UNIVERSITE KASDI MERBAH OUARGLA

Faculté des Sciences et de La Technologie et Sciences de La Matière

Département des Mathématiques et d'Informatique

Mémoire

MASTER ACADEMIQUE

Domaine : Mathématiques et Informatique

Filière : Informatique Académique

Spécialité : Informatique fondamentale

Présenté par : MESSALTI Ahmed Zoubir

HEMILA Abdelkader

Thème

Conception et réalisation d'un système de suivi et d'aide à la décision basé sur une ontologie dans le domaine de la mammographie

Soutenu publiquement

Le : Juin 2013

Devant le Jury :

Mr . MEFLAH Med Salim	Président	UKM OUARGLA
Mr. MAHDJOUB Med Bachir	Examinateur	UKM OUARGLA
Mme. BENKHROUROU Chafika	Examinateur	UKM OUARGLA
Dr. LAALLAM Fatima zohra	Rapporteur	UKM OUARGLA

Année Universitaire : 2012/2013

Remerciements

Louange à ﷲ Seigneur de l'univers

Nous remercions respectueusement notre promoteur, Dr.LAALLAM Fatima Zohra d'avoir accepté de diriger ce travail. Nous la remercions aussi pour sa gentillesse sa spontanéité. Nous reconnaissons sa grande implication dans ce travail par ses directives, ses remarques et suggestions, mais aussi par ses encouragements dans les moments clés de l'élaboration de ce mémoire. Nous tenons à la remercier aussi pour la liberté qu'elle nous a donnée pour accomplir ce travail sans imposer son point de vue. Cela nous a permis de monter avec elle une vraie collaboration. Cette liberté nous à permis d'apprendre à prendre des décisions et d'avoir confiance en soi même. Nous la saluons pour sa manière de penser et de procéder, sa manière d'être, bref toute sa personnalité.

Nos respects et notre gratitude vont également aux membres du jury : Mr MEFLAH Med Salim et Mr MAHDJOUB Med Bachir et Mme BENKHROUROU Chafika qui nous ont fait l'honneur de juger ce travail.

Nous remercions également toute l'équipe du Master informatique, nos enseignants, merci aussi à l'équipe de l'hôpital SAADNA ABD NOUR à Setif pour tous les conseils et encouragements dont nous avons bénéficiés tout au long de ce travail.

Enfin, un grand Merci à tous ceux qui nous ont soutenu de près ou de loin. Merci à nos camarades de la promotion 2012/2013

Zoubir et Abdelkader

Dédicaces

Grace à Dieu voilà notre travail terminé et il est temps pour moi de partager ma joie avec tous ceux qui m'ont soutenu et encouragé.

À travers ce modeste mémoire, je tiens à présenter mes sincères dédicaces à

Ma mère (رحمة الله عليها) qui m'a confié de suivre mes études et m'a donné l'espoir de réussir

Mon père qui a consacré sa vie à notre éducation et à faire notre bonheur et qui nous encourage toujours d'achever nos études tout en espérant de voir les fruits de ses sacrifices.

A ma chère sœur et à mes chers frères

A mes Oncles, Tantes et cousins et alliés de la famille.

A l'ensemble des amis que j'ai connu pendant mes études et à ceux qui m'ont prodigué leurs vifs conseils, encouragements et témoigné de leur amitié.

A mon binôme Abdelkader et sa famille.

Zoubir

Dédicaces

Grace à Dieu voilà notre travail terminé et il est temps pour moi de partager ma joie avec tous ceux qui m'ont soutenu et encouragé.

À travers ce modeste mémoire, je tiens à présenter mes sincères dédicaces à mes Parents qui ont consacré leurs vie à notre éducation et à faire notre bonheur et qui m'encouragent toujours pour achever mes études tout en espérant de voir les fruits de leurs sacrifices.

A mes chères sœurs et mes chers frères

A mes Oncles, Tantes et cousins et alliés de la famille.

A l'ensemble des amis que j'ai connu pendant mes études et à ceux qui m'ont prodigué leurs vifs conseils, encouragements et témoigné de leur amitié.

A mon binôme Zoubir et sa famille.

ABD EL KADER

Sommaire

Résumé……VI

Abstract……VI

الملخص……VII

Introduction générale VIII

Chapitre 1 L'imagerie médicale et le cancer du sein 1

I. Introduction 2

II. L'imagerie médicale 2

II.1. Définition d'imagerie médicale 2

II.2. Le but de l'imagerie médicale 3

II.3. Les modalités d'imagerie médicale 3

II.3.1. La radiographie 3

II.3.2. Le scanner 3

II.3.3. L'échographie 4

II.3.4. L'Imagerie par Résonance Magnétique (IRM) 4

II.3.5. La mammographie 4

III. Le système CAD en imagerie médicale 5

III.1. Le CAD en mammographie (état de l'art) 5

III.2. La détection automatique des cancers en mammographie 6

III.2.1. Les composants d'un système de détection automatique 6

IV. Le cancer du sein 11

IV.1. L'anatomie du sein 11

IV.1.1. Description générale 11

IV.1.2. Description détaillée 12

V. Pathologies mammaire 13

V.1. Les maladies bénignes 13

- Fibro-kystique : 14
- Fibroadénome : 14

V.2. Les maladies malignes 15

V.2.1. Les symptômes 15

VI. Les facteurs de risques 15

VII. Le dépistage 16

VII.1. Le diagnostic 18
VII.2. Les différents types de cancers 18
VII.3. Caractéristiques des différents types de cancer 19
VIII. La classification BI-RADS [45] 21
VIII.1. La classification proposée 22
VIII.2. Le traitement du cancer 23
Chirurgie 24
Hormonothérapie 24
IX. Conclusion 25
Chapitre 2 Les ontologies *26*
I. Introduction 27
II. Définition 27
III. Pourquoi les ontologies ? 28
IV. Les aspects d'une ontologie 28
IV.1. Formelle 28
IV.2. Consensuelle 29
IV.3. Référençable 29
V. Les composants d'une ontologie 29
V.1. concepts 29
V.2. Les relations 30
V.3. Les fonctions 30
V.4. Les axiomes 31
V.5. Les instances (individus) 31
VI. Les différents types d'ontologies 31
VI.1. Selon le degré de formalisme 31
VI.2. Selon les objets modélisés 32
VII. Construire une ontologie 34
VII.1. Étapes de construction d'une ontologie 34
VII.2. quelques méthodes de construction d'ontologies 35
VII.3. Usages des ontologies 36
VIII. Les langages de représentation 37
VIII.1. RDF 37
VIII.2. RDFS 37
VIII.3. DAML-OIL 38
VIII.4. OWL 38

IX. Conclusion 39

Chapitre 3 Modélisation du système du dépistage du cancer de sein.....*40*

I. Introduction.....*41*

II. Description générale du système.....*41*

II.1. Extraction des paramètres 41

II.2. Base de données mammographies.....42

II.2.1. escription de fichier ics.....43

II.2.2. description du fichier overlay 43

II.3. Classification par l'ontologie 44

III. Construction d'une ontologie pour la prise de décision des pathologies mammaires...*44*

III.1. Spécification.....44

III.2. Conceptualisation.....45

- *Construction du glossaire de termes*.....45
- *Classification des concepts en hiérarchies de concepts*.....48
- *Construction de la table des relations binaires* 48
- *Construction d'un dictionnaire de concepts*.....49

IV. *Conception d'une BD pour le stockage des informations :*.....*49*

V. Modélisation de l'inférence.....*50*

VI. Conclusion *50*

Chapitre 4 Implémentation et validation *51*

I. Introduction.....*52*

II. Environnement de développement.....*52*

II.1. Protégé 52

II.2. Le langage SWRL.....53

II.3. JESS (The JavaExpert System Shell).....53

II.4. Jfeechart 54

III. la base d'images utilisée *54*

IV. Structure de notre application.....*54*

V. Classification par l'ontologie *55*

VI. Utilisation de notre application *57*

VII. Conclusion *62*

Conclusion Générale.....*63*

Liste des figures

Figure I.1 Schéma général d'une chaîne de détection de microcalcifications ou d'opacités en mammographie. 7
Figure I.2 Schéma anatomique du sein. 12
Figure I.3 Vue des cellules myoépithéliales adossées à la membrane basale. 13
Figure I.4 Schéma détaillé des lobules. 13
Figure I.5 Différentes formes pour les microcalcification 19
Figure I.6 Différentes formes pour les opacités. 20
Figure I.7 Différents types de contours pour les opacités. 21
Figure I.8 Forme schématique d'une distorsion architecturale. 21

Figure II.1 Classification des ontologies selon Guarino……....32
Figure II.2 Schéma représentatif des relations entre les ontologies de domaine, de raisonnement, de représentation 33

Figure III. 1 Schéma général de l'approche.....41
Figure III. 2 Le fichier A-1141-1.ics…….……..43
Figure III. 3 Fichier A_1141_1.RIGHT_CC. Overlay……..……...44
Figure III. 4 diagramme de relations binaires……..48
Figure III.5 modèle conceptuel de données...……49

Figure IV.1 Un exemple de réglés SWLR ..... **Erreur ! Signet non défini.**
Figure IV.2 Diagramme de classes de notre application……55
Figure IV .3 OntologieCancer éditée sous l'outil PROTEGE. 56
Figure IV 4 Les règles SWRL. 57
Figure IV 5 Fenêtre principale du système. 59
Figure IV 6 chargement du dossier de la patient 60
Figure IV 7 Un résultat obtenu a partir de l'ontologie 60
Figure IV 8 L'intégration de JESS. 61
Figure IV 9 Menu de la démonstration statistique. 61
Figure IV 10 JFreechart(diagramme en barre ,diagramme circulaire)….....62

Liste des tableaux

Tableau I.1 Classification des masses. 22
Tableau I.2 Classification des calcifications. 23
Tableau I.3 Classification des désorganisations architecturales. 24

Tableau III. 1 Paramètres utilisés pour la classification. 42
Tableau III. 2 Glossaire de termes. 48
Tableau III. 3 Table des relations binaires. 48
Tableau III. 4 Construction d'un dictionnaire de concepts. 49

Résumé

Le cancer du sein est un véritable fléau de nos jours puisqu'il touche une grande population féminine et constitue à lui seul le plus grand pourcentage de mortalité chez les femmes. Cependant la détection de la tumeur à une étape précoce augmentera considérablement les chances de guérisons des patientes. Il apparait indispensable de développer de nouvelles méthodes ou de nouveaux protocoles pour le dépistage du cancer du sein. Nous proposons à cet effet une nouvelle méthode de classification des images de mammographies basée sur l'utilisation des ontologies.

Cette méthode consiste à rajouter l'aspect sémantique à l'information médicale des images de mammographie via une ontologie spécialisée développée à cet effet en se basant sur le système de caractérisation BI-RADS.

La méthode proposée a été testée sur plusieurs mammographies provenant de la base d'images DDSM (*Digital Database for Screening Mammography*).

Mots clés : CAD; ontologie; extraction de paramètres ; classification ; mammographie ; microcalcifications

Abstract

Breast cancer continues to be a significant public health problem in the world because it affects a large female population; it is also the largest percentage of mortality among women. Early detection is the key to improving breast cancer prognosis. It appears essential to develop new methods or new protocols for the detection of breast cancer. We propose for this purpose a new method of classifying images of mammograms based on the use of ontology.

This method is to add the semantic aspect to medical information of mammography images via a specialized ontology developed for this purpose based on the characterization system BI-RADS.

The proposed method was tested many mammograms from the image database DDSM (Digital Database for Screening Mammography).

Keywords: CAD; ontology; parameter extraction; classification;

mammography; microcalcifications.

الملخص

يشكل سرطان الثدي مشكلة رئيسية هذه الأيام لأنه يمس شريحة كبيرة من النساء، حيث تسبب هذا النوع من السرطان في أكبر عدد من الوفيات لدى النساء في الجزائر. مع ذلك فإن اكتشاف الورم في مرحلة مبكرة يزيد فرص الشفاء من المرض. لذالك يبدو من الضروري استحداث أساليب جديدة أو بروتوكولات جديدة للكشف عن سرطان الثدي. نقترح لهذا الغرض تقنية جديدة لتصنيف الصور الإشعاعية للثدي بالأشعة السينية تعتمد على انتلوجيا. هذه الطريقة تعتمد على اضافة الجانب الدلالي إلى المعلومات الطبية، استعملنا أنطولوجيا متخصصة قمنا أيضا بتصميمها لهذا الغرض على أساس نظام التوصيف (BI-RADS) جملة الاختبارات أجريت على قاعدة الصور الإشعاعية DSM و النتائج المتحصل عليها مرضية.

كلمات البحث: أنظمة المساعدة على تشخيص سرطان الثدي، الانطولوجيا، استخراج المعلمة، التصنيف. ،, التصوير الإشعاعي للثدي.

Introduction générale

Au cours des dernières années, le cancer du sein s'est largement imposé comme un problème majeur de santé dans le monde. C'est le cancer le plus fréquent chez la femme, avec plus de 800 000 nouveaux cas diagnostiqués chaque année dans le monde. Il représente la première cause de mortalité féminine dans la tranche d'âge de 35 à 55 ans, constituant ainsi un sérieux problème de santé publique Il est responsable de 20% des décès dus aux cancers. En Algérie, le cancer du sein représente 30% des cas de cancer recensés, et représente près de 50 % des cancers gynécologiques chez la femme, au cours de ces 15 dernières années l'incidence du cancer du sein a été multipliée par 3. Du fait de son diagnostic tardif, il en résulte souvent un traitement lourd, mutilant et coûteux qui s'accompagne d'un taux de mortalité élevé. L'incidence du cancer du sein reste croissante en Algérie et il n'existe toujours pas des programmes de dépistage organisé à l'instar de nos voisins Maghrébins ou Européens.

Dans le but d'éviter des traitements lourds et de réduire la morbidité et la mortalité par le cancer du sein, une détection précoce est nécessaire motivant ainsi des campagnes de dépistage chez les femmes à partir d'un certain âge variant entre 40 et 50 ans selon les pays. Le type d'imagerie actuellement utilisée pour cette tâche est la mammographie.

La mammographie est un examen radiologique consacré à la détection des pathologies du sein. Cet examen s'effectue avec un appareil dédie uniquement à cet usage c'est le mammographe. Cet appareillage utilise les rayons X pour produire des images de hautes résolutions du sein.

L'interprétation des clichés en un temps limité et la détection des pathologies est une tâche très difficile par un examen mammographie cela nécessite l'utilisation d'outils d'aide à la détection par ordinateur (CAD) a fin d'améliorer les performances du lecteur.

Notre travail consiste à présenter une aide à la décision au médecin responsable du dépistage du cancer des seins en développant toute une stratégie d'extraction de connaissance à partir de données biomédicales (images mammographiques) pour la détection d'éventuelles pathologies. A fin de rehausser le taux de classification nous proposons de modéliser les connaissances au sein d'une ontologie locale appropriée pour la classification des régions sélectionnées à partir des mammograms numériques, Pour chaque signe radiologique considéré, à savoir les opacités et les micro-calcifications.

En premier lieu, on fait une extraction des paramètres caractérisant les images issues de la base de données MARATHON. Ces images sont déjà diagnostiquées par des experts. Les paramètres extraits sont organisés sous forme d'une ontologie pour l'annotation des images. Après une étape de conceptualisation, l'ontologie est éditée en utilisant Protégé2000. Pour permettre l'inférence de nouvelles connaissances, partant des connaissances existantes dans l'ontologie, le Langage de Règles pour le Web Sémantique (SWRL) est utilisé.

Ce travail est organisé en 4 chapitres :

Le premier chapitre «L'imagerie médicale et le cancer du sein »: contient trois parties. Nous verrons, dans la première partie, la définition de l'imagerie médicale ensuite les différentes techniques de l'imagerie médicale: Radiographie, scanner, échographie, IRM et la mammographie. Dans la deuxième partie, nous rappellerons les notions fondamentales d'un système de détection puis nous étudierons les techniques existantes dans la littérature pour accomplir la détection des différents signes radiologiques suspects en mammographie conventionnelle. Dans la troisième partie, nous présentons le cancer de sein. Nous commençons par la définition de cette maladie puis nous donnons l'anatomie du sein, les différents types du cancer (maladies bénignes et malignes), les symptômes et en fin les différents traitements possibles.

Dans Le deuxième chapitre « Les ontologies », nous exposerons, en premier lieu les différentes caractéristiques d'une ontologie, en décrivant la manière de formaliser cette dernière. Ensuite, nous nous étalerons sur la typologie des ontologies ainsi que les diverses classifications existantes. Enfin, nous présentant les langages de spécification et les outils d'édition et d'interrogation des ontologies.

Le troisième chapitre «Modélisation du système de dépistage du cancer de sein » expose le développement d'une nouvelle approche d'extraction de connaissance à partir de données biomédicales, il s'agit d'une approche basé sur les ontologies pour donner à l'information un aspect sémantique.

Le quatrième chapitre «Implémentation et validation » présente les outils nécessaires pour l'implémentation de notre système, à savoir : la base d'images, l'ontologie conceptualisée, Protégé2000, JESS, SWRL et les règles d'inférence.

Une conclusion générale de ce travail est présentée à la fin du mémoire.

Chapitre 1

L'imagerie médicale et le cancer du sein

I. Introduction

L'imagerie médicale est certainement l’un des domaines de la médecine qui a connu une véritable révolution pendant ces vingt dernières années. Ces récentes découvertes permettent non seulement un meilleur diagnostic mais offrent aussi de nouveaux espoirs de traitement pour de nombreuses maladies (comme le cancer du sein).

Le cancer du sein est le premier néoplasie chez la femme dans le monde et dans notre pays chaque année, environ 7500 cas de cancer du sein sont enregistrés, tout en sachant que ce type de néoplasie en Algérie vient en tête des tumeurs malignes chez la femme et constitue la première cause de mortalité chez la gent féminine, avec environ 3500 décès enregistrés chaque année [1].

Le cancer du sein concerne une femme sur 11 dont l'âge se situe le plus souvent entre 50 et 60 ans. Uniquement 5 % des femmes présentant un cancer du sein ont moins de 35 ans. Il s'agit du cancer le plus fréquent de la femme et le nombre de cancer du sein croît régulièrement [2]. Ces chiffres montrent l'importance d'une détection précoce de cette maladie. La mammographie par rayon X reste la technique la plus fiable pour le diagnostic précoce du cancer du sein. Les masses et les microcalcifications sont les premiers signes d'alerte de cette maladie.

Dans ce chapitre, nous allons prendre un aperçu général sur l’imagerie médicale, ces modalités et un aperçu détaillé sur la mammographie et leurs systèmes d’aide à la détection, ainsi que l’aspect médicale du cancer de siens.

II. L’imagerie médicale

II.1.Définition d’imagerie médicale

[3], nous indique la définition : **Imagerie médicale : «** Spécialité médicale consistant à produire des images du corps humain vivant et à les interpréter à des fins diagnostiques, thérapeutiques (imagerie interventionnelle) ou de surveillance de l'évolution des pathologies. La discipline universitaire et hospitalière "radiologie et imagerie médicale" comporte deux spécialités médicales distinctes, intitulées : radiologie et imagerie médicale, d'une part, et médecine nucléaire, d'autre part. La radiologie utilise les rayons X, l'échographie, les ultrasons, l'imagerie par résonance magnétique (IRM.), le phénomène de résonance magnétique nucléaire, la médecine nucléaire des isotopes radioactifs ».

II.2.Le but de l’imagerie médicale

Le but de l'imagerie médicale est de fournir une représentation visuelle intelligible qui après analyse révèle une information à caractère médical. Ce procédé permet à un médecin d’examiner l'intérieur du corps d'un patient sans l'opérer. L'imagerie médicale peut être utilisée à des fins cliniques pour l'établissement d'un diagnostic ou pour le trai tement de pathologies mais également dans le cadre de travaux de recherche scientifique étudiant la physiologie des êtres vivants [4].

II.3.Les modalités d’imagerie médicale

Pour chaque modalité d’imagerie, le principe permettant d’accéder à l’information:

II.3.1. La radiographie

La radiographie est probablement l'un des examens médicaux les plus courants. Si ce n’est pas une technique nouvelle, elle se révèle être encore d’une grande actualité et d’une grande utilité. En effet, même si de nombreuses autres techniques d’imagerie médicale très sophistiquées sont apparues depuis quelques décennies (échographie, scanner, IRM, TEP etc…), la radiographie est loin d’être obsolète car elle a su évoluer. On citera pour exemple la tendance au remplacement du film photographique employé pour révéler les images radiographiques par des capteurs numériques, ce qui ouvre la porte au traitement d’image ainsi qu'au partage par le biais d’Internet [5].

Le principe de la radiologie est une impression photographique par les rayons X sur des capteurs plans (numérisation directe), la quantité de rayons X reçue dans les différentes zones du capteur dépendant de l'absorption des rayons X par les différents tissus traversés soit leur densité. Ainsi, les os très denses apparaîtront en clair, tandis que les autres organes tels que les poumons seront visibles sur la radiographie d'une couleur beaucoup plus sombre [6].

II.3.2. Le scanner

Le Scanner appelé aussi tomodensitométrie est un examen qui utilise les rayons X.

Le Scanner étudie le cerveau, la cage thoracique, l'abdomen ou encore les os. Il recherche des anomalies qui ne sont pas visibles sur des radiographies standard ou à l'échographie. Il permet de :

- Mettre en évidence des infections, une hémorragie, des kystes, des tumeurs, des ganglions,

- Localiser avec précision un organe par rapport à un autre, de définir le trajet d'un vaisseau.
- Guider des ponctions d'organes profonds évitant une intervention chirurgicale.

II.3.3. L'échographie

L'échographie est une technique permet l'étude de multiples organes de l'abdomen, du petit bassin, du cou (thyroïde, ganglions, foie, rate, pancréas, reins, vessie, organes génitaux) mais aussi les vaisseaux (artères et veines), les ligaments et le cœur. Elle recherche des anomalies qui pourraient les atteindre (tumeurs, infections, malformations) et peut parfois guider un prélèvement en profondeur. Cette technique basée sur les ultra-sons. Une sonde envoie un faisceau d'ultrasons dans la zone du corps à explorer. Selon la nature des tissus, ces ondes sonores sont réfléchies avec plus ou moins de puissance. Le traitement de ces échos permet une visualisation des organes observés [3].

Au cours d'une grossesse, elle permet d'étudier la vitalité et le développement du fœtus, de dépister des anomalies ou encore de déterminer le sexe de l'enfant [3].

C'est un examen rapide et indolore pour le patient, sans contre-indication et, sans danger pour le fœtus au cours de la grossesse. Il ne permet pas l'étude de tous les organes (os, poumon) et l'embonpoint peut rendre l'interprétation des images difficile.

II.3.4. L'Imagerie par Résonance Magnétique (IRM)

L'imagerie par résonance magnétique permet d'analyser à distance des organes de manière très précise. Cette technique permet de visualiser des détails invisibles sur la radiographie standard, l'échographie ou le scanner. Schématiquement, cette technique utilise un gros aimant et analyse la réaction des différents tissus du corps à ces champs magnétiques. Les données recueillies sont ensuite traitées informatiquement et la zone étudiée peut être restituée en deux ou trois dimensions [3].

II.3.5. La mammographie

Il s'agit d'un examen radiographique des seins. La mammographie permet grâce à de faibles doses de rayons X l'obtention de clichés permettant de détecter certaines anomalies de la glande mammaire. Son but principal est le dépistage du cancer du sein. Le pronostic du cancer du sein est lié à son stade au moment de sa découverte. La mammographie, examen simple et fiable, permet de dépister le cancer du sein à un stade précoce et donc d'améliorer les chances de guérison. L'examen permet de détecter des

anomalies non repérées par la palpation des seins. Elle est effectuée par un radiologue, à l’hôpital ou dans un cabinet de radiologie [7].

On doit distinguer la *mammographie de dépistage* « recherche la présence d'une tumeur chez des personnes non symptomatiques » de la *mammographie diagnostique* « réalisé lorsqu’une anomalie a été détectée au cours d’un examen de dépistage ou lorsqu’une anomalie dans le sein a été découverte par la patiente ou par son médecin traitant».

Quelles que soient les circonstances, deux clichés (photos) par sein sont réalisés, un cliché de face et un en oblique, ce qui permet de comparer les deux côtés de chaque sein. C’est ce que l’on appelle l’examen mammographique de base.

Il existe actuellement deux types de mammographes : les mammographes conventionnels qui permettent l'obtention des films radiologique habituels et les mammographes numériques qui utilisent un détecteur numérique permettant le traitement des images (zoom, modification de contraste, diagnostic assisté par ordinateur CAD).

III. Le système CAD en imagerie médicale

Un système d’aide à la prise de décision dédiée à l’imagerie médicale est un programme numérique qui analyse des clichés médicaux et consiste à assister le clinicien dans la tâche parfois délicate d’interprétation et de classement des images médicales. Cette technologie est représentée par deux applications principales. Les systèmes d’aide à la détection (computer aided detection system CADe) et les systèmes d’aide au diagnostic (computer aided diagnosis system CAD). Les CADe ont pour rôle de détecter et d’attribuer un marqueur à une ou plusieurs tumeurs dans une image. Les CAD différencient les tumeurs malignes et bénignes parmi celles détectées par les CADe ou fournis par un radiologue.

Le développement de systèmes CAD en imagerie médicale a débuté dans les années 80 [8, -9] et ciblait des applications à forts enjeux médicaux comme la mammographie.

III.1. Le CAD en mammographie (état de l’art)

La technologie de détection assistée par ordinateur ou CAD est une méthode complémentaire à la mammographie utilisée pour identifier les régions suspectes dont les caractéristiques physiques sont similaires à celles d’un cancer afin d’aider le radiologue à pratiquer le dépistage précoce du cancer du sein.

Le CAD est plus sensible pour les micro-calcifications que pour les masses, et les lésions détectées auraient pu l'être par un radiologue plus expérimenté que celui qui a lu la mammographie initiale.

Depuis quelques années, de nombreuses équipes de recherche tentent de développer des systèmes informatiques d'analyses des clichés de mammographie. Plusieurs voies sont explorées telles que la classification automatique des anomalies détectées, la reconnaissance de forme et la détection des pathologies mammaires. De nombreux travaux ont été effectués pour développer des logiciels destinés à améliorer la performance diagnostique des radiologues en procédant à un marquage des lésions suspectes de malignité. Parmi ces systèmes on trouve [10]:

- Image Checker M1000, de la société R2 technology (approuvé FDA en juin 1998, marqué CE).
- Second Look, de la société CADX Medical Systems (demande d'approbation FDA déposée, marqué CE).
- Mammex TR, de la société Scanis (demande d'approbation FDA déposée, marqué CE).

Le principe de ces trois systèmes est le même. Ils sont conçus pour identifier et marquer des « régions d'intérêts » ou des images suspectes. Concevoir un système de détection et de classification automatique des lésions en mammographie, ne doit pas être compris comme une tentative de remplacer le médecin. La question n'est pas de se substituer au médecin, mais de lui proposer des outils performants qui l'aident dans son travail d'analyse des clichés de mammographie.

III.2. La détection automatique des cancers en mammographie

Le passage à la mammographie numérique a ouvert la porte à de nouvelles possibilités en termes d'outils visant à extraire de manière optimale l'information dans l'image. Ainsi, des traitements visant à améliorer l'image, ou encore des outils d'aide à la détection sont couramment utilisés par les radiologues.

III.2.1. Les composants d'un système de détection automatique

De manière globale un système de détection automatique de cancers en mammographie se compose de deux branches : une dédiée à la détection des micro-calcifications et l'autre à la détection des opacités. Chacun de ces modules peut se décomposer comme une étape de marquage suivie d'une prise de décision.

Le marquage peut selon les cas être composé d'une détection rapide suivie d'une segmentation. La prise de décision se compose quant à elle d'une étape d'extraction de

caractéristiques suivie d'une étape de classification. Certains de ces éléments peuvent apparaître de manière plus ou moins implicite. La figure I.1 illustre la décomposition de haut niveau des processus de détection automatique. Dans certains cas, un prétraitement des données permettant de mettre en évidence les signes recherchés peut être utilisé. Néanmoins, cette étape montre généralement rapidement ses limitations dans la mesure où, pour vraiment mettre en évidence un motif, il faut être capable de le détecter, or c'est le but de cette étape de pré-traitement.

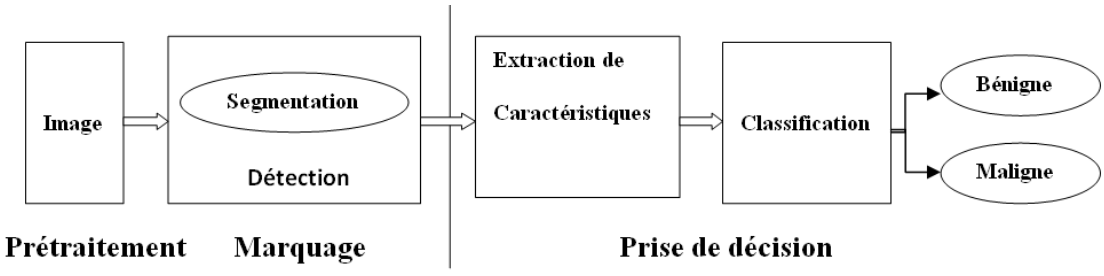

Figure I.1 : Schéma général d'une chaîne de détection de microcalcifications ou d'opacités en mammographie.

III.2.1.1. Pré-traitement des images

Une première étape souvent utilisée est donc la préparation des images avant la détection. En effet les structures que l'on recherche n'étant pas toujours facilement discernables, une étape de pré-traitement destinée à les mettre en évidence peut faciliter leur détection.

Une approche couramment utilisée en traitement d'images consiste à travailler sur l'histogramme dans le but de définir une fonction de transfert sur les niveaux de gris permettant de mettre en valeur les détails présents dans l'image.

D'autres méthodes ont aussi été proposées comme l'utilisation de filtres d'amélioration de la netteté [11], où encore la suppression du fond de l'image [12]. Cette dernière approche consiste en la soustraction d'une version filtrée passe bas de l'image originale à cette dernière.

L'idée d'un prétraitement, bien qu'intuitive, peut poser quelques problèmes. En effet, dans l'idéal on voudrait mettre en évidence seulement les zones potentiellement suspectes dans le but de faciliter leur détection ultérieurement. Or pour pouvoir accomplir cette tâche, il faudrait savoir quelles zones de l'image on doit améliorer, c'est-à-dire connaître les zones suspectes, ce qui est difficile puisque l'on cherche à améliorer l'image pour justement détecter ces structures.

III.2.1.2.Marquage

➢ *Détection des microcalcifications*

La première classe de signes radiologiques qui sont recherchés sont les foyers de microcalcifications. En effet les amas de calcium étant constitués objets de petite taille assez contrastés, ce genre d'approche semble avoir un réel potentiel. Par exemple, [13] proposent de modéliser le problème comme la détection d'objets gaussiens dans un bruit markovien en utilisant une décomposition par banc de filtres. Une autre approche [14] consiste à effectuer un filtrage avec un chapeau mexicain dans le but de mesurer le contraste des structures de dimension compatible avec le noyau du filtre. En ajoutant une contrainte dérivée du bruit de Poisson que l'on peut généralement estimer dans les mammographies, et en faisant varier l'échelle de l'ondelette, on peut obtenir des résultats de détection satisfaisants. D'autres formalismes sont aussi possibles, ainsi on peut trouver l'utilisation de la logique floue pour la détection des calcifications [15, -16] ou encore des approches utilisant des filtrages de la dynamique de l'image à l'aide d'outils de morphologie mathématique [17].

➢ *Détection des opacités*

La détection des opacités est une tâche un peu plus complexe que pour les microcalcifications, notamment à cause de la variabilité qui existe entre différents types de lésions.

➢ *Extraction de marqueurs*

La détection des zones de sur-densité est une première étape pour la détection de zones suspectes dans le sein. Ainsi [18] proposent d'utiliser une mesure de détection des sur-densités en calculant la proportion de pixels situés autour d'une lésion qui ont une intensité inférieure au minimum de l'intensité à l'intérieur. Dans la même optique de détection de sur-densités, d'autres approches tentent de détecter les zones contrastées à l'aide d'un chapeau mexicain de la même manière que pour les microcalcifications. Néanmoins ce genre d'approche ne semble pas être bien adapté à la taille des objets à détecter [19]. En effet, une opacité est généralement beaucoup plus étendue qu'une microcalcification, avec une variabilité en termes de forme plus importante.

[20] a proposé l'utilisation d'outils de morphologie mathématique pour l'extraction automatique d'opacités. L'idée utilisée est de se servir d'algorithmes de ligne de partage des eaux combiné à un critère de dynamique modélisant le contraste des structures contenues dans l'image.

D'autres approches reposent sur le fait que les lésions malignes présentent souvent une forme irrégulière voire des spicules. Ainsi, [21] propose l'utilisation de filtres en

quadrature pour la détection de ces spicules. [22] quant à eux proposent une décomposition en ondelettes sur deux niveaux permettant le calcul de mesures dérivées du flot de gradient combinées à d'autres mesures reposant sur l'analyse locale des histogrammes d'orientations de contour.

Les lésions variant aussi du point de vue de la taille, des approches multi-résolution sont souvent proposées [22,-23]. Ainsi, l'analyse fractale a été proposée comme première étape pour la détection de marqueurs [24]. La granulation qui consiste globalement à sous-quantifier les niveaux de gris pour extraire des composantes connexes emboîtées vérifiant un critère de concentricité a aussi été proposée comme étape de marquage par [25].

Dans tous les cas, les méthodes proposées permettent de sélectionner de manière grossière, c'est-à-dire avec beaucoup de faux positifs.

- *Segmentation*

Les approches précédentes sont généralement conçues pour la détection de zones potentiellement suspectes. Néanmoins bien souvent pour vérifier si ces dernières contiennent une lésion maligne, il est nécessaire d'extraire la forme de lésion potentiellement détectée. Les approches les plus simples reposent sur des seuillages globaux de l'image [26] Ont été développé un algorithme itératif pour la détection automatique des masses dans les mammographies. Plusieurs approches de segmentation utilisées nous verrons dans le chapitre suivant.

III.2.1.3.Prise de décision

La prise de décision se fait généralement après une étape de caractérisation. D'un point de vue haut niveau, on fait des mesures à partir des résultats de détection et/ou de segmentation (selon que l'on travaille sur des microcalcifications ou des opacités) dans le but de prendre dans un second temps une décision à l'aide de méthodes de classification standard.

III.2.1.4.Extraction de caractéristiques

Dans le cas des microcalcifications, une étape préliminaire d'extraction de caractéristiques est donc requise [27]. Parmi ces caractéristiques, on peut en trouver ceux qui portent sur la texture comme l'analyse des régions avoisinantes, la dépendance spatiale des niveaux de gris, les statistiques de longueurs sur les niveaux de gris ou encore la différence de niveaux de gris [28]. En ce qui concerne les opacités, la première classe de caractéristiques utilisées repose sur la forme de l'objet. On peut citer par exemple la compacité, le rapport de Ferret, ou encore le périmètre.

Des mesures sur l'intensité peuvent aussi être utilisées [29]. On peut par exemple analyser la distribution des niveaux de gris dans la lésion supposée, sur son contour ou

dans son voisinage. On peut aussi utiliser des mesures plus robustes qui sont invariantes par changement de contraste [30]. Ce type de mesures apporte une information complémentaire par rapport aux précédentes mesures.

Une caractérisation importante de certaines lésions est la présence de spicules au niveau de leur périphérie. Pour prendre cet élément en compte, une série de mesures permettant d'évaluer à quel point la lésion est spiculée ont aussi été proposées**.** Elles reposent essentiellement sur l'analyse des orientations locales des contours. Ces orientations peuvent par exemple être obtenues à partir d'une décomposition en ondelettes [31]. Sur le même principe, des travaux proposent d'évaluer la complexité d'un contour par analyse fractale [32].

L'information fournie par l'association du contour et du contenu de l'image peut aussi être exploitée en extrayant et en analysant une bande plus ou moins large qui suit le contour de la lésion [33]. Cette bande est généralement étirée de manière à rendre le contour linéaire avant analyse. L'analyse se fait par l'extraction de mesures sur cette bande, comme l'étude des orientations du contour [34]. Ce type d'approche est donc dépendant du contour, ainsi les mesures obtenues vont être fortement liées aux caractéristiques des segmentations considérées. Par exemple, si l'on utilise un contour qui suit bien les spicules, ces dernières apparaîtront de manière linéaire dans l'image de bande.

- **Sélection de caractéristiques**

Un trop grand nombre de caractéristiques pour les individus à classifier peut nuire à la performance du classifieur. En effet, cela peut augmenter de manière non nécessaire l'espace de travail empêchant par la même occasion une généralisation efficace lors de l'apprentissage de la partie classification. Des heuristiques comme l'aire sous la courbe ROC du classifieur [35], permettent de modéliser la qualité d'un ensemble de caractéristiques. Cela permet d'utiliser des méthodes d'optimisation comme les algorithmes génétiques pour obtenir l'ensemble optimal. Ce genre d'approche a été mis en œuvre avec succès en mammographie [36].

III.2.1.5.Classification des anomalies

L'étape de classification a pour but de donner la réponse finale sur ce qui est détecté et considéré comme un signe de lésion. L'idée est de combiner les informations extraites précédemment pour obtenir une décision. On peut voir cela de manière similaire à ce que fait le radiologue pour prendre sa décision. Le problème de classification n'est pas propre à la détection de lésions dans le sein, en effet c'est un champ de recherche à part entière qui peut être utilisé dans des applications très différentes.

La phase de classification fait appel aux outils de classification supervisés de l’intelligence artificielle. Les techniques telles que les réseaux de neurones (RN), la logique floue (FL), les machines à vecteurs de support (SVM) et les méthodes de type plus proches voisins sont les plus couramment utilisées.

IV. Le cancer du sein

Chacun des organes de notre corps est constitué de *cellules*, différentes les unes des autres par la forme et la fonction. Les cellules ont la propriété de se diviser normalement pour produire plus de cellules. Chaque type de cellule a un cycle cellulaire spécifique et les cellules normales possèdent en elles-mêmes le signal de leur propre mort c’est ce que l’on appelle la mort programmée ou apoptose. Ces processus de croissance et de mort contribuent à nous maintenir en bonne santé [37].

Parfois, les cellules deviennent anormales, se divisent hors de tout contrôle, et échappent au mécanisme de leur mort programmée. Quand les cellules prolifèrent au-delà de ce qui est nécessaire pour remplacer les cellules mortes, il se forme un excès de tissu. Cet excès de tissu est une tumeur. Elle peut être *bénigne* ou *maligne* [37].

Dans le cas du cancer du sein, les cellules peuvent rester dans le sein ou se répandre dans le corps par les vaisseaux sanguins ou lymphatiques. La plupart du temps, la progression d’un cancer du sein prend plusieurs mois et même quelques années [38].

IV.1. L’anatomie du sein

IV.1.1. Description générale

Les seins sont des organes glandulaires paires et symétriques ont un rôle important dans l'image corporelle de la femme et l'exercice de sa féminité dans les champs affectif et social. La principale fonction naturelle du sein est la lactation, autrement dit la production de lait : elle a pour but de nourrir le nouveau-né.

Sur le plan superficiel, le sein est constitué de

- **Peau**
- **L'aréole**
- **Le mamelon**

Sur le plan interne, le sein est essentiellement constitué de

- **Lobules**
- **Les canaux galactophores**
- **Tissu conjonctif adipeux**
- **Le tissu glandulaire**

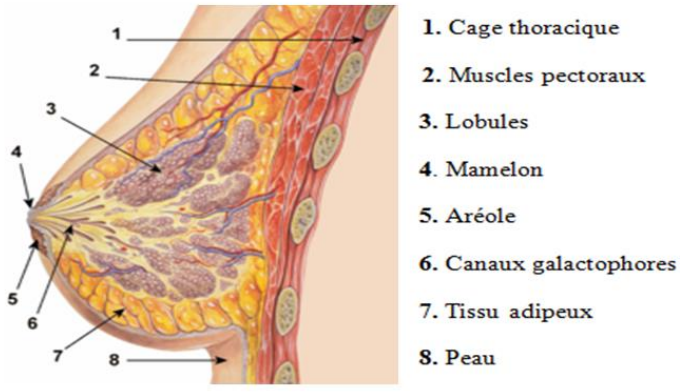

Figure I. 2 : Schéma anatomique du sein.

IV.1.2. Description détaillée

Le cancer du sein se développe essentiellement dans deux endroits : les lobules « *carcinome lobulaire* » et les canaux galactophores « *carcinome canalaire* ».

➢ **Canaux galactophores**

Les canaux galactophores, appelés aussi canaux mammaires, transportent le lait des lobules aux tétons. Ils sont formés de 2 couches de cellules entourant une cavité appelée lumina ou le lait et autres secrétions sont véhiculés. La couche interne est formée de cellules dites épithéliales tandis que la couche externe est formée de cellules myoépithéliales. Ces dernières sont adossées à une membrane appelée membrane basale qui sépare le tissu conjonctif des cellules épithéliales [39].

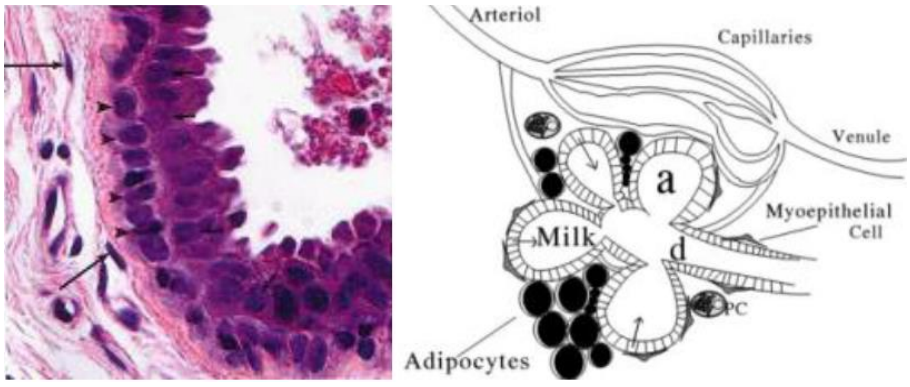

Figure I.3 : Vue des cellules myoépithéliales adossées à la membrane basale.

- **Les lobules**

Les lobules sont des glandes exocrines secrétant le lait. Ils sont composés de plusieurs alvéoles qui peuvent être décrites comme des sacs formés de cellules épithéliales. Chaque lobule sécrète du lait dans un canal commun, ce dernier rejoignant un canal galactophore principal [39].

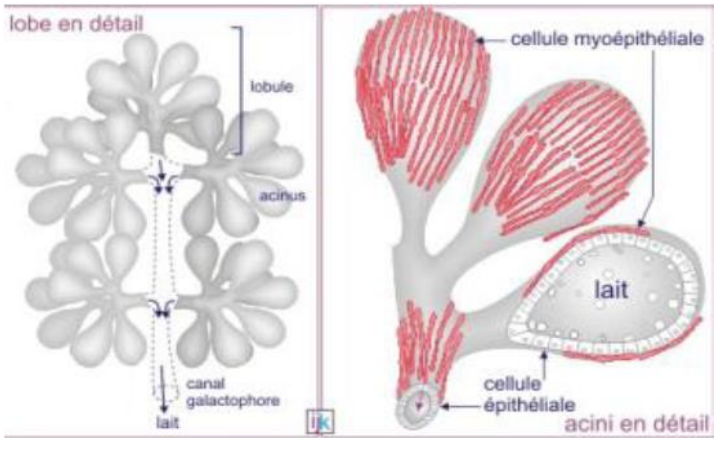

Figure I.4 : Schéma détaillé des lobules.

V. Pathologies mammaire

Les maladies des seins sont fréquentes et varient au cours de la vie génitale de la femme. Elles touchent l'organe lui-même (le sein et la glande mammaire qu'il contient), ou sa fonction (lactation, allaitement). Chez les hommes les maladies des seins sont exceptionnelles [40]. Comme nous l'avons vu précédemment les tumeurs peuvent être bénignes ou malignes.

V.1. Les maladies bénignes

Les tumeurs bénignes ne sont pas des cancers. On peut les enlever et, dans la grande majorité des cas, elles ne reviennent pas. Surtout les cellules des tumeurs bénignes

n'envahissent pas les autres tissus. Les tumeurs bénignes du sein ne présentent aucune menace vitale [37]. Parmi ces types de maladies nous mentionner quelques-uns:

- **Fibro-kystique** :

C'est une affection bénigne qui touche beaucoup de femmes de 30 à 50 ans. Des nodules et des kystes se forment à l'intérieur du sein, à la suite de la stimulation des hormones féminines.

Les symptômes peuvent être : un gonflement, de la douleur ou de la sensibilité aux deux seins, parfois un écoulement ; ces symptômes peuvent être particulièrement accentués durant la phase prémenstruelle et augmentés par l'ingestion de certains aliments (aliments comprenant de la caféine, du gras ou du sel) ou le port d'un soutien-gorge mal adapté [41].

A l'examen, on peut parfois trouver un ou plusieurs nodules plus gros. Les kystes peuvent apparaître et augmenter de volume rapidement. La maladie fibrokystique n'augmente pas le risque de développer un cancer du sein. Néanmoins, tout nouveau nodule devrait vous inciter à consulter un médecin qui verra si une investigation plus approfondie sera nécessaire. Pour soulager l'inconfort ou la douleur, des kystes peuvent être ponctionnés [41].

- **Fibroadénome :**

C'est une tumeur bénigne du sein à croissance lente qui touche en général les femmes de moins de 30 ans. Un nodule rond, unique ou multiple, ferme, bien délimité est formé, habituellement sans aucune douleur. Ce nodule peut grossir durant la phase prémenstruelle et peut alors être plus facilement palpable. En présence d'un tel nodule, une investigation doit être faite : examen clinique des seins, mammographie et/ou échographie. Pour les femmes de plus de 25 ans, on peut l'enlever. Pour les femmes plus jeunes, on peut simplement le surveiller. Il est rare qu'un fibroadénome mène au développement du cancer du sein.

- **Kyste :**

C'est une tumeur bénigne qui touche les femmes entre 30 et 50 ans. C'est un sac rempli de liquide, perçu comme une masse molle, parfois plus ferme, pseudo-solide, ou une succession de grains de plomb, parfois sensible [37]. Un kyste peut être unique ou multiple (ex : la maladie fibrokystique).

Il occasionne souvent malaise ou inconfort. Il peut apparaître et grossir rapidement. Un kyste peut être ponctionne par le médecin dans son cabinet, ce quii permettre un diagnostic rapide et la disparition du kyste [41].

Il existe d’autres types de tumeurs bénigne : lipome, adenolipome, tumeur phyllode, papillome intracanalaire, adénose et sclérose.

En présence d’un nodule découvert dans le sein, que ce soit par la femme elle-même ou par son médecin, une investigation est essentielle. Il faut d’abord procéder à un bon examen clinique des seins et réaliser une investigation radiologique [41].

V.2. Les maladies malignes

Les tumeurs malignes sont des cancers. Les cellules cancéreuses croissent et se divisent sans contrôles ; elles deviennent indifférencies, ce qui signifie qu’elles perdent les caractéristiques morphologiques et fonctionnelles du tissus d’origine. Elles peuvent envahir et détruire les tissus voisins et les organes. De plus, elles peuvent se mobiliser et pénétrer dans le sang ou le réseau lymphatique [37].

Le cancer du sein est une tumeur maligne de la glande mammaire. Autrement dit, c'est un cancer qui naît dans les unités cellulaires dont la fonction est de sécréter le lait, les unités ducto-lobulaires du sein, essentiellement chez la femme (le cancer du sein survient 200 fois moins souvent chez l'homme, qui possède lui aussi des seins, bien qu'atrophiés).

Le nombre de personnes atteintes a progressé légèrement, mais régulièrement, au cours des 3 dernières décennies. Par contre, le taux de mortalité a continuellement diminué au cours de la même période, grâce aux progrès réalisés en matière de dépistage, de diagnostic et de traitement [38].

V.2.1. Les symptômes

- La plus connue est la boule dans le sein.
- Un écoulement de liquide (marron ou jaune, parfois sanglant)
- Un changement de consistance du sein (plus dur, plus tendu).
- Les seins douloureux.
- Une fossette ou une ride creusant la surface du sein.
- Un aspect de "peau d'orange".
- Une déformation du mamelon le rétractant vers l’intérieur.
- Un aspect eczémateux du mamelon, qui devient rouge, croûteux ou érodé.

La constatation d’un de ces signes impose une consultation médicale sans retard. Le médecin jugera des examens complémentaires nécessaires.

VI. Les facteurs de risques

- ***L’âge***

L'âge moyen du diagnostic de cancer du sein se situe chez la femme de plus de 60 ans, ceci ne signifie pas qu'une femme plus jeune ne peut pas avoir un cancer du sein. Cela signifie tout simplement que, plus on avance en âge, plus on doit être attentif [37].

- ***Antécédent personnel***

Le fait d'avoir déjà eu un cancer à un sein, de cancer du l'ovaire, de l'utérus de colon… accroît le risque qu'un second cancer se forme.

- ***Antécédent familiaux***

L'histoire familiale et l'hérédité jouent aussi un rôle non négligeable. Une femme dont la mère, la sœur ou la fille est atteinte d'un cancer du sein a plus de risque d'en développer un à son tour.

- ***Autres facteurs***
  - Le sexe féminin (le sexe féminin, 99% des cancers du sein apparaissant chez la femme, 1% chez l'homme).
  - des premières règles avant 12 ans et une ménopause après 55 ans.
  - une absence de grossesse ou une première grossesse après 40 ans.
  - un traitement hormonal substitutif de la ménopause prescrit pendant plus de 10 ans.
  - une consommation exagérée d'alcool, de sucres et de graisses d'origine animale ainsi que l'obésité.

VII. Le dépistage

Il s'agit d'un ou de plusieurs processus permettant de déterminer l'existence infra clinique d'une tumeur au sein. Le dépistage par mammographie est considéré actuellement comme étant le moyen le plus fiable pour dépister le cancer du sein. Néanmoins, il ne faut pas négliger l'examen clinique des seins qui est une inspection mamelle et une palpation des seins ainsi que des aisselles par la femme elle-même. Ajoutant à cela l'ultrason et même la biopsie dans certains cas.

- *Autopalpation*

Elle regroupe un ensemble de gestes simples qui, pratiqués tous les mois, après les règles, vous permettront de bien connaître vos seins, de détecter toute anomalie et de la signaler à votre médecin.

L'autopalpation est une pratique très positive et importante pour toutes les femmes. Elle est d'abord d'une meilleure connaissance de son corps, de ses changements lors des diverses étapes de la vie.

- *Examen clinique des seins*

L’examen clinique des seins est un autre moyen de dépistage du cancer du sein. L’examen clinique du sein est recommandé tous les ans. Durant l’examen clinique, le médecin palpe les deux seins ainsi que la région des deux aisselles et des ganglions au-dessus et au-dessous de la clavicule. S'il percevait un nodule, c'est-à-dire une petite boule lui paraissant anormale, ou si différents éléments de votre état de santé l'y incitaient, il vous prescrirait alors une mammographie et/ou une échographie [42].

- *Mammographie*

La mammographie est le meilleur moyen de dépistage du cancer du sein. Elle permet de révéler des petites tumeurs qui ne sont pas toujours détectables par la seule palpation des seins. Elle est recommandée tous les deux ans pour les femmes de plus de 50 ans. Pour les femmes de moins de 50 ans, elle peut être recommandée dans certaines conditions. Demandez conseil à votre médecin. La mammographie n’est pas un examen infaillible mais elle permet tout de même de dépister 90% des lésions du cancer du sein. La mammographie détecte également des lésions autres que celles des cancers. C’est pour cela qu’il est nécessaire de passer d’autres examens pour s’assurer qu’il s’agit vraiment d’un cancer du sein [42].

- *L’échographie*

L’échographie est un examen complémentaire à celui de la mammographie, non utilisée en première intention pour le dépistage du cancer du sein mais seulement en cas d’anomalie mammographique. Il s’agit d’ultrasons qui permettent d’analyser plus en détails une tache « opacité » ou une anomalie architecturale qui a été détectée par une mammographie. L’échographie ne remplace pas la mammographie car il ne permet pas de détecter les micro-calcifications, très importantes en dépistage [43]. L’immense majorité des grosseurs identifiées en mammographie courante ne sont pas des tumeurs cancéreuses, mais des kystes bénins (apparaissant alors comme liquidiens en échographie) ou des lésions mammaires bénignes de type adéno-fibrome par exemple (apparaissant alors comme solides en échographie) [43].

- *Biopsie*

Biopsie est un prélèvement du tissu mammaire réalisé avec une aiguille ou un petit pistolet automatique. Cette aspiration se fait avec une simple anesthésie locale. Elle ne nécessite pas d’anesthésie générale et ne provoque pas d’effets secondaires. L’aiguille est introduite dans la tumeur qui a été détectée par un examen mammographique, échographique ou clinique [43]. L’analyse du prélèvement au microscope, après technique particulière nécessitant plusieurs heures

voire jours, permet de déterminer si le tissu est ou non cancéreux et donc de formuler un diagnostic définitif [43].

VII.1. Le diagnostic

La présence d'un cancer du sein doit être évoquée lors de l'autopalpation d'un nodule « boule » du sein ; lors de la découverte d'une anomalie à l'examen clinique annuel ou lors de la découverte d'une anomalie suspecte à la mammographie [44]. Le diagnostic d'un cancer du sein est porté lors de l'examen histologique d'un fragment du tissu tumoral. L'examen clinique, la mammographie ou l'échographie mammaire servent au dépistage du cancer du sein, mais seul l'examen histologique peut le confirmer avec certitude. L'examen clinique, l'échographie mammaire et la mammographie permettent également l'évaluation de la taille de la tumeur et donc de son stade [44].

Trois types principaux de prélèvements histologiques existent : micro-biopsie, macro-biopsie ou biopsie chirurgicale. La micro-biopsie (prélèvement d'un fragment de petit calibre) et la macro-biopsie (prélèvement d'un fragment de plus gros calibre) sont souvent pratiquées par un radiologue. Ces prélèvements sont habituellement réalisés sous guidage échographique ou mammographique (biopsie stéréotaxique), ce qui permet d'améliorer leur performance. Ils sont pratiqués sous anesthésie locale, sans hospitalisation. Les résultats sont communiqués quelques jours plus tard au médecin prescripteur qui les remet à son tour à la patiente [44].

VII.2. Les différents types de cancers

Différents types de cancer du sein existent. On peut les diviser en deux grades catégories les formes « invasives » et les formes « non invasives »

- *Les cancers non invasifs ou in situ*

Ce sont confines a l'intérieur des canaux ou des lobules mammaires ou ils ont pris naissance, et ne s'étendent pas dans les territoires voisins. On en connaît deux principaux types : le *carcinome canalaire in situ* (CCIS) et le *cancer lobulaire in situ* et le cancer lobulaire in situ (CLIS) [37].

Ces deux formes de cancers localisés présentent des taux de guérison très importants (supérieur a 90%) ils peuvent toute fois évoluer vers des formes plus sévères. Avec un tel diagnostic, vos chances de guérison définitive, sans récidive, sont très élevées à une condition : ces lésions doivent être traitées sans attendre [37].

- *Le cancer invasif ou infiltrant*

Le cancer invasif est un cancer qui a franchi les limites du canal ou lobule où il a pris naissance, et a envahi les tissus avoisinants. Beaucoup plus sévère dans son devenir que le cancer in situ, le cancer invasif peut disséminer dans tout l'organisme. Le traitement détaillé plus loin, est généralement long, faisant appel à différentes méthodes, et exposant à de dures épreuves pour les malades. Mais avec ces traitements, la guérison reste possible. Le plus fréquent se développe à partir des canaux excréteurs ou de lobules mammaires [37].

Le cancer invasif est le plus souvent découvert sous la forme d'un nodule mammaire visible sur un cliche de mammographie ou palpable lors de l'examen du sein par la femme elle-même ou son médecin [37].

Il existe aussi des formes rares de cancer du sein, comme la forme inflammatoire ou la maladie de Paget, qui se présentent différemment [37].

VII.3. Caractéristiques des différents types de cancer

Il existe plusieurs signes radiologiques associés à des cancers.

➢ *Les calcifications* **:** correspondent à de petits amas de calcium aussi connus sous le nom de **microcalcifications**. Elles ne sont pas forcément signe de malignité. Ces différentes formes sont illustrées de manière schématique à la figure suivante.

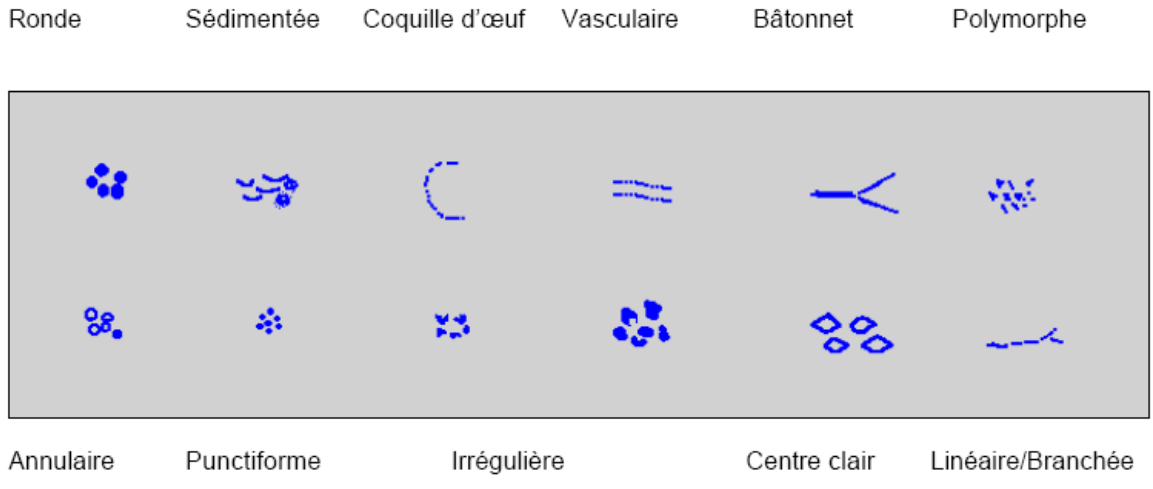

Figure I.5 : Différentes formes pour les microcalcification

Ainsi des grosses calcifications rondes à bord lisse sont bien souvent bénignes. Malheureusement, les microcalcifications malignes sont généralement moins visibles.

➢ *Les masses ou les opacités:* Une seconde famille de signes radiologiques se traduit par des sur-densités ou **opacités** dans les images. Ces dernières présentent une variabilité importante du point de vue de leur forme, de leur taille et de leur contour. Cette dernière caractéristique aide le radiologue dans son diagnostic différentiel. On distingue généralement quatre classes de formes pour les opacités : les opacités rondes, ovales, lobulaires ou irrégulières. Ces différentes formes sont illustrées de manière schématique à la figure suivante.

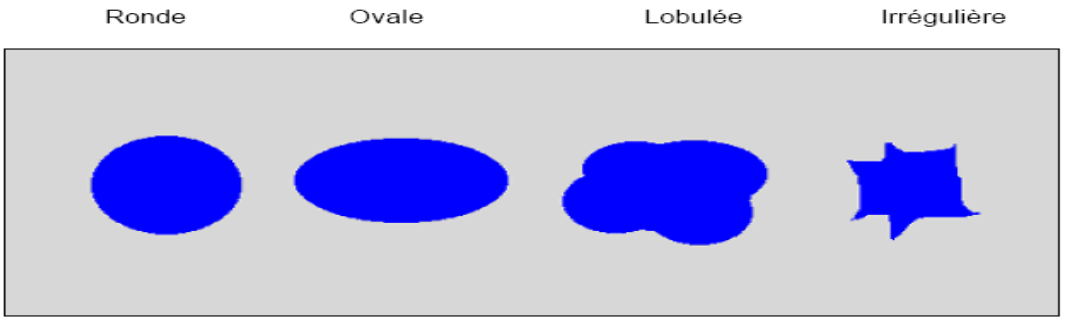

Figure I.6 : Différentes formes pour les opacités.

Comme indiqué précédemment, les opacités varient aussi au niveau de la définition de leurs contours. On dénombre cinq grandes classes de contours. La première correspond aux contours circonscrits qui sont des contours bien définis où la frontière entre lésion et fond est franche et qui correspondent en général à des lésions bénignes. Une seconde classe de contours est distinguée lorsqu'une partie du contour est masquée par la superposition ou l'adjacence de tissus normaux laissant penser à une lésion circonscrite dont une partie du contour est cachée. Une troisième possibilité pour le contour est d'être micro-lobulé, c'est-à-dire qu'il comporte de petites ondulations. Le quatrième type de contours regroupe les contours mal définis qui peuvent laisser penser à la présence d'infiltrations. Enfin une dernière classe regroupe les opacités dont les contours sont spiculés, c'est-à-dire comportant des structures filiformes qui rayonnent en s'éloignant du centre de l'opacité, et qui sont hautement suggestives de malignité.

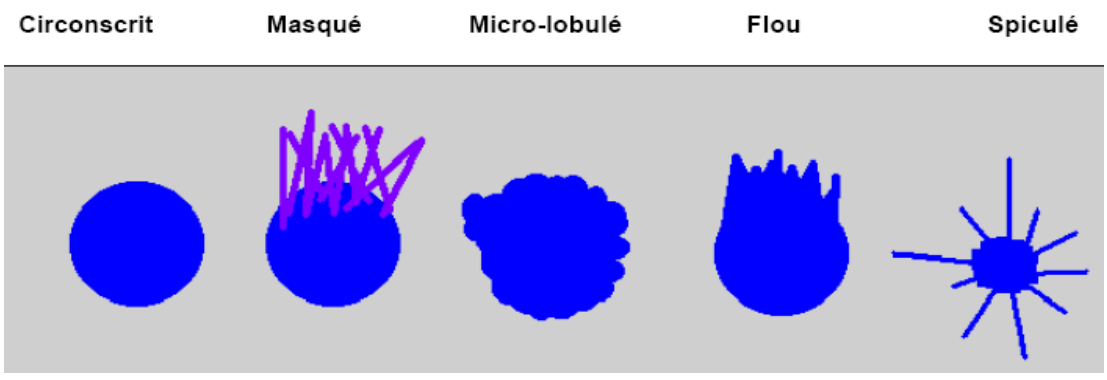

Figure I.7 : Différents types de contours pour les opacités.

➢ *distorsions architecturales*

La troisième forme de signes radiologiques traduisant la présence de cancer correspond aux **distorsions architecturales.** Ces dernières ne comportent pas de densité centrale comme c'est le cas pour les opacités précédemment décrites. Elles se traduisent par des structures linéaires ou des spicules qui convergent vers une même zone focale. Un exemple schématique de ce genre de signes, qui sont hautement suspects, est présenté à la figure suivante.

Figure I.8 : Forme schématique d'une distorsion architecturale.

VIII. La classification BI-RADS [45]

C'est une classification en cinq catégories des images mammographiques en fonction du degré de suspicion de leur caractère pathologique. D'après la classification BI-RADS (Breast Imaging Reporting and Data System) de l'ACR (American College of Radiology) adaptée par l'ANAES.

- **ACR 1** : Mammographie normale, rien à décrire.
- **ACR 2** : On signale des anomalies bénignes (fibroadénome, hamartome, lipomes, ganglions, cicatrices, calcifications bénignes) ne nécessitant ni surveillance ni examen complémentaire.

- **ACR 3** : Il existe une anomalie probablement bénigne dont la VPP est inférieure à 2% pour laquelle une surveillance à court terme est conseillée.
- **ACR 4** : Il existe une anomalie indéterminée ou suspecte, qui fait poser l'indication d'une vérification histologique. (la VPP est entre 3 et 94%).
- **ACR 5** : L'anomalie est évocatrice d'un cancer. (la VPP est supérieure à 95%).

VIII.1. La classification proposée

La classification des anomalies mammographiques proposée est inspirée du système BIRADS de l'ACR [45], des recommandations de l'ANAES (Agence nationale d'accrédiation et d'évaluation en santé) et d'une revue critique de la littérature.

Le but de la classification BIRAS de l'American College of Radiology et d'améliorer l'évaluation et la prise en charge des anomalies mammaires.il s'agit d'une méthode standardisée pour les comptes rendu.

➢ *Les opacités*

Classe	Opacité
2	1. Opacité ronde avec macro-calcifications 2. Ganglion intra mammaire 3. Images de densité graisseuse ou mixte 4. Opacité circonscrite et stable 5. Multiples* opacités circonscrites bilatérales 6. Opacité tubulaire ou branchée à contours nets évoquant un canal dilaté dans la région rétro-mamelonnaire
3	1. Opacité circonscrite, ronde / ovale 2. Opacité circonscrite à contours discrètement polylobés
4	1. Opacité circonscrite ayant augmenté de volume 2. Opacité circonscrite à contours microlobulés 3. Opacité d'allure circonscrite avec au moins un bord masqué
5	1. Opacité mal circonscrite : contours flous, irréguliers ou spiculée 2. Opacité + lésions dites associées

Tableau I.1 : Classification des masses.

✓ Les calcifications

Classe	Calcifications
2	1. Macrocalcif. 2. Calcif. de type vaculaire 3. Calcif. en aiguille brisées ou en bâtonnets (lithiase galactophorique) 4. Calcif. d'adiponécrose 5. Calcif. en pop-cor 6. Fils de suture calcifiés 7. Microcalcif. annulaires, arciforme semilunaires, sédimentées ou rhomboédriques 8. Calcif. cutanées 9. Microcalcif. Rondes et/ou punctiformes, régulières et diffuses 10. Foyer de microcalcif. de classe 3 stable depuis au moins 2 ans
3	1. Microcalcif. rondes et/ou punctiformes régulières, peu nombreuses et groupées en petit amas +/- rond. 2. Petit amas rond ou ovale de calcif. dystrophiques peu nombreuses évoquant un début de calcification post radique.
4	1. Microcalcif. rondes et/ou punctiformes régulières nombreuses et groupées 2. Microcalcif irrégulières, peu nombreuses et groupées. 3. Microcalcif de classe 3 en foyer évolutif
5	1. Microcalcif. vermiculaires, branchées et/ou coudées 2. Microcalcif irrégulières nombreuses et groupées 3. Microcalcif de topographie galactophorique 4. Microcalcif + désorganisation architecturale, opacité ou autres lésions dites associées

Tableau I.2 : Classification des calcifications.

✓ *Désorganisation architecturale*

Classe	Désorganisation architecturale
2	1. Cicatrice connue et stable (post traumatique, post chirurgie et/ou séquelle d'abcès)
3	1. Distorsion architecturale en dehors d'une cicatrice connue et stable 2. Image spiculée sans centre dense
4	1. Distorsion architecturale associée à une opacité, des microcalcif. ou une lésion dite associée

Tableau I.3 : Classification des désorganisations architecturales.

VIII.2. Le traitement du cancer

Le premier but du traitement du cancer du sein est d’enlever la tumeur et de supprimer toutes les cellules cancéreuses. Il existe différents types de traitements qui peuvent être réalisés seuls ou être associes entre eux.

- *Chirurgie*

La chirurgie est souvent le premier traitement entrepris. Elle sert à enlever la tumeur cancéreuse. Dans le cas du cancer du sein, elle se nomme *mastectomie*. La mastectomie est partielle ou segmentaire (une partie du sein est enlevée) ou totale (tout le sein est retiré). Le choix du type de mastectomie repose entre autres sur la taille de la tumeur, son type et son emplacement dans le sein. La préférence de la femme est aussi prise en compte [2].

- *Radiothérapie*

La radiothérapie est un traitement localise du cancer, réalisé après la chirurgie pour améliorer les chances de guérison. La radiothérapie consiste à utiliser des radiations ionisantes (rayons) qui vont détruire les éventuelles cellules cancéreuses qui auraient pu rester après la chirurgie du sein. Ces rayons sont de différentes énergies (photons, rayons X ou électrons) sont utilisés en fonction de la zone à traiter (profondeur, taille, radiosensibilité de la lésion). Ces différents rayons peuvent être associés entre eux au cours d'un même traitement.

- *Chimiothérapie*

La chimiothérapie a pour objectif de diminuer le risque de récidives et de métastases. Elle consiste à administrer des médicaments anticancéreux, à intervalle fixe, en général toutes les 3 semaines.

La chimiothérapie peut être « adjuvante » (réalisée après la chirurgie) ou néo-adjuvante (réalisée avant la chirurgie). Dans le cas de cancer du sein le traitement est adjuvant. En cas de tumeur avancée ou inflammatoire, il est parfois nécessaire de débuter le traitement par une chimiothérapie néo-adjuvante pour diminuer la taille tumorale et permettre éventuellement une chirurgie conservatrice.

Ses principaux effets secondaires sont la fatigue générale, les nausées, les vomissements et la chute temporaire des cheveux. Des produits adaptés sont administrés parallèlement à la chimiothérapie afin de réduire la pénibilité de ses effets

- *Hormonothérapie*

Dans le cas du cancer du sein métastatique, l'hormonothérapie peut être utilisée avant ou après la chimiothérapie selon les symptômes et l'endroit des métastases. Par rapport à la

chimiothérapie, l'hormonothérapie offre l'avantage de causer moins d'effets secondaires et d'être administrée sous forme de comprimés permettant un traitement à la maison. Pour pouvoir bénéficier de l'hormonothérapie, les cellules cancéreuses doivent être hormonosensibles. Cette information sur le statut hormonal de votre cancer a été obtenue lors de votre chirurgie et ne change pas au fil du temps.

Les médicaments peuvent diminuer la progression de la maladie et augmenter la survie. Différents critères peuvent guider votre médecin dans le choix du médicament à utiliser.

De nos jours le cancer du sein constitue un problème de santé publique tant dans les pays développés que dans les pays en développement. Cela à cause de sa fréquence, de sa morbidité et surtout de sa mortalité. Et pour cela la mammographie reste le moyen le plus utilise du dépistage précoce des cancers du sein car il permet, même le dépistage des tumeurs asymptomatiques et non palpables manuelle.

IX. Conclusion

Les images médicales sont pour les médecins une source d’informations précieuse et conséquente, permettant l’analyse et le suivi des pathologies. En effet, ce type d’imagerie fournit, de façon non invasive, des représentations contrastées et de haute résolution de l’anatomie des organes, permettant ainsi son observation in vivo. Cependant, l’étude systématique et « manuelle » de clichés est longue et fastidieuse, et souffre de la variabilité d’interprétation des spécialistes. Ainsi, de nombreuses recherches ont été menées afin d’automatiser l’analyse des images ; la segmentation en forme le point de départ. Dans notre travail, on s’intéresse aux étapes d’extraction des caractéristiques (ou de paramètres) et de classification des anomalies.

Chapitre 2

Les ontologies

I. Introduction

La rapidité de l'évolution de la masse d'informations dans tous les domaines a génère un besoin d'organisation et de structuration des contenus. Les ontologies servent à la représentation des données échangées dans un domaine particulier afin de faciliter la communication interne au système informatique et externe entre les différents acteurs du domaine. Leur utilisation peut varier de la représentation des données à la recherche d'informations.

L'utilisation d'ontologies en informatique vise à intégrer une couche de connaissances aux systèmes afin de permettre des traitements élaborés de l'information qu'ils manipulent.

La conception d'ontologies est une tâche difficile qui nécessite la mise en place de procédés élaborés afin d'extraire la connaissance d'un domaine, manipulable par les systèmes informatiques et interprétable par les êtres humains.

Dans cette section, nous effectuons une présentation générale des ontologies. Nous définissons d'abord la notion d'ontologie d'un point de vue philosophique puis informatique, et les raisons d'apparition de cette notion. Ensuite nous décrivons ses différents constituants ainsi que ses principaux types. Nous terminons cette présentation en évoquant les méthodes de construction d'ontologie.

II. Définition

➢ De point de vu philosophique (ou métaphysique)

Le terme ontologie vient du grec : ***ONTOS*** « l'être » et ***LOGOS*** « sciences, étude » c'est-à-dire la science ou la théorie de l'être. Il est originaire de la *philosophie aristotélicienne* il y a plus de 2000 ans.

Une autre définition est comme suite : « Science de l'être en tant qu'être, c'est-a-dire de l'être en général, de ses diverses espèces, de ses propriétés et de ses relations [46].

➢ De point de vu informatique

En informatique dans le domaine de l'Intelligence Artificielle, et plus particulièrement en Ingénierie des connaissances, les ontologies traitent de la représentation des connaissances d'un domaine particulier à travers l'utilisation de concepts et de rôles. L'intérêt de l'utilisation d'ontologies par des programmes et/ou des agents logiciels est qu'elle permet de partager la compréhension d'un domaine afin de pouvoir obtenir une interprétation sans équivoque ou ambiguïté de la signification des ressources présentent sur le web. Il existe plusieurs définitions de la notion d'ontologie en informatique :

La première définition est celle énoncée par ***Thomas Gruber*** [47] c'est la définition la plus communément admise, il définit « une ontologie comme une spécification explicite d'une conceptualisation », La conceptualisation fait référence à un modèle (au sens ensemble structuré) de concepts ce qui signifie que tous les types primitifs, les concepts et les contraintes qui sont utilisés dans les spécifications de l'ontologie sont explicitement définis.

Willem Nico Borst [48] affine cette définition en présentant « une ontologie comme une spécification explicite et formelle d'une conceptualisation partagée ». Dans une spécification formelle, l'ontologie est interprétable sans ambiguïté par une machine, L'ontologie doit donc être traduite dans un langage formel et opérationnel de représentation des connaissances, ce processus de traduction est appelé opérationnalisation.

III. Pourquoi les ontologies ?

Les facteurs pouvant menés au développement des ontologies sont divers, on peut citer entre autres [49] :

- Partager la compréhension commune de la structure des informations entre les personnes ou les fabricants de logiciels.
- Permettre la réutilisation du savoir sur un domaine.
- Capturer, modéliser, transformer une connaissance contextualisée.
- Présenter une organisation hiérarchique ou une taxonomie des concepts pertinents et des relations qui existent entre ces concepts.
- Traiter et automatiser les raisonnements de/sur cette connaissance.
- Expliciter ce qui est considère comme implicite sur un domaine.
- Distinguer le savoir sur un domaine du savoir opérationnel.

IV. Les aspects d'une ontologie

Une ontologie est une conceptualisation qui doit être :

IV.1. Formelle

Basée sur une représentation explicite, formelle, axiomatique et systématique des propriétés du modèle permettant un certain niveau de raisonnement automatique (l'ontologie doit être représentée sous une forme compréhensible et traitable par la machine) et doit adopter des principes rigoureux de modélisation des connaissances, ce qui

permettra d'assurer une distinction formelle entre les éléments d'un domaine, indépendamment de leur réalité.[50].

IV.2. Consensuelle

L'ontologie est une conceptualisation qui doit faire l'objet d'un consensus de la part d'une communauté plus large que celle qui a participé à son développement dans le cadre d'une application particulière. L'ontologie ainsi définie dans ce cadre est validée et ses concepts pourront être utilisés pour se comprendre [50].

IV.3. Référençable

Chaque concept défini dans le cadre d'une ontologie est associé à un identificateur permettant ainsi de le référencer à partir de n'importe quel autre environnement, indépendamment du modèle d'ontologie où ce concept a été défini. Ceci se fait dans un souci de réutilisabilité, l'ontologie doit être partagée pour pouvoir être exploitée par différentes applications ou des agents logiciels [50].

V. Les composants d'une ontologie

Les ontologies permettent de spécifier les connaissances d'un domaine particulier. Cette spécification des connaissances (ou modélisation des connaissances) se fonde sur un ensemble de constituants classes, relations entre concepts, fonctions, axiomes et instances [51].

V.1. concepts

Les concepts sont utilisés pour représenter les objets sur lesquels portent les connaissances à spécifier. Un concept décrit une notion et peut être décomposé en quatre parties : une intension, une extension, une terminologie et un identifiant.

- ✓ **L'intension** correspond à la définition formelle du concept. Elle présente un ensemble de caractéristiques (propriétés et attributs) qui expriment la sémantique du concept [52]
- ✓ **L'extension** correspond à l'ensemble des êtres que le concept englobe. Ces êtres possèdent en commun les caractéristiques définies par l'intension. Chacun de ces êtres correspond à une production du concept. Nous appellerons production lexicale du concept une représentation en corpus d'un élément de son extension ;
- ✓ **La terminologie d'un concept** correspond aux termes qui lexicalisent la notion décrite par le concept. Usuellement elle est constituée d'un terme vedette (celui qui est le plus employé ou qui doit être employé) et de ses synonymes et variantes. Notons qu'un

terme peut être ambigu, c'est-à-dire qu'il lexicalise plusieurs notions, chaque notion correspondant à un sens du terme. Dans ce cas, il convient au concepteur de l'ontologie de statuer sur le sens à retenir pour le terme dans le contexte de l'ontologie (c'est-à-dire en fonction des concepts présents dans celle-ci)

- ✓ **L'identifiant** (ou libellé formel) permet d'identifier le concept dans le modèle. Il correspond à une unité lexicale qui nomme le concept. L'identifiant est généralement choisi dans la terminologie du concept (il s'agit en général du terme vedette). Il est néanmoins préférable de distinguer l'identifiant du concept et le terme afin d'éviter les confusions entre les représentations conceptuelles et linguistiques d'une notion. Une solution consiste à utiliser une unité lexicale dérivée du terme vedette comme par exemple le terme vedette préfixé par « concept » ou « C_ ».

V.2. Les relations

Une relation correspond à un lien s'établissant entre des concepts et décrit un type d'interaction entre ces concepts. Dans les ontologies, les relations sont généralement binaires. Une relation peut se décomposer en plusieurs éléments : une intension, une extension, une terminologie, un identifiant et une signature [53].

- ✓ **L'intension** d'une relation exprime la nature de la relation (équivalente au type d'interaction qu'elle décrit). Elle s'exprime par l'ensemble des attributs et des propriétés communes à toutes les réalisations de cette relation
- ✓ **L'extension** d'une relation correspond à l'ensemble des réalisations de cette relation dans le domaine modélisé
- ✓ **La terminologie** d'une relation correspond aux termes qui lexicalisent la notion décrite par la relation. Comme pour les concepts un terme peut être ambigu c'est-à dire lexicaliser plusieurs relations (par exemple "appeler quelqu'un " peut signifier lui téléphoner, lui donner un nom ou le héler pour le faire venir). Dans ce cas, c'est au concepteur de déterminer, en fonction de l'ontologie, à quelle relation le terme doit faire référence
- ✓ **L'identifiant** identifie une relation par un terme traduisant la nature de la relation. Ce terme est généralement issu ou inspiré d'un ou plusieurs termes de la terminologie
- ✓ **La signature** d'une relation correspond aux concepts qui peuvent être liés par une réalisation de la relation dans le domaine modélisé. Elle s'exprime par un n-uplet de concepts.

V.3. Les fonctions

Elles constituent un cas spécial de relations dans lesquelles le nième élément de la relation est unique est défini en fonction des n-1 éléments précédents

V.4. Les axiomes

Les axiomes servent à modéliser des assertions toujours vraies dans le domaine (acceptées sans démonstration) qui se traduisent sous la forme de propriétés sur les concepts et/ou les relations. Les axiomes sont utilisés pour représenter des connaissances élémentaires du domaine qui ne peuvent être formellement définies par les autres constituants d'une ontologie. C'est le cas des fonctions qui imposent des contraintes sur le modèle. Par exemple, « mère-biologique-de » est une fonction contraignant que la mère biologique d'une personne soit toujours une femme. Les axiomes permettent ainsi de restreindre l'interprétation des concepts et des relations dans le modèle.

Ils peuvent être inclus dans une ontologie pour plusieurs buts, tels que définir la signification de composants d'ontologie, définir les contraintes complexes sur les valeurs des attributs, les arguments de relations, etc., vérifier l'exactitude d'informations indiquées dans l'ontologie ou déduire une nouvelle information.

V.5. Les instances (individus)

Elles regroupent à la fois les instances des concepts et celles des relations. Les instances d'un concept sont utilisées pour représenter les éléments spécifiques, singuliers. Une instance de relation est un n-uplet d'instances de concept (n le nombre de concepts de la signature). Par exemple « les individus *Mouhcen* et *Khadija* sont des instances du concept *Personne* ».

VI. Les différents types d'ontologies

Nous pouvons catégoriser les ontologies selon leur formalisation, leur type de conceptualisation, leur propos, leur niveau de complexité, et selon le nombre de points de vue des concepteurs, etc.

VI.1. Selon le degré de formalisme

Selon l'état de la formalisation Uschold et Gruninger [54] ont distingué quatre types d'ontologies selon le type de langage utilisé pour les implémenter. Selon eux, il existe :

- **des ontologies fortement informelles** si elles sont écrites dans un langage naturel
- **des ontologies semi-informelles** si elles sont exprimées dans une forme restreinte et structurée de langage naturel (c.-à-d. en utilisant des modèles)

- **des ontologies semi-formelles** si elles sont définies dans un langage artificiel et formellement
- **des ontologies rigoureusement formelles** si elles sont définies dans un langage avec la sémantique formelle, les théories et les preuves des propriétés telles que la solidité et la perfection.

VI.2. Selon les objets modélisés

Les ontologies ont été aussi regroupées dans [53] en se basant sur les objets modélisés par l'ontologie afin de répondre à un but précis (voir la **Figure II.1**)

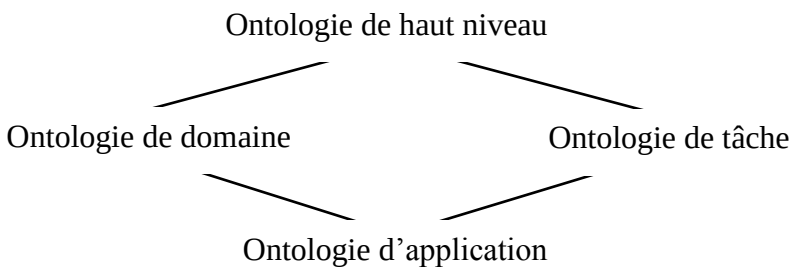

Figure II.1: Classification des ontologies selon Guarino.

- **Les ontologies supérieures**

Dites aussi de haut niveau, génériques, méta-ontologies. Elles sont universelles, réutilisables et référençable à partir des concepts des autres niveaux d'ontologies (voir la **Figure II.2**). Elles comportent des concepts abstraits (généraux) subsumant les concepts existant dans les différents domaines.

Une ontologie de haut niveau est généralement conçue afin de réduire les incohérences des termes définis plus bas dans la hiérarchie.

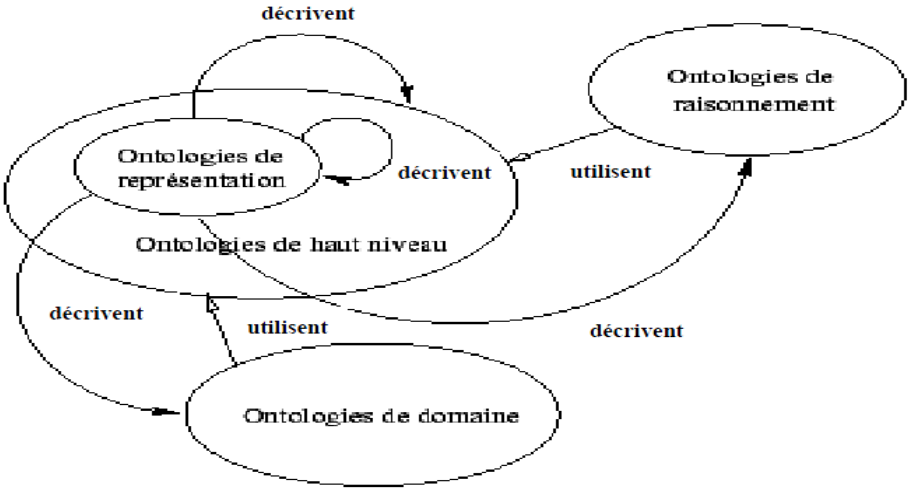

Figure II.2 : Schéma représentatif des relations entre les ontologies de domaine, de raisonnement, de représentation

- **Les ontologies de domaine**

Les ontologies de domaine décrivent les connaissances dans un domaine particulier (domaine de l'automobile, du monde hospitalier, etc.). Ces ontologies fournissent les vocabulaires sur les concepts et leurs relations dans un domaine, sur les activités qui ont lieu dans ce domaine, et sur les théories et les principes élémentaires régissant ce domaine.

Il y a une frontière claire entre les ontologies de domaine et les ontologies de haut niveau. Les concepts dans les ontologies de domaine sont habituellement des spécialisations des concepts déjà définis dans les ontologies de haut niveau, et le même principe pourrait se produire avec les relations.

De plus en plus d'ontologies de domaines sont disponibles : ontologie de la génétique, ontologie de la géométrie, ontologie pour les musées, ontologie médicale, ontologie pour l'enseignement, ontologie pour le bâtiment, ontologie de systèmes documentaires, ontologie pour la gestion, ontologie dans le secteur automobile, etc.

- **Les ontologies de tâche**

Ces ontologies décrivent le vocabulaire relié à une tâche générique ou à une activité (comme diagnostiquer, programmer, vendre, etc.) en spécialisant les termes dans les ontologies de haut niveau. Les ontologies de tâche fournissent un vocabulaire systématique des termes utilisés pour résoudre les problèmes liés aux tâches qui peuvent ou ne peuvent appartenir au même domaine. Ce sont des ontologies dépendantes des applications. Les

ontologies de tâche de domaine ("Domain-Task ontologies") sont des ontologies de tâche réutilisables dans un domaine donné, mais pas à travers des domaines. Elles se veulent indépendantes des applications.

- **Les ontologies d'application**

Elles sont spécifiques et non réutilisables. Elles contiennent des connaissances du domaine nécessaires à une application donnée [55]. Elles sont souvent des spécialisations des ontologies de domaine et des ontologies de tâche

- **Les ontologies de représentation**

Elles sont les ontologies les plus spécifiques. Ce type d'ontologies est un cas particulier d'ontologies supérieures qui regroupe des concepts déjà utilisés pour formaliser les connaissances. Indépendamment des domaines puisqu'elles décrivent des primitives cognitives communes. Elles utilisées pour formaliser un modèle de représentation des connaissances.

VII. Construire une ontologie

VII.1. Étapes de construction d'une ontologie

La construction d'une ontologie est un travail réalisé conjointement par un ou plusieurs ingénieurs (ontologues) et des experts du domaine, ainsi qu'éventuellement de futurs utilisateurs de l'ontologie. De nombreuses méthodes de construction d'ontologie existent mais, de manière générale, ce processus peut se décomposer en trois étapes successives : la conceptualisation, l'ontologisation et l'opérationnalisation.

- **L'étape de conceptualisation** consiste à identifier les connaissances du domaine à représenter dans l'ontologie. Il s'agit ici de choisir parmi l'ensemble des connaissances du domaine présentes dans les corpus ou chez les experts, celles qu'il convient de modéliser. Les connaissances choisies sont ensuite conceptualisées, c'est-à-dire exprimées sous forme de concepts, de relations, d'axiomes ou d'instances. Cette expression est réalisée en langue naturelle la plupart du temps
- **L'étape d'ontologisation** consiste à structurer et à formaliser le modèle issu de la conceptualisation et exprimé en général en langue naturelle. Le but de cette étape est d'obtenir une ontologie spécifiant formellement les connaissances à représenter. La formalisation est plus ou moins complète en fonction du formalisme choisi. Cette phase doit être réalisée par l'ontologue aidé des experts
- **L'étape d'opérationnalisation** a pour objectif de traduire le modèle dans un langage de représentation dit « opérationnel », c'est-à-dire un formalisme utilisable dans un processus

automatique. Cette étape est inutile si le formalisme utilisé dans la phase d'ontologisation est opérationnel.

VII.2.quelques méthodes de construction d'ontologies

Il n'existe pas qu'un seul mode ou qu'une seule méthode « correcte » pour développer des ontologies. Il existe plusieurs méthodes comme :

- **L'approche SENSUS**

C'est une méthode qui consiste à construire une ontologie de domaine à partir d'une plus grande ontologie, l'ontologie SENSUS [56]. La méthode propose de relier les termes spécifiques du domaine à cette ontologie et de tailler dans SENSUS, les termes qui ne sont pas relevant dans la nouvelle ontologie qu'on souhaite construire. Le résultat de ce processus est le squelette de cette nouvelle ontologie, qui est générée automatiquement en utilisant ce processus et l'outil OntoSaurus.

Conformément à cette méthode, la construction une ontologie dans un domaine spécifique doit suivre les étapes suivant :

- *Première étape* : identifier les termes clés du domaine.
- *Deuxième étape* : relier manuellement les termes clés à SENSUS.
- *Troisième étape* : inclure tous les concepts qui se trouvent sur le chemin depuis le terme clé jusqu'à la racine de SENSUS.
- *Quatrième étape* : ajouter les nouveaux termes de domaine. Dans cette étape on ajoute manuellement tous les termes qui sont relevantes pour le domaine et qui ne sont pas encore apparue. Ainsi les étapes 2 et 3 sont répétées pour inclure les concepts sur le chemin, des nouveaux concepts jusqu'à la racine de SENSUS.
- *Cinquième étape* : Ajouter le sous arbre entier. Dans cette étape, celui qui fait l'ontologie doit faire attention aux nœuds qui ont un grand nombre de chemins qui passe à travers le nouvel arbre généré.

- ***La methodology On-To-Knowledge (OTK)***

Est une méthodologie qui recommande un procède itératif de développement, elle comporte quatre phases principales : une phase de spécification de condition, une phase d'amélioration, une phase d'évaluation et une phase d'application et d'évolution.

Cette méthode propose l'acquisition des connaissances en spécialisant une ontologie générique, et finalement recommande la suite d'outils Onto Studio comme système de développement des ontologies. Le principal auteur est GRUNINGER **[53]**.

- ***METHONTOLOGY***

Cette méthodologie a été développée au sein du groupe d'ontologie à l'université polytechnique de Madrid. Dans cette méthodologie le cycle de vie d'une ontologie constitue de plusieurs étapes: la spécification, la conceptualisation, la formalisation, l'intégration, l'implémentation et la maintenance. Chaque étape peut intégrer une phase d'acquisition des connaissances, documentation et évaluation. Ce procédé cyclique rende cette méthode très attractive pour construire des ontologies dynamiques. METHONTOLOGY recommande la suite d'outils Protege 2000 et webODE. Le principal auteur est M. FERNANDEZ [57].

VII.3. Usages des ontologies

Les ontologies, au début, ont connu une large utilisation dans le domaine de l'intelligence artificielle. Aujourd'hui les ontologies sont largement utilisées pour différents buts (traitement de langage naturel, gestion de la connaissance, e-commerce, intégration intelligente d'information, le Web sémantique, etc.) dans différentes communautés (ingénierie de connaissance, bases de données, génie logiciel, etc).

Notons que la communauté de bases de données aussi bien que la communauté de conception orientée objet construisent également des modèles de domaine en utilisant les concepts, les relations, les propriétés, etc., mais la plupart du temps les deux communautés imposent des contraintes moins sémantiques que celles imposées dans les ontologies lourdes.

Dans le commerce électronique, les ontologies sont habituellement utilisées pour représenter les produits et les services qui sont offerts dans les systèmes du e-commerce et qui sont donnés aux utilisateurs dans les catalogues pour consultation [58].

La vision du Web sémantique est d'ajouter la sémantique compréhensible par la machine (méta- information) au World Wide Web en utilisant une ontologie pour définir et organiser cet espace de méta-information [58].

Le Web sémantique vise à réaliser l'intégration de toutes les sources d'informations sur le Web, permettant la réutilisation des données à travers les applications et rendant la recherche intelligente sur Internet possible [58].

Comme synthèse, nous pouvons dire que les ontologies possèdent un grand nombre d'applications et d'usages tels que :

- ✓ Les systèmes de Recherche d'informations (RI).
- ✓ Les inférences, pour découvrir les incohérences sur les connaissances ou les données.
- ✓ Les systèmes à base de connaissances dédiés à la résolution de problèmes.
- ✓ L'aide au diagnostic : des maladies, des problèmes techniques etc.
- ✓ L'interopérabilité des systèmes d'information.
- ✓ La coopération des logiciels.
- ✓ Les systèmes de traitement automatique du langage naturel.

Si les principaux projets utilisant des ontologies ne visent pour le moment que la gestion de connaissances au niveau sémantique, les ontologies pourraient permettre à terme la création de systèmes capables non seulement de gérer des connaissances mais aussi de raisonner sur ces connaissances et, pourquoi pas, d'en produire de nouvelles.

VIII. Les langages de représentation

Une des principales décisions à prendre dans le procédé de développement d'ontologies consiste à choisir le langage dans lequel l'ontologie sera exprimée et utilisé, parmi ces langages on peut citer :

VIII.1. RDF

RDF (Ressource Description Framework) développé et recommandé par le W3C, permet de décrire les ressources du web sémantique qui sont l'élément de base de RDF. Chaque ressource est pourvue d'un identifiant URI (Uniform Resource Identifier).

Tout document RDF est composé d'un ensemble de triplets (sujet, prédicat, objet) ou encore (ressource, propriété, valeur). Un ensemble de tels triplets est appelé un graphe RDF. Ceci peut être illustré par un diagramme composé de nœuds et d'arcs orientés, dans lequel chaque triplet est représenté par un lien nœud-arc-nœud (d'où le terme de "graphe").

A ce modèle est associée une syntaxe écrite en XML et basée sur les triplets :

- ➢ Ressource (Sujet) : une entité d'informations pouvant être référencée par un identificateur. Cet identificateur doit être une URI.
- ➢ Propriété (prédicat) : l'attribut ou la relation utilisé (e) pour décrire une ressource.
- ➢ VFaleur (objet) : la valeur d'une propriété associée à une ressource spécifique

VIII.2. RDFS

Afin de renforcer ce langage, RDF Schéma a été construit par W3C comme extension de RDF comportant des primitives basées sur des frames. RDF Schéma permet notamment

de déclarer les propriétés des ressources ainsi que le type des ressources. La combinaison de RDF et RDF Schéma est connue sous le nom RDF(S). Bien que relativement limités dans la mesure où ils ne sont pas très expressifs, les langages RDF(S) peuvent cependant spécifier des concepts, des taxonomies et des relations binaires.

VIII.3. DAML-OIL

DAML-OIL a été proposé par le W3C pour représenter des méta-données et des ontologies. DAML a été transformé en DAML+OIL en intégrant certaines propriétés d'OIL. Il repose sur RDF et RDF schéma et fournit en plus des primitives plus riches issues de la logique de description

VIII.4. OWL

OWL (Web Ontology Language) est le standard actuellement recommandé par W3C pour représenter les ontologies. C'est une extension du vocabulaire de RDF(S). Il est dérivé du langage d'ontologie DAML+OIL.

OWL a été fractionné en fait en trois sous-langages distincts offrants une complexité croissante et destinés à des communautés différentes d'utilisateurs. Chacun est une extension par rapport à son prédécesseur plus simple. Ils sont caractérisés par l'expressivité croissante (d'OWL Lite à OWL FULL) et la calculabilité des inférences décroissantes (d'OWL Lite à OWL FULL).

- **OWL LITE:** est d'expressivité faible par apport aux autres sous langages, mais qui reste quand même suffisant pour des utilisateurs qui ont principalement besoin d'une hiérarchie de classification et de contraintes simples. Par exemple, malgré qu'il permette d'exprimer des contraintes de cardinalité, il limite leurs valeurs à 0 ou 1. Une cardinalité 0 ou 1 correspond à des relations fonctionnelles, par exemple, une personne a une adresse. Toutefois, cette personne peut avoir un ou plusieurs prénoms, OWL Lite ne suffit donc pas pour cette situation. Ainsi, le langage offre une calculabilité maximale (temps de calcul acceptables), ce qui compense sa faible expressivité.
- **OWL DL:** nommé DL car il correspond à la logique descriptive. Il est d'expressivité maximale sans perte de calculabilité. Il convient aux utilisateurs qui veulent une expressivité maximale tout en maintenant la complétude de calcul (toutes les inférences sont garanties calculables) et la décidabilité (tous les calculs s'effectuent dans un temps fini). OWL DL inclut tous les constructeurs du langage OWL, mais ils sont utilisables seulement sous certaines restrictions pour garantir la décidabilité des calculs. Par exemple,

lorsqu'une classe est sous classe de plusieurs classes, elle ne peut pas être instance d'une autre classe.

- **OWL FULL:** offre un maximum d'expressivité. Il a l'avantage de la compatibilité complète avec RDF/RDFS, mais aucune garantie concernant la calculabilité (la complétude et la décidabilité des calculs liés à l'ontologie) n'est offerte par ce langage.

IX. Conclusion

La diversité et la puissance des applications potentielles des ontologies laissent à penser que leur place au sein des systèmes d'information ne peut que croître.

Plusieurs chercheurs se sont intéressés au problème de représentation d'un domaine par une ontologie, et cela en développant des approches et des outils pour sa construction et son enrichissement. Dans le chapitre suivant, nous allons construire une ontologie dans le domaine médical qui va servir à la classification des tumeurs du cancer de seins.

Chapitre 3

Modélisation du système du dépistage du cancer de sein

I. Introduction

La stratégie de classification automatique du cancer du sein reste une stratégie difficile à résoudre, malgré les efforts entrepris par les chercheurs. La difficulté revient de plusieurs paramètres (la variabilité des images mammographiques, la variété des types de cancers qui existe, la variété de leur format, etc).

Dans ce chapitre, nous exposons l'architecture de notre système dédié à la classification automatique des microcalcifications et d'opacités à partir d'images mammographiques, nous présentons notre approche qui se base sur les ontologies.

II. Description générale du système

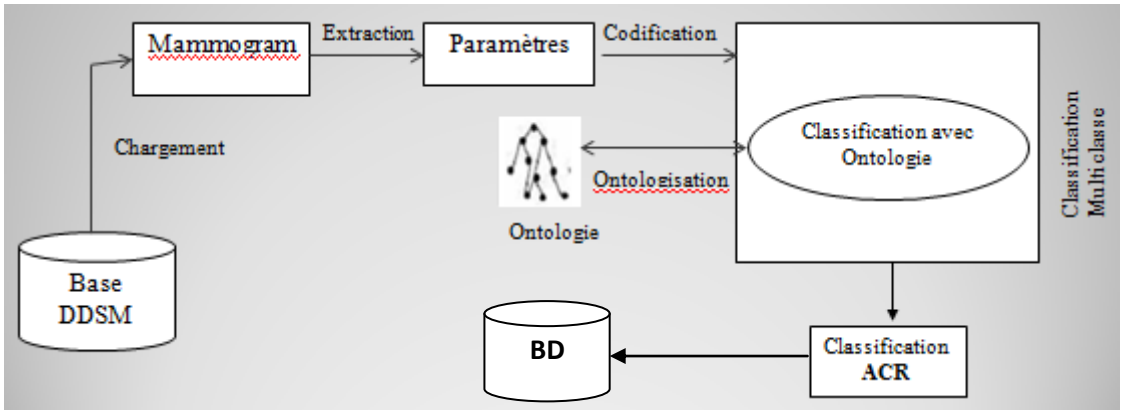

Figure III. 4 : Schéma général de l'approche.

Un ensemble de paramètres est extrait de l'image qui est sauvegardée dans les fichiers ***ics*** et ***Overlay*** *(Base DDSM)*. Ces paramètres extraits sont chargés dans l'ontologie. Cet ensemble de paramètres constitue l'entrée du système. Partant de ces paramètres, le système infère de nouvelles connaissances. Ces connaissances inférées concernent une classification des tumeurs. A la fin du traitement, des informations sont stockées sur une base de données, pour alléger l'ontologie et servir pour les statistiques.

II.1. Extraction des paramètres

Après la phase de chargement, l'étape suivante est l'extraction des paramètres qui décrivent les régions de l'image. Les paramètres à extraire dépend du problème traité et des méthodes utilisées. Dans notre cas, nous décrirons les critères qui seront considérés dans le diagnostic pour différencier les foyers de micro-calcifications et les masses malignes de bénins.

Les paramètres choisis pour le procédé de classification devraient refléter la spécificité des micro-calcifications et les masses apparaissant dans la mammographie aussi précise que possible. Le choix des paramètres est important pour effectuer une classification efficace et distinguer les masses et les foyers de micro calcifications suspectes. Notre choix est guidé par un expert du domaine. C'est un médecin spécialiste.

Il existe plusieurs paramètres pour décrire les masses et les foyers de micro-calcifications. Dans ce travail, nous avons opté pour les paramètres les plus pertinents illustrés dans le tableau suivant.

Paramètres	Signification
Forme de masse	le type de masse
Contour de masse	les bords de la tumeur
Morphologie de calcification	le type de calcification
Distribution de calcification	la répartition de calcification
La densité	la densité de la région
Assessment	le degré de suspicion
Subtilité	le degré de subtilité
Age de la patiente	l'âge de la patiente

Tableau III. 1 : paramètres utilisés pour la classification.

II.2. Base de données mammographies

Pour mesurer les performances, tester et évaluer notre système de classification on a besoin d'un référentiel qui s'appelle « Benchmark ». Dans le domaine d'aide au diagnostic du cancer du sein, il existe plusieurs bases d'images : MIAS (Mammographic Image Analysis Society), DDSM (Digital Database for Screening Mammography), AMDI (Indexed Atlas of Digital Mammograms) et WDBC (Wisconsin Diagnosis Breast Cancer). Les mammographies étudiées proviennent de la base de données DDSM.

- **DDSM**: La base de données Marathon de l'université de la Floride du Sud. Une description de cette base a été effectuée par « American college of Radiology » dans le lexique de BI-RADS (Breast Imaging Reporting and Data System). La base correspondante contient 2620 dossiers de patientes classés en trois groupes : normaux, bénins et malins. Chaque dossier contient 4 fichiers dont les plus importants sont les fichiers ics et Overlay.

II.2.1. Description de fichier ics

Le fichier ics contient des informations importantes telles que la date de l'étude, l'âge du patient, la date de la numérisation des films, le type de numériseur utilisé, une densité du tissu mammaire, la taille de chaque fichier image, le nombre de bits par pixel, la résolution de la numérisation, etc. La figure suivante illustre un exemple de ce fichier.

```
ics_version 1.0
filename A-1141-1
DATE_OF_STUDY 19 4 1991
PATIENT_AGE 53
FILM
FILM_TYPE REGULAR
DENSITY 4
DATE_DIGITIZED 17 6 1998
DIGITIZER HOWTEK 43.5
SEQUENCE
LEFT_CC LINES 5641 PIXELS_PER_LINE 2671 BITS_PER_PIXEL 12
RESOLUTION 43.5 NON_OVERLAY
LEFT_MLO LINES 6001 PIXELS_PER_LINE 2821 BITS_PER_PIXEL 12
RESOLUTION 43.5 NON_OVERLAY
RIGHT_CC LINES 5896 PIXELS_PER_LINE 2761 BITS_PER_PIXEL 12
RESOLUTION 43.5 OVERLAY
RIGHT_MLO LINES 5971 PIXELS_PER_LINE 3256 BITS_PER_PIXEL 12
RESOLUTION 43.5 OVERLAY
```

Figure III. 5: Le fichier A-1141-1.ics.

II.2.2. Description du fichier overlay

Les cas anormaux ont entre un et quatre fichiers overlay, ceux ci dépendent du nombre d'images que le radiologue marque comme sans anormalités. Pour chaque anormalité, on a des informations sur le nombre d'anormalités, le type, la forme et les bords de la tumeur présente, le degré de suspicion, le degré de subtilité, le type de pathologie et enfin la description se termine par le code du contour de l'anormalité.

Pour toutes les mammographies contenant une tumeur, le spécialiste a tracé un contour autour de la région tumorale, ce contour est analysé sous forme de chaîne à l'aide du code de Freeman. Les deux premiers chiffres de la chaîne représentent les coordonnées d'un pixel du contour respectivement la colonne et la ligne, sur lesquels on se base pour déterminer les coordonnées des autres pixels du contour.

```
TOTAL_ABNORMALITIES 1
ABNORMALITY 1
LESION_TYPE CALCIFICATION TYPE PLEOMORPHIC DISTRIBUTION
CLUSTERED
ASSESSMENT 4
SUBTLETY 2
PATHOLOGY MALIGNANT
TOTAL_OUTLINES 1
BOUNDARY
850 2007  0 0 0 0 0 2 2 2 2 0 0 0 0 1 1 1 1 1 0 0 0 0 0 2 2 2
... 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 #
```

Figure III. 6: Fichier A_1141_1.RIGHT_CC. Overlay.

II.3. Classification par l'ontologie

Nous avons eu recours aux ontologies, pour la manipulation automatique de l'information au niveau sémantique avec l'utilisation du langage SWRL qui enrichit la sémantique d'une ontologie définit en OWL en se basant sur le système de caractérisation BI-RADS (Breast Imaging Reporting and Data System). Il s'agit d'une classification, en cinq catégories.

III. Construction d'une ontologie pour la prise de décision des pathologies mammaires

Dans cette section, nous construisons notre ontologie que nous avons appelée *OntologieCancer*, selon les critères de classification étudiés dans chapitre ontologie.

Le corpus utilisé est les fichiers ics et Overlay. Avec l'aide de l'expert, on a extrait les concepts nécessaires à la conceptualisation de l'ontologie.

III.1. Spécification

Une ontologie ne peut être construite qu'après la phase de spécification. Il s'agit d'établir un document informel de spécification des besoins. Au niveau de ce document, nous décrivons l'ontologie à construire à travers les cinq aspects suivants :

- **Le domaine de connaissance**

  Domaine de prise de décision de dépistage du cancer de sein.

- **L'objectif**

  Améliorer les évaluations et la prise en charge des anomalies mammaires, il s'agit d'une méthode standardisée pour les comptes rendu.

- **Les utilisateurs futurs de l'ontologie**

  Patientes, Professionnels de santé.

- **Les sources d'informations**

La classification du système BIRADS de l'ACR, des recommandations de l'ANAES.

- **La portée de l'ontologie**

Patientes, Professionnels de santé, médecin, ACR, anomalie mammographique…

III.2. Conceptualisation

Dans cette étape on distingue les principales tâches suivantes :

- *Construction du glossaire de termes*

Construire un glossaire de termes est la première tâche à effectuer dans l'étape de conceptualisation. Il recueille et décrit tous les termes (concepts, instances, attributs, relations entre les concepts, etc.) qui sont utiles et potentiellement utilisables dans le domaine où nous spécifions leurs descriptions détaillées et non ambiguës dans un langage naturel. **Le tableau 3** illustre le glossaire des termes d'OntologieCancer. Les termes recensés sont ceux en commun avec le domaine médicale et ceux propre au domaine considéré.

Nom	**Description**
Patiente	est une personne souffrant d'une maladie. Elle subit plusieurs activités et il y a plusieurs documents qui la concernent.
Mammographie	La **mammographie** est un examen radiologique des seins qui permet de révéler des lésions de très petite taille, à partir de quelques millimètres. Ces lésions sont souvent alors indétectables à la palpation.
Microcalcifications	Les macro-calcifications sont des images calciques de grande taille, habituellement faciles à distinguer des microcalcifications de macro-calcifications .Elles sont facilement vues en mammographie du fait de leur forte densité : elles apparaissent comme des ponctuations blanches sur le gris des tissus.
Microcalcifications	Les microcalcifications mammaires sont des dépôts calciques dans un tissu. en grande majorité leur diamètre est compris entre 0,2 et 0,5 mm. Elles sont facilement vues

	en mammographie du fait de leur forte densité : elles apparaissent comme des ponctuations blanches sur le gris des tissus.
Masse	une lésion occupant un volume, vue sous deux incidences différentes
Opacité	une lésion occupant un volume n'est vue que sur une seule incidence
ACR1	Mammographie normale
ACR2	Anomalie typiquement bénigne
ACR3	Anomalie probablement bénigne
ACR4	Anomalie suspecte
ACR5	Anomalie évocatrice d'un cancer
Forme Ronde « ROUND »	sphérique, circulaire ou globuleuse
Forme Ovale « OVAL »	elliptique ou en forme d'oeuf.
Forme Lobulée « LOBULATED »	légère ondulation
Forme Irrégulière « IRREGULAR »	non géométrique, pas de description Simple
ContourCirconscrits «CIRCUMSCRIBED »	bords nets, bien définies
ContourMicrolobulés «MICROLOBULATED »	petites ondulations courtes
ContourMasqués «OBSCURED »	cachés par les tissus voisins
ContourIndistincts «ILL_DEFINED »	mal définis, flous
ContourSpiculés «SPICULATED »	opacités radiaires des bords
MorphologieVermiculaire « DYSTROPHIC »	calcifications en forme de bâtonnets, parfois ramifiés, moulant une partie de l'arbre galactophorique et dessinant les dernières lettres de l'alphabet (V, W, X, Y; Z).
MorphologieRonde	ces calcifications ont une taille variable et sont

« ROUND »	habituellement considérées comme bénignes. Lorsqu'elles sont de petite taille (< 1 mm), elles se forment généralement dans des acini. Si elles mesurent moins de 0,5 mm, le terme **punctiforme** est utilisé.
MorphologiePunctiforme « PUNCTATE »	sont rondes ou ovales et mesurent moins de 0,5 mm de diamètre. Leurs contours sont bien définis. Ces calcifications sont souvent bénignes mais peuvent parfois être rencontrées en association avec un cancer.
MorphologieVasculaire « VASCULAR »	calcifications linéaires ou en rails se superposant à l'opacité tubulaire correspondant à une artère.
MorphologiePoussiéreuse Poudreuses ou amorphes « AMORPHOUS »	ces calcifications sont souvent plus ou moins rondes ou en forme de flocons. Sont trop fines pour que l'on puisse apprécier leur forme
MorphologieIrrégulière Polymorphes « PLEOMORPHIC »	avec des formes variables mesurant moins de 0,5 mm ainsi que les calcifications fines
Morphologie Linéaire ou branchée «FINE-LINEAR BRANCHING»	Elles mesurent moins de 0,5 mm d'épaisseur. Elles sont irrégulières et de forme parfois linéaire généralement discontinues, coudées ou branchées. Ces calcifications suggèrent fortement la présence d'un processus néoplasique comblant la lumière d'un canal galactophore.
Distribution Groupées ou en foyer(s) « CLUSTERED »	ces termes sont utilisés lorsque de multiples calcifications (au moins 5) occupent un petit volume (2 cm3)
Distribution linéaire « LINEAR »	suggère une distribution dans un canal galactophore et ses branches. Cette distribution évoque parfois un triangle avec sommet dirigé vers le mamelon.
Distribution segmentaire « SEGMENTAL »	une distribution dans un ou plusieurs canaux galactophores et ses branches. Bien que des calcifications d'origine sécrétoire puissent adopter cette distribution, elle augmente la suspicion de malignité en particulier pour les calcifications rondes, punctiformes ou amorphes.
Distribution régionale « REGIONAL»	calcifications éparses dans un large volume (>2 cm3) et n'ayant pas une distribution galactophorique.

Calcifications diffuses / éparses «DIFFUSE »	répartition aléatoire dans l'ensemble du sein.

Tableau III. 2: Glossaire de termes.

➢ *Classification des concepts en hiérarchies de concepts*

Cette tâche consiste à définir les relations taxonomiques entre les concepts définis dans le glossaire de termes.

Plusieurs types de relations peuvent exister : généralisation/spécialisation, association, composition, agrégation…

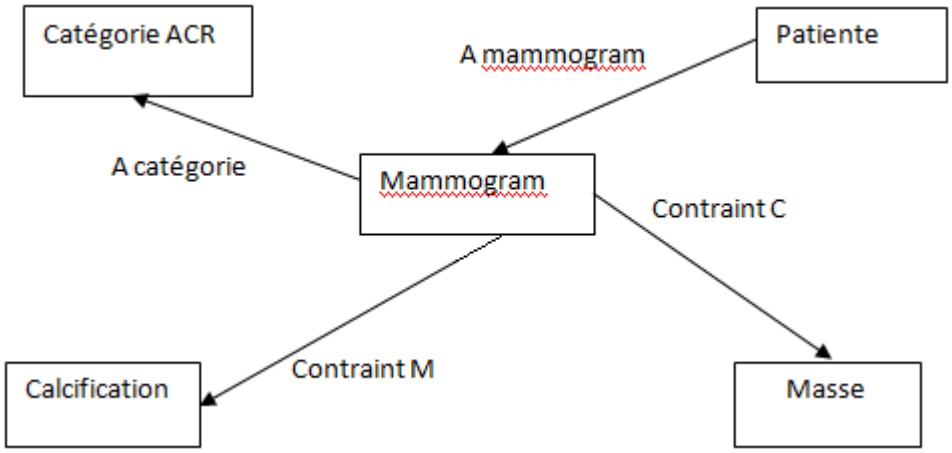

Figure III. 7: diagramme de relations binaires.

➢ *Construction de la table des relations binaires*

Le but de cette tâche consiste à construire une table de relations binaires décrites en détaille. Pour chaque relation utilisé dans le diagramme des relations binaires, nous définissons le nom de la relation, le nom des concepts sources et cibles et les cardinalités source et cible **(voir Tableau 4).**

Nom de la relation	**Concept source**	**Concept cible**	**Card source**	**Card cible**
mammogramCategorieACR	Mammogram	Catégorie ACR	1..n	1..n
mammogramCalcification	Mammogram	Calcification	0..n	0..n
mammogramMasse	Mammogram	Masse	0..n	0..n
patienteMammogram	Patiente	Mammogram	1..n	1..1

Tableau III. 3: table des relations binaires.

➢ *Construction d'un dictionnaire de concepts*

Un dictionnaire de concepts contient tous les concepts du domaine, leurs synonymes, leurs acronymes, leurs attributs et leurs relations (voir **Tableau III.4**).

Nom de concept	**Attributs**	**Les relations**	**Synonymes**
Mammogram	Assessment Densité Subtilité	PatienteMammogram mammogramMasse mammogramCalcification mammogramCategorieACR	Cliché mammographique
Patiente	AgePatiente	PatienteMammogram	malade
CatégorieACR	nomCatégorie	mammogramCategorieACR	
Masse	contourMasse formeMasse	mammogramMasse	Opacité
Calcification	distributionCalci MorphologieCalci	mammogramCalcification	

Tableau III. 4: Construction d'un dictionnaire de concepts.

IV. Conception d'une BD pour le stockage des informations :

Pour alléger l'ontologie et permettre un traitement rapide, on a besoin de stocker les informations dans une BD. Ces informations vont servir pour faire un suivi des patientes et faire des statistiques.

Cette BD est conçue comme suit ***:***

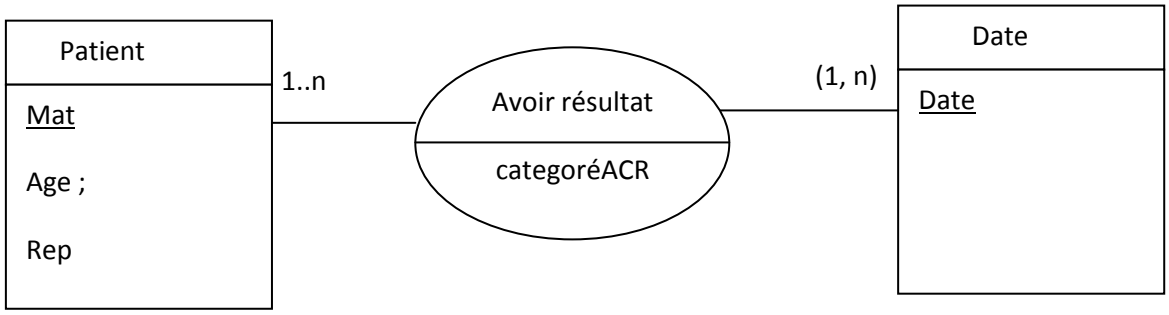

Figure III. 8: Modèle Conceptuel de données.

V. Modélisation de l'inférence

Le formalisme choisi pour la modélisation des connaissances liées aux traitements est les règles de production. Ce formalisme est le plus approprié pour notre problème. Ces règles sont élaborées selon le schéma :

Si condition1 et condition2 et … et conditionN ALORS conclusion1 et conclusion2 et… et conclusionN.

La partie droite de la règle spécifie les conditions qui doivent être vérifiées et la partie gauche spécifie les conclusions déduites en cas de satisfaction des conditions de la règle.

VI. Conclusion

Au cours de ce chapitre, nous nous sommes étalés sur l'architecture de notre système. Nous avons commencé par donner une vue générale de ses principaux composants. L'approche sur laquelle se base notre système est une approche ontologique, afin de classifier les images mammographiques.

Dans le chapitre suivant, nous allons passer à l'implémentation de notre système.

Chapitre 4

Implémentation et validation

I. Introduction

Après l'étape de conception, nous allons présenter dans ce chapitre les outils utilisés dans le développement de notre système. Ensuite, nous passons à la présentation de notre application, à travers des captures d'écrans

II. Environnement de développement

- ✓ l'environnement Windows,
- ✓ le langage de programmation java avec l'IDE NetBeans.
- ✓ pour l'édition de notre ontologie, nous avons utilisé le logiciel PROTEGE2000 version 3.4.4,
- ✓ afin d'effectuer des raisonnements et déduire de nouvelles connaissances, on a utilisé le langage SWRL avec le moteur de règle JESS.

II.1. Protégé

Protégé est le plus connu et le plus utilisé des éditeurs d'ontologies car il est gratuit et téléchargeable. Il est open-source, développé par l'Université de Stanford, il a évolué depuis ses premières versions (Protégé-2000) pour intégrer, à partir de 2003, les standards du web sémantique RDF (Resource Description Framework), RDFS (Resource Description Framework Schéma) et notamment OWL. Il offre de nombreux composants optionnels : raisonneurs, interfaces graphiques…

Protégé est également une librairie Java qui peut être étendue pour créer des applications indépendantes qui accèdent en lecture et en écriture aux ontologies créées. Il s'appuie sur la notion de pont Bridge qui permet d'échanger les informations stockées dans l'ontologie avec des moteurs d'inférence externes pour effectuer des raisonnements et pour déduire de nouvelles connaissances en exécutant les règles SWRL liées aux instances de l'ontologie. Parmi les moteurs existants, nous avons choisi JESS (JavaMC Expert System Shell) vu sa facilité d'intégration dans l'éditeur Protégé et dans des applications Java indépendantes. Ce moteur d'inférence efficace est entièrement écrit en Java avec une documentation et des forums de discussions en ligne pour faciliter son utilisation. Il permet aussi d'invoquer directement à partir des règles SWRL.

II.2. Le langage SWRL

SWRL (Semantic Web Rule Language ou langage de règles du Web sémantique) est un langage standardisé par W3C qui permet aux utilisateurs de rédiger des règles exprimées en termes de concepts OWL pour fournir des capacités de raisonnement déductif. SWRL définit plusieurs opérateurs mathématiques et assure des garanties formelles fortes lors de l'exécution de l'inférence. Il est construit sur le même fondement logique de description OWL. Ce langage enrichit la sémantique d'une ontologie définie en OWL en utilisant des mécanismes de raisonnement sur les connaissances exprimées en OWL. Ces mécanismes sont implantés dans des moteurs d'inférence qui assurent l'exécution des règles SWRL à travers des ponts connus par le nom bridge afin d'automatiser le processus de déduction sur les règles.

Les règles SWRL sont élaborées selon le schéma : antécédent (body)→ conséquent (head). La partie corps (swrl : body) spécifie les conditions qui doivent être vérifiées et la partie entête (swrl: head) spécifie les actions à prendre en cas de satisfaction des conditions de la règle.

```
Patiente(?p) ? Radio(?r) ? CategorieACR(?cat) ? nomCategorie(?cat, "ACR4") ?
Masse(?m) ? contourMasse(?m, "ILL-DEFINED") ? formeMasse(?m, "IIRREGULARY") ?
radioMasse(?r, ?m) ? patienteRadio(?p, ?r) ?
radioCategorieACR(?r, ?cat)
```

Figure IV.1: Un exemple de Réglé SWLR

II.3. JESS (The JavaExpert System Shell)

JESS est un environnement écrit en Java, est un langage à base de règles qui fonctionne selon le modèle des moteurs d'inférences : un ensemble de règles s'appliquent sur une base de faits permettant d'inférer de nouveaux faits. C'est un moteur d'inférence fonctionnant essentiellement en chainage avant et chainage arrière. Parmi ces avantages on cite :

- Expressivité du langage de représentation.
- Facile à apprendre et à utiliser.
- Il existe plusieurs outils supplémentaires pour faciliter la programmation avec Jess : Editeur de règle pour JESS, et des plugin de protégé pour OWL et Jess (Jess_Tab).
- Portabilité et rapidité.

II.4. Jfeechart

API java utilisé pour la description statistique**.**

III. la base d'images utilisée

Nous avons utilisé un sous-ensemble de la base DDSM sélectionné manuellement. Les types de tumeurs pris en compte dans notre système sont : Bénign, Cancer et Benign without callback. Nous avons aussi utilisé des images mammographiques variées : des images contenant des micro-calcifications et des images contenant des masses.

IV. Structure de notre application

Le but de notre application est d'extraire des caractéristiques de l'image mammographique en introduisant l'aspect sémantiques à travers une ontologie. Nous avons organisé notre application sous forme de classes java dont le diagramme de classe simplifié est présenté dans la figure **IV**.2.

Dans la classe *patiente.java* nous extrayons les attributs de l'image mammographique existant dans les fichiers *OVERLAY* et *ICS.*

*L*a classe *JavaSWRL.java* prend en charge la lecture de l'ontologie OntologieCancer, elle permet la lecture de tous les concepts et les relations et la classe.

MaFenetre contient l'interface de notre application.

DBacces est pour le suivi de la maladie (historique).

JFReechart est utilisé pour faire des statistiques sur la distribution des catégories ACR pour une population donnée.

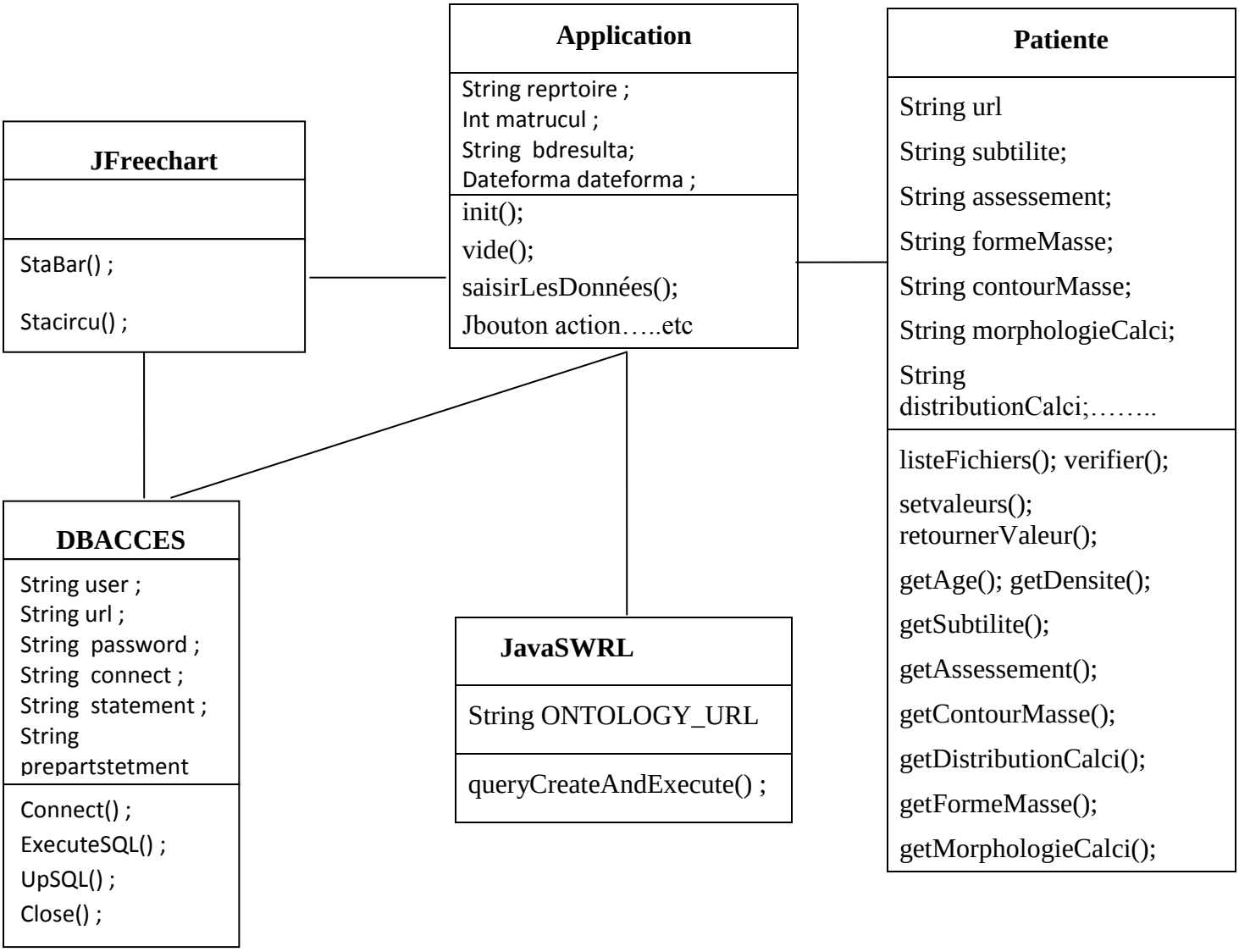

Figure IV.2: Diagramme de classes de notre application

V. Classification par l'ontologie

Notre système commence par l'extraction des connaissances à partir des données biomédicales existant dans les fichiers ICS et OVERLAY. Une ontologie, pour la manipulation automatique de l'information au niveau sémantique, est utilisée. L'utilisation du langage SWRL permet d'enrichir la sémantique de cette ontologie définit en OWL, en se basant sur le système de caractérisation BI-RADS (classification multi classe).

La construction de l'ontologie est une tâche complexe, même pour implémenter une petite ontologie, celle-ci va prendre plusieurs lignes de code et nécessite un grand effort et du temps. Notamment, si cette ontologie est codée directement par le développeur en

langage d'ontologie sans faire recours à aucun outil. Pour cela, PROTEGE est conçu pour libérer l'ontologiste de cette complexité et générer automatiquement la structure de l'ontologie créée en OWL-DL.

Nous donnons, dans la figure suivante, un aperçu de l'ontologie *OntologieCancer* éditée sous PROTEGE.

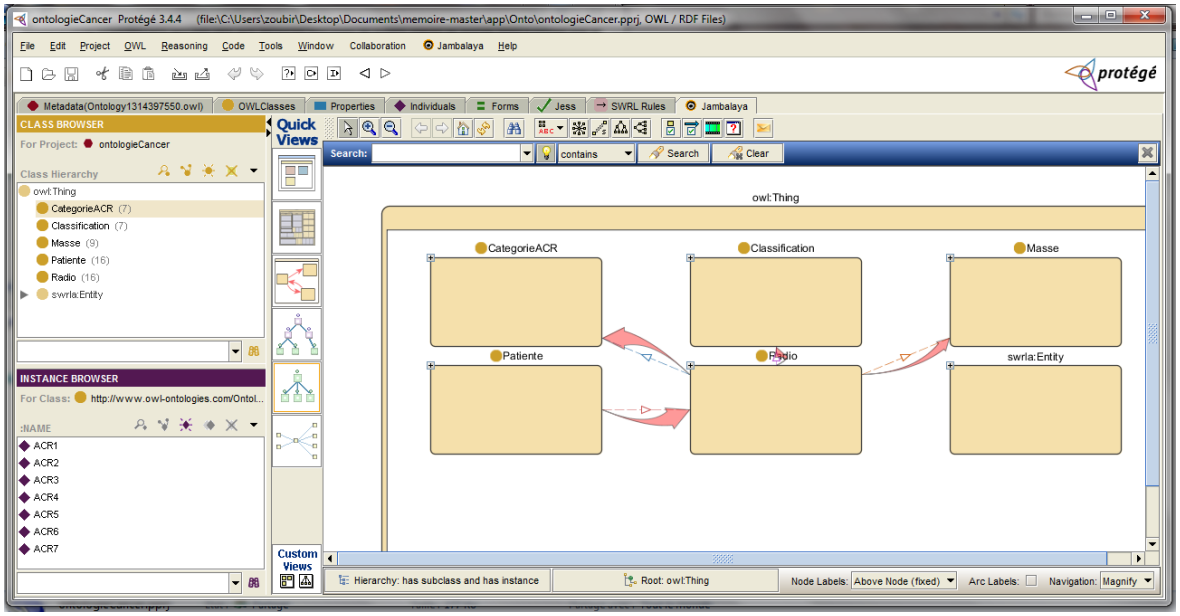

Figure IV.3: OntologieCancer éditée sous l'outil PROTEGE.

Afin d'effectuer des raisonnements et de déduire de nouvelles connaissances en partant de celle exécutantes, les règles d'inférence sont écrites en SWRL. La figure suivante illustre ces règles SWRL.

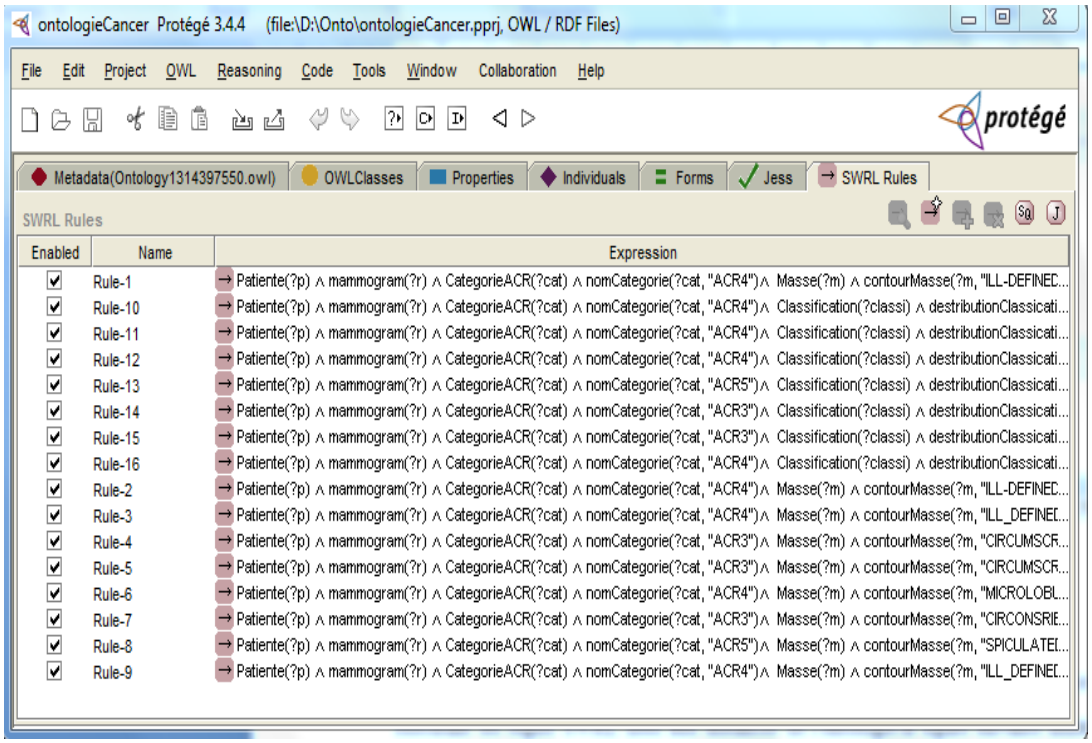

Figure IV 4: Les règles SWRL.

On utilise notre ontologie OntologieCancer comme classifieur qui se base sur le raisonnement sémantique. Notre ontologie est enrichie par des règles SWRL qui ont permit d'augmente son expressivité à donnée une meilleure classification.

VI. Utilisation de notre application

Notons que chaque patiente peut avoir plusieurs dossiers. Chaque dossier contient tous les fichiers nécessaires pour l'analyse à savoir: une image radio et les deux fichiers ICS et OVERLAY. Tout d'abord, on accorde à chaque patiente un Matricule dés son première arrivée au système.

La fenêtre principale :

Figure IV 5: Fenêtre principale du système**.**

La fenêtre principale permet :

- ✓ l'insertion d'un nouveau dossier, en accordant à une nouvelle patiente un matricule. Après l'insertion, le dossier peut être analysé, en cliquant sur le bouton « Analyse des radios».
- ✓ d'analyser un dossier qui existe déjà, en introduisant un matricule.
- ✓ de faire des statistiques sur les dossiers existants.

En cliquant sur « patient », la fenêtre suivante permet l'insertion d'un nouveau dossier d'une patiente qui existe déjà. Après l'insertion, le dossier sera analysé, en cliquant sur le bouton « Analyse des radios » qui va permettre d'afficher la fenêtre de la figure IV 7.

Si une patiente existe déjà, il faut cliquer sur « patiente », dans la fenêtre principale. La fenêtre suivante apparaitra :

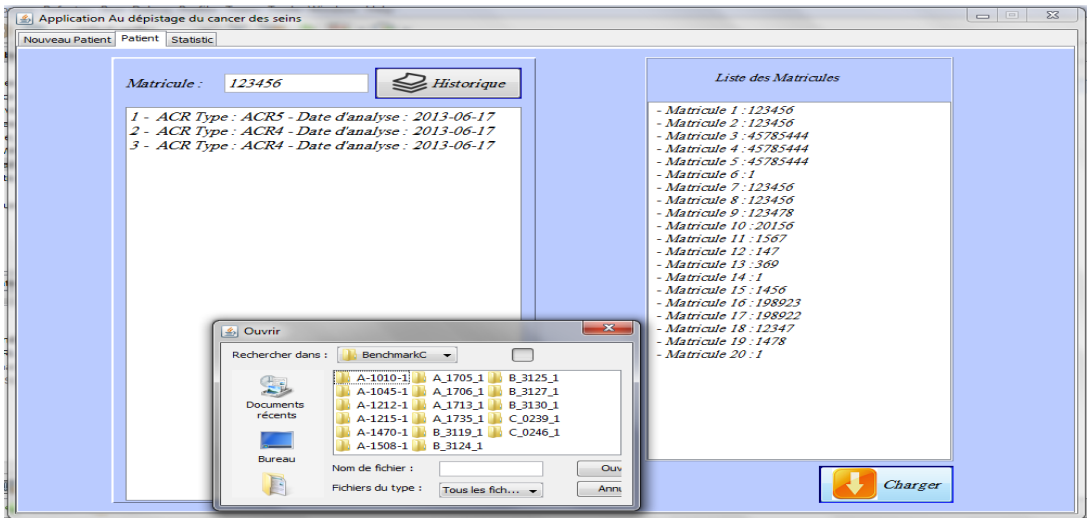

Figure IV 6: Chargement du dossier de la patiente.

Pour analyser le dossier, il faut d'abord le chargé. En cliquant sur le bouton « charger ». Pour l'analyser, il faut cliquer sur « analyse des radios » alors la fenêtre suivante (figure IV 7) apparait.

Cette même fenêtre (IV 6) permet aussi d'effectuer un suivi sur la maladie d'une patiente. Il faut donner son matricule puis cliquer sur le bouton « historique ».

En cliquant sur le bouton « analyse des radios », la fenêtre suivante apparait. Elle permet d'afficher le radio et les informations relatives. En cliquant sur le bouton « analyser », le système affiche la catégorie d'ACR. Pour retourner au menu principal, il faut cliquer sur le bouton « Retour ». Le bouton « quitter » permet de quitter l'application.

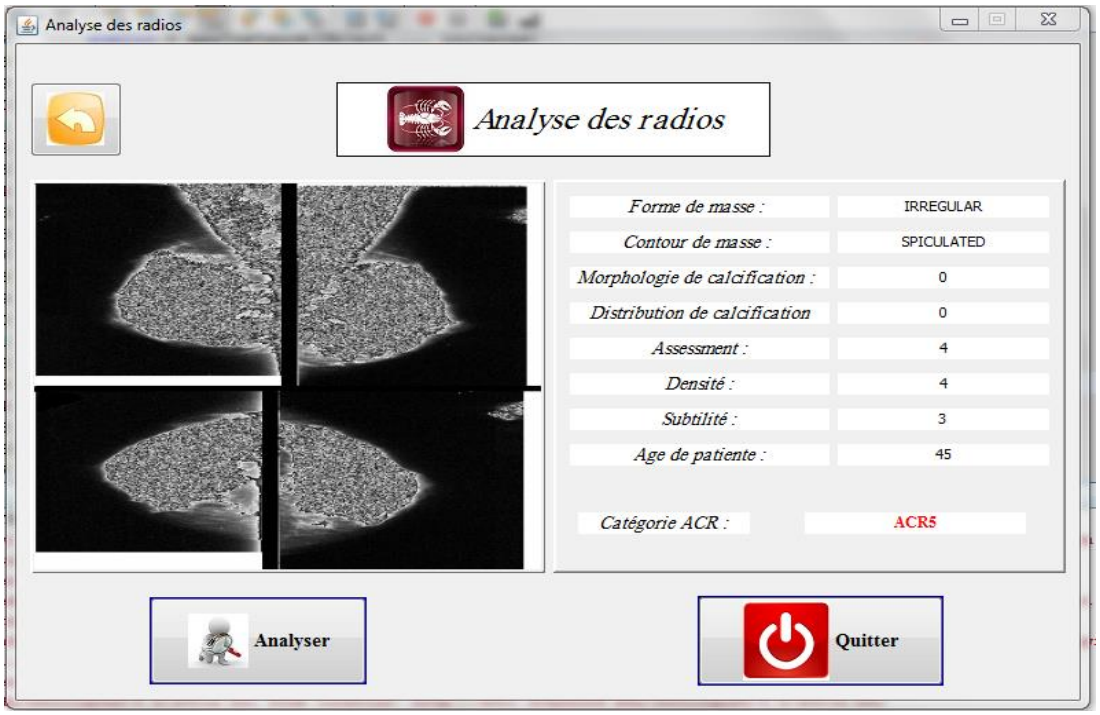

Figure IV 7: Un résultat obtenu à partir de l'ontologie

La figure suivante montre, juste à titre indicatif, une intégration de Jess (JavaMC Expert System Shell) dans Protégé. JESS propose un environnement pour la création et l'édition de systèmes à base de règles.

On commence par copier le contenu d'OWL et SWRL dans Jess, puis on l'exécute. Il indique le nombre d'individus inférés (prédicats d'appartenance à une classe), et le nombre d'axiomes inférés (propriétés sur des instances). Ces nouvelles informations peuvent être intégrées dans la base OWL.

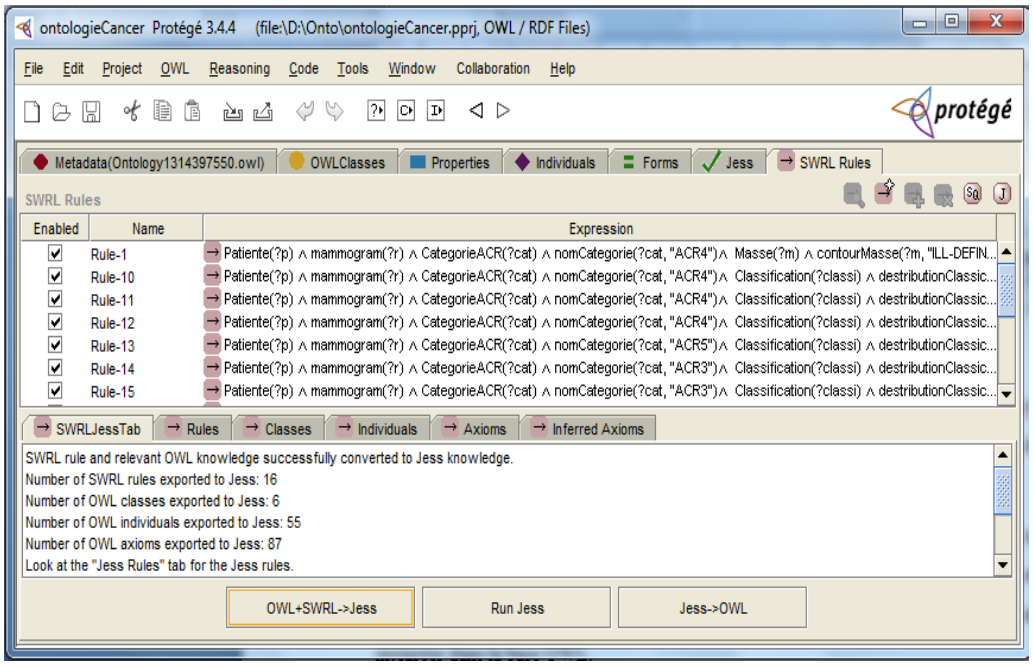

Figure IV 8: L'intégration de JESS.

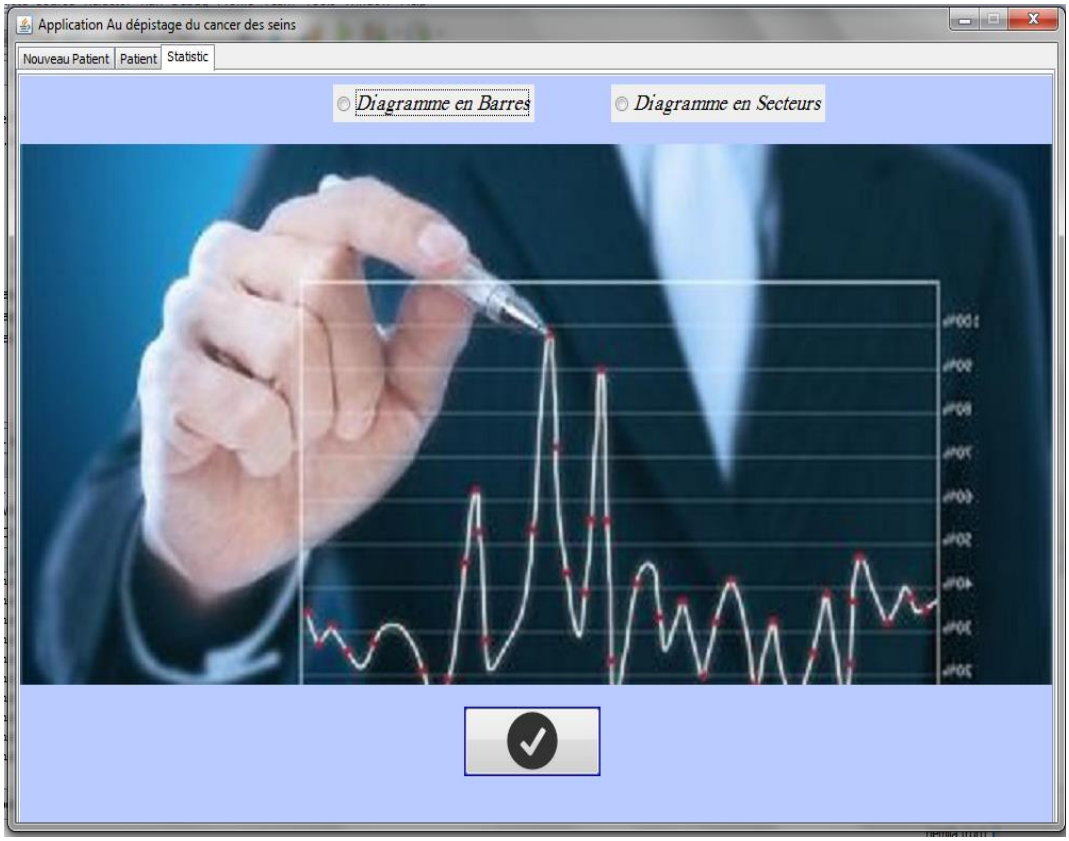

Figure IV 9: Menu de la démonstration statistique.

La figure suivante montre La distribution de l'ACR sur l'ensemble d'une population. Pour l'obtenir, il suffit juste de faire un clique sur le radiobutton (diagramme en barre ou bien diagramme en cercle)

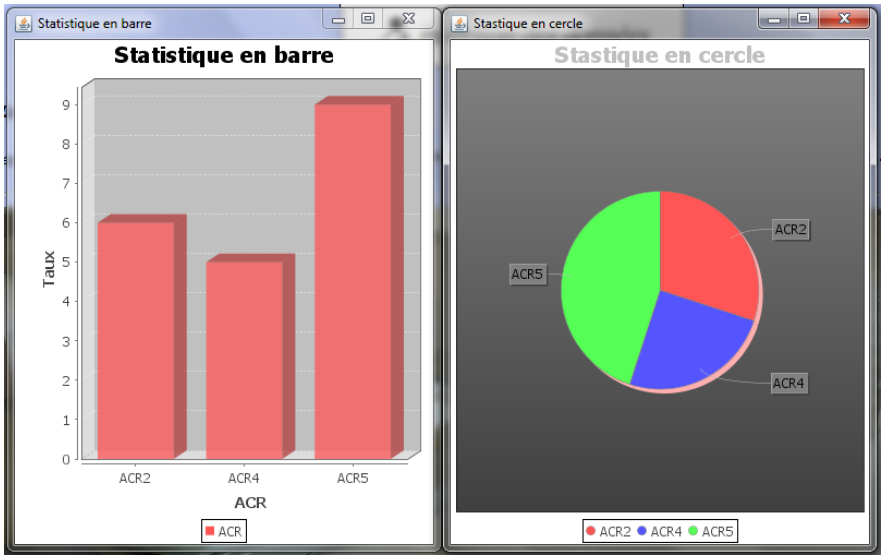

Figure IV 19: JFreechart(diagramme en barre etdiagramme circulaire).

VII. Conclusion

Dans ce chapitre nous avons présenté les outils utilisés pour l'implémentation de notre système qui est basé sur une ontologie. Les résultats de classification des données obtenus par l'approche ont été très prometteurs.

Conclusion Générale

L'objectif que nous avons fixé, au début de ce mémoire, était de proposer une approche d'extraction de connaissance à partir de données biomédicale pour la classification des régions sélectionnées à partir des mammograms numériques, pour chaque signe radiologique considéré, à savoir les masses et les microcalcifications. Ainsi, nous avons développé une approche basée sur l'utilisation d'ontologies pour l'annotation des images et la tache de classification.

L'ensemble du travail que nous avons réalisé au cours de ce mémoire a montré l'apport des ontologies pour ce genre de problèmes. L'ontologie nous a permit la manipulation automatique de l'information au niveau sémantique et nous a permit l'utilisation du langage SWRL qui enrichit la sémantique d'une ontologie définit en OWL.

Pour cela nous avons utilisé la base d'images DDSM (Digital Database for Screening Mammography) pour évaluer notre approche.

Nous avons en premier lieu extrait des paramètres caractérisant les images issues de la base de données MARATHON de l'université de la Floride, elles sont déjà diagnostiquées par des experts. Ensuite, nous avons modélisé les connaissances sous forme d'une ontologie pour donner à l'information un aspect sémantique, en se basant sur le système de caractérisation BI-RADS (classification multi classe).

Pour l'édition de notre ontologie, on a utilisé Protègè2000 qui est un éditeur très utilisé. Pour permettre le raisonnement sur l'ontologie, on a utilisé qui est un moteur d'inférence très puissant. Le langage SWRL est utilisé pour facilité l'expression des règles.

Les résultats de classification des données obtenues par l'approche ont été très prometteurs.

Bibliographie

[1]http://www.algerielle.com/sante-de-la-femme/le-cancer-du-sein-en-algerie 1110365.html

[2]http://www.vulgaris-medical.com/encyclopedie/cancer-du-sein-chez-la-femme-8417.html

[3] J.-P. WAINSTEN et coll. dans le Larousse médical,Edition (2009)

[4] http://www.imanord.fr/

[5]http://culturesciences.chimie.ens.fr/content/la-radiographie-iii-applications-de-la-radiographie-en-imagerie-medicale-991

[6] http://www.cim-monaco.com

[7] http://www.docteur-benchimol.com/mammographie.html

[8] H. P CHAN et al (1987). "Image feature analysis and computer-aided diagnosis in digital radiography. 1. Automated detection of microcalcifications in mammography." *Medical Physics* vol. 14, pp. 538-548.

[9] M. L.GIGER et al(1987). "Computerized detection of lung nodules in digital chest radiographs." *Proceeding SPIE* vol. 767, pp. 384-386.

[10] CJ. VYBORNY 'Can computers help radiologists read mammograms ?

[11] H.-P. CHAN et al(1987). Digital mammography. ROC studies of the effects of pixel size and unsharp-mask filtering on the detection of subtle microcalcifications. *Investigative Radiology*, 22(7):581–589.

[12] X. H. ZHOU et R. GORDON (1989). Detection of early breast cancer: an overview and future prospects. *Critical Reviews in Biomedical Engineering (CRBE)*, 17(3):203–255.

[13] R. STRICKLAND, et H. I. HAHN, (1997). Wavelet transform methods for object detection and recovery. *IEEE Transactions on Medical Imaging*, 6(5):724–735.

[14] S. BERNARD et al (2006). Microcalcification detection on simple backprojection reconstructed slices using wavelets. *Computer Assisted Radiology and Surgery (CARS)*, 1:84– 86.

[15] H.-D. CHENG et al (1998). A novel approach to microcalcification detection using fuzzy logic technique. *IEEE Transactions on Medical Imaging*, 17(3):442–450.

[16] A. RICK, (1999). *Représentation de la variabilité dans le traitement d'images flou*. Thèse de doctorat, université Paris VI.

[17] M. GRIMAUD (1991). *La géodésie numérique en morphologie mathématique. Application à la détection automatique des microcalcifications*. Thèse de doctorat, Ecole nationale supérieure des mines de Paris.

[18] M. D. HEATH et , K. W. BOWYER (2000). Mass detection by relative image intensity. *International Workshop on Digital Mammography (IWDM)*. Medical Physics Publishing (Madison, WI).

[19] G. PETERS (2007). *Computer-aided Detection for Digital Breast Tomosynthesis*. Thèse de doctorat, Ecole Nationale Supérieure des Télécommunications.

[20] C. VACHIER (1995). *Extraction de caracteristiques, segmentation d'image et morphologie mathematique*. Thèse de doctorat, Ecole nationale supérieure des mines de Paris.

[21] H. BORNEFALK (2005). Use of quadrature filters for detection of stellate lesions in mammograms. *Scandinavian Conference on Image Analysis (SCIA)*, pages 649–658.

[22] F. ZOU et al(2008). Gradient vector flow fields and spiculated mass detection in digital mammography images. *International Workshop on Digital Mammography (IWDM)*, pages 299–306, Berlin, Heidelberg. Springer-Verlag.

[23] S.LIU et al (2001). Multiresolution detection of spiculated lesions in digital mammograms. *IEEE Transactions on Medical Imaging*, 10(6):874–884.

[24] L. ZHENET et A.CHAN, (2001). An artificial intelligent algorithm for tumor detection in screening mammogram. *IEEE Transactions on Medical Imaging*, 20(7):559–567.

[25] N. ELTONSY et al (2007). A concentric morphology model for the detection of masses in mammography. *IEEE Transactions on Medical Imaging*, 26(6):880–889.

[26] V. G. MARTINEZ et al (1999). Iterative method for automatic detection of masses in digital mammograms for computer-aided diagnosis. *SPIE Symposium on Medical Imaging*, volume 3661, pages 1086–1093.

[27] H. D.CHENG et al (2003). Computer-aided detection and classification ofmicrocalcifications in mammograms : a survey. *Pattern Recognition*, 36(12):2967 – 2991.

[28] J. K. KIM et H.W. PARK (1999). Statistical textural features for detection of microcalcifications in digitized mammograms. *IEEE Transactions on Medical Imaging*, 18(3):231–238.

[29] Z. HUO, M. L.GIGER et al (1998). Automated computerized classification of malignant and benign masses on digitized mammograms. *Academic Radiology*, 5(3):155–168.

[30] M. BRUYNOOGHE, (2006). Mammographic mass detection using unsupervised clustering in synergy with a parcimonious supervised rule-based classifier. *International Workshop on Digital Mammography (IWDM)*, pages 68–75.

[31] Li, L.,MAO, F, W. QIAN, et L. CLARKE (1997). Wavelet transform for directional feature extraction in medical imaging. *IEEE International Conference on Image Processing*, volume 3, page 500, Los Alamitos, CA, USA. IEEE Computer Society.

[32] H. KIM, et W. KIM, (2005). Automatic detection of spiculated masses using fractal analysis in digital mammography. *Computer Analysis of Images and Patterns (CAIP)*, pages 256–263.

[33] N. MUDIGONDA, R. RANGAYYAN, et J. LEO DESAUTELS, (2001). Detection of breast masses in mammograms by density slicing and texture flow-field analysis. *IEEE Transactions on Medical Imaging*, 20(12):1215– 1227.

[34] A. R. DOMINGUEZ, et , A. K. NANDI (2009). Toward breast cancer diagnosis based on automated segmentation of masses in mammograms. *Pattern Recognition*, 42(6):1138 – 1148. Digital Image Processing and Pattern Recognition Techniques for the Detection of Cancer.

[35] M. KUPINSKI, et M. GIGER, (1999). Feature selection with limited datasets. *Medical Physics*, 26(10): 2176–2182.

[36] M. KUPINSKI, et M. GIGER, (1997). Feature selection and classifiers for the computerized detection of mass lesions in digital mammography. *International Conference on Neural Networks*, 4:2460–2463.

[37] Z. BROWN et al. « 100 Question –réponses le cancer du sein »copyright 2007.

[38] http://www.passeportsante.net/fr/Maux/Problemes/Fiche.aspx?doc=cancer_sein_pm

[39] A. OTHMANI « ARTICULATION ENTRE ONTOLOGIE ET TRAITEMENT D'IMAGES » RAPPORT DE MEMOIRE DE MASTERE université EL MANAR Tunisie

[40] http://www.docteurclic.com/encyclopedie/maladies-des-seins.aspx

[41] http://www.ass.nc/themes/cancer-du-sein/maladies-benignes-du-sein

[42] http://www.ass.nc/themes/cancer-du-sein/moyens-de-depistage

[43] http://www.igr.fr/fr/page/questions-reponses-sur-le-depistage_1301#echographie

[44] http://www.docteur-benchimol.com/cancer_du_sein.html

[45] ACR (2003). *BI-RADS Breast Imaging Reporting and Data System*. American College of Radiology.

[46] L. ABROUK « Annotation de documents par le contexte de citation basée sur une ontologie » thèse doctorat en informatique de l'université de Montpellier 2.

[47] T. GRUBER, A translation approch to portable ontology specifications, Knowledge Acquisition, Vol. 5 (2), pp. 199-220, 1993.

[48] W. N. BORST, Construction of Engineering Ontologies, PhD Thesis, University Of Twente, Netherlands, 1997.

[49] S. ADICHE et Y. NOUREDDINE « PROFIL UTILISATEUR D'UNE PLATE-FORME E-LEARNING »mémoire de fin d'étude d'ingénieur d'état en informatique

[50] A. GÓMEZ-PÉREZ, M. FERNANDEZ-LOPEZ & O. CORCHO," Ontological Engineering". Advanced Information and Knowledge Processing. Madrid, Spain: Springer, 2 editions. 2004.

[51] E. BOZSAK et al - towards a large scale semantic web, Proceedings of the Third International Conference on E-Commerce and Web Technologies (EC-Web 2002), Aix-en-Provence, France, Lecture Notes in Computer Science, LNCS 2455, Springer, pp. 304-313, 2002.

[52] N. GUARINO et W. WELTY, A Formal Ontology of Properties, Proceedings of the 12th European Workshop on Knowledge Acquisition, Modeling and Management (KAW'99), Banff, Canada, LNCS 1937, pp. 97-112, 2000.

[53] A. GÓMEZ-PÉREZ, M. FERNÁNDEZ-LÓPEZ et O. CORCHO, Ontological Engineering, Advanced Information and Knowledge Processing (Series), Springer-Verlag, ISBN 185233551-3, 2004.

[54] M. USCHOLD et M. GRÜNINGER, "ONTOLOGIES: Principles, Methods and applications".Knowledge engineering review, vol.11, N.2, 1996

[55] D. MONTICOLO «Une approche organisationnelle pour la conception d'un système de gestion des connaissances fondé sur le paradigme agent » thèse doctorat de l'université de technologie de Belfort Montbéliard et université de franche Comté.

[56] B. SWARTOUT, P. RAMESH, K. KNIGHT, T. RUSS, “Towards Distributed Use of Large- Scale Ontologies,” Proceedings of 10th Knowledge Acquisition Workshop (KAW'96).

[57] M. FERNANDEZ, A. COMEZ-PEREZ et al. METHONTOLOGY: from ontological art towardsontological engineering ≫ in Proceedings of the Spring Symposium Series on Ontological Engineering (AAAI’97), AAAI Press, 1997

[58] A. KEITA « Conception Coopérative d’Ontologies Pré-Consensuelles : Application au domaine de l’Urbanisme » thèse de doctorat de L’Institut National des Sciences Appliquées de Lyon.