'approche avec une évaluation des résultats, plus précisément la phase de la transformation des images en ontologie locale, la phase de fusion des ontologies locales et la phase du raisonnement à base de cas. Nous présentons, à la fin, une évaluation globale basée sur l'appréciation de 4 experts du domaine médical.

5.1 Bases d'images

Pour valider notre approche, nous utilisons trois base de données suivant le type du traitement effectué :

- Pour valider le traitement des images mammographiques, nous utilisons la base de mammographies gratuite : mini-MIAS. Cette base contient 322 images dont 209 représentent des cas normaux et 113 images représentant des cas pathologiques de différents degrés de malignité.
- Pour valider le traitement des images IRM, nous utilisons la base d'IRM mammaire publique : TCIA. Cette base contient 429 volumes répartis sur 89 sujets.
- Pour valider la fusion et le raisonnement à base de cas, nous disposons de 35 images mammographiques relatives à cinq patientes et de 134 volumes IRM correspondants aux mêmes patientes. Les images sont fournies au service de radiologie de l'Hôpital régional de Ben Arous - Tunisie. Les cinq cas sont pathologiques à différents stades cliniques.

Nous pourrions justifier le manque de la dernière base par le fait que les couples (IRM, Mammographie) ne sont pas tout le temps disponibles ensemble. En effet, les médecins procèdent dans les cas les plus critiques à un examen IRM. D'une part, parce que l'IRM est un examen cher et peu accessible à tout le monde et nécessite un matériel sophistiqué, qui, aussi, n'est pas disponible dans tous les hôpitaux. D'autre part, les rayons magnétiques sont graves pour la santé. Dans la majorité des cas, l'examen mammographique est le plus utilisé pour diagnostiquer un cas.

5.2 Description du scénario d'application

L'objectif de l'approche proposée est de fusionner les connaissances issues de deux images de modalités différentes représentant le sein : une mammographie et une IRM. La figure 5.1 présente la procédure générale adoptée pour le test et l'évaluation de l'approche proposée. Trois principales évaluations sont effectuées :

- L'évaluation du traitement des images mammographiques et IRM consistant à évaluer le résultat obtenu de l'étape de la transformation de chaque image en une ontologie locale.
- L'évaluation de la fusion des deux images consistant principalement à évaluer d'une part la détection d'objets similaires vus dans les deux images et d'autre part l'ontologie résultante de la combinaison de différents objets.
- L'évaluation de l'étape du raisonnement à base de cas consistant à expertiser le degré d'efficacité des diagnostics possibles retenus.

Les sections suivantes présentent les différents résultats obtenus à chaque étape et les différents tests effectués pour chaque évaluation.

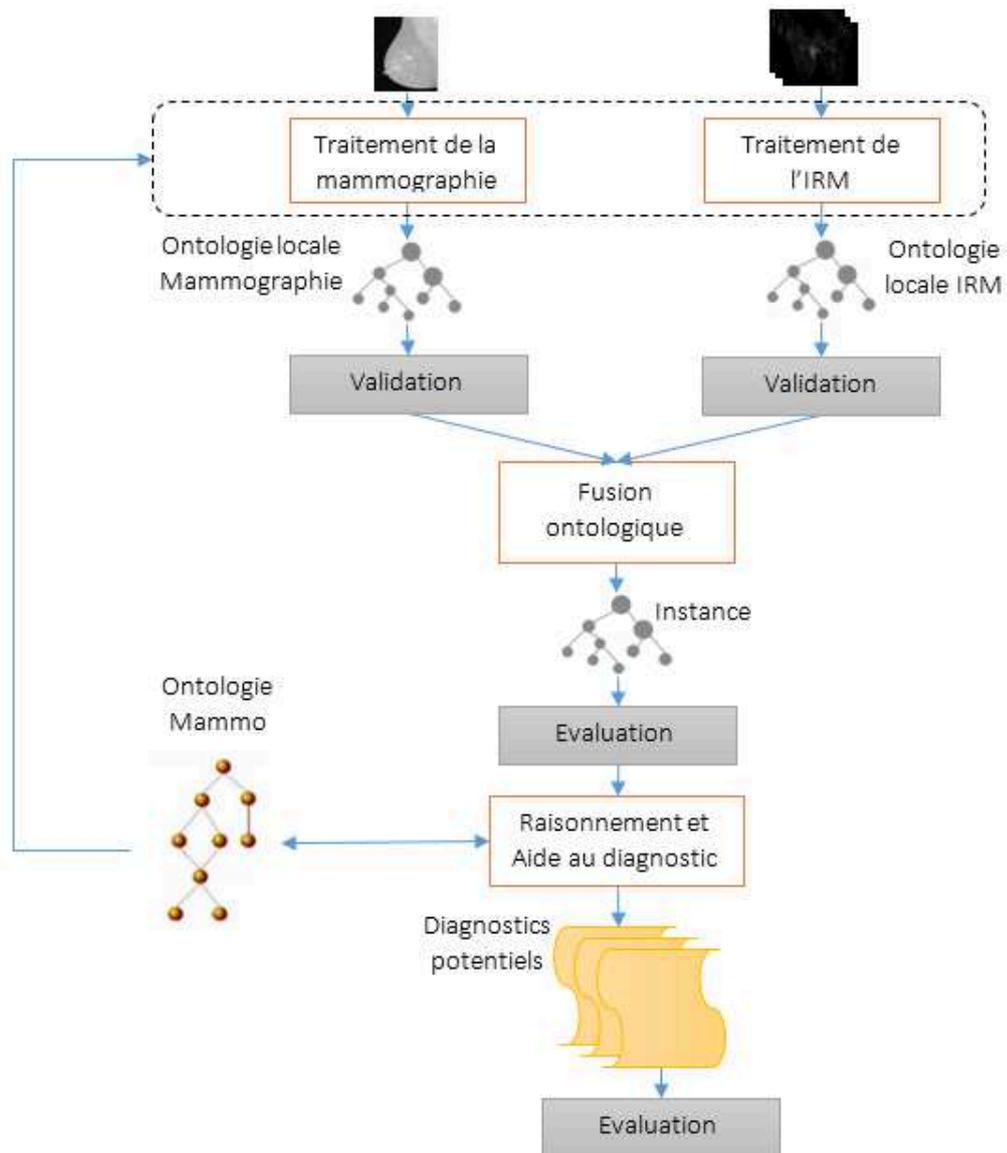


FIGURE 5.1 – Procédure globale de validation de l'approche proposée

5.3 Résultats de l'étape du traitement des images

Comme nous l'avons expliqué dans le chapitre précédent, notre approche est traitée suivant différents niveaux. Cette première phase, qui est le traitement des images vise à transformer une image en une ontologie locale. Cette phase comprend donc quatre niveaux comme le montre la figure 5.2. Elle permet de passer de l'espace pixellique (respectivement voxillique) de l'image mammographique (respectivement IRM) à l'espace objet des deux images.

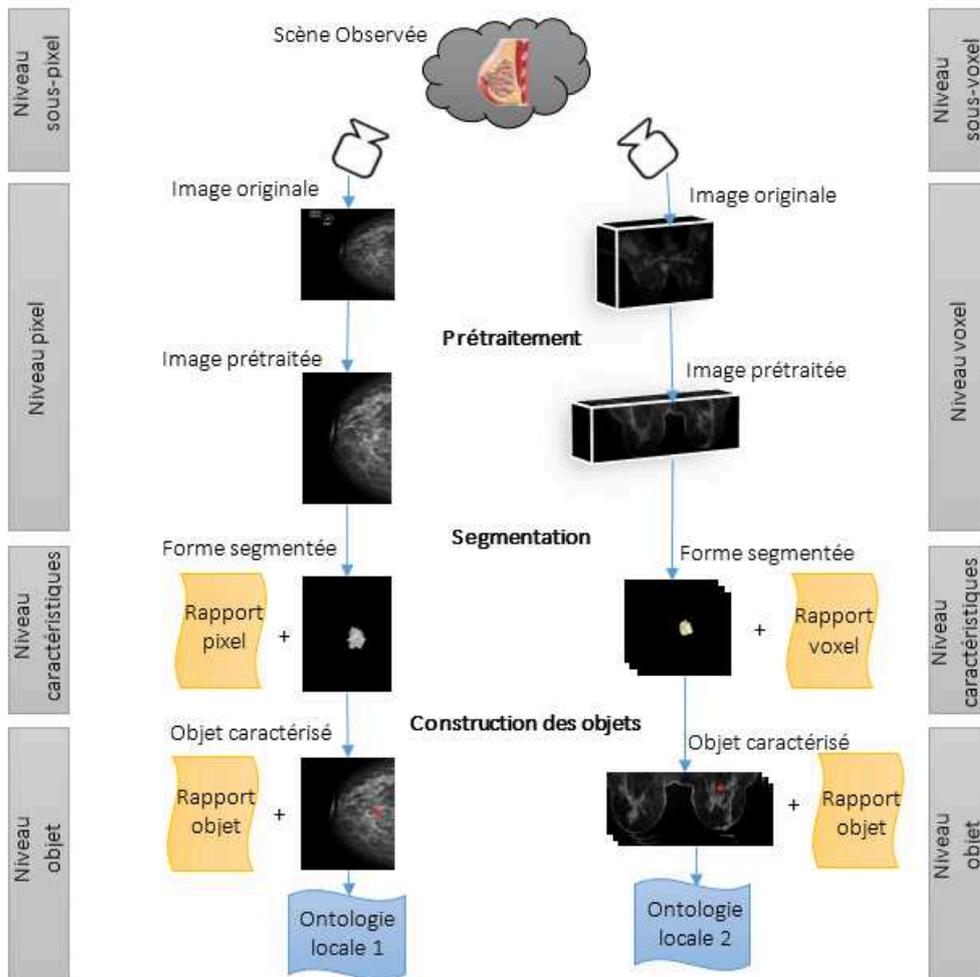


FIGURE 5.2 – Procédure du traitement d'images mammaires

TABLE 5.1 – Exemple d'évaluation de caractéristiques visuelles extraites d'une forme pathologique

Caractéristiques	Valeur trouvée	Valeur calculée par l'expert	Erreur
Diamètre minimum	2.5	2.8	0.3
Diamètre maximum	4	4	0
Coordonnées du centre	X = 2951.106 Y = 2328.277	X= 2950 Y= 2328	1.106 0.277

5.3.1 Résultats du prétraitement et de la segmentation

Nous rappelons que cette étape vise la correction de certaines imperfections liées à la qualité des données visuelles sur l'image (rehaussement, filtrage). Nous indiquons que cette étape ne fait pas partie réellement de notre contribution dans le travail de recherche que nous avons mené. Cependant, une évaluation des résultats obtenus est effectuée par un expert pour guider le choix de la technique de filtrage adoptée. L'annexe A présente les différents résultats obtenus et l'évaluation effectuée.

5.3.2 Résultats de l'extraction de caractéristiques

La procédure d'extraction de caractéristiques utilise certaines formes régulières connues (l'ellipse circonscrite, cercle inscrit, etc...), pour effectuer certains calculs permettant d'identifier la forme pathologique détectée suivant le principe détaillée dans le chapitre précédent.

Afin d'évaluer l'exactitude des valeurs calculées nous avons opté pour une comparaison entre ces valeurs et des valeurs spécifiées par l'expert pour chaque image. Nous calculons l'erreur quadratique moyenne pour chaque valeur calculée et ensuite la moyenne pour un objet pathologique. Le Tableau 5.1 présente un exemple d'évaluation des caractéristiques extraites à partir d'un objet pathologique d'une image mammographique.

Le processus de la validation de l'étape de l'extraction de connaissances est évalué suivant le principe décrit par la figure 5.3.

Une base d'objets est construite à partir des différents rapports écrits par les experts et associés aux différentes images. Une mesure de similarité est effectuée entre chaque objet pathologique décrit par la signature sémantique et l'objet qui lui correspond dans la base d'objets. La mise en correspondance des différents objets implique un taux d'erreur de 11.504% pour le traitement des images mammaires et un taux de 9.988% pour les images IRM. Le taux d'erreur trouvé est dû principalement au taux élevé de la densité mammaire dans certaines images mammographiques ou

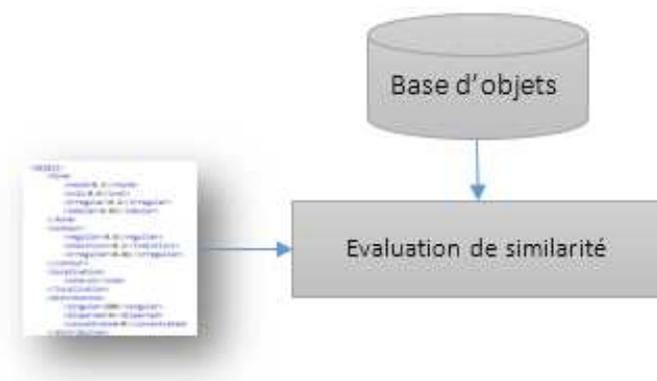


FIGURE 5.3 – Principe de l'évaluation de la spécification des objets pathologiques extraits

aux imperfections liées au contraste de l'image.

5.3.3 Validation des ontologies locales résultantes

Pour obtenir l'ontologie locale relative à une image, les différentes caractéristiques sémantiques sont mappées avec les concepts de l'ontologie de domaine.

Afin de valider les performances des ontologies construites à partir de chaque image, nous adoptons une procédure à deux étapes :

- Une étape de vérification de la contradiction et de la cohérence syntaxique et sémantique de l'ontologie (Fig. 5.4)
- Une étape de validation des résultats

La contradiction désigne la présence d'un désaccord au niveau des valeurs attribuées aux concepts de l'ontologie. Par exemple : "La distribution d'un objet pathologique est estimée à 80% singulière et à 80% dispersée." implique une contradiction au niveau des connaissances. Quant à la cohérence, elle est exprimée en terme d'accord logique ou de non présence de conflits entre les différentes connaissances. Par exemple : "Une forme est estimée à irrégulière et le contour est estimé à régulier." est une caractérisation incohérente.

Afin de valider cette étape, une première vérification de la cohérence de l'ontologie résultante est effectuée par le raisonneur de l'éditeur "Protégé"¹. Cette vérification permet de valider de manière globale la cohérence de l'ontologie.

Comme l'ontologie est construite avec des concepts ayant des poids possibilistes, une deuxième vérification est effectuée pour vérifier la contradiction et la cohérence

1. Editeur d'ontologies gratuit : <http://protege.stanford.edu/>

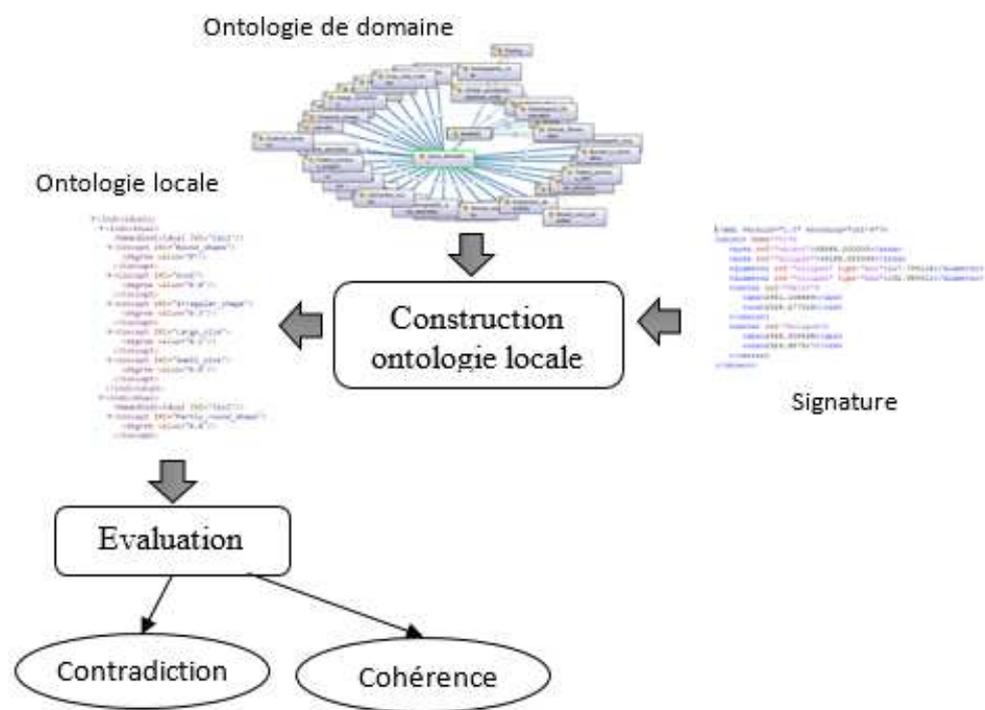


FIGURE 5.4 – Processus de validation d’une ontologie locale

de ces concepts. Pour ce faire, nous utilisons une base de règles créée au préalable à partir des prestations des experts. Ci-après quelques exemples de règles.

a. Contradiction Sémantique

- *"Si la forme d'un objet est ronde ou ovale, elle ne peut pas être irrégulière ou distorsion."*
- *"Un objet ne peut pas être de distribution singulière et dispersée simultanément ou singulière et concentrée simultanément."*
- *"Le diamètre minimum de l'objet ne peut pas être plus grand que le diamètre maximum."*

b. Cohérence Sémantique

- *"Si un objet a une forme ronde ou ovale, il ne peut pas avoir un contour irrégulier."*
- *"Si un objet a une forme irrégulière il ne peut pas avoir un contour régulier."*
- *"Si un objet a une forme lobulaire il ne peut pas avoir une distribution dispersée ou concentrée."*
- *"Si un objet a une forme ronde, son diamètre minimum est égal ou presque égal à son diamètre maximum."*

Une matrice binaire de validation de l'ontologie (Tab.5.2) est générée vérifiant sa contradiction et sa cohérence sémantiques de l'ontologie selon les conditions suivantes :

- Si un objet contient une contradiction, la case correspondante est évaluée à 1, sinon à 0 ;
- Si l'objet contient une incohérence, la case correspondante est évaluée à 0 sinon à 1.

Le tableau 5.2 montre la consistance d'une ontologie locale n'ayant aucune contradiction ni incohérence sémantique.

TABLE 5.2 – Exemple de d'une matrice de validation d'une ontologie locale

	Contradiction	Cohérence
Objet1	0	1
Objet2	0	1

Le tableau 5.3 présente les performances de notre méthode de transformation d'une image en ontologie locale pour les cas étudiés.

La modélisation des connaissances extraites de l'image par une ontologie locale a permis le passage du niveau des données brutes de l'image au niveau de sa sémantique. Par ailleurs, le caractère possibiliste de l'ontologie construite, augmente la performance des résultats. En effet, le degré de possibilité attribué à certains attributs permet d'augmenter la vérité des valeurs estimées. Les résultats exposés

TABLE 5.3 – Mesures de performance des ontologies locales construites à partir des images mammaires

Mesure	Image Mammographique	IRM
Ontologies	98	89
Ontologies ayant des contradictions	4	6
Ontologies ayant des incohérences	2	4
Efficacité	93.877%	88.764%

dans le tableau 5.3 prouvent l'efficacité de la procédure de cette transformation. Cependant, jusqu'ici, ils ne montrent pas la précision et l'exactitude des résultats.

Dans la littérature, les travaux menés sur la transformation des images en ontologies locales ne présentent pas une évaluation des ontologies à ce niveau là. Ils se contentent de l'évaluation de la précision, la spécificité et l'exactitude des résultats obtenus en utilisant l'ontologie construite (dans la recherche, dans l'interprétation, etc...). Il est, donc, difficile, à ce niveau, de comparer nos résultats avec ceux décrits dans d'autres travaux.

5.4 Résultats et évaluation de la fusion des ontologies locales

La phase de fusion commence, comme expliqué précédemment, par le calcul d'une matrice de similarité afin de déterminer les objets similaires des objets différents pour exécuter la fusion. Le résultat de la fusion de deux ontologies locales construites à partir des images est une ontologie globale représentant une instance de la scène observée dans sa globalité.

Pour valider l'opération de la fusion de deux ontologies locales, nous effectuons une vérification sur deux niveaux :

- Au niveau des mesures de similarité effectuées entre les objets constituant les deux ontologies locales.
- Au niveau de la contradiction et de la cohérence de l'ontologie globale par le même principe adopté pour la vérification des ontologies locales ;

a. Evaluation de la mesure de similarité

Afin d'évaluer l'exactitude des mesures de similarité calculées pour les différents objets des ontologies locales, nous calculons une matrice d'appariement (Tab.5.4) des objets vérifiant l'exactitude de la similarité calculée automatiquement par rapport à une base d'objets créée au préalable. Cette base est construite à partir des rapports

diagnostics des différents cas étudiés, . Cette matrice est déterminée suivant les règles suivantes :

- Si les deux objets sont bien appariés (estimés similaires dans les deux cas, ou estimés non similaires dans les deux cas), la case correspondante est évaluée à 1.
- Si les deux objets sont mal appariés (estimés similaires dans un seul cas) la case correspondante est évaluée à 0.

TABLE 5.4 – Exemple de matrice d'appariement des objets des ontologies locales

	Appariement automatique		Appariement à partir de la base	
	obj11	obj12	obj11	obj12
obj21	0	1	0	1

TABLE 5.5 – Matrice d'évaluation de l'appariement des objets

	obj11	obj12
obj21	1	1

La matrice représentée par le tableau 5.5 montre que l'objet21 de l'ontologie locale 2 est bien apparié avec les objets de l'ontologie 1. Ce qui revient à dire que l'estimation trouvée coïncide avec la spécification extraite de la base.

b. Evaluation de la fusion des ontologies

La validation de l'étape de fusion est effectuée à travers l'exactitude de l'appariement des différents objets pathologiques suivant la mesure de similarité calculée. Afin de valider les performances de cette étape, nous avons calculé les mesures suivantes :

- **Les Faux Positifs (FP)** : Deux objets sont considérés comme similaires alors qu'ils sont réellement distincts.
- **Les Vrais Positifs (VP)** : Deux objets sont considérés comme similaires et ils le sont réellement.
- **Les Vrais Négatifs (VN)** : Deux objets sont considérés comme distincts et ils le sont réellement.
- **Les Faux Négatifs (FN)** : Deux objets sont considérés comme distincts alors qu'ils sont réellement similaires.

On considère deux objets comme similaires si la mesure de possibilité correspondante est supérieure ou égale à 0.8. Cette valeur est fixée selon l'avis d'un expert.

Les différentes mesures décrites précédemment sont utilisées pour calculer les mesures de précision, spécificité, et exactitude. Le tableau 5.6 décrit les différentes valeurs obtenues pour l'étude des cas disponibles.

Les résultats obtenus montrent la performance de notre méthode appliquée sur les cas étudiés. Cependant, le faible nombre de cas de test reste la limite majeure de

TABLE 5.6 – Évaluation de la fusion des ontologies locales

Objets - mammographie	3
Objets - IRM	7
FP	1
TP	2
TN	2
FN	0
Précision	66.666%
Spécificité	66.666%
Exactitude	80%

cette expérimentation. En outre, le faible nombre d'objets pathologiques présents sur les images mammographiques ne complique pas la mesure de similarité, contrairement à un autre type de scènes observées. Prenons l'exemple d'une scène naturelle contenant des arbres, des fleurs, des nuages, des oiseaux, des lapins, etc... Le grand nombre et la diversité des objets constituant la scène compliquent la mesure de similarité et par la suite le processus de fusion.

5.5 Résultats et évaluation du raisonnement médical basé sur l'ontologie globale

Cette partie vise à présenter au médecin une assistance au diagnostic tirant avantage des cas précédemment diagnostiqués.

Le test de cette partie est effectué sur les images mammographiques issues de la base MIAS. En effet, 56 cas sont utilisés pour l'apprentissage de la base des cas, et 62 sont utilisés pour le test. La validation du raisonnement et de la décision proposée au médecin est effectuée suivant deux étapes :

- La validation des cas similaires sélectionnés à travers la construction d'une matrice de vérification. Chaque case de la matrice contient une valeur exprimant le degré de possibilité qu'un diagnostic associé à un cas similaire est en accord avec le cas cible traité. Nous retenons à la fin les cas similaires validés avec une valeur de possibilité supérieure ou égale à 0.7.
- La vérification des diagnostics proposés : effectuée afin de s'assurer de la cohérence et de l'adéquation du couple description - diagnostic. En effet, deux caractéristiques sont évaluées :
 - La contradiction sémantique : une caractéristique dans la description du cas est en désaccord avec le diagnostic proposé. Exemple : Dans la description la forme d'un objet pathologique est estimée "irrégulière" et le diagnostic indique une masse bénigne.

- L'incohérence du diagnostic : une caractéristique dans la description du cas implique un diagnostic non logique. Par exemple : Dans la description du cas la répartition est estimée "dispersée" ou "concentrée" et le diagnostic indique une masse.

5.5.1 Validation de cas similaires

Afin d'évaluer les différents cas similaires détectés, nous utilisons une matrice d'évaluation (Tab. 5.8) permettant d'indiquer les degrés de similitudes entre le cas requête et les différents cas possibles.

Le tableau 5.7 présente l'algorithme adopté pour l'évaluation de la similarité du cas requête aux N cas constituant la base.

TABLE 5.7 – Algorithme d'évaluation des différents cas similaires

```

Pour i de 1 à n faire
    s = Similarité(cas_r.description, cas[i].description);
    matrice_Evaluation[i][1] = s;
    d = Verification_Diagnostic(cas_r, cas[i].diagnostic);
    matrice_Evaluation[i][1] = d;
    Ordonner(matrice_Evaluation);
    Extrait(matrice_Evaluation, valeur_seuil);
Fin Pour.

```

Chaque diagnostic associé à chaque cas similaire est évalué par rapport à la description du cas requête par une mesure de possibilité. La matrice d'évaluation est ordonnée par rapport à la colonne diagnostic (du plus possible aux moins possible). La dernière étape consiste à extraire les diagnostics les plus possibles au cas requête en sélectionnant ceux ayant un degré d'adéquation supérieur ou égal à une valeur seuil indiquée par l'expert (nous retenons dans ce cas la valeur 0.7).

TABLE 5.8 – Extrait d'une matrice d'évaluation

Cas possibles	Description	Diagnostic
cas1	0.1	0
cas2	0.8	0.7
cas3	0	0
...
casn	0.8	0.8

La matrice d'évaluation (Tab. 5.8) indique que le premier cas de la base est similaire à 10% au cas requête et son diagnostic n'est pas adéquat à la situation

étudiée. La description du deuxième cas est similaire à celle du cas requête à 80%. Son diagnostic est possible à 70%.

5.5.2 Vérification des diagnostics proposés

Les différents diagnostics retenus font l'objet d'une évaluation en terme de contradiction et d'incohérence afin d'évaluer l'exactitude des diagnostics proposés. Pour cela, une autre matrice de vérification est construite pour cette fin. Il s'agit d'une matrice binaire dont la colonne contradiction est évaluée à 1 si un diagnostic est estimé contradictoire ou à 0 dans le cas contraire. De la même manière, la colonne cohérence est évaluée à 1 si le diagnostic est cohérent avec la description du cas et à 0 sinon. Un extrait d'une telle matrice est représentée par le tableau 5.9 évaluant les différents diagnostics proposés.

TABLE 5.9 – Extrait d'une matrice de vérification de la contradiction et de l'incohérence des diagnostics potentiels

Diagnosics possibles	Contradiction	Cohérence
cas1	0	1
cas2	0	1
cas3	1	0
...
cask	0	1

A partir de cette matrice, deux mesures sont calculées :

- Le taux de diagnostics contradictoires constituant la moyenne des valeurs de la colonne contradiction de la matrice.
- Le taux de diagnostics cohérents constituant la moyenne de la colonne cohérence de la matrice.

Une réorganisation des résultats est effectuée ordonnant les k diagnostics retenus par rapport à leurs contradiction et cohérence. Une autre évaluation est faite par l'expert pendant son diagnostic. Il attribue une valeur binaire à chaque diagnostic potentiel proposé évaluant son exactitude : une valeur 1 est attribuée à un diagnostic estimé exact et une valeur 0 est attribuée dans le cas contraire. Un taux d'exactitude est calculé comme suit :

$$\tau_e x = (\text{Nombre des diagnostics estimés exacts})/N$$

Où : N est le nombre total de diagnostics potentiels retenus.

Le taux d'erreur est calculé ensuite par : $\tau_e r r = 1 - \tau_e x$ La figure 5.5 présente les différentes mesures calculées afin d'évaluer l'approche proposée pour le raisonnement à base des cas similaires.

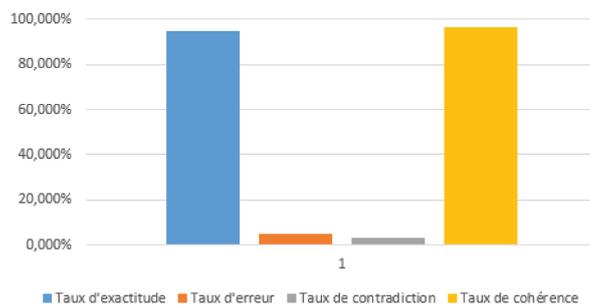


FIGURE 5.5 – Évaluation de performance de la phase du raisonnement

5.6 Evaluation globale

Une évaluation empirique globale est effectuée par trois médecins spécialistes (un médecin spécialiste en cancérologie et deux médecins gynécologues) et trois médecins radiologues. Ces experts du domaine médical présentent leurs avis concernant l'approche dans sa globalité et une appréciation entre 0 et 3 selon l'utilité et l'efficacité de la solution. Un 0 indique une mauvaise solution inutile pour le médecin. 3 réfère à une excellente solution pouvant apporter une aide considérable au médecin.

Les appréciations des médecins sont :

- **Gynécologue 1** : *"La solution est utile dans certains cas seulement ; j'ai une autre lecture des données radiologiques"*
- **Spécialiste en cancérologie** : *"L'étude est très intéressante surtout pour les radiologues. La modélisation symbolique est très utile la lecture du cas. Nous espérons que les travaux seront étendus pour les autres moyens d'investigation."*
- **Radiologue 1** : *"Le travail est très intéressant et utile. Il reproduit le raisonnement effectué par le médecin. Nous souhaitons avoir les causes possibles d'une pathologie observée avec une catégorisation par fréquence (causes fréquentes, causes moins fréquentes, causes rares)"*.
- **Radiologue 2** : *"Travail bien fait et utile pour le diagnostic. L'étude doit être plus approfondie sur plusieurs autres aspects radiologiques."*

Le tableau 5.10 présente les notes attribuées par les experts à notre proposition d'un système d'aide au diagnostic du cancer du sein basé sur la fusion des images multimodales. La moyenne des notes attribuées est de 3.25 sur 5. Cette note prouve l'utilité et l'apport de notre solution pour l'aide au diagnostic pour les médecins. Cependant, selon les experts du domaine, une amélioration de certains aspects est nécessaires (Ex. prise en considération d'autres aspects radiologiques, prise en considération d'autres moyens d'investigation, etc...).

TABLE 5.10 – Évaluation de l’approche proposée par les experts du domaine médical

Expert	Note attribuée
Médecin Gynécologue	3.5
Spécialiste en cancérologie	3.5
Radiologue 1	4
Radiologue 2	2

5.7 Conclusion

Ce chapitre avait pour objectif de présenter les différents tests effectués pour évaluer chaque module de l’approche que nous avons proposée au chapitre précédent. Cette évaluation est effectuée en se servant de ressources réelles : images traitées et ontologie de domaine utilisée.

Nous avons utilisé les mesures d’évaluation les plus utilisées dans la littérature dans la plupart des travaux traitant le traitement des images, la fusion, et le raisonnement : le taux d’erreur, la précision, la spécification, etc... Une évaluation globale a été effectuée auprès des experts du domaine de radiologie et des médecins spécialistes qui donnaient un avis très favorable pour la solution proposée.

Conclusion générale et perspectives

L'avènement des nouvelles technologies de l'imagerie et l'émergence de la radiologie médicale, ont contribué à l'évolution du diagnostic médical par l'exploration des parties cachées du corps humain, afin de faire apparaître les causes pathologiques d'une maladie. Cependant, la multitude et la diversité des sources d'information médicales rend la tâche de la prise de décision fastidieuse à l'expert. Cette difficulté est due, d'une part à la complexité de la situation étudiée, et d'autre part à la contamination des informations par différentes formes d'imperfections (redondance, ambiguïté, imprécision, etc...). Pour ces raisons, les médecins ont recours à des systèmes d'analyse et d'interprétation d'images médicales et d'aide au diagnostic, pouvant leur apporter l'assistance requise dans leur processus de prise de décision.

Dans ce contexte, et à la croisée des domaines : de l'imagerie médicale du sein, du traitement des images et de l'ingénierie des connaissances, nous nous sommes intéressés, dans cette thèse, à la problématique de la fusion d'images mammaires multimodales dans l'objectif de trouver une combinaison possible des différentes informations utiles dans les deux supports d'images en tenant compte de leur hétérogénéité, en vue de proposer au médecin une aide au diagnostic et à la prise de décision.

Dans cet objectif, nous avons proposé une approche multi-niveaux de traitement permettant de passer de l'espace bas niveau de la représentation des données sous forme de pixels sur les supports des images, à un espace haut niveau de représentation des connaissances proposant des interprétations de la scène observée (structure interne du sein). Nos contributions sont réparties sur cinq volets :

- **Volet de modélisation conceptuelle des connaissances** : par la transformation d'une image mammaire vers une ontologie locale possibiliste.
- **Volet de fusion de connaissances** : par l'évaluation de la similarité entre deux ontologies locales possibilistes et la fusion des différents concepts et relations.
- **Volet d'enrichissement d'une ontologie de domaine** : par le peuplement d'une ontologie de domaine mammaire à travers des instances possibilistes.
- **Volet de raisonnement et aide au diagnostic** : par la proposition d'un raisonnement possibiliste à base des cas permettant de fournir les diagnostics les plus possibles pour un certain cas requête.
- **Volet de la modélisation virtuelle des connaissances** : par la produc-

tion d'un modèle virtuel reproduisant une scène observée de façon symbolique afin de présenter au médecin une vision simplifiée de la situation étudiée.

La complémentarité du contenu informationnel dans les différents supports d'imagerie captive les experts et les amène à utiliser plusieurs sources simultanément. La nécessité de fusionner les différentes connaissances est, alors, d'une forte importance. Leur hétérogénéité requiert une modélisation conceptuelle exhaustive et fiable afin de mener à bien leur combinaison. L'ontologie est un paradigme, pratiquement récent, appliqué dans plusieurs domaines afin de fournir une spécification conceptuelle de leurs connaissances. C'est un outil puissant permettant de combler le fossé entre la description insuffisante des caractéristiques de bas niveau extraites de l'image et la richesse de l'interprétation des interprétations sémantiques effectuées par les experts des domaines. Les différentes imperfections affectant l'information médicale représentent la complexité majeure affrontée dans un tel traitement. Faire face à cette situation, évoquant un embarras notable, est une tâche cruciale. La théorie des possibilités constitue un cadre mathématique vigoureux dans la modélisation des connaissances ambiguës dans un cadre incertain. La puissance de cet outil réside dans sa capacité à modéliser plusieurs types d'imperfections.

Ainsi, coupler le paradigme d'ontologie et la théorie des possibilités tirant profit de leurs avantages pour modéliser le contenu informationnel des images mammaires représente une contribution très puissante dans l'objectif de trouver une représentation normalisée des connaissances dans le but de leur fusion. Cette solution permet de dépasser la limite du recalage des images et des formes géométriques : chose très complexe pour les images mammaires multimodales.

Dans nos travaux nous avons opté pour une solution basée sur l'exploitation d'une ontologie de domaine existante afin de construire l'ontologie locale à l'image. Le manque d'ontologies instanciées du domaine mammographique nous a incité à proposer une méthode pour l'enrichissement de l'ontologie de domaine utilisée en vue de son peuplement à partir de l'ontologie résultante de la fusion des ontologies locales.

Nous tirons profit de l'ontologie de domaine, d'une part, pour guider la construction de l'ontologie locale à partir des caractéristiques extraites des images ; et d'autre part pour guider un raisonnement à base de cas appliqué afin de présenter l'aide requise au médecin. Le raisonnement à base de cas constitue le processus cognitif basique suivi par l'expert pour sa prise de décision. Pour cette raison, nous avons opté pour une solution permettant de présenter les différents diagnostics possibles pour un cas étudié en se basant sur les cas qui lui sont similaires. Finalement, la modélisation virtuelle simplifiée de la scène réelle sert de complément au rapport diagnostic issu du raisonnement effectué. Cette représentation virtuelle est utile dans la mesure où elle forme un aperçu réduit de la situation ce qui peut assister le médecin dans d'éventuels traitements (chirurgie, biopsie, etc...)

Plusieurs perspectives sont abordables à partir de ces travaux. D'un aspect spécifique, ces perspectives se rapportent à la méthodologie de la mise en oeuvre de cette approche dans le domaine mammographique. D'un aspect plus général, elles se rapportent à l'approche dans sa globalité vue d'un niveau d'abstraction élevé.

Ceci revient à grouper ces perspectives en deux catégories : des perspectives à court terme et des perspectives à long terme.

A court terme, notre méthode peut envisager d'autres sources d'information : d'autres supports d'images, des supports textuels relatifs aux rapports des données asymptotiques d'un patient. Ceci nous amène à complètement automatiser la segmentation des informations intéressantes à partir des différents supports. A un autre niveau, il est intéressant d'automatiser la spécification des seuils utilisés pour l'application de la théorie des possibilités afin de définir l'ontologie locale ou spécifier le diagnostic possible.

A long terme, notre approche n'est pas limitée au domaine du traitement des images mammaires, elle peut être étendue pour être appliquée dans d'autres domaines d'application avec d'autres types d'images ou même d'autres types de supports. Ceci revient à dire que l'approche proposée est extensible pour être appliquée dans un cadre d'interprétation de scènes observées par multiples capteurs.

Liste des publications

I-Articles dans des conférences

Yosra BEN SALEM, Kamel HAMROUNI, and Basel SOLAIMAN, *Image Fusion Models and Techniques at pixel level*, IEEE International Image Processing Applications And Systems Conference (2016), (**Publié**)

BEN SALEM Yosra, IDOUDI Rihab, HAMROUNI Kamel, SOLEIMAN Basel and BOUSETTA Sana, *Image Based Mammographic Ontology Learning*, 11th International Conference on Intelligent Systems : Theories and Applications (2016), (**Publié**)

Yosra BEN SALEM , Rihab IDOUDI, Karim SAHEB ETTABAA, Kamel HAMROUNI, Basel SOLEIMAN, *Ontology based possibilistic reasoning for breast cancer aided diagnosis*, European, Mediterranean and Middle Eastern Conference on Information Systems (EMCIS) (Conférence Classe B) (**Publié**)

II-Articles dans des journaux

Yosra Ben Salem, Rihab Idoudi, Kamel Hamrouni, Karim Saheb Ettabaa, and Basel Soulaïman, *High Level Mammographic Information Fusion For Real World Ontology Population*, Journal of Digital Information Management (JDIM) (indexé scopus) (**Publié**)

Bibliographie

- [A. Branici, 2010] A. Branici, A. (2010). *Représentation et raisonnement formels pour le pronostic basé sur l'imagerie médicale microscopique. Application à la graduation du cancer du sein*. PhD thesis, Université de Franche-Comté. (Cité en page 71.)
- [Aamodt and Plaza, 1994] Aamodt, A. and Plaza, E. (1994). Case-based reasoning : Foundational issues, methodological variations, and system approaches. *AI communications*, 7(1) :39–59. (Cité en page 54.)
- [Aarthy and Prabu, 2016] Aarthy, S. and Prabu, S. (2016). A computerized approach on breast cancer detection and classification. *IIOAB JOURNAL*, 7(5) :157–169. (Cité en page 18.)
- [Abdrabou and Salem, 2010] Abdrabou, E. A. M. L. and Salem, A.-B. M. (2010). A breast cancer classifier based on a combination of case-based reasoning and ontology approach. In *Computer science and information technology (imcsit), proceedings of the 2010 international multiconference on*, pages 3–10. IEEE. (Cité en page 56.)
- [Abdul et al., 2014] Abdul, M., Muhammad, A. M., Mustapha, N., Muhammad, S., and Ahmad, N. (2014). Database workload management through cbr and fuzzy based characterization. *Applied Soft Computing*, 22 :605–621. (Cité en page 57.)
- [Abirami et al., 2016] Abirami, C., Harikumar, R., and Chakravarthy, S. S. (2016). Performance analysis and detection of micro calcification in digital mammograms using wavelet features. In *Wireless Communications, Signal Processing and Networking (WiSPNET), International Conference on*, pages 2327–2331. IEEE. (Cité en page 17.)
- [Ackoff, 1989] Ackoff, R. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, 16 :3–9. (Cité en pages vii, 30 et 31.)
- [Al-Faris et al., 2015] Al-Faris, A. Q., Isa, N. A. M., Ngah, U. K., and Shuaib, I. L. (2015). Automatic exclusion of skin border regions from breast mri using proposed combined approach. In *Biomedical Engineering (ICoBE), 2015 2nd International Conference on*, pages 1–6. IEEE. (Cité en page 23.)
- [Alsun et al., 2010] Alsun, M. H., Lecornu, L., Solaiman, B., Le Guillou, C., and Cauvin, J. M. (2010). Medical diagnosis by possibilistic classification reasoning. In *Information Fusion (FUSION), 2010 13th Conference on*, pages 1–7. IEEE. (Cité en page 57.)

- [Alunni, 2012] Alunni, J.-P. (2012). Imagerie du cancer du sein inflammatoire. *Journal de Radiologie diagnostique et interventionnelle*, 93(2) :103–111. (Cité en page 20.)
- [Andaloussi et al., 2009] Andaloussi, S., Lamard, M., Cazuguel, G., Tairi, H., Meknassi, M., Cochener, B., and Roux, C. (2009). Recherche d’images médicales par leur contenu numérique : utilisation de signatures construites à partir de la bemd et la transforme de hilbert. In *Journées de Recherche en Imagerie et Technologies de la Santé*. (Cité en page 38.)
- [A.Umameswari and G.R.Suresh, 2014] A.Umameswari and G.R.Suresh (2014). A new cryptographic digital signature for secure medical image communication in telemedicine. *International Journal of Computer Applications*, 86(11) :4–9. (Cité en page 38.)
- [Azar et al., 2000] Azar, F. S., Metaxas, D. N., and Schnall, M. D. (2000). A finite element model of the breast for predicting mechanical deformations during biopsy procedures. In *Mathematical Methods in Biomedical Image Analysis, 2000. Proceedings. IEEE Workshop on*, pages 38–45. IEEE. (Cité en page 44.)
- [Baharuddin et al., 2016] Baharuddin, W. N. A., Abdullah, S. N. H. S., Sahran, S., Qasem, A., bin Abdullah, A., Iqbal, R., and Ismail, F. (2016). Type 2 fuzzy logic for mammogram breast tissue classification. In *Industrial Informatics and Computer Systems (CIICS), 2016 International Conference on*, pages 1–6. IEEE. (Cité en page 16.)
- [Baum et al., 2011] Baum, K. G., Schmidt, E., Rafferty, K., Król, A., and Helguera, M. (2011). Evaluation of novel genetic algorithm generated schemes for positron emission tomography (pet)/magnetic resonance imaging (mri) image fusion. *Journal of digital imaging*, 24(6) :1031–1043. (Cité en page 53.)
- [Bekker et al., 2016] Bekker, A. J., Greenspan, H., and Goldberger, J. (2016). A multi-view deep learning architecture for classification of breast microcalcifications. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pages 726–730. IEEE. (Cité en page 16.)
- [Bekker et al., 2015] Bekker, A. J., Shalhon, M., Greenspan, H., and Goldberger, J. (2015). Learning to combine decisions from multiple mammography views. In *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, pages 97–100. IEEE. (Cité en page 16.)
- [Bierly et al., 2000] Bierly, P., Kessle, E., and Christensen, E. (2000). Organizational learning, knowledge and wisdom. *Journal of Organizational Change Management*, 13(6) :595–618. (Cité en page 30.)
- [Bliznakova et al., 2012] Bliznakova, K., Sechopoulos, I., Buliev, I., and Pallikarakis, N. (2012). Breastsimulator : A software platform for breast x-ray imaging research. *Journal of Biomedical Graphics and Computing*, 2(1) :1. (Cité en pages vii et 45.)
- [Bloche and Maitre, 2004] Bloche, I. and Maitre, H. (2004). Les méthodes de raisonnement dans les images. (Cité en pages 31 et 37.)

- [Bonnie et al., 2016] Bonnie, J., Elissa, P., Brett, P., and the Breast Screening Leadership Group (2016). Screening in the 40-49 age group. In *SBI News*, number 1, pages 12–14. 45. The Society of Breast Imaging. (Cité en page 9.)
- [Bossé et al., 2005] Bossé, E., Joussetme, A., and Maupin, P. (2005). Knowledge, uncertainty and belief in information fusion and situation analysis. In *Data Fusion for Situation Monitoring, Incident Detection, Alert and Response Management*, 198(3) :61–80. (Cité en page 30.)
- [Boström et al., 2007] Boström, H., Andler, S. F., Brohede, M., Johansson, R., Karlsson, A., Van Laere, J., Niklasson, L., Nilsson, M., Persson, A., and Ziemke, T. (2007). On the definition of information fusion as a field of research. (Cité en page 48.)
- [Brown, 1981] Brown, C. M. (1981). Some mathematical and representational aspects of solid modeling. *IEEE transactions on pattern analysis and machine intelligence*, (4) :444–453. (Cité en page 84.)
- [Bulu et al., 2012] Bulu, H., Alpkocak, A., and Balci, P. (2012). Ontology-based mammography annotation and case-based retrieval of breast masses. *Expert Systems with Applications*, 39(12) :11194–11202. (Cité en page 71.)
- [Bulu et al., 2013] Bulu, H., Alpkocak, A., and Balci, P. (2013). Uncertainty modeling for ontology-based mammography annotation with intelligent bi-rads scoring. *Computers in Biology and Medicine*, 43(4) :301 – 311. (Cité en page 41.)
- [Carter, 1999] Carter, J. H. (1999). *Clinical decision support systems*. Springer. (Cité en page 37.)
- [Castro et al., 2011] Castro, J. L., Navarro, M., Sánchez, J. M., and Zurita, J. M. (2011). Introducing attribute risk for retrieval in case-based reasoning. *Knowledge-Based Systems*, 24(2) :257–268. (Cité en page 53.)
- [Chaffey and Wood, 2005] Chaffey, D. and Wood, S. (2005). *Business Information Management : Improving Performance Using Information Systems*. Pearson Education Limited. (Cité en page 30.)
- [Chevret, 2007] Chevret, S. (2007). Modèles mathématiques utilisés en médecine. *Réanimation*, 16(3) :240–244. (Cité en pages 37 et 42.)
- [Comas, 2005] Comas, L. (2005). *Modèles et algorithmes pour la scintigraphie cardiaque*. PhD thesis, Besançon. (Cité en page 43.)
- [Darzi et al., 2011] Darzi, M., AsgharLiaei, A., Hosseini, M., et al. (2011). Feature selection for breast cancer diagnosis : a case-based wrapper approach. *World Academy of Science, Engineering and Technology, International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering*, 5(5) :220–223. (Cité en page 55.)
- [Dasarathy, 1997] Dasarathy, B. V. (1997). Sensor fusion potential exploitation-innovative architectures and illustrative applications. *Proceedings of the IEEE*, 85(1) :24–38. (Cité en page 49.)

- [Davenport and Prusack, 1998] Davenport, T. and Prusack, L. (1998). *Working Knowledge : How Organizations Manage What They Know*. Harvard Business School Press. (Cité en page 31.)
- [D.B., 1996] D.B., L. (1996). Case-based reasoning : experiences, lessons, and future directions. (Cité en page 54.)
- [de Cea and Yang, 2016] de Cea, M. V. S. and Yang, Y. (2016). Case-based decision strategy using outlier probability in detection of microcalcifications in mammographic lesions. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 3409–3413. IEEE. (Cité en page 15.)
- [Dempster, 1967] Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *The annals of mathematical statistics*, pages 325–339. (Cité en page 34.)
- [Denizot, 2007] Denizot, M. (2007). La théorie de la signature des plantes et ses implications. *Bulletin. Ac. Science. et Lettres de Montpellier*, 37 :205–216. (Cité en page 38.)
- [Desodt-Lebrun, 1996] Desodt-Lebrun, A. (1996). Fusion de données. *Techniques de l'ingénieur. Informatique industrielle*, 12(R7226) :1–9. (Cité en page 49.)
- [Dhanalekshmi and Phadke, 2013] Dhanalekshmi, P. S. M. and Phadke, A. C. (2013). Classification of circular and lobulated masses in mammograms using fractal analysis. pages 1219–1221. (Cité en page 39.)
- [Dubois et al., 2007] Dubois, D., Esteva, F., Godo, L., and Prade, H. (2007). Fuzzy-set based logics-an history-oriented presentation of their main developments. *The Many Valued and Nonmonotonic Turn in Logic*, 8 :325–449. (Cité en page 36.)
- [Dubois and Kalfsbeek, 1990] Dubois, D. and Kalfsbeek, H. (1990). Elicitation, assessment and pooling of expert judgement using possibility theory. In *Proc. of the 8th Inter. Congress of Cybernetics and Systems*, pages 11–15. New York. (Cité en page 34.)
- [Dubois and Prade, 1988] Dubois, D. and Prade, H. (1988). Theory of possibility an approach to computerized processing of uncertainty. (Cité en page 35.)
- [Dubois and Prade, 1994] Dubois, D. and Prade, H. (1994). La fusion d'informations imprécises. *TS. Traitement du signal*, 11(6) :447–458. (Cité en page 78.)
- [Dutang and Contamine, 2005] Dutang, M. and Contamine, J. (2005). La modélisation un outil au service de la recherche. 0 :1–6. (Cité en page 37.)
- [Ekong et al., 2012] Ekong, V. E., Inyang, U. G., and Onibere, E. A. (2012). Intelligent decision support system for depression diagnosis based on neuro-fuzzy-cbr hybrid. *Modern Applied Science*, 6(7) :79. (Cité en page 57.)
- [El-Sappagh et al., 2015] El-Sappagh, S., Elmogy, M., and Riad, A. (2015). A fuzzy-ontology-oriented case-based reasoning framework for semantic diabetes diagnosis. *Artificial intelligence in medicine*, 65(3) :179–208. (Cité en pages 54 et 58.)
- [Ertosun and Rubin, 2015] Ertosun, M. G. and Rubin, D. L. (2015). Probabilistic visual search for masses within mammography images using deep learning. In

- Bioinformatics and Biomedicine (BIBM)*, 2015 IEEE International Conference on, pages 1310–1315. IEEE. (Cité en page 16.)
- [Fabre, 1999] Fabre, S. (1999). Apport de l'information contextuelle à la fusion multicapteurs : application à la fusion pixel. (Cité en page 49.)
- [Freund and Schapire, 1997] Freund, Y. and Schapire, R. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1) :119 – 139. (Cité en page 18.)
- [Garrido et al., 2008] Garrido, J. L., Hurtado, M. V., Noguera, M., and Zurita, J. M. (2008). Using a cbr approach based on ontologies for recommendation and reuse of knowledge sharing in decision making. In *Hybrid Intelligent Systems, 2008. HIS'08. Eighth International Conference on*, pages 837–842. IEEE. (Cité en page 56.)
- [Gnonnou and Smaoui, 2014] Gnonnou, C. and Smaoui, N. (2014). Segmentation and 3d reconstruction of mri images for breast cancer detection. In *Image Processing, Applications and Systems Conference (IPAS), 2014 First International*, pages 1–6. IEEE. (Cité en page 24.)
- [Gong et al., 2013] Gong, B., Liu, J., Wang, X., and Tang, X. (2013). Learning semantic signatures for 3D object retrieval. *IEEE Transactions on Multimedia*, 15(2) :369–377. (Cité en page 39.)
- [Gosselin, 2005] Gosselin, P. (2005). *Méthodes d'apprentissage pour la recherche de catégories dans des bases d'images*. phdthesis, Université de Cergy-Pontoise. (Cité en page 38.)
- [Goubalan et al., 2016] Goubalan, S. R., Goussard, Y., and Maaref, H. (2016). Un-supervised malignant mammographic breast mass segmentation algorithm based on pickard markov random field. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 2653–2657. IEEE. (Cité en page 14.)
- [Gowsikhaa et al., 2012] Gowsikhaa, D., Abirami, S., and Baskaran, R. (2012). Construction of image ontology using low-level features for image retrieval. In *International Conference on Computer Communication and Informatics (ICCCI)*. (Cité en page 40.)
- [Guarino, 1997] Guarino, N. (1997). Some organizing principles for a unified top-level ontology. In *AAAI Spring Symposium on Ontological Engineerin*, pages 57–63. (Cité en page 40.)
- [Gubern-Merida et al., 2015] Gubern-Merida, A., Kallenberg, M., Mann, R. M., Marti, R., and Karssemeijer, N. (2015). Breast segmentation and density estimation in breast mri : a fully automatic framework. *ieee journal of biomedical and health informatics*, 19(1) :349–357. (Cité en page 24.)
- [H. Briand, 2008] H. Briand, S. L. (2008). *Modélisation des connaissances*. Revue des Nouvelles Technologies de l'Information. (Cité en page 37.)
- [Han et al., 2014] Han, L., Hipwell, J. H., Eiben, B., Barratt, D., Modat, M., Ourselin, S., and Hawkes, D. J. (2014). A nonlinear biomechanical model based

- registration method for aligning prone and supine mr breast images. *IEEE transactions on medical imaging*, 33(3) :682–694. (Cité en page 46.)
- [Heriana and Soesanti, 2015] Heriana, O. and Soesanti, I. (2015). Tumor size classification of breast thermal image using fuzzy c-means algorithm. In *Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET), 2015 International Conference on*, pages 98–103. IEEE. (Cité en page 16.)
- [Heron, 2009] Heron, J. (2009). La mammographie. techreport, Faculté de Médecine de Caen - France. (Cité en pages 10 et 11.)
- [Hertz, 2006] Hertz, T. (2006). Learning distance functions : Algorithms and applications. (Cité en page 75.)
- [Hong et al., 2014] Hong, S., Qian, Z., Feifei, S., Jiandong, W., Quansheng, W., Jingdan, Q., and Deqiang, K. (2014). Breast tissue segmentation on mr images using kfcM with spatial constraints. In *Granular Computing (GrC), 2014 IEEE International Conference on*, pages 254–258. IEEE. (Cité en page 25.)
- [Hopp et al., 2012a] Hopp, T., Baltzer, P., Dietzel, M., Kaiser, W. A., and Ruiter, N. V. (2012a). 2d/3d image fusion of x-ray mammograms with breast mri : visualizing dynamic contrast enhancement in mammograms. *International journal of computer assisted radiology and surgery*, 7(3) :339–348. (Cité en page 50.)
- [Hopp et al., 2012b] Hopp, T., Duric, N., and Ruiter, N. (2012b). Image fusion of ultrasound computer tomography volumes with x-ray mammograms using a biomechanical model based 2d/3d registration. *Computerized Medical Imaging and Graphics*, 40 :170 – 181. (Cité en page 51.)
- [Hu et al., 2003] Hu, B., Dasmahapatra, S., and Shadbolt, N. (2003). From lexicon to mammographic ontology : Experiences and lessons. In *International Workshop of Description Logics-DL*. (Cité en page 71.)
- [Hullermeier et al., 2002] Hullermeier, E., Dubois, D., and Prade, H. (2002). Model adaptation in possibilistic instance-based reasoning. *IEEE Transactions on Fuzzy Systems*, 10(3) :333–339. (Cité en page 57.)
- [Idoudi, 2017] Idoudi, R. (2017). Fouille de connaissances en diagnostic mammographique. (Cité en page 39.)
- [Idoudi et al., 2016] Idoudi, R., Ettabaa, K. S., Solaiman, B., and Mnif, N. (2016). Association rules based ontology enrichment. *International journal of web applications*, 8(1) :16–25. (Cité en page 71.)
- [Idoudi et al., 2014] Idoudi, R., Hamrouni, K., and Solaiman, B. (2014). Ontological approach to mammographic knowledge representation. In *Advanced Technologies for Signal and Image Processing (ATSIP), 2014 1st International Conference on*, pages 31–34. IEEE. (Cité en page 71.)
- [Ilyass et al., 2015] Ilyass, H., Mohamed, F., Diop, I., and Tarik, J. (2015). Ontology-based mammography annotation for breast cancer diagnosis. In *2nd World Symposium on Web Applications and Networking (WSWAN)*, pages 1–6. (Cité en page 41.)

- [Iskandar et al., 2007a] Iskandar, D., Thom, J., and Tahaghoghi, S. (2007a). Content-based image retrieval using image regions as query examples. In *Proceedings of the Nineteenth Conference on Australasian Database*, volume 75 of *ADC '08*, pages 38–46, Darlinghurst, Australia, Australia. Australian Computer Society, Inc. (Cité en page 41.)
- [Iskandar et al., 2007b] Iskandar, D., Thom, J., and Tahaghoghi, S. (2007b). Querying image ontology. In *Australasian Document Computing Symposium*. (Cité en page 41.)
- [Jacod and Protter, 2003] Jacod, J. and Protter, P. (2003). *L'essentiel en théorie des probabilités*. Cassini. (Cité en page 33.)
- [Jalaguier-Coudray, 2009] Jalaguier-Coudray, A. (2009). L'irm mammaire. Technical report, Institut Curie- Paris - France. (Cité en page 21.)
- [Janssens, 2013] Janssens, J. H. M. (2013). Outlier selection and one-class classification. (Cité en page 15.)
- [Jenhani et al., 2007] Jenhani, I., Amor, N. B., Elouedi, Z., Benferhat, S., and Mellouli, K. (2007). Information affinity : A new similarity measure for possibilistic uncertain information. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 840–852. Springer. (Cité en page 75.)
- [Jouirou et al., 2015] Jouirou, A., Baâzaoui, A., Barhoumi, W., and Zagrouba, E. (2015). Curvelet-based locality sensitive hashing for mammogram retrieval in large-scale datasets. In *Proc. IEEE/ACS 12th Int. Conf. of Computer Systems and Applications (AICCSA)*, pages 1–8. (Cité en page 38.)
- [Kolmogorov, 1956] Kolmogorov, A. (1956). *Foundations of the Theory of Probability*. Chelsea Publishing Company - New York, 2 edition. (Cité en page 33.)
- [Kong et al., 2008] Kong, G., Xu, D. L., and Yang, J. B. (2008). Clinical decision support systems : a review on knowledge representation and inference under uncertainties. 1(2) :159–167. (Cité en page 37.)
- [Krishnamurthy and Kesavan, 2015] Krishnamurthy, L. and Kesavan, J. (2015). An improved content based image retrieval using three region colour and straight line signatures of the image. In *Proc. Int. Conf. Communications and Signal Processing (ICCSP)*, pages 1919–1924. (Cité en page 38.)
- [KW A, 2011] KW A, A. A. (2011). Development of a realistic computational breast phantom for dosimetric simulations. *Nucl Sci Tech*, 2 :147–52. (Cité en pages vii, 44 et 45.)
- [Lamine et al., 2014] Lamine, F. B., Kalti, K., and Romdhane, L. B. (2014). A content-based digital mammography retrieval using inexact graph matching. In *Proc. Applications and Systems Conf. Int. Image Processing*, pages 1–4. (Cité en page 38.)
- [Lee et al., 1992] Lee, S., Yang, M., and Chen, J. (1992). Signature file as a spatial filter for iconic image database. 3 :373–397. (Cité en page 38.)

- [Lesniak et al., 2012] Lesniak, J., van Schie, G., Tanner, C., Platel, B., Huisman, H., Karssemeijer, N., and Székely, G. (2012). Multimodal classification of breast masses in mammography and mri using unimodal feature selection and decision fusion. *Breast Imaging*, pages 88–95. (Cité en page 52.)
- [Li et al., 2015] Li, S., Liu, H., Guo, D., Zheng, Y., Gee, J., and Zhao, Z. (2015). Constraint split bregman method combined with total variation for structural decomposition of breast dce-mri. In *Computer Science and Network Technology (ICCSNT), 2015 4th International Conference on*, volume 1, pages 1306–1310. IEEE. (Cité en page 24.)
- [Liaqat, 2013] Liaqat, M. (2013). Image classification and retrieval based on crisp and fuzzy ontology. In *International Conference on Computer, Control & Communication (IC4)*. (Cité en page 41.)
- [Liu et al., 2010] Liu, X., Shao, Z., and Liu, J. (2010). Ontology-based image retrieval with sift features. In *2010 First International Conference on Pervasive Computing, Signal Processing and Applications*, pages 464–467. (Cité en page 41.)
- [Luc, 2004] Luc, D. (2004). La classification bi-rads de l’acr. Technical report. (Cité en page 11.)
- [Ma et al., 2009] Ma, A. K., Gunn, S., and Darambara, D. G. (2009). Introducing debra : A detailed breast model for radiological studies. *Physics in medicine and biology*, 54(14) :4533. (Cité en pages 43 et 44.)
- [Makandar and Halalli, 2015] Makandar, A. and Halalli, B. (2015). Combined segmentation technique for suspicious mass detection in mammography. In *Trends in Automation, Communications and Computing Technology (I-TACT-15), 2015 International Conference on*, volume 1, pages 1–5. IEEE. (Cité en page 13.)
- [Makandar and Halalli, 2016] Makandar, A. and Halalli, B. (2016). Pre-processing of mammography image for early detection of breast cancer. *Int J Comput Appl*, pages 0975–8887. (Cité en page 14.)
- [Maken and Bradley, 2015] Maken, F. A. and Bradley, A. P. (2015). Multiple instance learning for breast mri based on generic spatio-temporal features. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pages 902–906. IEEE. (Cité en page 25.)
- [Mangolini, 1994] Mangolini, M. (1994). Apport de la fusion d’images satellitaires multicapteurs au niveau pixel en télédétection et photo-interprétation. (Cité en page 49.)
- [Melese, 1968] Melese, J. (1968). *La gestion par les systèmes : essai de praxéologie*. Editions hommes et techniques. (Cité en page 35.)
- [Mina and Isa, 2015] Mina, L. M. and Isa, N. A. M. (2015). Breast abnormality detection in mammograms using artificial neural network. In *Computer, Communications, and Control Technology (I4CT), 2015 International Conference on*, pages 258–263. IEEE. (Cité en page 18.)

- [Minavathi et al., 2013] Minavathi, Murali, S., and Dinesh, M. (2013). Information fusion from mammogram and ultrasound images for better classification of breast mass. In *Computerized Medical Imaging and Graphics*, pages 943–953. Springer India. (Cité en page 51.)
- [Moron, 1996] Moron, V. (1996). *Mise en correspondance de données 3D avec un modèle CAO : application à l'inspection automatique*. PhD thesis, Lyon, INSA. (Cité en page 84.)
- [Mouchaweh et al., 2006] Mouchaweh, M., Bouguelid, M., Billaudel, P., and Riera, B. (2006). Variable probability-possibility transformation. In *25th European Annual Conference on Human Decision-Making and Manual Control (EAM'06), September*, pages 27–29. Citeseer. (Cité en page 34.)
- [N. Angayarkanni, 2016] N. Angayarkanni, D. Kumar, G. A. (2016). The application of image processing techniques for detection and classification of cancerous tissue in digital mammograms. *Journal of Pharmaceutical Sciences and Research*, 8(10) :1179–1183. (Cité en page 15.)
- [Osher and Sethian, 1988] Osher, S. and Sethian, J. A. (1988). Fronts propagating with curvature-dependent speed : algorithms based on hamilton-jacobi formulations. *Journal of computational physics*, 79(1) :12–49. (Cité en page 65.)
- [Pamart et al., 2015] Pamart, A., Buglio, D. L., and Luca, L. D. (2015). Morphological analysis of shape semantics from curvature-based signatures. In *Proc. Digital Heritage*, volume 2, pages 105–108. (Cité en page 38.)
- [Pan et al., 2010] Pan, W., Coatrieux, G., Cuppens-Boulahia, N., Cuppens, F., and Roux, C. (2010). Medical image integrity control combining digital signature and lossless watermarking. *Data Privacy Management and Autonomous Spontaneous Security*, pages 153–162. (Cité en page 38.)
- [Paramkusham et al., 2015] Paramkusham, S., Shivakshit, P., Rao, K. M. M., and Rao, B. V. V. S. N. P. (2015). A new features extraction method based on polynomial regression for the assessment of breast lesion contours. In *Proc. Int. Conf. Industrial Instrumentation and Control (ICIC)*, pages 579–583. (Cité en page 39.)
- [Patil et al., 2016] Patil, R. N., Kadam, A. D., Vanjale, S. B., Thakore, D., and Joshi, S. (2016). A semi-supervised research approach to web-image re-ranking : Semantic image search engine. In *Proc. and Optimization Techniques (ICEEOT) 2016 Int. Conf. Electrical, Electronics*, pages 1883–1891. (Cité en page 39.)
- [Quellec et al., 2016] Quellec, G., Lamard, M., Cozic, M., Coatrieux, G., and Cazuguel, G. (2016). Multiple-instance learning for anomaly detection in digital mammography. *IEEE transactions on medical imaging*, 35(7) :1604–1614. (Cité en page 18.)
- [Raddaoui, 2013] Raddaoui, B. (2013). *Contributions aux approches logiques de l'argumentation en intelligence artificielle*. PhD thesis, Artois. (Cité en page 68.)

- [Rajagopal et al., 2007] Rajagopal, V. et al. (2007). *Modelling breast tissue mechanics under gravity loading*. PhD thesis, ResearchSpace Auckland. (Cité en page 44.)
- [Raufaste et al., 2003] Raufaste, E., da Silva Neves, R., and Mariné, C. (2003). Testing the descriptive validity of possibility theory in human judgments of uncertainty. *Artificial Intelligence*, 148(1-2) :197–218. (Cité en page 34.)
- [Richardson and Domingos, 2002] Richardson, M. and Domingos, P. (2002). The intelligent surfer : Probabilistic combination of link and content information in pagerank. *Advances in neural information processing systems*, 2 :1441–1448. (Cité en page 52.)
- [Riesbeck and Schank, 2013] Riesbeck, C. K. and Schank, R. C. (2013). *Inside case-based reasoning*. Psychology Press. (Cité en page 54.)
- [Ross et al., 2006] Ross, A. A., Jain, A., and Nandakumar, K. (2006). *Score Level Fusion*, volume 6, pages 91–142. Springer USv, Boston, MA. (Cité en page 49.)
- [Salis et al., 2010] Salis, J. R., Lagouarde, P., and Darmoni, S. (2010). Etude des systèmes d’aide à la décision médicale : Etat des lieux - partie 1. (Cité en page 60.)
- [Samani et al., 2001] Samani, A., Bishop, J., Yaffe, M. J., and Plewes, D. B. (2001). Biomechanical 3-d finite element modeling of the human breast using mri data. *IEEE Transactions on Medical Imaging*, 20(4) :271–279. (Cité en page 44.)
- [Samwald et al., 2015] Samwald, M., Giménez, J. A. M., Boyce, R. D., Freimuth, R. R., Adlassnig, K.-P., and Dumontier, M. (2015). Pharmacogenomic knowledge representation, reasoning and genome-based clinical decision support based on owl 2 dl ontologies. *BMC medical informatics and decision making*, 15(1) :12. (Cité en page 56.)
- [Saraswathi et al., 2016] Saraswathi, D., Dharani, D., and Srinivasan, E. (2016). An efficient feature extraction technique for breast cancer diagnosis using curvelet transform and swarm intelligence. In *Wireless Communications, Signal Processing and Networking (WiSPNET), International Conference on*, pages 441–445. IEEE. (Cité en page 17.)
- [Schank, 1983] Schank, R. C. (1983). *Dynamic memory : A theory of reminding and learning in computers and people*. cambridge university press. (Cité en page 53.)
- [Shafer, 1976] Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press, Princeton. (Cité en page 34.)
- [Shirazi and Rashedi, 2016] Shirazi, F. and Rashedi, E. (2016). Detection of cancer tumors in mammography images using support vector machine and mixed gravitational search algorithm. In *Swarm Intelligence and Evolutionary Computation (CSIEC), 2016 1st Conference on*, pages 98–101. IEEE. (Cité en page 16.)
- [Simand and Jolion, 2005] Simand, I. and Jolion, J. (2005). Représentation d’images par chaînes de symboles : Application à la recherche par le contenu. In *Colloque sur le traitement du signal et des images*, pages 37–40. GRETSI, Groupe d’Etudes du Traitement du Signal et des Images. (Cité en page 38.)

- [Singh et al., 2016] Singh, S. P., Urooj, S., and Lay-Ekuakille, A. (2016). Breast cancer detection using pcpct and adewnn : A geometric invariant approach to medical x-ray image sensors. *IEEE Sensors Journal*, 16(12) :4847–4855. (Cit  en page 17.)
- [Smaoui and Hlima, 2016] Smaoui, N. and Hlima, A. (2016). Designing a new approach for the segmentation of the cancerous breast mass. In *Systems, Signals & Devices (SSD), 2016 13th International Multi-Conference on*, pages 313–317. IEEE. (Cit  en page 15.)
- [Society, 2016] Society, A. C. (2016). Mammograms and other breast imaging tests. Technical report, American Cancer Society. (Cit  en page 12.)
- [Solaiman, 2001] Solaiman, B. (2001). Information fusion concepts : from information elements definition to the application of fusion approaches. In *Aerospace/Defense Sensing, Simulation, and Controls*, pages 205–212. International Society for Optics and Photonics. (Cit  en pages 32 et 49.)
- [Sournia, 1995] Sournia, J. (1995). *Histoire du diagnostic en m decine*.  ditions de Sant . (Cit  en page 59.)
- [Sushma and Kumar, 2015] Sushma, S. and Kumar, S. P. (2015). Image enhancement using bio-inspired algorithms on mammogram for cancer detection. In *Emerging Research in Electronics, Computer Science and Technology (ICERECT), 2015 International Conference on*, pages 11–16. IEEE. (Cit  en page 14.)
- [Swapnil et al., 2016] Swapnil, P., Pandey, E., Yathav, J. R., Baig, A., and Bailur, A. (2016). Region marking and grid based textural analysis for early identification of breast cancer in digital mammography. In *Advanced Computing (IACC), 2016 IEEE 6th International Conference on*, pages 426–429. IEEE. (Cit  en page 17.)
- [Szafarz, 1985] Szafarz, A. (1985). L’ volution du concept de probabilit  math matique de pascal   laplace. *Technologia*, 8(3) :67–75. (Cit  en page 29.)
- [Taylor and Toujilov, 2011] Taylor, P. and Toujilov, I. (2011). Mammographic knowledge representation in description logic. In *International Workshop on Knowledge Representation for Health Care*, pages 158–169. Springer. (Cit  en page 71.)
- [T.R.Gruber, 1993] T.R.Gruber (1993). A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2) :199–220. (Cit  en page 39.)
- [Tutac et al., 2008] Tutac, A. E., Racoceanu, D., Putti, T., Xiong, W., Leow, W. K., and Cretu, V. (2008). Knowledge-guided semantic indexing of breast cancer histopathology images. In *2008 International Conference on BioMedical Engineering and Informatics*, volume 2, pages 107–112. (Cit  en page 41.)
- [Valois, 1994] Valois, R. (1994). La d finition et la d monstration dans la logique d’ ristote. *Laval th ologique et philosophique*, 50(2) :349–361. (Cit  en page 48.)
- [Vedanarayanan and Nandhitha, 2016] Vedanarayanan, V. and Nandhitha, N. (2016). Suitability of gabor filter for denoising and contrast enhancement of digital mammograms. In *Science Technology Engineering and Management (ICONS-TEM), Second International Conference on*, pages 568–571. IEEE. (Cit  en page 13.)

- [Veltz, 2004] Veltz, F. (2004). Image mining : voir plus vite. *Defense Nationale-Paris-Etudes Politiques Strategiques Militaires Economiques Scientifiques*, 60 :101–112. (Cité en page 38.)
- [Verlinde, 1999] Verlinde, P. (1999). Contribution à la vérification multi-modale de l'identité en utilisant la fusion de décisions. (Cité en page 49.)
- [Wald, 2002] Wald, L. (2002). *Data Fusion - Definitions and architectures - Fusion of images of different spacial resolutions*. Presses de l'Ecole, Ecole des Mines de Paris. (Cité en page 49.)
- [Walley, 1996] Walley, P. (1996). Measures of uncertainty in expert systems. *Artificial intelligence*, 83(1) :1–58. (Cité en page 34.)
- [Wang et al., 2008] Wang, H., Jiang, X., Chia, L., and Tan, A. (2008). Ontology enhanced web image retrieval : Aided by wikipedia & spreading activation theory. In *Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval*, MIR '08, pages 195–201, New York, NY, USA. ACM. (Cité en page 40.)
- [Wang and Yang, 2016] Wang, J. and Yang, Y. (2016). Boosted classification of breast cancer by retrieval of cases having similar disease likelihood. In *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*, pages 908–911. IEEE. (Cité en page 17.)
- [Wang et al., 2016] Wang, J., Yang, Y., and Nishikawa, R. M. (2016). Quantitative study of image features of clustered microcalcifications in for-presentation mammograms. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 3404–3408. (Cité en page 17.)
- [Wang et al., 2014] Wang, X., Qiu, S., Liu, K., and Tang, X. (2014). Web image re-ranking usingquery-specific semantic signatures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(4) :810–823. (Cité en page 39.)
- [Wang and Yang, 2011] Wang, Y. and Yang, X. (2011). Image domain ontology fusion approach using multi-level inference mechanism. In *International Conference on Computer Science and Network Technology*. (Cité en page 41.)
- [Wessel et al., 2013] Wessel, C., Schnabel, J. A., and Brady, M. (2013). Realistic biomechanical model of a cancerous breast for the registration of prone to supine deformations. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 7249–7252. IEEE. (Cité en page 45.)
- [Wu and Sullivan, 2001] Wu, Z. and Sullivan, J. (2001). Automatic finite element mesh generation from mri scans for breast cancer investigations. In *Bioengineering Conference, 2001. Proceedings of the IEEE 27th Annual Northeast*, pages 59–60. IEEE. (Cité en page 44.)
- [Yang et al., 2016] Yang, S. C., He, Y. J., and Wun, Y. J. (2016). Designated target enhancement and segmentation in multi-spectral mr images. In *2016 International Symposium on Computer, Consumer and Control (IS3C)*, pages 1059–1062. (Cité en page 24.)

- [Yeh and Chang, 2006] Yeh, W. and Chang, Y. (2006). An efficient signature extraction method for image similarity retrieval. *22(1)* :63–94. (Cit  en page 38.)
- [Zadeh, 1965] Zadeh, L. (1965). Fuzzy sets. *Information and control*, 8(3) :338–353. (Cit  en page 35.)
- [Zadeh, 1978] Zadeh, L. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1(1) :3 – 28. (Cit  en page 35.)
- [Zeleny, 1987] Zeleny, M. (1987). Management support systems : towards integrated knowledge management. *Human Systems Management*, 7(1) :59–70. (Cit  en pages 30 et 31.)
- [Zhang et al., 2015a] Zhang, Bo, Peeters, and Hans (2015a). Efficient mr inhomogeneity correction by regularized entropy minimization and proximal alternations. In *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, pages 1069–1072. IEEE. (Cit  en page 23.)
- [Zhang et al., 2015b] Zhang, X., Dou, H., Ju, T., and Zhang, S. (2015b). Fusing heterogeneous features for the image-guided diagnosis of intraductal breast lesions. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 1288–1291. (Cit  en page 52.)
- [Zlatoff, 2006] Zlatoff, N. (2006). Indexation d’images 2d. vers une reconnaissance d’objets multi-crit res. (Cit  en page 38.)
- [Zyout and Togneri, 2016] Zyout, I. and Togneri, R. (2016). A new approach for the detection of architectural distortions using textural analysis of surrounding tissue. In *Proc. 38th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3965–3968. (Cit  en page 39.)



Résultats du pré-traitement des images mammaires

L'étape du prétraitement est une étape préparatrice de l'image au traitement ayant pour objectif de corriger certaines imperfections liées à la qualité des données visuelles sur l'image. Tout d'abord, une sélection automatique de la région d'intérêt est effectuée afin de supprimer tout objet externe au sein (Bruit, étiquette) et de considérer seulement la zone du sein. Ensuite, un rehaussement et un filtrage sont opérés afin d'augmenter le contraste de l'image et réduire le bruit. Cette étape prépare l'image à l'étape de la segmentation. Il est à noter que cette démarche est adoptée pour le traitement des deux images mammographie et IRM.

A.1 Prétraitement et segmentation des Mammographies

L'examen mammographique résulte généralement de deux images une image pour chaque côté. Cette image est caractérisée par un faible contraste et certaines imperfections liées principalement aux conditions de prise de l'image suite à la compression du sein et à la sensibilité du capteur utilisé. Généralement les images mammographiques sont étiquetées et présentent parfois des formes intruses au fond de l'image. Par conséquent une sélection préalable de la zone du sein, qui représente notre zone d'intérêt, est nécessaire pour la bonne conduite du reste du traitement. Cette tâche est effectuée en décomposant l'espace de l'image en des composantes compactes puis en sélectionnant la région de plus grande taille (le sein) et en éliminant les autres zones de l'arrière plan de l'image. Afin d'améliorer la qualité de l'image, nous avons essayé les techniques de filtrage les plus utilisés dans la littérature pour diminuer le bruit dans l'image et augmenter son contraste. La figure A.1 illustre la procédure adoptée pour le prétraitement des images mammographiques.

Les tableaux : Tableau A.1 et Tableau A.2 présentent les résultats des différents algorithmes appliqués pour l'étape du filtrage. En premier temps nous avons appliqué les filtres seuls sur l'image originale (Tableau A.1), ensuite nous avons appliqué une

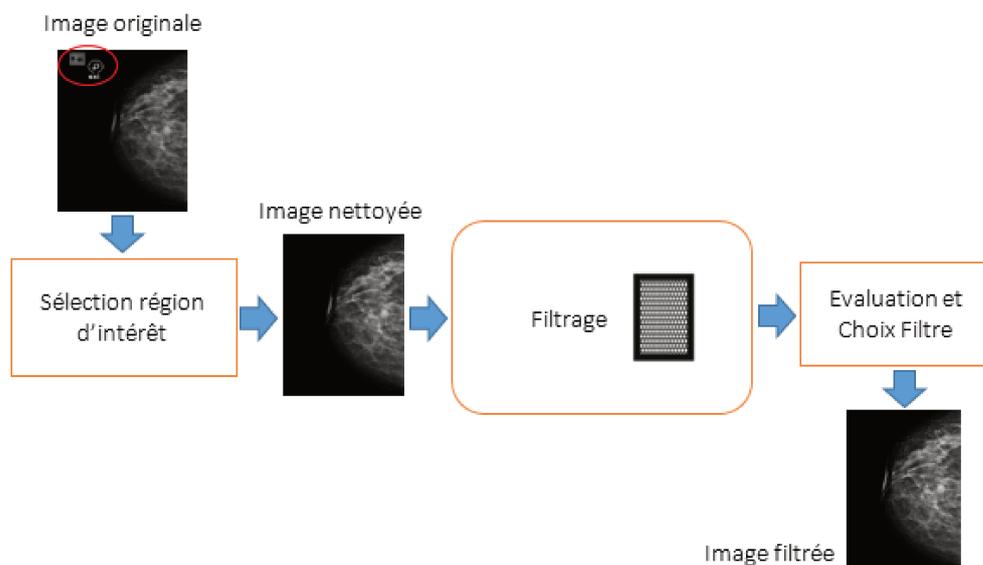


FIGURE A.1 – Procédure du prétraitement des images mammographiques

TABLE A.1 – Résultat des filtres les plus utilisés sur une image mammographique

Anisotropic Diffusion (AD)	Filtre Bilateral	Filtre Gaussian	Filtre Moyen	Filtre Median	Nonlinear Noise Reduction	Regularised Isotropic Diffusion

minimisation du caractère flou (Fuzzy Minimization (FM)) sur l'image originale puis nous avons appliqué les filtres sur l'image résultante (Tableau A.2).

Pour évaluer les différents résultats nous mesurons l'erreur quadratique moyenne (EQM) pour chaque filtre appliqué. Les différentes valeurs sont présentées dans les tableaux suivants : Pour les deux filtres Anisotropic Diffusion (AD) et Regularised Isotropic Diffusion (RID), l'algorithme met un temps très long pour effectuer un filtrage sur une image à laquelle nous avons déjà appliqué une minimisation du flou.

D'après les mesures trouvées, l'algorithme RID (Regularised Isotropic Diffusion) présente l'EQM la plus petite. En analyse d'évaluation quantitative cette technique est la meilleure pour le traitement de l'image. Une analyse d'évaluation qualitative est également effectuée auprès d'un expert afin d'évaluer la qualité de l'image. Pour

TABLE A.2 – Résultats du filtrage par certaines techniques avec une étape de minimisation du caractère flou

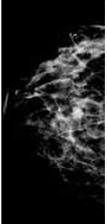
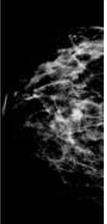
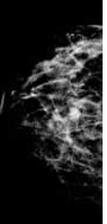
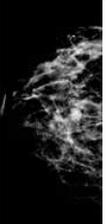
Anisotropic Diffusion (AD)	Filtre Bilateral	Filtre Gaussian	Filtre Moyen	Filtre Median	Nonlinear Noise Reduction	Regularised Isotropic Diffusion
-						-

TABLE A.3 – EQM des résultats issus des filtres simples

Filtre	EQM
Anisotropic Diffusion (AD)	3094,684
Bilateral Filter (BF)	3513,933
Gaussian Blur (GB)	4538,786
Mean Filter (MnF)	7223,201
Median Filter (MdF)	6546,624
Non linear Noise reduction (NLNR)	4541,252
Regularised Isotropic Diffusion (RID)	2950,279

TABLE A.4 – EQM des résultats issus des filtres avec étape de minimisation du flou

Filtre	EQM
FM + AD	-
FM + GB	222878,320
FM + BF	221267,683
FM + MnF	210894,145
FM + MdF	218384,866
FM + NLNR	224848,862
FM + RID	-

évaluer la qualité de l'image au niveau de la zone suspecte le médecin donne une note de -1 à 3 pour chaque image filtrée pour désigner la qualité des données. Ces valeurs sont attribuées comme suit :

- -1 : Qualité médiocre
- 0 : Même qualité que l'image originale
- 1 : Qualité moyenne
- 2 : Qualité bonne
- 3 : Qualité excellente

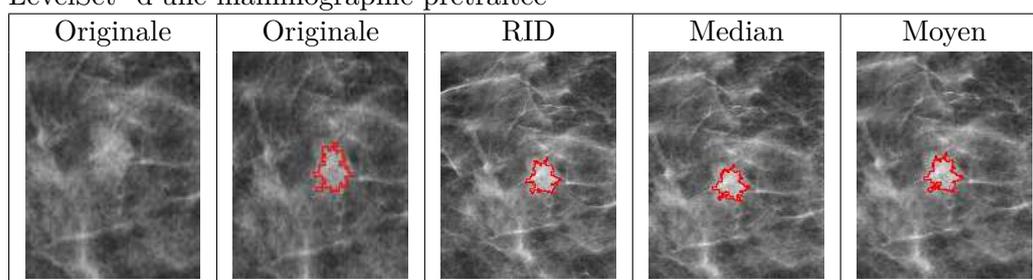
D'après ces résultats nous constatons que les médecins optent plutôt pour une solution de filtrage qui met en évidence la forme pathologique présente.

La deuxième étape du processus de traitement des images est la segmentation des formes pathologiques. Afin d'assurer cette étape, la méthode "Level Set" est appliquée sur une image mammographique filtrée en utilisant le filtre RID. Le résultat obtenu est présenté par la tableau A.6.

TABLE A.5 – Évaluation du médecin des résultats issus des filtres simples

Filtre	Appréciation pour filtre simple	Appréciation pour FM + Filtre
Anisotropic Diffusion (AD)	1	-
Bilateral Filter (BF)	1	1
Gaussian Blur (GB)	1	2
Mean Filter (MnF)	1	2
Median Filter (MdF)	1	2
Non linear Noise reduction (NLNR)	1	2
Regularised Isotropic Diffusion (RID)	1	-

TABLE A.6 – Résultats de la segmentation semi-automatique avec l'algorithme "LevelSet" d'une mammographie prétraitée



A.2 Prétraitement et segmentation des IRM

Les mêmes étapes expliquées, pour le prétraitement et la segmentation des images mammographiques sont suivies aussi pour le prétraitement et la segmentation des images IRM (Fig.A.2). La première étape est la sélection de la région d'étude afin de garder seule la zone du sein. Cette étape est effectuée en détectant la longueur maximale du sein en lui ajoutant un espace de tolérance, ensuite une étape de filtrage est effectuée afin d'améliorer le contraste de l'image et augmenter sa netteté.

L'étape de filtrage est effectuée, comme traité au niveau image mammographique, en appliquant les filtres les plus utilisés en littérature afin de mener une comparaison entre les résultats trouvés et choisir le meilleur filtrage. Ayant un volume IRM (Un exemple d'un volume IRM mammaire est donnée par la figure A.3), deux cas sont pris en considération : le premier test est effectué sur le volume tout entier ; le second test est effectué en considérant l'image coupe par coupe.

L'image IRM utilisée dans ce qui suit comme exemple illustratif contient 94 coupes. Sur la coupe 54 une petite lésion est présente dans le sein gauche. Sur la coupe 56 une masse est présente dans le sein droit (masse observée sur la mammographie correspondante)(Fig. A.3). Les résultats de l'étape de filtrage de l'IRM effectué sur

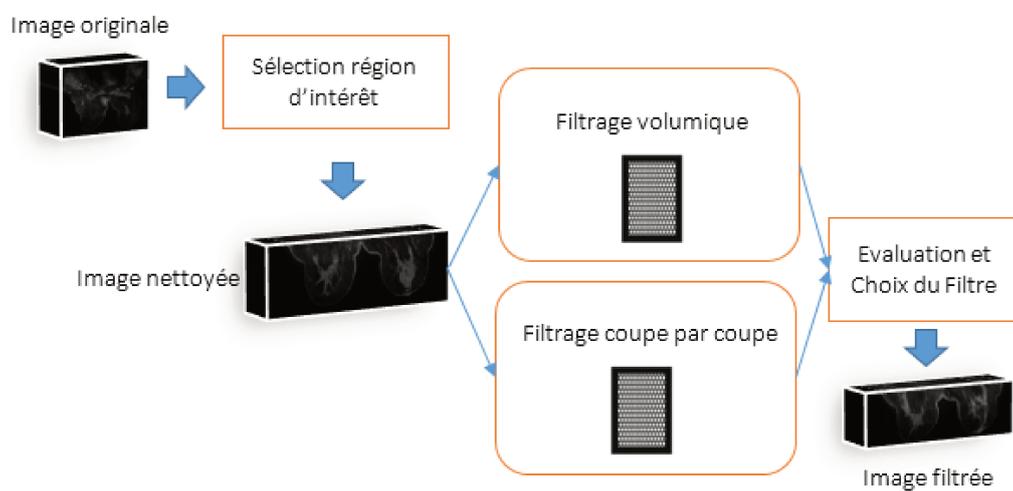


FIGURE A.2 – Procédure du prétraitement des images IRM

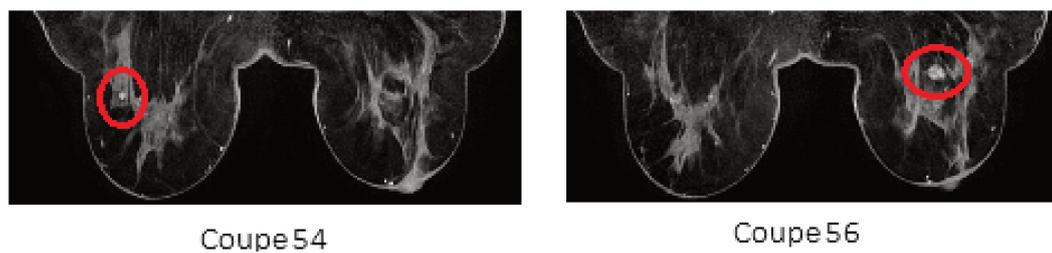
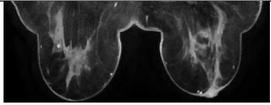
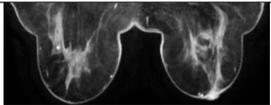
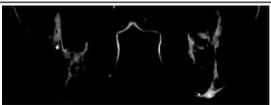
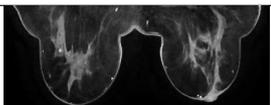
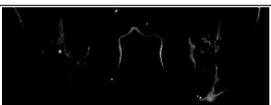
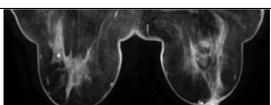
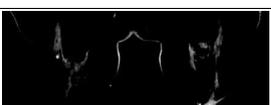
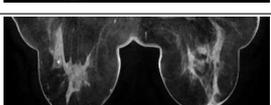
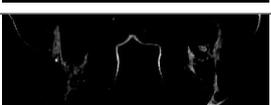
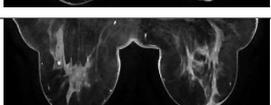
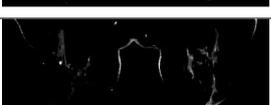
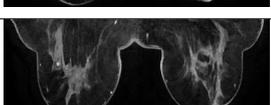


FIGURE A.3 – Objets pathologiques observées sur une image IRM

TABLE A.7 – Résultats de l'étape de filtrage appliqué sur le volume entier (coupe 54)

Type du filtre	Filtre Seul	FM + Filtre
Anisotropic Diffusion (AD)		-
Gaussian blur		
Bilateral filter		
Filtre moyen		
Filtre median		
Nonlinear Noise Reduction		
Regularised Isotropic Diffusion		-

le volume entier sont donnés par les tableaux Tableau A.7 et A.8.

Une évaluation quantitative est effectuée en calculant l'erreur quadratique moyenne (EQM) pour chaque filtre appliqué. Les valeurs sont présentées dans les tableaux Tableau A.9 et Tableau A.10.

D'après les mesures trouvées, le filtre RID (Regularised Isotropic Diffusion) est, aussi, le meilleur choisir. Une évaluation qualitative est aussi envisagée similairement à ce qui est présenté plus haut pour la mammographie. Le résultat de cette évaluation pour l'exemple illustratif utilisé est donné par les tableaux A.11.

Le deuxième cas envisagé est le filtrage du volume IRM coupe par coupe. Les résultats de cette opération sont présentés dans les tableaux A.12 et A.13.

L'évaluation quantitative en termes d'Erreur Quadratique Moyenne donne les valeurs présentée dans les tableaux A.14 et A.15. Les mesures montrent que le filtre "Anisotropic Diffusion" représente le meilleur filtrage. Les résultats de l'évaluation qualitative effectuée par l'expert sont présentés dans le tableau A.16.

Selon les mesures quantitatives effectuées, une comparaison des résultats pour les deux cas, montre que le filtrage coupe par coupe est meilleur que le filtrage volumique. Comme mentionné pour le cas de la mammographie, la segmentation est

TABLE A.8 – Résultats de l'étape de filtrage appliqué sur le volume entier (coupe 56)

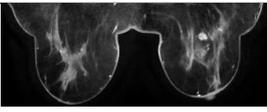
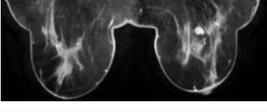
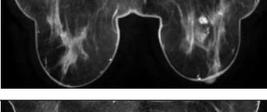
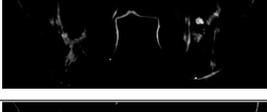
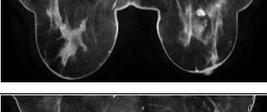
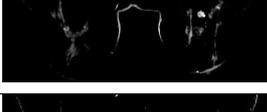
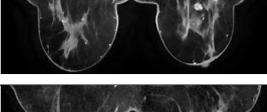
Type du filtre	Filtre Seul	FM + Filtre ω
Anisotropic Diffusion (AD)		-
Gaussian Blur		
Bilateral Filter		
Filtre Moyen		
Filtre Median		
Nonlinear Noise Reduction		
Regularised Isotropic Diffusion		-

TABLE A.9 – EQM des résultats issus des filtres appliqués sur une IRM traité en volume entier

Filtre	EQM
Anisotropic Diffusion (AD)	433,379
Bilateral Filter (BF)	301,156
Gaussian Blur (GB)	772,007
Mean Filter (MnF)	988,686
Median Filter (MdF)	738,323
Non linear Noise reduction (NLNR)	301,602
Regularised Isotropic Diffusion (RID)	127,498

TABLE A.10 – EQM des résultats issus des filtres appliqués sur une IRM traitée en volume entier avec étape de minimisation du flou

Filtre	EQM
FM + AD	-
FM + BF	15242,182
FM + GB	13551,930
FM + MnF	13664,276
FM + MdF	13528,815
FM + NLNR	13516,598
FM + RID	-

TABLE A.11 – Évaluation du médecin des résultats issus des filtres simples appliqués sur une IRM traitée en volume entier

Filtre	Appréciation pour filtre simple	Appréciation pour FM + Filtre
Anisotropic Diffusion (AD)	-1	-
Gaussian Blur (GB)	2	2
Bilateral Filter (BF)	0	-1
Mean Filter (MnF)	2	0
Median Filter (MdF)	2	1
Non linear Noise reduction (NLNR)	2	0
Regularised Isotropic Diffusion (RID)	1	-

effectuée aussi en utilisant la méthode "Level Set".

TABLE A.12 – Résultats de l'étape de filtrage appliquée sur une IRM coupe par coupe (coupe 54)

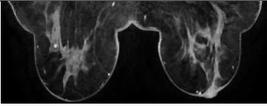
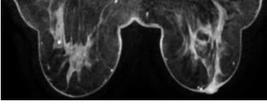
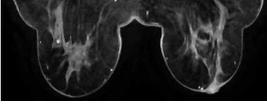
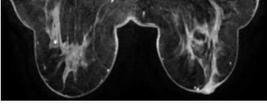
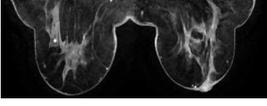
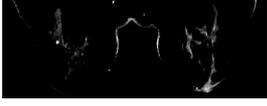
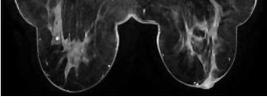
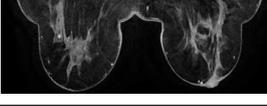
Type du filtre	Filtre Seul	FM + Filtre ω
Anisotropic Diffusion (AD)		-
Gaussian Blur		
Bilateral Filter		
Filtre moyen		
Filtre median		
Nonlinear Noise Reduction		
Regularised Isotropic Diffusion		-

TABLE A.13 – Résultats de l'étape de filtrage appliquée sur une IRM coupe par coupe (coupe 56)

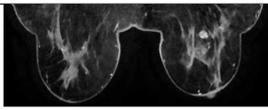
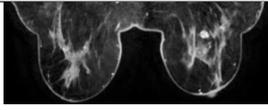
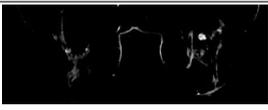
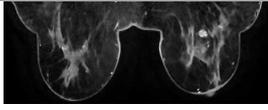
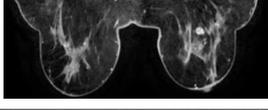
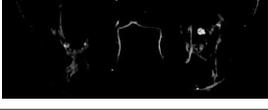
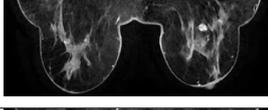
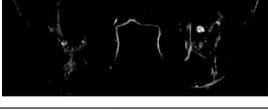
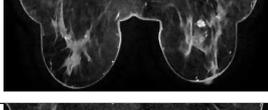
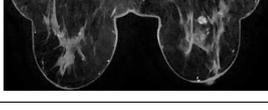
Type du filtre	Filtre Seul	FM + Filtre ω
Anisotropic Diffusion (AD)		-
Gaussian blur		
Bilateral filter		
Filtre moyen		
Filtre median		
Nonlinear Noise Reduction		
Regularised Isotropic Diffusion		-

TABLE A.14 – EQM des résultats issus des filtres appliqués sur une IRM traitée coupe par coupe

Filtre	EQM
Anisotropic Diffusion (AD)	142,531
Bilateral Filter (BF)	404,518
Gaussian Blur (GB)	174,153
Mean Filter (MnF)	327,259
Median Filter (MdF)	189,334
Non linear Noise reduction (NLNR)	213,042
Regularised Isotropic Diffusion (RID)	42,078

TABLE A.15 – EQM des résultats issus des filtres appliqués sur une IRM traitée coupe par coupe avec étape de minimisation du flou

Filtre	EQM
FM + AD	-
FM + GB	13349,534
FM + BF	14402,576
FM + MnF	13276,576
FM + MdF	12967,977
FM + NLNR	14630,078
FM + RID	-

TABLE A.16 – Évaluation du médecin des résultats d'une IRM traitée coupe par coupe

Filtre	Appréciation pour filtre simple	Appréciation pour FM + Filtre
Anisotropic Diffusion (AD)	-1	-
Gaussian Blur (GB)	1	1
Bilateral Filter (BF)	1	-1
Mean Filter (MnF)	1	1
Median Filter (MdF)	1	2
Non linear Noise reduction (NLNR)	1	1
Regularised Isotropic Diffusion (RID)	1	-

Le cancer du sein est le cancer le plus répandu chez les femmes de plus de 40 ans. En effet, des études ont montré qu'une détection précoce et un traitement approprié du cancer du sein augmentent de manière significative les chances de survie. La mammographie constitue le moyen d'investigation le plus utilisé dans le diagnostic des lésions mammaires. Cependant, cette technique peut être insuffisante pour montrer les structures du sein et faire apparaître les anomalies présentes et le médecin peut faire appel à d'autres modalités d'imagerie telle que l'imagerie IRM. Ces modalités sont généralement complémentaires. Par conséquent, le médecin procède à une fusion mentale des différentes informations sur les deux images dans le but d'effectuer le diagnostic adéquat. Pour assister le médecin et l'aider dans ce processus, nous proposons une solution permettant de fusionner les deux images. Bien que l'idée de la fusion paraisse simple, sa mise en œuvre pose de nombreux problèmes liés non seulement au problème de fusion en général mais aussi à la nature des images médicales qui sont généralement des images mal contrastées et présentant des données hétérogènes, imprécises et ambiguës. Notons que les images mammographiques et les images IRM présentent des représentations très différentes des informations, étant donné qu'elles sont prises dans des conditions distinctes. Ce qui nous amène à poser la question suivante: Comment passer de la représentation hétérogène des informations dans l'espace image, à un autre espace de représentation uniforme.

Afin de traiter cette problématique, nous optons pour une approche de traitement multi-niveaux : niveau pixel, niveau primitives, niveau objet et niveau scène. Nous modélisons les objets pathologiques extraits des différentes images par des ontologies locales. La fusion est ensuite effectuée sur ces ontologies locales et résulte en une ontologie globale contenant les différentes connaissances sur les objets pathologiques du cas étudié. Cette ontologie globale sert à instancier une ontologie de référence modélisant les connaissances du diagnostic médical des lésions mammaires. Un raisonnement à base de cas est exploité pour fournir les rapports diagnostic des cas les plus similaires pouvant aider le médecin à prendre la meilleure décision.

Dans le but de modéliser l'imperfection des informations traitées, nous utilisons la théorie des possibilités avec les différentes ontologies. Le résultat fourni est présenté sous forme de rapports diagnostic comportant les cas les plus similaires au cas étudié avec des degrés de similarité exprimés en mesures de possibilité. Un modèle virtuel 3D complète le rapport diagnostic par un aperçu simplifié de la scène étudiée.

Mots clef : Fusion de connaissances, Images multimodales, Représentation Conceptuelle des connaissances, Ontologies, Raisonnement à Base de cas, Représentation virtuelle 3D des connaissances

The breast cancer is the most prevalent cancer among women over 40 years old. Indeed, studies evinced that an early detection and an appropriate treatment of breast cancer increases significantly the chances of survival. The mammography is the most tool used in the diagnosis of breast lesions. However, this technique may be insufficient to evince the structures of the breast and reveal the anomalies present. The doctor can use additional imaging modalities such as MRI (Magnetic Reasoning Image). Therefore, the doctor proceeds to a mental fusion of the different information on the two images in order to make the adequate diagnosis. To assist the doctor in this process, we propose a solution to merge the two images. Although the idea of the fusion seems simple, its implementation poses many problems not only related to the paradigm of fusion in general but also to the nature of medical images that are generally poorly contrasted images, and presenting heterogeneous, inaccurate and ambiguous data. Mammography images and IRM images present very different information representations, since they are taken under different conditions. Which leads us to pose the following question: How to pass from the heterogeneous representation of information in the image space, to another space of uniform representation from the two modalities?

In order to treat this problem, we opt a multilevel processing approach : the pixel level, the primitive level, the object level and the scene level. We model the pathological objects extracted from the different images by local ontologies. The fusion is then performed on these local ontologies and results in a global ontology containing the different knowledge on the pathological objects of the studied case. This global ontology serves to instantiate a reference ontology modeling knowledge of the medical diagnosis of breast lesions. Case-based reasoning (CBR) is used to provide the diagnostic reports of the most similar cases that can help the doctor to make the best decision.

In order to model the imperfection of the treated information, we use the possibility theory with the ontologies. The final result is a diagnostic reports containing the most similar cases to the studied case with similarity degrees expressed with possibility measures. A 3D symbolic model complete the diagnostic report with a simplified overview of the studied scene.

Keywords: Knowledge Fusion, Multimodal Images, Conceptual knowledge representation, Ontologies, Case Base Reasoning, Symbolic Knowledge Representation