

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

وزارة التعليم العالي والبحث العلمي

BADJI MOKHTAR-ANNABA UNIVERSITY  
UNIVERSITE BADJI MOKHTAR-ANNABA



جامعة باجي مختار - عنابة

Faculté des Sciences de l'ingénieur

Année : 2014/2015

Département d'informatique

**THESE**

Présentée en vue de l'obtention du diplôme de

**Doctorat 3<sup>ème</sup> Cycle LMD**

**INDEXATION DES IMAGES BASEES  
AUTOMATES CELLULAIRES ET AGENTS**

Option

Traitement d'images

Par

**Benmazou Sarah**

**DIRECTEUR DE THESE:** H. F. Merouani

Professeure

U. Badji-Mokhtar Annaba

**DEVANT LE JURY**

**PRESIDENT:** M. T. Kimour

Professeur

U. Badji-Mokhtar Annaba (UBMA)

**EXAMINATEURS:** H. Seridi

Professeur

U. 08 Mai 1945 Guelma

N. Benamrane

Professeure

U. Mohamed Boudiaf Oran (USTO)

T. Bouhadada

MCA

U. Badji-Mokhtar Annaba (UBMA)

*A mon père*

*A ma mère*

*A Mon Frère Mohamed Zakaria*

*Mes Sœurs : Maha,*

*Abir et Insaf.*

# Remerciements

Merci, mon Dieu.

Je tiens à exprimer mes plus vifs remerciements à Madame Hayet Farida Merouani qui fut pour moi une directrice de thèse attentive et disponible, malgré ses nombreuses charges. Sa compétence, sa rigueur scientifique et sa clairvoyance m'ont beaucoup apprises. Professeure Merouani a été et restera un exemple de travail de recherche.

Je remercie chaleureusement le Professeur Mohamed Taher Kimour de l'université d'Annaba pour m'avoir fait le grand bonheur de présider le jury.

J'adresse mes plus vifs remerciements aux membres du jury Nacéra Benamran professeure à l'université d'Oran, Seridi Abdelhamid professeur à l'université de Guelma, Taher Bouhadada Maître de Conférence à l'université d'Annaba pour avoir accepté la lourde charge de lire et faire le rôle de rapporteurs de cette thèse.

Madame Soumia Layachi pour toutes ses remarques, les suggestions et les conseils, cela m'a beaucoup encouragé; en un mot, Merci.

Je remercie toutes les personnes rencontrées au cours des voyages de recherche, réalisés durant cette thèse, à l'occasion de conférences, d'ateliers de travail ou de stage, ces échanges ont beaucoup influencé mon travail.

Qu'il me soit permis de remercier plus particulièrement Nazim Fatès chargé de recherches à l'INRIA, Nadjemedine Belledjhem docteur en radiologie à l'hôpital Ibn Rochd d'Annaba, Nour el Houda Djilani docteur spécialiste en radiologie à l'hôpital Ibn Zohr de Guelma.

Je remercie les différentes personnes, étudiants et enseignants, rencontrées au cours de mon cursus universitaire, au sein du département d'informatique Badji Mokhtar Annaba et qui ont influencé mes choix en matière de thème de recherche. Je remercie en particulier Madame Ahlem Melouah pour conseils et avis.

Je remercie enfin ma famille pour sa présence et son soutien tout au long de ma scolarité.

Je ne saurais clore cette page sans remercier mes amis, surtout Fadia pour sa présence. . .

Merci.

---

## المخلص

---

من خلال البحث في هذا الموضوع نهدف إلى عرض نظام البحث و الفهرسة عن طريق محتوى الصورة، في هذا السياق نعتد على الهيكلة الكلاسيكية لنظام الفهرسة التقليدية وأنظمة البحث التفاعلية، بحيث يقوم المستخدم بتعليم النتيجة التي تم الحصول عليها من قبل النظام بغرض صقل نتيجة البحث بما يسمى التغذية العكسية. يعتبر نظام البحث و الفهرسة عن طريق محتوى الصورة من التطبيقات الهامة في المجال الطبي و هذا لما تحتله الصور الطبية من أهمية في العمل داخل المستشفيات.

نعتد في هذا البحث على نظام العملاء الأذكاء لغرض البحث عن الصور، هذه الأخيرة تمت فهرستها بالاعتماد على خلايا ذاتية السلوك والتي تلعب دور البيئة التي يتعامل من خلالها العملاء . البحث عن الصور يتم عن طريق نظام العملاء الأذكاء. يقوم العملاء بتعليم الخلايا و هذا لغرض وضع مسارات للتوجيه، هذه الطريقة تسمح بتسهيل عملية البحث عن الخلايا التي تحوي صوراً ذات صلة في مرحلة التغذية العكسية كما ان ذلك يعمل على توفير الوقت من خلال العمل المتوازي و التفاعل بين العملاء بهدف تحقيق نتيجة جيدة. جودة نتائج البحث عن الصور للتحقق من صحة النظام المقترح حددت من قبل مختص في المجال.

**الكلمات المفتاحية:** خلايا ذاتية السلوك، عرض نظام البحث و الفهرسة عن طريق محتوى الصورة ، نظام العملاء الأذكاء، التصوير الطبي، تصنيف الصورة.

---

# Abstract

---

In this thesis, we focus on content based image retrieval (CBIR) in medical images databases, we consider classical structure of content based indexing and retrieval, the images retrieved are shown to the user who labels them, following the classical relevant feedback scheme.

The development of computer power and the rise of information technologies have strongly supported the numerical medical image production attached to various methods, this images are stored in medical databases. Their processing requires the development of effective procedures making it possible to structure them, and to retrieve a given image among structured data.

In our work, multi-agent systems (MAS) are used, malignant and benign mammographic images are organized with cellular automata (CA) to build a network of cells that plays the role of an environment of MAS.

Agents traverse the network in search of relevant images and mark the relevant paths to guide agents to the interesting cells in the relevant feedback phase.

This learning allows agents to more easily find cells that contain relevant images in the feedback phase and provides significant time reduction through parallelism and interaction of agents with the aim of a good result.

The quality of results to research pathological images validates the proposed system performed by physician.

**Key-words:**

Cellular Automata, Content based Image Retrieval, CBIR system, Medical Imaging, Multi-agent systems, Images clustering.

---

# Résumé

---

Cette thèse a pour objet d'indexation et de recherche d'images par le contenu (CBIR) dans un contexte de base d'images médicales. Nous nous plaçons dans la structure classique d'un système d'indexation et de recherche classique des systèmes interactifs, c'est-à-dire que l'utilisateur à la performance d'annoter le résultat obtenu par le système afin d'affiner la recherche, c'est ce l'on appelle *bouclage de pertinence*.

Les systèmes CBIR demeurent des problèmes importants qui trouvent des applications potentielles en médecine, étant donné la place que l'image numérique médicale occupe à présent dans le quotidien des hôpitaux.

Dans notre travail, la recherche est réalisée à l'aide d'un système multi-agents (SMA), les images mammographiques malignes et bénignes sont organisées avec les automates cellulaires (AC) afin de construire un réseau de cellules qui joue le rôle d'un environnement des SMA. Les agents parcourent le réseau à la recherche des images pertinentes et marquent les chemins pertinents afin de guider les agents vers les cellules intéressantes dans la phase de bouclage. Cet apprentissage permet aux agents de trouver plus facilement les cellules qui contiennent des images pertinentes dans la phase de bouclage et offre une importante réduction du temps grâce au parallélisme et l'interaction des agents dans le but d'un bon résultat.

La qualité des résultats obtenus pour la recherche d'images pathologiques valide le système proposé aussi évalué par un médecin du service.

## Mots-Clés

Automates cellulaires, Recherche et indexation, Système CBIR, systèmes multi-agents, Imagerie médicales, Classification d'images.

---

# Table des Matières

---

ملخص.....	I
Abstract.....	II
Résumé.....	III
Table des Matière.....	IV
Liste des Figures.....	X
Liste des Tableau.....	XV
Liste des Abréviations.....	XVI
Introduction et Motivation.....	1
Objectifs et guide de thèse.....	4
<b>Chapitre I : Les Automates Cellulaires.....</b>	<b>5</b>
<b>1.1. Introduction.....</b>	<b>5</b>
<b>1.2. Historique.....</b>	<b>6</b>
<b>1.3. Définition.....</b>	<b>7</b>
<b>1.4. Propriétés.....</b>	<b>8</b>
<b>1.4.1. L'espace cellulaire.....</b>	<b>8</b>
<b>1.4.2. Voisinage.....</b>	<b>9</b>
<b>1.4.3. Ensemble d'états.....</b>	<b>11</b>
<b>1.4.4. Règle locale de transition.....</b>	<b>11</b>
<b>1.4.4.1. Règles déterministes.....</b>	<b>11</b>
<b>1.4.4.2. Règles probabilistes.....</b>	<b>11</b>
<b>1.4.4.3. Les conditions ou limites périodiques.....</b>	<b>13</b>
<b>1.4.4.4. Les conditions aux limites réflexives.....</b>	<b>13</b>
<b>1.4.4.5. Les conditions aux limites fixes.....</b>	<b>13</b>

1.1.4.6. Configuration initiale.....	14
1.5. Classification des automates cellulaires.....	14
1.5.1. Autres classes d'automates cellulaires.....	16
1.5.1.1. Automates cellulaires asynchrones.....	16
1.5.1.2. Automates cellulaires à mémoires.....	17
1.5.1.3. Automates cellulaires non-stationnaires.....	17
1.5.1.4. Automates cellulaires flous.....	17
1.5.1.5. Automates cellulaires hiérarchiques.....	17
1.5.1.6. Jeu de la vie.....	18
1.5.1.7. L'automate de Fredkin.....	20
1.5.1.8. L'automate de Brian's Brain.....	21
1.5.1.9. L'automate stochastique.....	21
1.5.1.10. L'automate cellulaire continu.....	22
1.5.1.11. L'automate cellulaire totalistique.....	22
1.5.1.12. Automate cellulaire à une dimension.....	23
1.6. Caractéristiques d'automates cellulaires.....	24
1.6.1. Auto-organisation.....	24
1.6.2. Auto- reproduction.....	24
1.6.3. Auto- similarité.....	25
1.6.4. Inversibilité.....	25
1.6.5. Indécidabilité.....	25
1.6.6. Discrétisation.....	25
1.6.7. La non linéarité.....	25
1.7. Utilisation et applications des automates cellulaires.....	25
1.8. Conclusion.....	27
<b>Chapitre II : les Système Multi-agents.....</b>	<b>29</b>
2.1. Introduction.....	29
2.2. Le système multi agent et le paradigme agent.....	29
2.2.1. Définition système multi-agents.....	30
2.2.2. Les caractéristiques d'un système multi-agents.....	31
2.2.3. Ouverture et perturbation du système.....	32
2.2.4. Décentralisation.....	33
2.2.5. Environnement.....	33
2.3. Architecture interne d'un agent.....	35

2.3.1. Agent cognitive .....	36
2.3.2. Agent réactive.....	36
2.3.3. Agent hybride.....	36
2.4. Le comportement .....	37
2.5. Interaction entre agents.....	37
2.5.1. Situation d'inférence.....	39
2.5.2. Situation de coopération.....	40
2.5.2.1. Collaboration simple.....	40
2.5.2.2. Encombrement.....	40
2.5.2.3. Collaboration coordonnées .....	40
2.5.3. Situations d'antagonisme.....	40
2.5.3.1. Compétition individuelle pure.....	40
2.5.3.2. Conflit individuelle pour les ressources.....	40
2.5.3.3. Compétition collective pure.....	41
2.5.3.4. Conflits collectifs pour des ressources.....	41
6.2. Communication entre agents.....	41
2.6.1. Communication indirecte.....	41
2.6.2. Communication directe.....	41
2.6.2.1. Par partage d'information.....	41
2.6.2.2. Par envoi de messages.....	42
6.3. Actes de langage.....	42
6.4. L'auto-organisation dans les SMA.....	43
6.5. Emergence.....	43
6.5.1. L'émergence au sein des SMA.....	44
6.5.2. Caractéristiques d'un phénomène émergent.....	45
6.5.3. Emergence dans les phénomènes naturels.....	46
6.6. Agents d'inspiration éthologique.....	46
2.10.1. Algorithme d'optimisation.....	47
6.7. Utilisation et application des SMA.....	47
6.8. Conclusion.....	52
<b>Chapitre III : Indexation et Recherche d'images.....</b>	<b>53</b>
3.1. Introduction.....	53
3.2. Méthodes d'indexation.....	53
3.2.1. Indexation selon le partitionnement de l'espace.....	54

3.2.2. Indexation selon le caractère statique et dynamique.....	55
3.2.3. Méthode hiérarchique dynamique.....	57
3.3. Recherche d'images.....	57
3.3.1. Recherche par cible.....	57
3.3.2. Recherche par catégorie.....	58
3.3.3. Recherche par association.....	58
3.4. Architecture du système.....	58
3.4.1. Segmentation.....	59
3.4.2. Caractéristique visuelles.....	62
3.4.2.1. Descripteur globale.....	62
3.4.2.2. Descripteur locale.....	63
3.4.3. Représentation des données.....	70
3.4.3.1. La Signature.....	70
3.4.3.2. La similarité.....	70
3.4.4. Requête.....	73
3.4.4.1. Requête textuelle.....	73
3.4.4.2. Requête image.....	74
3.4.5. Affichage des résultats.....	74
3.4.6. Le bouclage de pertinence.....	74
3.4.6.1. Bouclage de pertinence par modèle d'apprentissage.....	75
3.4.6.2. Bouclage de pertinence par modèle vectoriel.....	75
3.4.6.3. Bouclage de pertinence par modèle probabilistique.....	76
3.4.7. Quelques systèmes de recherche existants.....	76
3.3. Classification.....	79
3.3.1. Méthodes de classification supervisée.....	79
3.3.2. Méthodes de classification non supervisées.....	81
3.4. Evaluation des résultants.....	82
3.5. Conclusion.....	84
<b>Chapitre IV : L'imagerie de Mammographie.....</b>	<b>85</b>
4.1. Introduction.....	85
4.2. Anatomie du sein.....	86
4.3. Cancer du sein.....	86
4.4. Facteur du risqué.....	90
4.4.1.1. Facteurs endogènes.....	90

4.4.1.2. Facteurs endogènes.....	90
4.5. L'imagerie médicale.....	90
4.5.1. L'échographie.....	91
4.5.2. L'imagerie par résonance magnétique.....	91
4.5.3. La mammographie.....	92
4.6. Relation entre l'imagerie médicale et l'anatomie du sein.....	93
4.7. Les pathologies mammaires.....	93
4.7.1. Les microcalcifications (Mcs).....	93
4.7.1.1. La distribution.....	96
4.8. Les masses.....	97
4.8.1. La forme.....	97
4.8.2. Le contour.....	98
4.9. Classification des anomalies de mammographie.....	99
4.9.1. La classification de Legal.....	99
4.9.2. Classification de Birads.....	100
4.10. Classifications des seins normaux.....	101
4.11. Conclusion.....	103
<b>Chapitre V: Extraction des caractéristiques.....</b>	<b>104</b>
5.1. Introduction.....	104
5.2. Architecture du système.....	105
5.2.1. Acquisition des images.....	105
5.2.2. Prétraitement et filtrage.....	107
5.2.3. La segmentation.....	109
5.2.4. Extraction des caractéristiques.....	113
5.2.4.1. Les moments de Hu.....	113
5.2.4.2. Les descripteurs géométriques.....	113
5.2.4.3. Les descripteurs spécifiques.....	116
5.2.4.4. La matrice GLCM (Gray-Level Co-Occurrence Matrix).....	118
5.3. Conclusion.....	124
<b>Chapitre VI : Classification Basée AC.....</b>	<b>125</b>
6.1. Introduction.....	125
6.2. Principe de l'automate cellulaire.....	126
6.2.1. Fonctionnement de l'automate cellulaire.....	127

<b>6.3.</b>	Résultats et évaluations de la performance de la classification proposée.....	130
<b>6.3.1.</b>	Aire sous la courbe ROC.....	130
<b>6.3.2.</b>	Résultats et évaluations.....	131
<b>6.4.</b>	Conclusion.....	150
<b>Chapitre VII : Indexation et Recherche des Images Par le Contenu Basée Agents.....</b>		
	.....	152
<b>7.1.</b>	Introduction.....	152
<b>7.2.</b>	Application de la recherche d'images en médecine.....	153
<b>7.2.1.</b>	Description du contenu des images.....	153
<b>7.2.2.</b>	Indexation.....	153
<b>7.2.3.</b>	Similarité.....	154
<b>7.2.4.</b>	Interactivité.....	154
<b>7.3.</b>	Notre système et nos choix.....	154
<b>7.3.1.</b>	Contenu de la base.....	155
<b>7.3.2.</b>	Indices visuels.....	155
<b>7.3.3.</b>	Présentation du système.....	155
<b>7.3.4.</b>	Schéma général de la recherche.....	157
<b>7.3.5.</b>	Fonctionnement de l'ensemble du système.....	163
<b>7.4.</b>	Résultats, discussion et évaluation.....	164
<b>7.5.</b>	Conclusion.....	170
<b>Conclusion et perspectives.....</b>		171
<b>Annexes.....</b>		174
<b>Référence bibliographiques.....</b>		156

---

# Liste des Figures

---

## Chapitre I : Les Automates Cellulaires

<b>Figure 1.1.</b> Forme typique du réseau cellulaire.....	8
<b>Figure 1.2.</b> Voisinages Von Neumann et Moore dans le cas 1D.....	10
<b>Figure 1.3.</b> Formes de voisinages les plus fréquentes.....	10
<b>Figure 1.4.</b> Voisinages de Smith, Cole et Custom.....	10
<b>Figure 1.5.</b> Représentation des conditions aux limites.....	13
<b>Figure 1.6.</b> Les classes de Wolfram.....	15
<b>Figure 1.7.</b> Une exécution du jeu de la vie sur une configuration torique.....	18
<b>Figure 1.8.</b> Planeur (glider).....	19
<b>Figure 1.9.</b> Boucle de Langton.....	20
<b>Figure 1.10.</b> Automate cellulaire à deux dimensions (Fredkin).....	21
<b>Figure 1.11.</b> Automate cellulaire de Brian's Brain.....	21
<b>Figure 1.12.</b> Automate Stochastique.....	22
<b>Figure 1.13.</b> Automate continu.....	22
<b>Figure 1.14.</b> Automate Totalistic.....	23
<b>Figure 1.15.</b> Automate cellulaire à une dimension (Triangle de Pascal).....	23

## Chapitre II : Les Système Multi-Agents

<b>Figure 2.1.</b> Représentation schématique d'un système multi-agents.....	32
<b>Figure 2.2.</b> Représentation d'un agent et de son environnement.....	34
<b>Figure 2.3.</b> L'agent, un processus cyclique à trois phases : perception, délibération et action.....	36
<b>Figure 2.4.</b> Déplacement linéaire des fourmis.....	46
<b>Figure 2.5.</b> Déplacement à la recherche du chemin le plus court.....	46
<b>Figure 2.6.</b> Les fourmis trouvent le chemin le plus court.....	47

### **Chapitre III : Indexation et Recherche d'images**

<b>Figure 3.1.</b> La structure Kd-tree, partitionnement d'un ensemble de cinq données représentées en deux dimensions, image reprise de.....	54
<b>Figure 3.2.</b> Comparaison de trios types différent de réseaux: nearest neighbour graph, thresholded graph et NNk network.....	57
<b>Figure 3.3.</b> Déroulement de la recherche d'images par le contenu.....	59
<b>Figure 3.4.</b> Aperçu général des techniques de segmentation d'image.....	62
<b>Figure 3.5.</b> Schéma présentant les notions de : fermeture, intérieur, contour et complémentarité.....	66
<b>Figure 3.6.</b> Les différents types de points dans un squelette .....	69
<b>Figure 3.7.</b> Types de modélisation : a) Intrinsèque et b) Discriminante.....	80

### **Chapitre IV : L'image de Mammographie**

<b>Figure 4.1.</b> Anatomie du sein :1) Cage thoracique, 2) Muscles pectoraux, 3) Lobules,4) Surface du mamelon, 5) Aréole, 6) Conduit galactophore, 7) Tissu adipeux, 8) Peau.....	86
<b>Figure 4.2.</b> L'incidence et la mortalité dans les pays développés.....	87
<b>Figure 4.3.</b> L'incidence et la mortalité dans les pays en voie de développement.....	88
<b>Figure 4.4.</b> Le pourcentage du cancer du sein par rapport à neuf types du cancer.....	89
<b>Figure 4.5.</b> Un exemple d'une image du sein avec la technique d'écographie.....	91
<b>Figure 4.6.</b> Exemple d'une imagerie par résonance magnétique du sein.....	92
<b>Figure 4.7.</b> Les distributions des calcifications : A) calcification diffusées ou éparses, B) régionale, C) groupées, D) Linéaire et E) Segmentaire.....	96
<b>Figure 4.8.</b> Asymétrie de densité (concave, convexe).....	97
<b>Figure 4.9.</b> Les formes d'une masse : A) Ronde, B) Ovale, C) Lobulée D) Irrégulière.....	98
<b>Figure 4.10.</b> Les contours d'une masse : A) Circonscrit, B) masqué, C) micro-lobulé, D) indistinct, E) spiculé.....	98
<b>Figure 4.11.</b> Densité mammaire, le lexique BI-RADS de l'American College of Radiology A) type 1, B) type 2, C) type 3, D) type 4.....	102

### **Chapitre V : Extraction des Caractéristiques**

<b>Figure 5.1.</b> Schéma du système de classification d'images de mammographie.....	105
--	-----

<b>Figure 5.2.</b> Echantillons de la base DDM d'Annaba utilisés pour l'évaluation. La première ligne contient des images de mammographie à masse malignes et la deuxième ligne contient des masses bénignes.....	106
<b>Figure 5.3.</b> Application du filtre Médian.....	107
<b>Figure 5.4.</b> Echantillons de la base MIAS utilisés pour l'évaluation. La première ligne contient des images de mammographie à masse bénignes et la deuxième ligne contient des masses malignes.....	108
<b>Figure 5.5.</b> Voisinage d'un agent à une position (i,j).....	110
<b>Figure 5.6.</b> L'évolution des agents pour la recherche et l'étiquetage d'une région homogène.....	111
<b>Figure 5.7.</b> Segmentation d'une masse circonscrite et une masse spiculée utilisant l'algorithme proposé par (Li et al., 2008), a) Initialisation, b) Propagation et c) Convergence vers la limite de la masse.....	112
<b>Figure 5.8.</b> Calcul de la mesure de rectangularité.....	114
<b>Figure 5.9.</b> Processus utilisé pour la création d'une matrice de co-occurrence.....	119
<b>Figure 5.10.</b> Exemple d'application de la matrice GLCM sur une image de masse maligne et une image de masse bénigne.....	119
<b>Figure 5.11.</b> La valeur du contraste avec $\theta=0^\circ$ utilisant une fenêtre 32x32 pixels.....	122
<b>Figure 5.12.</b> La valeur de la corrélation avec $\theta=0^\circ$ utilisant une fenêtre 32x32 pixels.....	122
<b>Figure 5.13.</b> La valeur de la variance avec $\theta=0^\circ$ utilisant une fenêtre 32x32 pixels.....	123
<b>Figure 5.14.</b> La valeur de du moment avec $\theta=0^\circ$ utilisant une fenêtre 32x32 pixels.....	123
<b>Figure 5.15.</b> La valeur de l'énergie avec $\theta=0^\circ$ utilisant une fenêtre 32x32 pixels.....	124
<b>Chapitre VI : Classification Basée AC</b>	
<b>Figure 6.1.</b> Voisinage hybride de l'AC.....	127
<b>Figure 6.2.</b> Exemple de l'automate cellulaire. L'état Alive (-1), l'état Dead (0), l'état Isolated (2) et l'état Dynamic (contient le vecteur de l'image).....	128
<b>Figure 6.3.</b> Schéma illustrative des étapes de la classification.....	129
<b>Figure 6.4.</b> Courbes ROC pour: trait en boules (un test diagnostique parfait assurant une séparation idéale entre les sujets malins et bénins), trait interrompu (un test diagnostique d'apport nul ne produisant aucune discrimination entre les individus), trait continu et trait mixte fin (des tests diagnostiques très informatifs).....	131
<b>Figure 6.5.</b> Carte navigable de l'AC.....	132
<b>Figure 6.6.</b> Courbe ROC obtenue avec classification AC et le descripteur Aire.....	133
<b>Figure 6.7.</b> Courbe ROC obtenue avec classification AC et le descripteur Périmètre.....	134

<b>Figure 6.8.</b> Courbe ROC obtenue avec classification AC et le descripteur Compacité....	134
<b>Figure 6.9.</b> Courbe ROC obtenue avec classification AC et le descripteur Excentricité...	135
<b>Figure 6.10.</b> Courbe ROC obtenue avec classification AC et le descripteur Rectangularité.....	135
<b>Figure 6.11.</b> Courbe ROC obtenue avec classification AC et le descripteur Rectangularité modifiée.....	136
<b>Figure 6.12.</b> Courbe ROC obtenue avec classification AC et le descripteur Haralick.....	137
<b>Figure 6.13.</b> Courbe ROC obtenue avec classification AC et le descripteur LRN.....	137
<b>Figure 6.14.</b> Courbe ROC obtenue avec classification AC et les descripteurs LRN/Hu...	138
<b>Figure 6.15.</b> Courbe ROC obtenue avec classification AC et le descripteur Haralick/LRN.....	139
<b>Figure 6.16.</b> Courbes ROC obtenue avec classification AC et les descripteurs Haralick/LRN et la combinaison Haralick\LRN.....	139
<b>Figure 6.17.</b> Courbe ROC obtenue avec classification AC et le descripteur d'aire.....	140
<b>Figure 6.18.</b> Courbe ROC obtenue avec classification AC et le descripteur Périmètre.....	141
<b>Figure 6.19.</b> Courbe ROC obtenue avec classification AC et le descripteur Excentricité.....	141
<b>Figure 6.20.</b> Courbe ROC obtenue avec classification AC et le descripteur Compacité...	142
<b>Figure 6.21.</b> Courbe ROC obtenue avec classification AC et le descripteur Rectangularité.....	143
<b>Figure 6.22.</b> Courbe ROC obtenue avec classification AC et le descripteur Rectangularité modifiée.....	143
<b>Figure 6.23.</b> Courbe ROC obtenue avec classification AC et les descripteurs d'Haralick.....	144
<b>Figure 6.24.</b> Courbe ROC obtenue avec classification AC et les descripteurs LRN.....	145
<b>Figure 6.25.</b> Courbe ROC obtenue avec classification AC et les descripteurs Hu/LRN.....	146
<b>Figure 6.26.</b> Courbe ROC obtenue avec classification AC et les descripteurs Haralick/LRN.....	146

<b>Figure 6.27.</b> Courbes ROC obtenues avec classification AC et les descripteurs Haralick/LRN et la combinaison Haralick\LRN.....	147
<b>Figure 6.28.</b> Courbes ROC obtenues avec classification AC et les descripteurs et tous les descripteurs utilisés sur la base DDMA (a) et MIAS (b).....	148
<b>Figure 6.29.</b> Comparaison avec quelques méthodes de la littérature.....	150
<b>Chapitre VII : Indexation et Recherche des Images Par le Contenu Basée Agents</b>	
<b>Figure 7.1.</b> Pathologies mammaires : a) Kystes mammaires bénins, b) Cancer.....	155
<b>Figure 7.2.</b> Schéma descriptif de la méthode d'indexation.....	156
<b>Figure 7.3.</b> Schéma illustrant l'étape de raffinement.....	157
<b>Figure 7.4.</b> Architecture du système proposé.....	158
<b>Figure 7.5.</b> Environnement du système.....	159
<b>Figure 7.6.</b> Architecture d'un agent Incoordonné.....	160
<b>Figure 7.7.</b> Fonctionnement du système.....	163
<b>Figure 7.8.</b> Résultat d'une recherche d'images mammographies malignes dans la base DDMA par la texture.....	165
<b>Figure 7.9.</b> Résultat d'une recherche d'images mammographies malignes dans DDMA et MIAS (combinaison Haralick et LRN).....	166
<b>Figure 7.10.</b> Courbe moyenne de rappel et précision du descripteur LRN, a) la base DDMA, b) la base MIAS.....	167
<b>Figure 7.11.</b> Courbe moyenne de rappel et précision du descripteur Haralick, a) la base DDMA, b) la base MIAS.....	168
<b>Figure 7.12.</b> Courbe moyenne de rappel et précision des descripteurs combinés Haralick et LRN, a) la base MIAS, b) la base DDMA.....	168
<b>Figure 7.13.</b> Courbe moyenne de rappel et précision des descripteurs sur DDMA et MIAS.....	169
<b>Figure 7.14.</b> Exemple des images à annoter.....	170

---

# Liste des Tableaux

---

<b>Tableau 1.1.</b> Quelques voisinages classiques.....	9
<b>Tableau 2.1.</b> Classification des situations d'interaction.....	39
<b>Tableau 4.1.</b> Les différents types de calcifications.....	94
<b>Tableau 4.2.</b> Classification de Le Gal.....	99
<b>Tableau 4.3.</b> Critères de malignité et bénignité.....	103
<b>Tableau 6.1.</b> Aire sous la courbe.....	133
<b>Tableau 7.1.</b> MAP pour une requête maligne de la base DDMA.....	165
<b>Tableau 7.2.</b> MAP des descripteurs utilisés sur les deux bases MIAS et DDMA.....	166

---

## Liste des Abréviations

---

AC	Automates Cellulaires
ACC	Agent Communication Chanel
ACL	Agent communication Language
ACO	Ant Colony Optimisation
ACR	American college of Radiology
AFUM	Average Fraction Under Minimum
AIS	Administrative Information System
AMS	Agent Management System
AND	Acide Désoxyribonucléique
ANN	Artificiel Neural Networks
BI-RADS	Breast Imaging Reporting and Data System
BSP-trees	Binary Space Partionning Trees
BUS	Breast Ultrasound
CAST	Cluster Affinity Search Technique
CBIR	Content Based Image Retrieval
CDDL	Common Development and distribution licence
CDS	Système de Diagnostic Clinique
CMJ	Cyan, Magenta, Jaune.
CMJN	Cyan, Magenta, Jaune, Noir pur
DDMA	Digital Database for Mammography d'Annaba
DF	Directory Facilitator
DICOM	Digital Imaging and Communication in Medicine
EM	Expectation- Maximisation
EMR	Electronic Medical Record Information System
ENS	Elliptic Normalized Skeleton
ES	Système Expert
FCL	Feature and Content Layer
FT	First Tier

GHDO	Generic Human Disease Ontology
GLCM	Gray-Level Co-Occurrence Matrix
HIS	Healthcare Information System
IA	Agent Intelligent
IAD	Intelligence Artificielle Distribuée
IGDS	Image Guided Decision Support
JVM	Java virtual Machine
IRM	Imagerie par Résonance Magnétique
KDD	Knowledge Discovery from Databases
KQML	Knowledge Query and Manipulation
LI	Lobulation Index
LRN	Normalized Radial Length
LRNM	Longueur Radiale Normalisée Modifiée
MADHS	Multi- Agent Diagnosis Helping System
MIAS	Mini Mammographic Database
MIL	Multiple Instance Learning
MIS	Medical Information Systems
MLP	Multi Layer Perceptron
MSIS	Medical Support Information System
NBMS	Naive Bayesian Machine systems
NIS	Nursing Support Information System
NSPD	Number of Substantial Protuberances and Depressions
PS	Points Simples
PT	points Terminaux
QBIC	Query by Image Content
RBF	Radial Basis Functions
RDL	Raw Data Layer
RETIN	Recherche et traque Interactive
RVB	Rouge, Vert, Bleu
SBE	Score Bull-Eye
SeSAm	Shell for Simulated Agent Systems
SIMPLIcity	Semantics-Sensitive Integrated Matching for Picture Libraries
SL	Schema Layer
SMA	Système Multi-Agent
SOM	Self Organizing Map
SVM	Support Vector Machine
TSL	Teinte, Saturation, Luminance

---

# Introduction générale

---

Les sources d'images numériques appareils photos, téléphone ou webcams sont présentés dans les foyers et produisent des quantités importantes de nouvelles collections.

Le développement des ordinateurs puissants dans le calcul et le stockage des informations a supprimé le besoin du tri et d'élimination.

Ces données volumineuses rendent l'humain incapable de les traiter manuellement, si on cherche une photo alors on est obligé de parcourir aléatoirement des dossiers contenant des milliers d'images. Cette tâche est devenue difficile pour l'être humain alors l'outil informatique s'impose pour l'automatiser.

L'indexation d'une image consiste à stocker l'information qui la décrit, cette information est appelée index, une fois la collection des images est indexée alors se pose la question de la recherche d'images.

Dans la suite de cette thèse, nous nous focaliserons sur l'indexation basée contenu, regardant le processus d'indexation, classiquement il se base sur ces étapes majeures :

- Le traitement d'image qui se base sur l'extraction des caractéristiques de l'image comme la couleur ou la texture afin d'extraire la signature de l'image.
- Une comparaison de la signature fournie par l'utilisateur avec l'ensemble des signatures de la base afin de trouver une liste d'images jugées pertinentes par rapport à ce que recherche l'utilisateur du système.
- Une étape optionnelle qui consiste à une interaction avec l'utilisateur, par exemple si l'utilisateur n'est pas satisfait de la réponse du système alors il peut raffiner le résultat par une interaction afin d'avoir une réponse pertinente.

Le développement des ordinateurs puissants dans le calcul et le stockage des informations a favorisé fortement la production d'images médicales numériques attachées à diverses modalités comme l'IRM, la mammographie...

Dans ce cas plusieurs problèmes se posent alors : la structuration des bases d'images *organisation*, *indexation*, la recherche d'une image particulière *fouille d'images* et la compression, transfert des images entre utilisateurs *le codage et la sécurisation*.

Le médecin souhaite un enrichissement de son diagnostic en consultant des cas correspondant à des images semblables. Le système de fouille d'images est un outil d'amélioration de la connaissance.

L'imagerie médicale est très importante dans le dépistage du cancer du sein, ce type de cancer est la localisation cancéreuse la plus fréquente et la première cause de mortalité par cancer chez la femme dans le monde, l'Algérie se situe parmi les pays du nord Afrique à plus forte incidence pour le cancer du sein et présente environ 50.70% du cancer comparativement avec 9 autres cancers les plus répandus.

## Cadre du travail

Le but est de produire un démonstrateur d'un système interactif reliant bases de données d'images médicales et utilisateurs *médecins* dans le but d'aide au diagnostic.

Selon notre recherche et discussion avec les spécialistes du domaine, on a conclu que généralement le travail des médecins, en particulier celui des radiologues, est fortement basée

sur la recherche d'indices visuels dans les images. Ainsi notre travail se base sur la recherche d'images par le contenu, particulièrement sur la recherche dans des bases d'images identiques à une image requête, il existe un nombre d'applications qui dépendent de la nature de l'objectif par exemple :

- On prend l'exemple d'une base d'images expertisées, un étudiant peut formuler un diagnostic, avec ce type de système les images semblables retournées peuvent l'aider pour vérifier son diagnostic et lui donner des informations grâce aux commentaires attachés aux images.
- On a aussi l'exemple d'une base qui contient un ensemble d'images des patients, le médecin peut faire sa recherche pour récupérer tous les cas similaires au cas de la requête qu'il lui aide dans le diagnostic et améliore son suivi thérapeutique.

## Problématique

Nous avons orienté notre travail vers l'application médicale avec l'aide des médecins du service de radiologie Ibn Rochd Annaba et Ibn Zohr Guelma. Après maintes discussions avec ces experts sur l'interprétation des pathologies dans les images de mammographie, les médecins indiquent que la recherche et la comparaison des images similaires de l'image en question aide mieux le médecin dans son diagnostic.

Notre objectif alors est de construire un système de recherche par le contenu des images de mammographie, bien que cet examen ne soit pas suffisant pour établir un diagnostic, il permet de dépister les lésions suspectes qui seront ensuite investiguées pour poser un diagnostic.

Partant de ce contexte d'utilisation du système, nous avons cherché à exploiter l'information donnée par l'utilisateur grâce aux annotations du système, nous avons cherché à exploiter l'information donnée par l'utilisateur grâce aux annotations afin d'améliorer la sélection des sites contenant des images intéressantes ainsi que la mesure de pertinence basée sur le contenu des images.

Nous avons utilisé les systèmes multi-agents (SMA), ces entités autonomes ayant la faculté de se déplacer sur l'automate cellulaire (AC), afin d'effectuer la recherche d'images.

Dans le système ainsi obtenu, les agents parcourent l'AC à la recherche d'images pertinentes relatives à la requête de l'utilisateur. Dès qu'ils ont trouvé des images pertinentes, le résultat est affiché. Les agents marquent l'AC collectivement afin de faire apparaître les chemins

pertinents. Nous avons effectué les expériences sur une base avec des images collectées de l'hôpital Ibn Rochd et la base MIAS.

Ensuite, le travail s'est focalisé sur la réutilisation des informations obtenues lors de l'utilisation du système dans la première phase afin d'améliorer la recherche.

## Organisation du mémoire

Dans ce mémoire, notre contribution se décompose en deux grandes parties : la première comprend quatre chapitres présentant les outils indispensables pour développer l'idée, état de l'art, des définitions des notions, principes et approches dans les automates cellulaires (AC), les systèmes multi-agents (SMA), l'indexation et recherche d'images et l'imagerie médicale. Une seconde partie expose trois chapitres sur l'organisation des images médicales pour faire une recherche rapide et pertinente.

La première partie est composée de quatre chapitres, un premier chapitre décrit les AC et ses propriétés, les types des AC et les domaines d'application. Le deuxième chapitre fait le point sur les systèmes multi-agents, agents, les caractéristiques, l'émergence et l'auto-organisation avec une brève présentation de l'application des SMA dans le domaine médical. Le troisième chapitre est consacré à l'indexation des images, les méthodes, les différents systèmes de recherche existant, les mesures d'évaluation et état de l'art sur la classification. Dans le chapitre quatre nous présentons des statistiques sur le cancer du sein, les techniques utilisées pour diagnostiquer la maladie, les types et les classifications du cancer du sein.

Une deuxième partie est formée de trois chapitres: A part la structuration des bases d'images médicales qui est indispensable, la recherche d'images par le contenu en médecine trouve son intérêt, le premier chapitre contient la méthode utilisée pour faire extraire les zones importantes dans les images et sur lesquels nous appliquant les descripteurs choisis. Le deuxième chapitre explique une méthode de classification basée sur les AC, cette méthode est avec laquelle nous avons organisé l'indexe. Le troisième chapitre présente la méthode de recherche d'images par le contenu, ensuite, nous évaluons notre système d'indexation et de recherche par l'intermédiaire d'une base d'images créée par les médecins et nous-mêmes ainsi que la base MIAS. Enfin ce mémoire s'achève par une conclusion et des perspectives.

# **Première Partie**

## **Les outils utilisés pour la réalisation d'un système CBIR appliqué aux images médicales**

<b>Chapitre I : Les Automates Cellulaires.....</b>	<b>5</b>
<b>Chapitre II : Les Systèmes Multi-Agents.....</b>	<b>29</b>
<b>Chapitre III : Indexation et Recherche d'images.....</b>	<b>53</b>
<b>Chapitre IV : L'image de Mammographie.....</b>	<b>85</b>

# Chapitre I

---

## Les Automates Cellulaires

---

### 1.1. Introduction

C'est en s'intéressant à l'évolution de construction graphique produite avec des règles simples que le mathématicien Stanislas Ulam<sup>1</sup> donna naissance aux AC (*Automates Cellulaires*).

Le point de départ est de construire un réseau régulier avec deux états finis (allumé et éteint), pour construire la configuration de départ Ulam commence à allumer aléatoirement quelques cellules. A partir de cette configuration dans le temps ( $t_0$ ), la configuration qui suit ( $t + 1$ ) va être déterminée en fonction du voisinage de cette cellule générée par les règles locales.

Par exemple si on prend une cellule éteint à côté de deux cellules allumées dans  $t_0$  alors cette cellule s'allumera dans ( $t + 1$ ) sinon elle s'éteint, à partir de ça Ulam constate que a partir des règles très simples on peut construire des structures complexes et que ce mécanisme simple peu conduire à générer des figures très complexe qui peuvent être dans certains cas répliqués.

Les premiers automates cellulaires bidimensionnels ont été construits sur des grilles carrées avec des graphes de 4 à 8 voisins. Chaque région est constituée de cellules binaires et chaque cellule correspond à un emplacement possible constituant la grille, le contenu de la cellule prend une valeur 1 lorsqu'elle est occupée et prend 0 lorsqu'elle est vide (libre) [Heudin, 1994]. Ce contenu est libre par exemple de présenter des molécules en cours de réaction, le

---

<sup>1</sup> Stanislaw Marcin Ulam (1909- 1984) Mathématicien américain d'origine polonaise. Il a employé la méthode de Monte-Carlo pour évaluer les intégrales mathématiques, cela conduisit au développement de la méthode de Monte-Carlo par Von Neumann, Metropolis,... Il a montré aussi que le modèle de la bombe à hydrogène inexact alors il a proposé une meilleure méthode par lui-même. Il a travaillé à la théorie des ensembles, la topologie, la théorie ergodique, et d'autres domaines et il a utilisé les mathématiques à la physique et à la biologie.

développement d'une épidémie, une population biologique, des particules élémentaires d'un liquide, d'un gaz ou encore des données de différents types.

La structure des automates cellulaires permet d'accéder directement à un emplacement bien défini dans cette espace à deux dimensions, ce point constitue le principal avantage des automates cellulaires. L'évolution des automates cellulaires est basée sur une application de transformation locale entre les cellules. A partir d'un état initial les opérateurs de l'automate sont appliqués séquentiellement dans un processus itératif. Toutes les cellules du réseau changent leurs états avec un traitement simultané et une application du même nombre des règles d'évolution.

Il en résulte qu'un automate cellulaire présente un système dynamique discret avec une conception simple basée à une géométrie uniforme.

### 1.2. Historique

La théorie des automates cellulaires est née en 1948 de la tentative de J. Von Neumann<sup>1</sup>, il a pensé à concevoir une machine pouvant générer des comportements complexes et variés à partir de lois simples de fonctionnement et d'imiter le fonctionnement du cerveau humain en s'inspirant de la biologie.

Tout d'abord, Von Neumann a pensé à la machine auto-reproductrice, cette machine est conçue pour puiser des pièces dans un bac de composants simples, il suffit de monter ces composants d'une manière à former une machine identique à elle-même une fois la mise en marche de la machine-fille [Neumann, 1966].

Von Neumann a abordé le problème par des moyens techniques mais ce projet été irréalisable, avec une collaboration avec son collègue S. Ulam lui suggéra de travailler sur un monde abstrait pour s'affranchir des contraintes posées par les lois de la physique, Von Neumann décide alors d'examiner la question dans un espace mathématique simple qui serait une métaphore du monde réel. Cette espace est représenté par une grille des cellules et chaque cellule a un état choisi dans un ensemble des états finis. Chaque cellule mise à jour son état selon le voisinage perçu, dans le modèle Von Neumann la mise à jour est limité par l'état de la cellule elle-même et à celui des quatre cellules adjacentes (Nord, Sud, Est, Ouest).

Ensuite le concept d'automate a été utilisé sous des terminologies différentes, dans des contextes disciplinaires divers (ingénierie, l'analyse d'images, etc.).

---

<sup>1</sup> John Von Neumann (1903-1957) Scientifique qui a contribué de façon très significative à de nombreux domaines scientifiques. Il a donné son nom à « l'architecture de Von Neumann » qui constitue encore aujourd'hui l'ossature des ordinateurs. Il a en outre contribué à la création de l'EDVAC, l'un des premiers ordinateurs électroniques américains.

La popularisation de ces modèles trouve son origine dans le jeu de la vie introduit par le mathématicien J. Conway<sup>1</sup> en 1970 [Gardner, 1970]. La motivation pour cette automate cellulaire est d'exhiber une règle simple engendrant un comportement complexe.

La valeur 0 désigne une cellule morte, la valeur 1 une cellule vivante, le voisinage est le site lui-même et ses 8 voisins. Ce jeu prend sa place sur un réseau bidimensionnel, à chaque pas de temps chaque cellule change son état en fonction du sien et de celui des huit cellules adjacentes.

Cette modélisation de l'évolution d'une population des cellules (vivantes ou mortes) par la règle locale simple a d'une manière inattendue, révélé des comportements divers et remarquables.

Dans les années 80 Stephen Wolfram a utilisé les automates cellulaires mettant en jeu les règles simples pour étudier la complexité dans la nature. Dans son livre *A New Kind of Science*, systématise son approche de la complexité et a contribué beaucoup au domaine des automates cellulaires.

Actuellement, les automates cellulaires sont généralisés dans de nombreux domaines, on l'utilise dans la modélisation des systèmes biologique, physiques et informatique, notamment la reconnaissance de formes et la cryptographie.

### 1.3. Définition

Un automate cellulaire est un système dynamique discret caractérisé par des entités élémentaires ou cellules organisées en réseau régulier et évolua dans le temps. Chaque cellule est caractérisée par une valeur qui appartient à un ensemble finis d'état. L'évolution de ces valeurs au cours du temps se fait de manière parallèle, chaque cellule voyant sa valeur actualisée selon une règle locale dans laquelle interviennent les valeurs des cellules voisines.

Formellement,  $A$  est un quadruplet  $(Z^d, \varepsilon, V, \delta)$  où  $Z^d = R$  est le réseau de cellules,  $\varepsilon$  est l'ensemble finit d'états, la partie finie ordonnée  $V = (z_1, z_2, \dots, z_{|V_A|})$  de  $Z^d$  est le voisinage et l'application  $\delta$  est la règle locale de transition de l'automate cellulaire  $A$ .

Etant donné un automate cellulaire  $A$ , on désignera par la suite par  $\varepsilon_A, V_A$  et  $\delta_A$  respectivement l'ensemble d'états, le voisinage et la règle locale de transition de  $A$ .

---

<sup>1</sup> John Horton Conway (1937) Il est principalement connu pour sa théorie des jeux combinatoires et pour avoir inventé le jeu de la vie, un automate cellulaire. Il a écrit en 1976 le premier livre traitant du sujet, *On Numbers and Games*, puis *Winning Ways for your Mathematical Plays*. Il est également l'un des inventeurs du jeu *Sprouts*, *Phutball*. Il a développé des jeux et casse-tête le *Cube Soma*, le *solitaire*, et les *Soldats de Conway*. Il est aussi à l'origine du problème de l'ange finalement résolu en 2006 ainsi que de la suite de Conway. Il a des travaux dans la géométrie, la théorie des groupes, la théorie des nombre, l'algèbre et la physique théorie.

Une configuration  $c$  de l'automate cellulaire est l'attribution pour chaque cellule  $z \in R$  un état  $c(z)$ , étant donnée une configuration  $c_0$ , l'application de la règle de transition locale donnée par :

$$c_1(\varepsilon) = \delta_A(c_0(\varepsilon + \varepsilon_1), \dots, c_0(\varepsilon + \varepsilon_v)) \tag{1.1}$$

### 1.4. Propriétés

Nous présentons dans ce qui suit quelques variantes et propriétés rattachées à cette définition.

#### 1.4.1. L'espace cellulaire

L'espace cellulaire  $Z^d$  est un réseau de dimension  $d$  formé par des cellules  $c$  arrangé selon la dimension de l'espace et de la forme choisie pour les cellules. Cette espace est présenté par une grille rectangulaire à une dimension ou  $Z^d$  se limite à une corde linéaire ou les éléments peuvent être disposés en ligne droite ou autour d'un tore, ou deux dimension dans ce cas le réseau peut être triangulaire hexagonal (figure (1.1)) aussi. Pour la dimension trois le réseau prend une géométrie cubique.

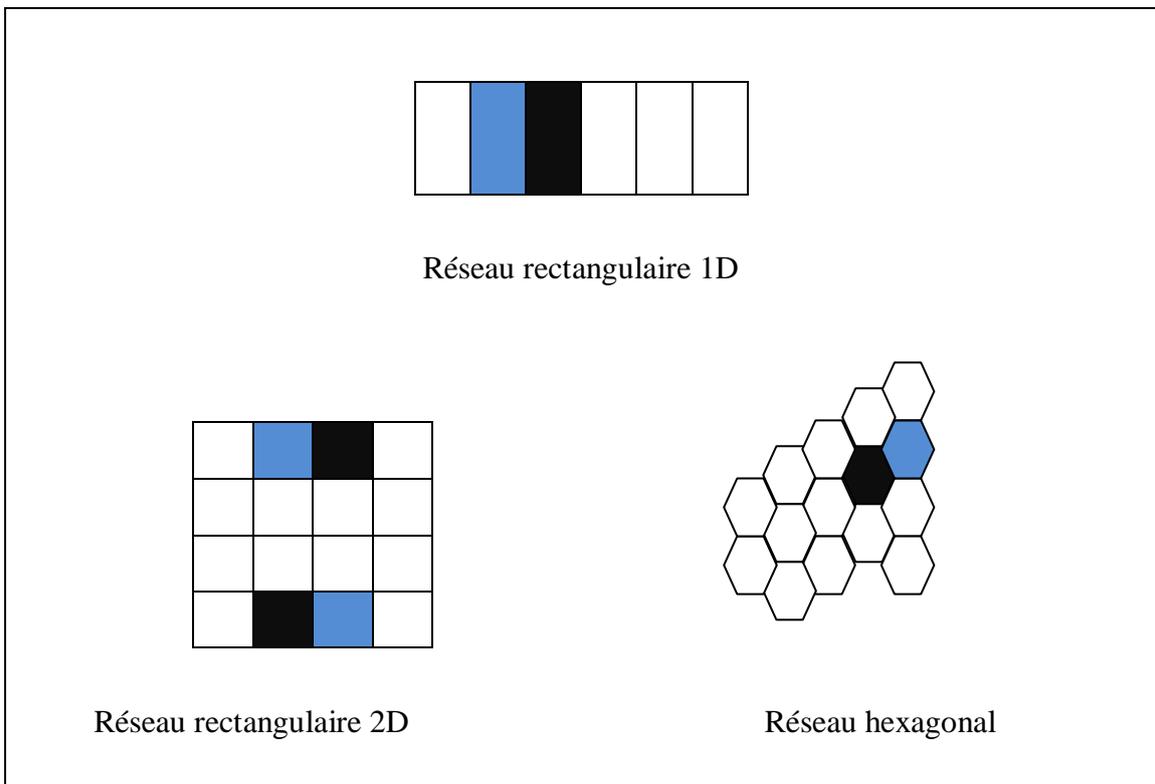


Figure1.1. Forme typique du réseau cellulaire

### 1.4.2. Voisinage

Le voisinage est un ensemble ordonné de vecteurs de déplacement sur la grille  $Z^d$ . Le choix de cet ordre présente un outil pour définir l'application de la règle locale mais n'a pas d'incidence sur l'évolution de l'automate cellulaire.

En effet : soit  $A$  notre automate cellulaire de dimension  $d$  et de voisinage  $(Z_1, Z_2, \dots, Z_{|V_A|})$ , et soit  $\tau$  une permutation de  $[1, V]$ .

L'automate cellulaire  $A_\tau$  définit par  $(Z^d, \varepsilon_A, \dots, V_{\tau(z)}, \delta_\tau)$ .

On travaillera toujours avec le ré ordonnément des voisins près mais on désigne par le mot voisinage l'ensemble des voisins indépendamment de l'ordre.

Dans le cas monodimensionnel le voisinage peut être représenté par (figure (1.2)). en deux dimensions quelque voisinages sont représenté dans (figure (1.3)). Cependant les voisinages les plus utilisés sont les voisinages de Moore et Von Neumann.

**Tableau 1.1.** Quelques voisinages classiques

Numéro	Nom	Voisinages
1	Voisinage unidirectionnel	$\{-1, 0\}^d$
2	Voisinage de Moore	$\{V \mid \ V\ _\infty \leq 1\}$
3	Voisinage de Von Neumann	$\{V \mid \ V\ _1 \leq 1\}$
4	Voisinage carrée de rayon $n$	$\{V \mid \ V\ _\infty \leq n\}$

Si on prend le voisinage de Moore et en ajoute les 4 voisins les plus proches dans le voisinage de Von Neumann qui ne sont pas inclus dans le voisinage de Moore, on obtient un voisinage de Von Neumann modifié (figure (1.3)).

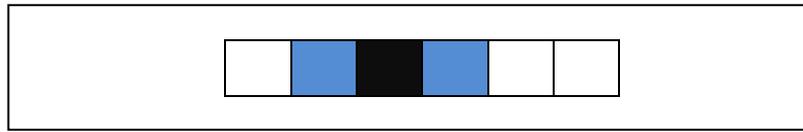


Figure 1.2. Voisinages Von Neumann et Moore dans le cas 1D

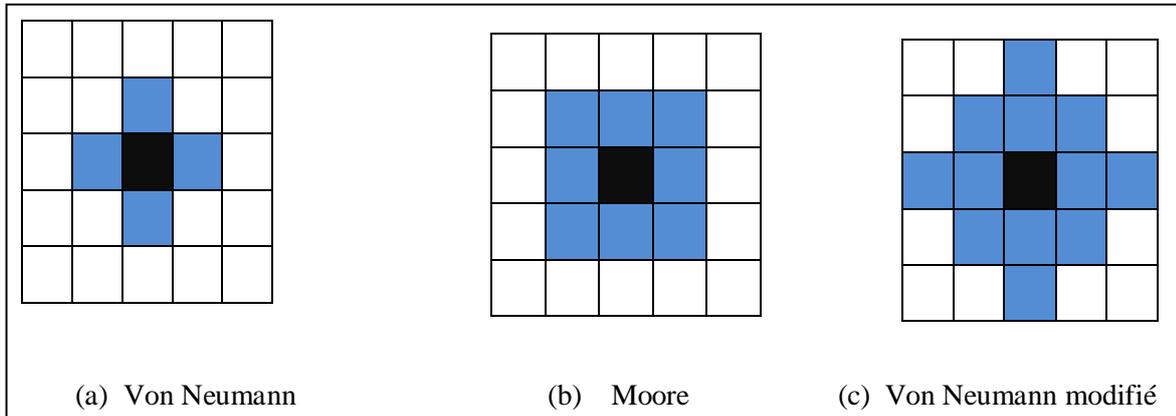


Figure 1.3. Formes de voisinages les plus fréquentes

D'autres types de voisinages peuvent être définis de façon arbitraire. On peut citer les voisinages de type Smith, Cole et Custom.

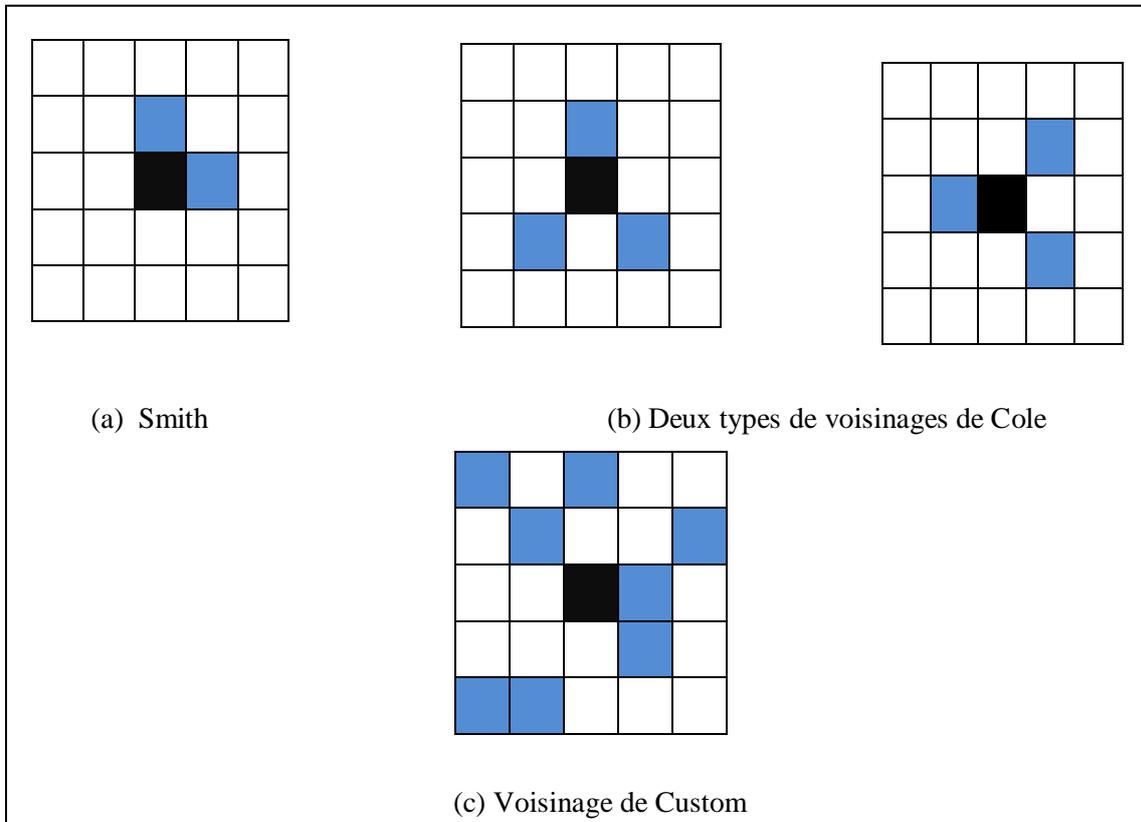


Figure 1.4. Voisinages de Smith, Cole et Custom

### 1.4.3. Ensemble d'états

Le renommage de l'ensemble d'état n'est pas essentiel du point de vue de l'évolution de l'automate cellulaire car les états peuvent être renommés pour une adaptation à la règle locale. Formellement on parle de la configuration d'un automate cellulaire donné par l'application  $e : \tau \rightarrow \varepsilon$  qui fait correspondre à chaque cellule  $c$  de  $\tau$  une valeur prise dans  $\varepsilon$  qui constituera l'état de la cellule  $c$  à l'instant  $t$ .

Où :  $\varepsilon$  la liste des valeurs prise par l'état d'une cellule, cette liste d'états peut être désignée par des valeurs entières, des couleurs, définies par des conditions et peut contenir autant d'éléments nécessaire pour l'étude [Theyssier, 2005].

Exemple :

$\varepsilon = \{0, 1, 2, \dots, k_{\varepsilon-1}\}$  avec  $k_{\varepsilon} \in \mathbb{N}$

Où :  $\varepsilon = \{\text{bleu, blanc, } \dots\}$

### 1.4.4. Règle locale de transition

Le choix des règles de transition est important pour la construction des automates cellulaires car à partir de ces règles nous pouvons définir la dynamique locale ou globale de l'automate cellulaire, elles permettent de mettre à jour l'automate à chaque itération, en ajoutant des contraintes sur la règle et on obtient des familles classiques d'automates cellulaires.

Ces règles dépendent de la géométrie de réseau, du voisinage, de l'ensemble des états et peut être déterministes ou probabilistes [Abdellaoui, 2003].

#### 1.4.4.1. Règles déterministes

Soit  $Z^d$  un réseau régulier,  $v$  un type quelconque de voisinage. Soit  $g$  la règle déterministe définie par :

$$\forall t \in I, \exists! e_i \in Q \text{ tel que } g(e_t(v)) = e_i \quad (1.2)$$

Où :  $Q = \{e_1, e_2, \dots, e_k\}$  est l'ensemble des états et  $I$  l'intervalle de temps discret.

Si  $v$  est de taille  $n$ ,  $e_t(v)$  est un élément de  $\varepsilon^n$ ,  $\forall t \in I$ . Donc  $g$  est définie dans  $\varepsilon^n$  et est à valeur dans  $\varepsilon$ .

#### 1.4.4.2. Règles probabilistes

Dans ce cas la fonction  $g$  peut fournir plusieurs valeurs possibles avec des probabilités associées non seulement un seul résultat pour chaque type de voisinage, on peut la définir

Par  $F$  :

$$F: \varepsilon^n \times \varepsilon \rightarrow [0, 1]$$

$$(e_t(v), e_i) \rightarrow p_i \quad (1.3)$$

$F$  fonction de transition

Où  $p_i$  est la probabilité d'obtenir l'état  $e_i$  partant d'une configuration de voisinage  $v$  de taille  $n$  donnée par  $e_t(v) \in \varepsilon^n$  avec

$$\sum_{i=1}^{k'} F(e_t(v), e_i) = 1 \quad (1.4)$$

Etant le nombre de tous les états envisagés.

### A. Définition Fonction de transition globale

« Soit  $C$  l'ensemble de toutes les configurations pour un espace cellulaire donné. Alors la fonction de transition globale  $F: C \rightarrow C$  est définie par

$$F(c)(\alpha) = f(h(\alpha)) \forall \alpha \in N \times N \text{ »}$$

Si on prend  $c_0$  comme une configuration initiale de cette automate, alors la fonction  $F$  va définir ce qu'on appelle une propagation, qu'est une séquence de configurations qui décrit l'évolution de l'automate dans le temps.

$$c_0, c_1, \dots, c_t, \dots \quad \text{avec } c_{t+1} = F(c_t) \forall t$$

### B. Propriétés des configurations

« -Deux configuration  $c$  et  $c'$  sont dite disjointes, lorsque  $\text{supp}(c) \cap \text{supp}(c') = \emptyset$ . Une configuration  $c$  et une zone  $U$  sont disjointes, si et seulement si,  $\text{supp}(c) \cap U = \emptyset$ .

-Soit une paire de configurations disjointes  $c$  et  $c'$ . L'union de  $c$  et  $c'$  est définie par :

$$(c \cup c')(\alpha) = \begin{cases} c(\alpha) & \text{si } \alpha \in \text{supp}(c) \\ c'(\alpha) & \text{si } \alpha \in \text{supp}(c') \\ v_0 & \text{sinon} \end{cases}$$

-Une configuration  $c$  est dite passive, si  $F(c) = c$  et complètement passive si toute sous-configuration  $c'$  de  $c$  est passive<sup>1</sup>.

-Une configuration  $c$  est dite stable, s'il existe un instant  $t$  tel que  $F^t(c)$  est passive.

---

<sup>1</sup> La passivité n'implique pas la passivité complète, par définition d'une sous configuration. L'inverse est vrai.

-Une configuration  $c_\delta$  est une translation de la configuration  $c$ , s'il existe un élément  $\delta \in \mathbb{N} \times \mathbb{N}$  tel que  $c_\delta(\alpha) = c(\alpha - \delta)$  où  $-$  est la soustraction produit sur  $\mathbb{N} \times \mathbb{N}$ .

-Soient  $c$  et  $c'$  deux configurations disjointes. La configuration  $c$  fait passer de l'information à la configuration  $c'$  s'il existe un instant  $t$  tel que ;

$$F^t(c \cup c') \upharpoonright Q \neq F^t(c') \upharpoonright Q$$

Où

$$Q = \text{supp}(F^t(c')). \gg$$

### 1.4.5. Conditions aux limites

Dans la théorie l'espace cellulaire est non borné, ainsi les automates cellulaires présente un système évolutionnaire ou les conditions aux limites doivent être précisées à l'avance pour la simulation, celles-ci peuvent être périodiques, réflexives, ou fixes.

#### 1.4.5.1. Les conditions ou limites périodiques

C'est les conditions aux limites par défaut, nous considérons ici les automates cellulaires à une dimension, les cellules gauche jouent le rôle de voisines des cellules droites et vice-versa, les cellules du haut jouent le rôle de voisines des cellules du bas et vice-versa.

#### 1.4.5.2. Les conditions aux limites réflexives

C'est les conditions aux limites ou l'automate cellulaire utilise un ensemble des cellules virtuelles rangées sur les bords.

#### 1.4.5.3. Les conditions aux limites fixes

C'est les conditions aux limites ou la solution se base sur l'ajout des cellules virtuelles qui comprennent une valeur d'état constante aux bords de l'automate cellulaire [Abdellaoui, 2003].

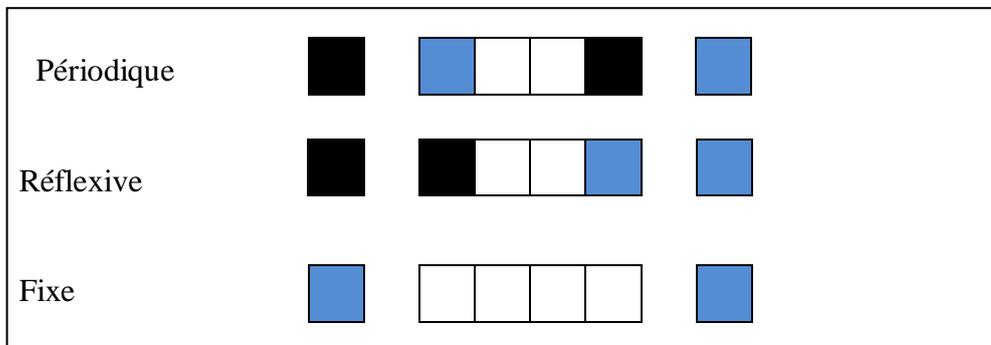


Figure 1.5.Représentation des conditions aux limites

### 1.1.4.6. Configuration initiale

La configuration désigne l'état dans lequel se trouve chaque cellule à un instant donné.

Soit l'état  $e_0(c)$  d'une cellule  $c$  à l'instant  $t = 0$ . Pour la cellule  $c$  la configuration initiale est considérée d'une façon aléatoire ou par une fonction [Theyssier, 2005], on peut la définir par :

$$\begin{aligned} e_0 : \tau &\rightarrow \varepsilon \\ c &\rightarrow e_0(c) \end{aligned} \tag{1.5}$$

#### A. Définition Configuration d'un automate cellulaire

« Une configuration (ou état générale réalisable du modèle cellulaire) est une fonction

$c: N \times N \rightarrow V$  telle que

$$\text{supp}(c) = \{\alpha \in N \times N \mid c(\alpha) \neq v_0\}$$

Soit fini.

On appelle  $c'$  une sous-configuration de la configuration  $c$  si

$$c \mid \text{supp}(c') = c' \mid \text{supp}(c')$$

Où  $\mid$  désigne la restriction fonctionnelle<sup>1</sup>.

Avec :  $v_0$  est l'état de repos.

La fonction  $c$  est dite de support fini relativement à l'état  $v_0$ .

Le symbole  $\alpha$  présente une cellule.

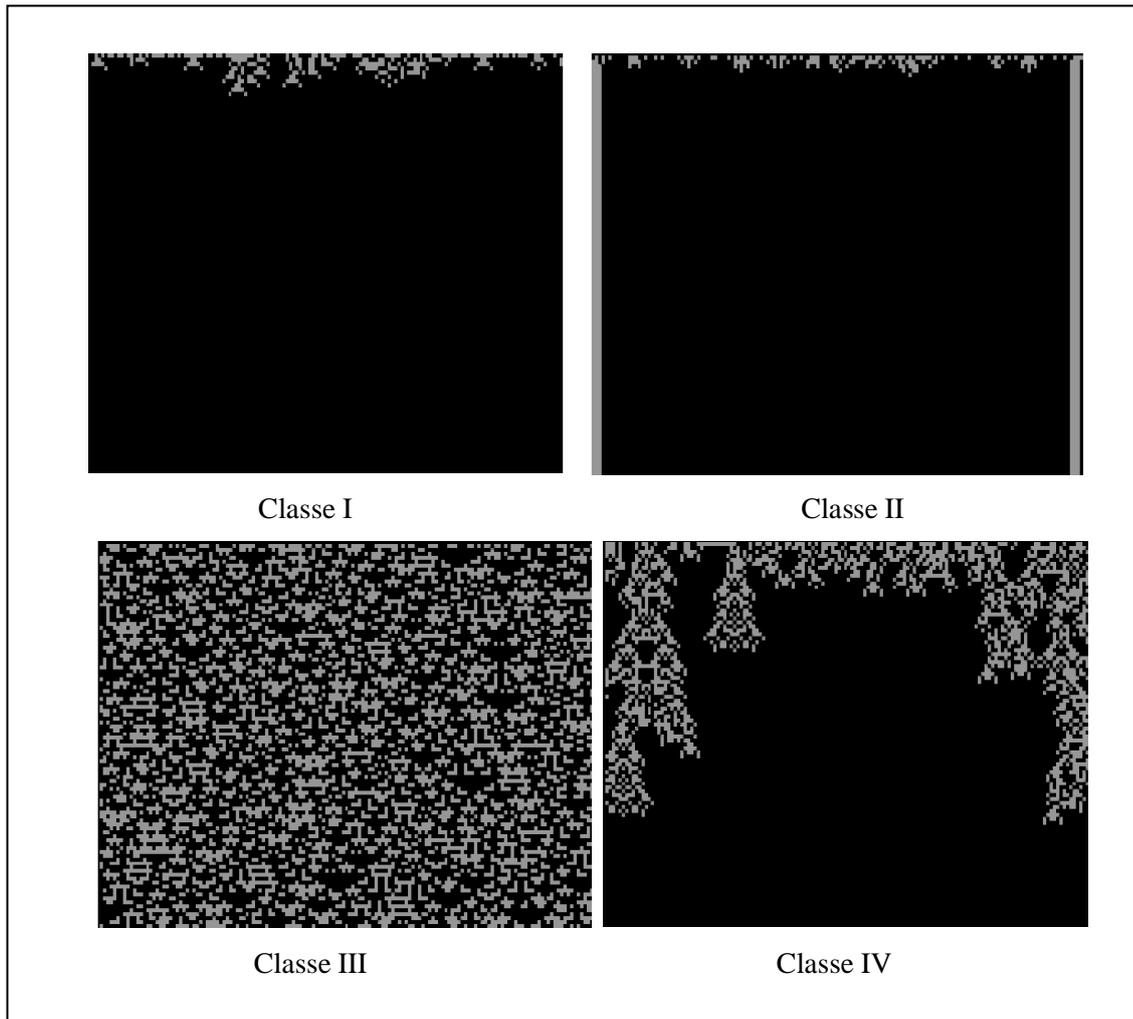
## 1.5. Classification des automates cellulaires

La motivation principale pour l'étude mathématique des automates cellulaires est de séparer les automates cellulaires simple des automates cellulaire complexe selon leurs propriétés afin d'utiliser cet outil pertinent dans un cadre de modélisation.

Il y a des tentatives pour classer les automates cellulaires en fonction de diagramme espace-temps, la plus connue est celle de Wolfram [Wolfram, 1984]. S. Wolfram exhibe un partitionnement des automates cellulaires en quatre classes, voici comment l'auteur présente cette classification dans un récent ouvrage [Wolfram, 2002]:

---

<sup>1</sup> Plus précisément  $c \mid A = \{(\alpha, c(\alpha)) \mid \alpha \in A\}$  pour un sous-ensemble  $A$



**Figure 1.6.** Les classes de Wolfram

**A. Classe I :** toutes les conditions initiales convergent en temps fini vers un même état uniforme;

**B. Classe II :** toutes les conditions initiales convergent en temps fini vers des états éventuellement distincts mais tous périodiques dans le temps;

**C. Classe III :** elle regroupe les automates cellulaires chaotiques qui sont aperiodiques et ne présentent pas de structures particulières;

**D. Classe IV :** rapidement des structures simples et localisées se mettent en place, qui interagissent entre elles de manière compliquée.

Cette classification est peu satisfaisante car la classe III regroupe les automates cellulaires chaotiques dont le comportement est complexe car difficilement prévisible et la classe IV comprend les automates cellulaires complexes dont le comportement est dit complexe car il en émerge des interactions très structurées qui semblent algorithmiquement complexes.

Ainsi, cette classification se base sur deux paires : un automate cellulaire et d'une configuration initiale qui sont classifiées et non les automates cellulaires eux-mêmes. D'autre part, le manque de formalisation de cette classification la rend inutilisable en pratique : S. Wolfram lui-même classe certains automates cellulaires dans plusieurs classes dans ses publications.

Pour la formalisation de cette classification K. Culik II et S. Yu ont travaillé sur cette classification et ont montré que l'appartenance aux classes est indécidable. Cette classification présente des inconvénients et d'autres propositions ont été faites pour formaliser la classification de S. Wolfram [Culik et al., 1988].

L'idée de R. H. Gilman [Gilman, 1987] est de reprendre les résultats obtenus en théorie des systèmes dynamiques et de les adapter au cas des automates cellulaires. On peut citer aussi une classification purement topologique par P. Kurka [Kurka, 1997].

Le principal problème de ces classifications dans le cadre de la topologie est la classification des classes correspondant aux automates cellulaires les plus complexes, beaucoup de travaux sont menés pour résoudre ce problème [Gattaneo et al., 1999, Formenti, 1998, Martin, 2001]

I. Rapaport a travaillé sur la classe IV de la classification de S. Wolfram qu'est considérée comme la classe des automates complexes dans les évolutions desquels apparaissent des structures simples et localisées qui interagissent entre elles de manière compliquée. C'est la seule tentative de classification qui se focalise sur ce type d'automates cellulaires [Rapaport, 1998].

Les classifications présentées ci-dessus reposent sur des notions de complexité variables : complexité visuelle des diagrammes espace-temps pour Wolfram, propriétés topologiques et de mesure ou encore propriétés géométriques et algébriques pour la classification.

### **1.5.1. Autres classes d'automates cellulaires**

Il existe autres classes d'automates cellulaires qui se varient dans leurs évolutions selon le mécanisme de transition, la mise à jour des cellules ou le voisinage, comme exemple on peut mentionner :

#### **1.5.1.1. Automates cellulaires asynchrones**

C'est l'automate cellulaire ou la mise à jour des cellules n'est pas simultanée [Abdellaoui, 2003].

### 1.5.1.2. Automates cellulaires à mémoires

Dans ce type d'automates cellulaires nous gardons en mémoire les cellules mise à jour, leur fonction de transition d'état est du type :

$$e_{t+1}(c) = f \left( e_t(v(c)), e_{t-1}(v(c)), \dots, e_{t-m}(v(c)) \right) \quad (1.6)$$

Où  $m$  la capacité de mémoire de la cellule  $c$  [Abdellaoui, 2003].

### 1.5.1.3. Automates cellulaires non-stationnaires

C'est l'automate cellulaire où les cellules peuvent mettre à jours leurs états suivant la règle :

$$e_{t+1}(c) = g_t \left( e_t(v_t(c)) \right) \quad (1.7)$$

### 1.5.1.4. Automates cellulaires flous

C'est la recherche des comportements complexes et l'étude d'impact des perturbations (sources de bruit,...) sur l'évolution d'automates cellulaire discret, on peut le définir un automate cellulaire flou par :  $A = (\tau, [0, 1], v, f)$  où :  $\tau$  est un réseau de dimension  $d$ , les états des cellules inclus dans l'intervalle  $[0, 1]$ ,  $v$  est le voisinage et  $f$  est la fonction de transition, cette fonction comprend des valeurs réelles représentant la fonction discrète en se basant sur des principes de la logique flou [Abdellaoui, 2003].

### 1.5.1.5. Automates cellulaires hiérarchiques

Soit un automate cellulaire monodimensionnels, on peut définir un automate cellulaire hiérarchique par :  $A_i = (\tau_i, \varepsilon, v_i, f_i)$  où  $0 \leq i \leq p, p \in \mathbb{N}$ ,  $\tau_i$  est une grille avec  $p_i$  cellules,  $\varepsilon$  est l'ensemble des états,  $v_i$  est le voisinages de taille  $n_i$ ,  $f_i$  est une fonction de transition.

La construction de l'automate cellulaire hiérarchique  $A = (A_0, A_1, \dots, A_p)$  est obtenue par le couplage de  $A_i$  avec  $A_{i+1}$ . Soit  $V_i$  est identifié à une configuration global de  $A_{i+1}$ . Chaque cellule  $c_i$  de  $A_i$  calcule son état à l'instant  $t + 1$  selon :

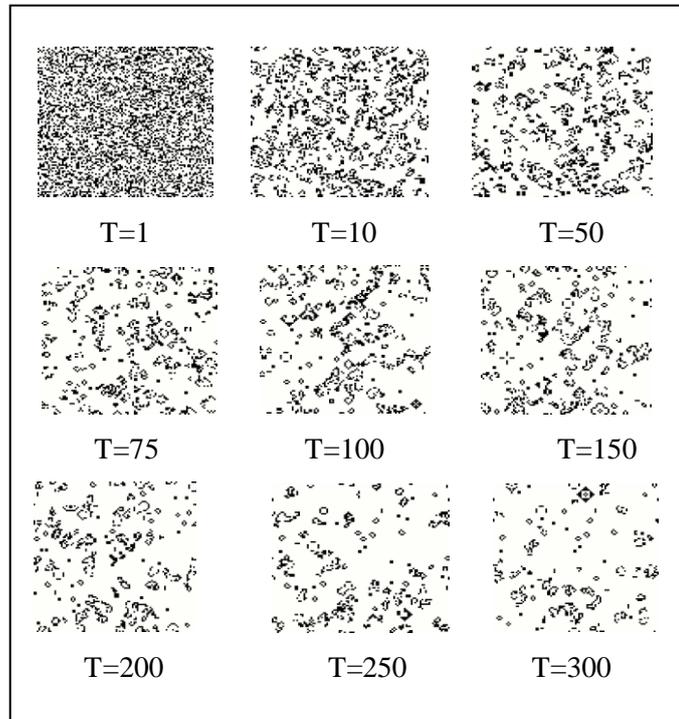
$$e_{t+1}(c_i) = f \left( e_t(v_i(c_i)), e_{i+1}^t(\tau) \right) \quad (1.8)$$

Où  $i = 0, \dots, p - 1$ , et  $e_{i+1}^t(\tau)$  désigne la configuration de  $A_{i+1}$  à l'instant  $t$ .

Qui mit à jours les états de ces non simultanément [Adamatzky, 1994]. Dans ce qui suit on va citer quelques automates cellulaires célèbres :

### 1.5.1.6. Jeu de la vie

Depuis quelques années certains automates cellulaires sortirent du lot à cause de leurs comportements particuliers. Dans les années 70 apparut l'automate le jeu de la vie par J.Conway [Conway, 1976].



**Figure 1.7.** Une exécution du jeu de la vie sur une configuration torique<sup>1</sup>

#### A. Emergence

« [. . . ] On peut appeler émergence les qualités ou propriétés d'un système qui présentent un caractère de nouveauté par rapport aux qualités ou propriétés des composants considérés isolément ou agencés différemment dans un autre type de système« [. . . ] » [Wolfram, 2002].

L'association complexe d'éléments aboutisse à des phénomènes et de nouveaux mécanismes, cette propriété est liée à la complexité à cause de la diversité et le croisement des éléments ainsi que liaisons et l'interaction non linéaires qui mènent à des comportements vraiment difficile à prédire [Alt et al., 1997].

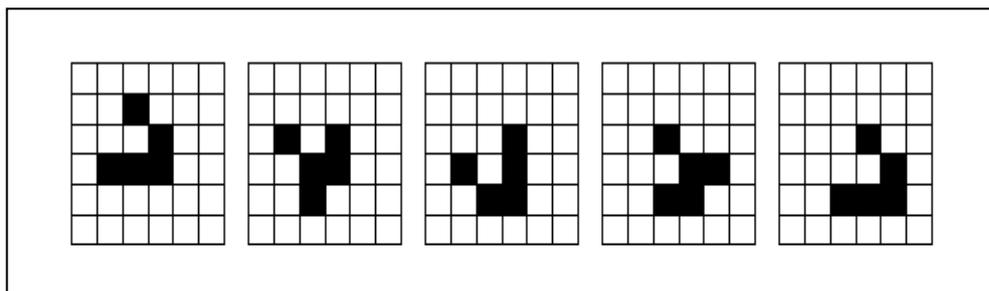
On dit que l'émergence est globale si elle caractérise les propriétés d'un système qui sont nouvelles par rapport aux propriétés de ses éléments organisés d'une manière isolé ou différente. Ce processus se base sur le parallélisme des interactions multiples entre les éléments indépendants, les automates cellulaires fonctionnent selon ce principe

---

<sup>1</sup> Simulation de l'automate cellulaire « Jeu de la Vie » sur la plateforme Netlogo

Les règles du Jeu de la vie sont fixées d'une manière à engendrer une diversité des structures imprévisibles.

Les spécialistes ont recensé plusieurs configurations, l'une des plus connue est le planeur (*glider*) qui apparaît après un remplissage aléatoire comme un signal qui se déplaçant dans la configuration, au bout de 4 itérations, le motif s'est déplacé d'une case vers le bas et d'une case vers la droite.



**Figure 1.8.** Planeur (*glider*)

On peut citer également le canon à planeurs de *Gosper*. Il s'agit d'un ensemble de cellules engendrant des planeurs.

### B. Le canon à planeurs

Le canon à planeurs de Gosper : tous les 30 pas de temps, un nouveau planeur est émis.

Pour l'automate jeu de la vie on a :  $S = \{\text{noir, blanc}\}$  l'ensemble des états, et  $G = Z^2$  l'automate cellulaire à deux dimensions, l'état noir désigne une cellule morte et l'état blanc désigne une cellule vivante. Les voisinages contiennent la cellule elle-même avec un voisinage de 8 connexités. La règle locale d'évolution est établie par :

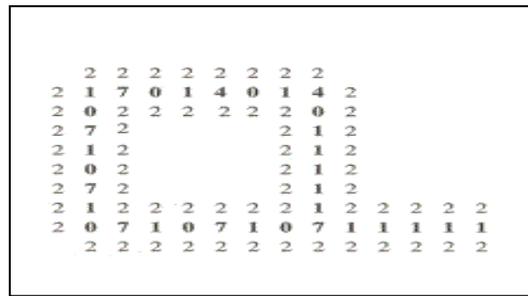
**A. Naissance:** si une cellule morte (état noire) à l'instant  $t$ , elle devient vivante (état blanc) à l'instant  $t+1$  si elle est entourée seulement par 3 cellule vivante.

**B. Mort par isolement ou surpopulation:** si une cellule vivante à l'instant  $t$ , elle meurt à l'instant suivant  $t+1$  si elle est entourée par 2 ou plus de 3 cellule vivante [Regnault, 2008].

### C. La boucle auto-répliquante de Langton

Dans les années 80, Christopher Langton<sup>1</sup> c'est inspiré des travaux d'Edgar Codd, il a réalisé un automate cellulaire avec des composants qui sont similaires à celles de acide désoxyribonucléique (ADN), il constitue l'information nécessaire à sa propre réplication.

<sup>1</sup> Christopher Langton (né 1948/1949) Informaticien américain et l'un des fondateurs du domaine de la vie artificielle. Il a développé plusieurs concepts et il a joué un rôle très important dans l'élaboration de systèmes complexes, notamment en biologie. Il a créé la fourmi de Langton et boucle de Langton, deux simples simulations de vie artificielle, en plus de son paramètre Lambda.



**Figure 1.9.** Boucle de Langton

L'automate de Langton est un automate à 8 états, g er  par 29 r gles et form  de 86 cellules : une part de ces cellule prend l' tat 2 (couleur bleu) forme une boucle et constitue un espace de circulation de l'information, cette boucle est finit par une tige o  ce d roule la duplication, une copie renvoy e vers la boucle et l'autre ce propage le long du bras, c'est le m canisme de transcription<sup>1</sup> dans l'ADN.

Au bout de ce bras les signaux se transforment en instruction, c'est l' quivalent du m canisme traduction<sup>2</sup> dans l'ADN. Le bras s' tend selon les instructions pour former des structures similaire   la structure initiale, cette structure se d tache par la suite avec une information initiale pour qu'elle puisse r pliquer. Apr s certain nombre de g n rations une r gle de st rilisation peut bloquer l' volution de l'automate cellulaire, cristallise les boucles et construit enfin une sorte de corail [Langton, 1986].

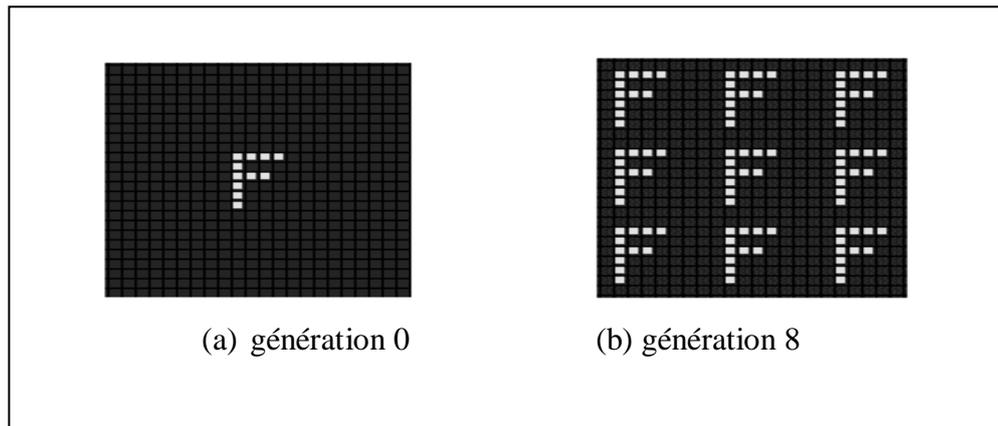
### 1.5.1.7. L'automate de Fredkin

Fredkin utilise un voisinage de Moore, c'est un type d'automate cellulaire sommatif ou l' tat de la cellule d pend du nombre des voisins actifs ind pendamment de leur position. La reproduction se cr e sauf si la valeur du voisinage est impaire [Kramm, 2005].

---

<sup>1</sup>La transcription: l'ADN du g ne   exprimer est copi  sous forme d'une mol cule d'ARN.

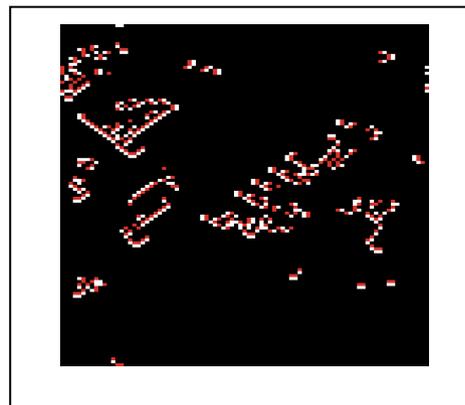
<sup>2</sup>La traduction:   partir des instructions pr sent es sur l'ARN la cellule construit une prot ine.



**Figure 1.10.** Automate cellulaire à deux dimensions (Fredkin)

### 1.5.1.8. L'automate de Brian's Brain

C'est un exemple d'automate cellulaire à deux dimensions. Typiquement ces automates cellulaires à deux dimensions utilisent deux états (vie/mort), mais Brian's Brain utilise trois états à savoir l'état vie, l'état fantôme, l'état mort.



**Figure 1.11.** Automate cellulaire de Brian's Brain<sup>1</sup>

### 1.5.1.9. L'automate stochastique

Pour engendrer une diversité au sein d'une configuration, on peut par exemple utiliser une fonction de probabilité pour construire des automates stochastiques (figure (1.12)).

---

<sup>1</sup> Simulation de l'automate cellulaire « Brian's Brain » sur la plateforme Netlogo

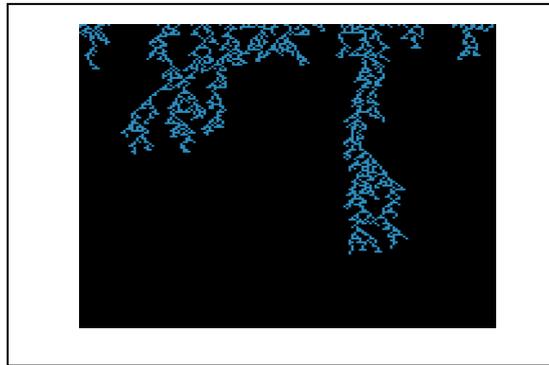


Figure 1.12. Automate Stochastique<sup>1</sup>

#### 1.5.1.10. L'automate cellulaire continu

La particularité de cet automate est les valeurs des cellules étendues allant de (0.0) vers (1.0). L'état (0.0) prend la couleur noir et l'état (1.0) prend la couleur verte.

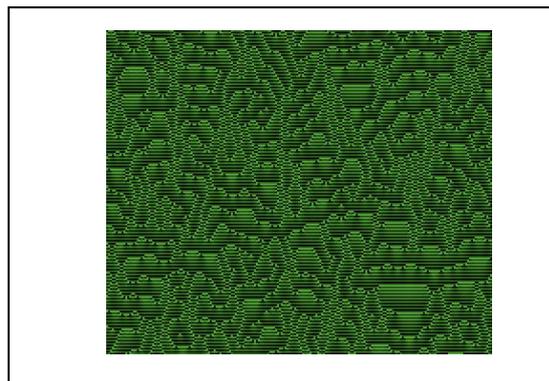


Figure 1.13. Automate continu<sup>2</sup>

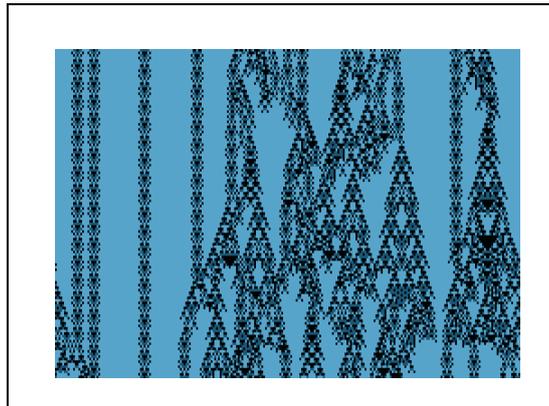
#### 1.5.1.11. L'automate cellulaire totalistique

Un autre type d'automate cellulaire, l'automate cellulaire totalistique. Dans cet automate l'état ultérieur d'une cellule ne dépend que de la somme des états de ses voisines

---

<sup>1</sup> Simulation de l'automate cellulaire « Stochastique» sur la plateforme Netlogo

<sup>2</sup> Simulation de l'automate cellulaire « Continu» sur la plateforme Netlogo

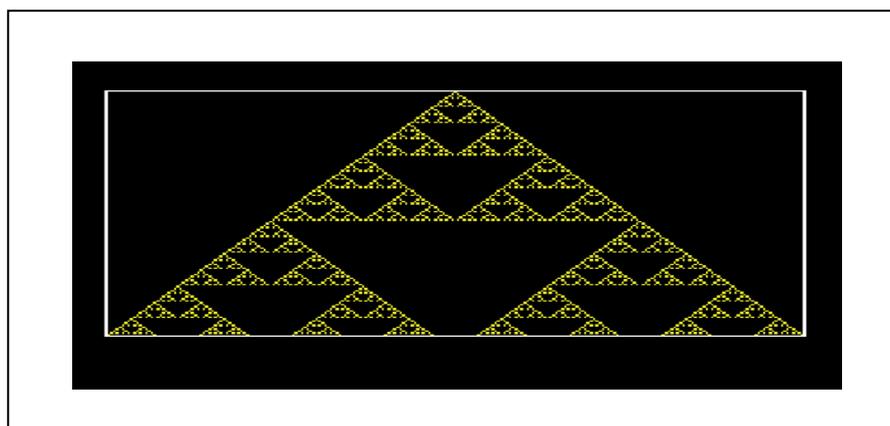


**Figure 1.14.** Automate Totalistique<sup>1</sup>

### 1.5.1.12. Automate cellulaire à une dimension

Pour ce type d'automate, on considère le cas simple d'un automate cellulaire, ou une cellule noire avec un voisinage à droite et à gauche de couleur jaune; donc l'automate à une dimension et deux états.

L'automate à une dimension va être une ligne, la deuxième dimension représente le temps. A chaque génération la nouvelle ligne est ajoutée au-dessous de la précédente, (figure (1.15)) présente ce type d'automate.



**Figure 1.15.** Automate cellulaire à une dimension (Triangle de Pascal)<sup>2</sup>

Finalement nous pouvons indiquer les variantes importantes pour une classification des automates cellulaires :

---

<sup>1</sup> Simulation de l'automate cellulaire « Totalistique » sur la plateforme Netlogo

<sup>2</sup> Simulation d'un automate cellulaire à une dimension « Triangle de Pascal » sur la plateforme Netlogo

**A.** Fonctionnement de l'automate cellulaire : Synchrones lorsque chaque cellule de l'automate cellulaire est mise à jour à chaque itération. Par opposition, asynchrone signifie que les cellules déclenchent leur mécanisme de transition les unes après les autres.

**B.** Dimension de l'automate : Grille utilisée peuvent être une grille à une, deux ou trois dimensions.

**C.** Structure du maillage. Dans les automates cellulaires bidimensionnels il s'agit de carré, mais on peut trouver d'autres formes comme l'hexagone

**D.** Mécanisme de transition : si le changement de l'état d'une cellule dépend de l'état du voisinage de cette cellule on dit que l'automate est déterministe, dans le cas du changement de l'état selon une fonction de probabilité on dit que notre automate stochastique.

## **1.6. Caractéristiques d'automates cellulaires**

La structure et la dynamique des automates cellulaires révèlent des propriétés fondamentales, quelques propriétés sont présentées ci-dessous.

### **1.6.1. Auto-organisation**

Est un phénomène permettant l'émergence d'une structure, d'un motif, d'un arrangement dans un système dynamique. Elle permet l'organisation d'un système sans qu'une contrainte extérieure ou globale au système impose cette organisation, ainsi l'évolution d'automate cellulaire conduit à des structures complexes.

### **1.6.2. Auto- reproduction**

L'auto- reproduction est à l'origine de la vie...le but de Von Neumann été de réaliser un mécanisme copiant la vie dans le sens où il pourrait se reproduire de lui même. Ainsi, cette propriété particulière permet pour certains automates cellulaires de produire des copies d'eux-mêmes. Ce point permet de classer les automates cellulaires en deux grandes catégories :

#### **A. Les automates cellulaires actifs**

Les automates cellulaires actifs sont des auto- reproducteurs dans le sens où ils contiennent une sous-configuration qui se comporte en *copieur universel* dirigeant activement la réplication par l'intermédiaire d'une fonction de transition [Abdellaoui, 2003].

#### **B. Les automates cellulaires passifs**

La reproduction des automates cellulaires passifs est provoqué par la règle de transition et non pas par les caractéristiques de la configuration initiale, nous pouvons citer *Le Jeu de la Vie* comme un type d'automate cellulaire passif.

### 1.6.3. Auto- similarité

On peut voir l'auto- similarité dans un automate cellulaire lorsque ses portions agrandies d'une manière similaire à celle de la totalité, les cellules sont soumises à une même topologie et changent leurs états selon la même règle locale [Abdellaoui, 2003].

### 1.6.4. Inversibilité

La propriété d'inversibilité permet à l'automate cellulaire de revenir aux états précédents. L'automate *déplacement Est* illustre cette propriété, son inverse est l'automate *déplacement Ouest*. Pour que l'automate cellulaire n'a pas d'inverse il suffit de trouver deux configurations différentes qui s'achèvent à la même configuration.

### 1.6.5. Indécidabilité

La fin du développement d'un automate cellulaire est indécidable. On ne peut pas déterminer si l'automate ne va pas s'éteindre au bout d'un certain nombre de générations ou s'il va se stabiliser [Abdellaoui, 2003].

### 1.6.6. Discrétisation

Les problèmes sociaux par exemple sont très difficile à modéliser, avec ces propriétés temps et espace ce type de problème trouve sa solution avec une modélisation.

### 1.6.7. La non linéarité

Les liens entre les phénomènes sont non linéaires, cependant modéliser avec les automates cellulaires le problème devient soit simple pour un traitement ultérieur.

## 1.7. Utilisation et applications des automates cellulaires

Wolfram dans son livre (*A New Kind of Science*) [Wolfram, 2002], a exploré la raison derrière l'utilisation des automates cellulaires dans un grand nombre de domaines d'application. Les automates cellulaires présentent un concept important pour développer des algorithmes pouvant être mis en œuvre sur des architectures parallèle. D'un autre côté les Automates Cellulaires constituent un objet mathématique fécond pour la modélisation de divers phénomènes.

- En sciences sociales, James M. Sakoda était la première personne qui a développé un modèle basé CA en sciences sociales, Sakoda a publié l'article (*The Checkerboard Model of Social Interaction*) en 1971[Sakoda, 1971]. Ainsi, la représentation d'un nombre important des individus dans l'espace ou leurs comportements est codé par une valeur et une interaction entre les individus proche permet l'utilisation des automates cellulaires. Thomas Schelling a analysé le processus de ségrégation entre les individus appartenant à deux classes différentes: noir et blanc [Schelling, 1971].

- AC ont été utilisés, par exemple, pour modéliser la cinétique des systèmes moléculaires et de la croissance cristalline en chimie [Packard, 1986].
  - Ainsi, le concept d'automates cellulaires a été utilisé pour modéliser les systèmes biologiques du niveau de l'activité intracellulaire jusqu'aux niveaux d'amas de cellules dans l'organisme [Nicholas and Paulien, 1997].
  - En physique, Maxwell avait déjà évoqué l'emploi de tels modèles discrets pour étudier les gaz et les écoulements de fluide [Maxwell, 1890]. Aussi, des applications couvrent l'étude des systèmes dynamiques à partir de l'interaction des particules à la classification des galaxies [Schonsch, 2002]. Ils sont également utilisés comme alternative aux équations différentielles de l'hydrodynamique, car leur implémentation est plus simple.
  - En économie, Axelrod a fait la première étape de l'analyse de la dynamique de coopération dans le cadre des AC [Axelrod, 1984]. Nowak et May ont développé l'idée et ont étudié la dynamique de la coopération au moyen d'un AC à deux dimensions avec des jeux pour deux personnes en tant que blocs de construction [Nowak and May, 1992]. Keenan et O'Brien a introduit un AC unidimensionnel pour modéliser et analyser les prix dans un contexte spatial [Keenan and Obrien, 1993, Maes et al., 1993, Föllmer, 1994].
  - En imagerie, les automates cellulaires unidimensionnels (1D) et a deux dimensions (2D), ont été appliqué avec un grand succès dans les systèmes d'information géographique, l'astronomie, l'imagerie médicale et dans l'amélioration des performances des systèmes qui les utilisent.
- L'AC a été utilisé dans la détection de contours, cette méthode a montrée ses résultats : une connectivité de pointe au niveau des jonctions, pas de faux bord et une similarité qualitative avec ceux produits par d'autres méthodes [Martin et al., 1984]. Douglas et Craig ont utilisé l'AC dans la détection des contours qui est indépendant des caractéristiques de l'image, les règles de transition sont très simples et fournissent un processus d'accélération de détection de bord parce que les règles sont déclenchées par des pixels noir seulement [Douglas and Craig, 1994].
- Dans l'amélioration des images, les AC sont utilisés pour améliorer les caractéristiques d'une image sans information primitive. Ils apparaissent comme un outil de traitement d'image en raison de leur caractère local et le simple parallélisme. La méthode étudiée par Hernandez et Herrmann présente une convergence très rapide à des points fixes, la stabilité de bruit, et des améliorations sur des images réelles [Hernandez and Herrmann, 1996]

-Dans la segmentation des images médicales et pour les retouches photos, cette idée a donné de bons résultats [Boykov and Jolly, 2001].

-L'AC est utilisé pour éliminer le bruit qui présente un problème dans le traitement d'image. Une image numérique est un tableau bi-dimensionnel de  $n \times n$  pixels. Chaque pixel est caractérisé par un triplet  $(i, j, k)$ , où  $(j, i)$  représente ses positions dans le tableau et,  $k$  présente la couleur associée. L'image est considérée comme un état d'une configuration particulière de l'automate cellulaire, défini par l'image. Une règle dynamique est utilisée, l'image bruitée présente une configuration initiale, le résultat, une image moins bruitée [Popovici and Popovici, 1990, Gonzales and Woods, 2002].

-Dans le même contexte, les caractères de texte incorporés dans les images représentent une source très riche d'informations pour l'indexation basée sur le contenu. Ces caractères de texte sont difficiles à détecter et reconnus en raison de leurs différentes tailles, un ensemble d'algorithmes séquentiels pour l'extraction de texte est appliqué en utilisant les automates cellulaires. L'algorithme proposé est simple et facile à utiliser, et ne nécessite qu'une image binaire (échantillon de texture) comme une entrée [Sahoo et al., 2009].

-L'automate cellulaire est utilisé aussi pour détecter des lésions dans des images récupérées par la technique d'échographie. La fonction de transfert d'énergie est formulée en se basant sur la différence entre les informations globales et locales de l'image en utilisant différentes stratégies de transfert d'énergie, deux systèmes de voisinage sont utilisés (systèmes de voisinage de Von Neumann et Moore) et intégrés comme l'environnement d'évolution, un critère de similitude est utilisé pour éliminer le bruit et de réduire la complexité de calcul. L'automate cellulaire a donné de bon résultat sur des images BUS (Breast Ultrasound) avec un faible contraste et des frontières floues [Yan et al., 2012].

Ding et al, a combiné le MIL (*Multiple Instance Learning*) et les automates cellulaires pour une segmentation des images d'échographie du sein, dans le but de détecter des tumeurs. Une carte auto-organisatrice est constituée, une distribution des instances de chaque sac pour construire le sac de vecteur caractéristique et enfin le SVM (*Support Vector Machine*) est utilisé pour classer les tumeurs, la méthode donne un bon résultat [Ding et al., 2012].

### 1.8. Conclusion

Dans ce chapitre nous avons survolé ce grand domaine « les automates cellulaires », qui sont des modèles mathématiques, ils constituent une méthode extrêmement utilisée pour résoudre différents problèmes. Nous avons détaillé les points clés de ce domaine, à savoir l'historique, les propriétés et les classes des automates cellulaires ; puis les caractéristiques générales, ainsi

que la notion d'émergence, sur laquelle les automates cellulaires fonctionnent. Nous avons indiqué qu'un simple changement dans la configuration initiale peut conduire à un changement radical dans la classe d'automate cellulaire. Une simulation de quelques automates est faite sur la plate forme Netlogo. Nous avons ensuite présenté quelques domaines qui utilisent l'AC pour résoudre certains problèmes, et qu'avec ces derniers les résultats sont assez satisfaisants. A notre connaissance, il n'existe pas un système de classification des images de mammographie utilisant les automates cellulaires, c'est pour cela que nous envisageons des les utilisés dans la classification non supervisée orientée vers le classement des images mammographie malignes et bénignes, d'une part. Et d'autre part, nous exposons une technique qui a donné de bons résultats dans la classification des données textuelles. Dans le chapitre suivant nous présentons les systèmes multi-agents, ce type de systèmes trouve sa force dans la simplicité de modélisation et la capacité de résoudre des problèmes très complexes.

## Chapitre II

---

# Les Systèmes Multi-Agents

---

### 2.1. Introduction

Dans ce chapitre nous nous intéressons au paradigme *Système Multi-Agent* (SMA), nous décrirons les concepts liés aux systèmes multi-agents dans la première partie. Elle portera sur des notions spécifiques tels que l'émergence et l'auto organisation, ces deux concepts sont primordiales pour la conception des systèmes, cette conception ne repose pas seulement sur les agents qui composent le système mais aussi sur l'environnement dans lequel les agents évoluent, interagissent, s'organisent pour atteindre le but attendu. Comme nous verrons aussi que les interactions et l'organisation des agents font la différence entre un système multi-agents et un ensemble d'agents indépendants.

Enfin nous concluons par une partie sur les agents d'inspiration éthologique, c'est-à-dire des agents avec un comportement animal, ce qui permet l'émergence du comportement des groupes sur lesquels nos travaux s'appuient.

### 2.2. Le système multi agent et le paradigme agent

Le système multi-agent est un paradigme qui est constitué d'un ensemble d'entités informatiques appelées agents. Ces agents se déplacent et s'interagissent entre eux dans un environnement commun avec une certaine dynamique au lieu d'évoluer parallèlement.

Dans notre travail nous montrons l'importance du système réactif et les définitions qui suivent sont orientées vers ce type de systèmes multi-agents.

Demazeau propose l'approche voyelles, elle décompose un SMA en quatre dimensions qui correspondent aux quatre voyelles (A, E, I et O) [Demazeau, 1995, Demazeau, 2001].

**A. Agent :** définition des architectures qui compose le système.

**B. Environnement** : c'est l'espace où les agents se déplacent, il se compose des objets perçus et manipulés par des agents.

**C. Interactions** : c'est le moyen d'échange entre les agents (ensemble langages et protocoles d'interaction).

**D. Organisation** : la structure d'organisation établit un cadre pour les relations et interactions entre les agents, hiérarchies, holarchies, etc.

Utilisateur, présente la cinquième dimension. C'est l'être humain qui a une observation et une perception du système de l'extérieur et il possède une influence sur lui.

On peut distinguer trois familles d'applications [Boissier et al., 2004] :

La simulation des systèmes complexes, par exemple dans les systèmes biologiques [Thomas et al., 2002] ou sociaux [Amblard, 2003, Gaud et al., 2008], il permet d'expérimenter des scénarios qui sont difficiles de les réaliser sur des populations réelles, de les paramétrer et même d'ajouter des contraintes pour arriver à un résultat.

La résolution des problèmes, où chaque agent est caractérisé par une anatomie qui le permet d'exercer un contrôle sur son comportement, et à partir de ce principe l'interaction va être entre des petits programmes (les agents) et un seul gros programme monolithique.

L'intégration de l'informatique par exemple avec les êtres humains ou les systèmes mécaniques, c'est-à-dire des agents autonomes qui peuvent se déplacer, prendre des décisions, des initiatives, accepter ou refuser des requêtes.

En conséquence, les caractéristiques d'un système multi agents se diffèrent en fonction de la famille du système étudié.

### 2.2.1. Définition système multi-agents

On peut considérer un système multi-agents comme un ensemble d'agents qui partagent un environnement commun, selon Ferber « *on appelle système multi-agent (SMA), un système composé des éléments suivants:*

- A. *Un environnement  $E$ , c'est-à-dire un espace disposant généralement d'une métrique.*
- B. *Un ensemble d'objets  $O$ . Ces objets sont situés, c'est à dire que, pour tout objet, il est possible, à un moment donné, d'associer une position dans  $E$ . Ces objets sont passifs, c'est à dire qu'ils peuvent être perçus, créés, détruits et modifiés par les agents.*
- C. *Un ensemble  $A$  d'agents, qui sont des objets particuliers ( $A \subset O$ ), lesquels représentent les entités actives du système.*
- D. *Un ensemble de relations  $R$  qui unissent des objets (et donc des agents) entre eux.*
- E. *Un ensemble d'opérations  $Op$  permettant aux agents de  $A$  de percevoir, produire, consommer, transformer et manipuler des objets de  $O$ .*

*F. Des opérateurs chargés de représenter l'application de ces opérations et la réaction du monde à cette tentative de modification, que l'on appellera les lois de l'univers. » [Ferber, 1995].*

### 2.2.2. Les caractéristiques d'un système multi-agents

Contrairement aux systèmes informatiques, il existe un ensemble des caractéristiques propre aux systèmes multi agents.

Selon Boissier, un système multi-agent possède les caractéristiques suivantes :

**A. La distribution:** le système est composé d'un ensemble de modules, l'élément de base est un agent.

**B. L'autonomie :** chaque agent est active, il prend ses décisions en fonction de ses objectives et des ses connaissances afin de trouver une solution.

**C. La décentralisation :** les agents dans cette organisation sont indépendants dans leurs décisions.

**D. L'échange de connaissances :** les agents échangent les informations entre eux selon des actes du langage.

**E. L'interaction :** les agents s'interagissent entre eux alors chaque agent dans cette structure à une influence locale sur un autre agent.

**F. L'organisation :** un réseau est crée a partir des interactions et des relations entre agents qui peuvent évoluer au cours du temps.

**G. La situation dans l'environnement :** l'évolution d'un système multi agent nécessite l'évolution d'une combinaison, les agents et l'environnement (lieu d'interactions, ressources, action et influence entre agents).

**H. Ouverture :** le système est ouvert à l'extérieur, les agents peuvent échanger des informations, entrée, sortie ou se modifiée au cours du temps.

**I. Emergence :** « *Dans tous les SMA, une fonction globale est attendue à partir d'un ensemble de spécifications au niveau local de chacune des entités. Cette propriété du niveau global n'est pas programmée dans les agents et n'existe que par leurs interactions conduisant à des processus permanents de réorganisation. »*

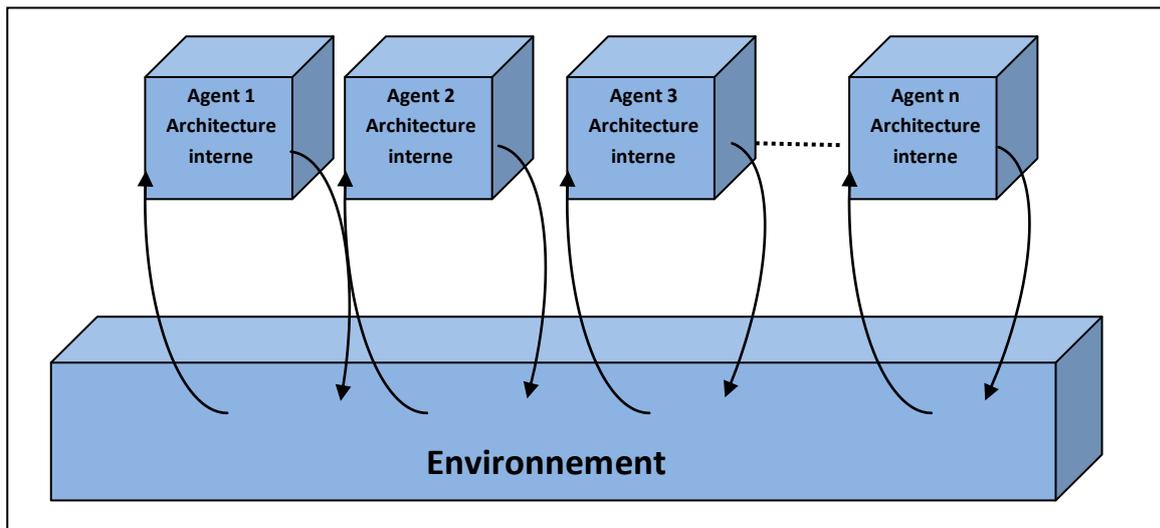
**J. L'adaptation :** il est un peut loin de dire que le système réalise une fonction globale adéquate au sein d'une organisation et de spécifier le but globale et organiser les agents pour l'atteindre.

**K. La délégation :** l'utilisateur délègue seulement une partie du contrôle de l'application globale aux agents.

**L. La personnalisation :** l'agent se présente comme utilisateur et il s'adapte typiquement dans les systèmes multi agents qui appartient à la famille des systèmes intégrés dans un contexte plus large.

**M. L'intelligibilité :** les systèmes multi agents proposent la modélisation des systèmes ou l'exécution des applications, se qui rend simple à appréhender à un utilisateur de l'extérieur [Boissier et al., 2004].

Schématiquement on peut représenter un système multi agents par la figure (2.1) suivante.



**Figure 2.1.** Représentation schématique d'un système multi-agents [Michel, 2004]

Ce schéma présente un système multi-agents et fait clairement apparaître le rôle important de l'environnement.

En effet, si la notion d'environnement joue un rôle fondamental dans la définition de ce qu'est un agent, elle prend une dimension supplémentaire dans le cas de plusieurs individus. A ce propos, nous détaillerons cette notion dans la section (cf. sec.2.2.5).

### 2.2.3. Ouverture et perturbation du système

Dans un système physique ou informatique la notion d'ouverture présente la possibilité d'échanger de l'information ou de la matière avec le monde extérieur dans un environnement possédant une dynamique à des évolutions inattendues [Sichman, 1995].

Dans les systèmes multi-agents, la notion d'ouverture présente la possibilité d'ajouter ou de retirer des fonctionnalités et des services aux agents [Vercouter, 2001, Vercouter, 2004].

D'après Wooldridge un système ouvert désigne l'impossibilité de connaître les composants d'un système et ses interactions lors d'une conception [Wooldridge and Ciancarini, 2001] et

que l'utilisateur joue un rôle dans la perturbation du fonctionnement d'un système multi-agents [Boissier et al., 2004].

### 2.2.4. Décentralisation

La dynamique dans un système multi-agent vient du comportement individuel des agents et de leurs interactions. Le système attribue une robustesse dans la résolution des problèmes, le changement du comportement d'un agent par exemple une panne n'a pas une influence sur le fonctionnement du système dans son ensemble.

Avec la décentralisation on s'efforce d'éliminer tout contrôle central. Les tâches à réaliser et les compétences pour le faire sont distribuées sur les agents. Cette décentralisation pose aussi quelques contraintes, elle peut être vue comme le fait qu'un intervenant extérieur du système est soumis à des contraintes de localité similaires à celle des agents donc il ne peut percevoir et agir que sur une partie du système. De la même manière, la synchronisation des actions (par exemple une action destinée pour influencer sur le système elle risque de ne pas s'appliquer à l'ensemble du système). On retrouve ce problème souvent dans les systèmes intégrés mais ne pose aucun problème dans les systèmes destinés vers la simulation ou la résolution de problème car il est possible d'observer et interagir à tout instant [Boissier et al., 2004].

### 2.2.5. Environnement

L'environnement définit les conditions d'existence d'un agent. L'agent ne peut pas exister sans système multi agent, la réaction est mutuelle entre les deux. D'un autre côté, il y a des travaux qui négligent l'importance de l'environnement car ils voient que les agents échangent les messages où il n'existe pas à proprement parler un environnement physique, mais les échanges entre agents nécessitent un médium de communication donc l'architecture interne des agents est liée à l'environnement [Odell et al., 2002].

En conséquence, l'environnement a souvent été considéré comme une unité centrale, cette entité active à sa propre dynamique, par exemple dans les travaux des fourmis, l'environnement modélise la dynamique des phéromones et la dynamique d'agents se fonde sur celle de l'environnement.

Muller propose plusieurs dimensions [Muller, 2010] :

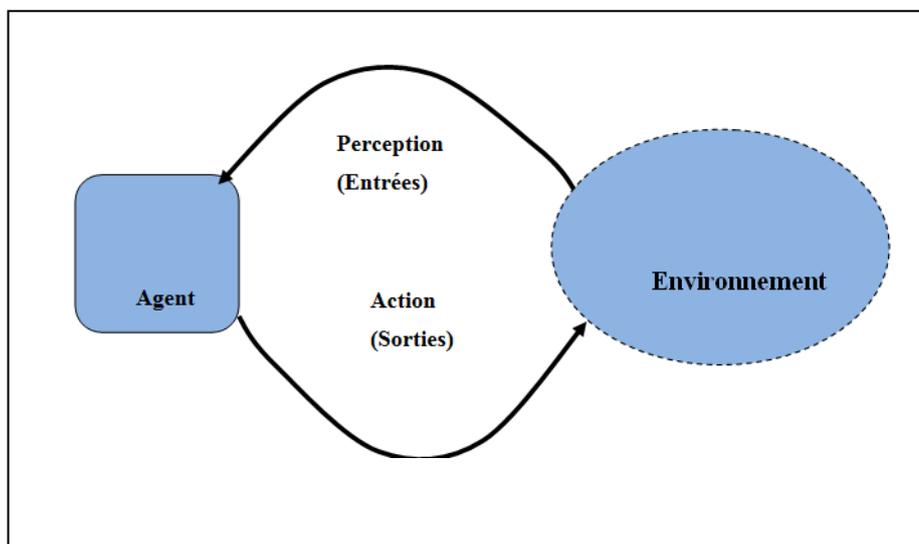
*« L'environnement comme médium d'interaction : dans ce cas les interactions entre agents sont premières (il faut donc les spécifier avant l'environnement) et l'environnement apparaît comme le choix de conception de ces interactions ce qui peut aller du canal de communication à l'objet » ;*

« L'environnement comme lieu d'interaction entre le local et le global : dans ce cas, plus que le medium de l'interaction, il doit être également une mémoire structurée et perceptible par les agents » ;

« L'environnement comme contrainte : dans la dynamique des agents, l'environnement limite alors les possibilités et, partant, dirige la dynamique locale soit pour rendre compte de ce qui n'est pas possible (modélisation), soit pour contraindre la recherche (résolution de problèmes). Dans cette perspective, les contraintes pourraient être codées par l'environnement. Faut-il distinguer contraintes internes (entre agents) et contraintes externes, locales et globales ? » ;

« L'environnement comme stabilisateur : si on veut que la dynamique du système multi-agents se stabilise dans son comportement ou sa structure, il peut être nécessaire d'introduire des frottements qui doivent être issus d'un environnement » ;

« L'environnement comme résultat de la distinction passif/actif : est environnement ce qui est passif contrairement aux agents qui sont actifs. C'est probablement davantage la distinction entre intentionnel et non intentionnel ».



**Figure 2.2.** Représentation d'un agent et de son environnement [Michel, 2004]

Nous présentons dans ce qui suit des définitions utilisées par la communauté des chercheurs en SMA, qui considèrent les agents comme des entités autonomes et actives qui composent un système multi-agents.

Maes [Maes, 1995] propose la définition du terme agent: «An agent is a computational system that inhabits a complex, dynamic environment. The agent can sense, and act on, its

*environment, and has a set of goals or motivations that it tries to achieve through these actions. »*

Erceau [Erceau, 1993] définit le concept comme : « *Les agents sont des entités physiques (capteurs, processeurs,...) ou abstraites (tâches à réaliser, déplacements,...), qui sont capables d'agir sur leur environnement et sur elles-mêmes, c'est-à-dire de modifier leur propre comportement. Elles disposent, pour ce faire, d'une représentation partielle de cet environnement et de moyens de perception et de communication. »*

L'une des définitions d'un agent qui fait consensus dans la communauté des SMA est donnée par Ferber [Ferber, 1995] : « *On appelle agent une entité physique ou virtuelle*

*a. qui est capable d'agir dans un environnement ;*

*b. qui peut communiquer directement avec d'autres agents ;*

*c. qui est mue par un ensemble de tendances (sous la forme d'objectifs individuels ou d'une fonction de satisfaction, voire de survie, qu'elle cherche à optimiser) ;*

*d. qui possède des ressources propres ;*

*e. qui est capable de percevoir (mais de manière limitée) son environnement ;*

*f. qui ne dispose que d'une représentation partielle de cet environnement (et éventuellement aucune) ;*

*g. qui possède des compétences et offre des services ;*

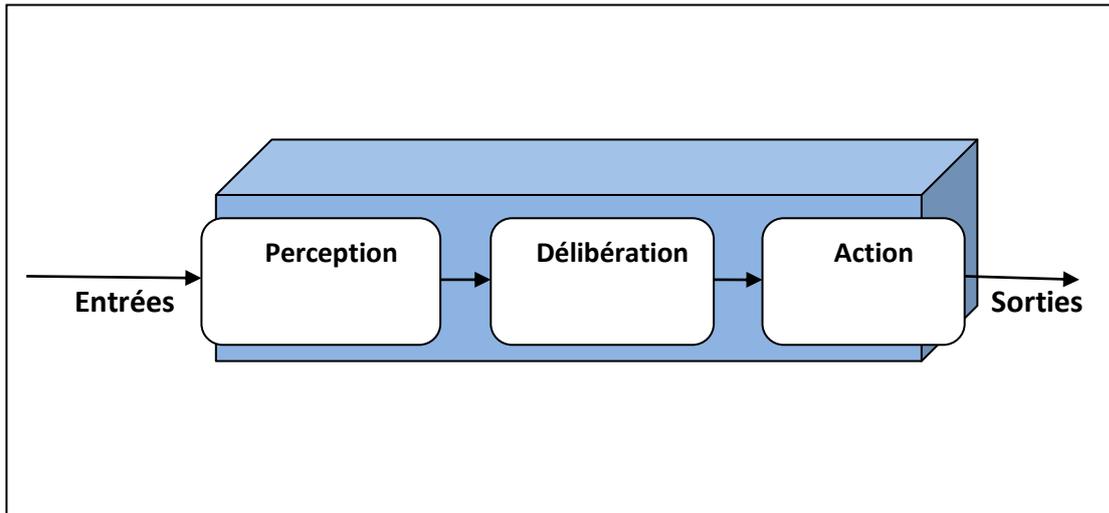
*h. qui peut éventuellement se reproduire ;*

*i. dont le comportement tend à satisfaire ses objectifs, en tenant compte des ressources et des compétences dont elle dispose, et en fonction de sa perception, de ses représentations et des communications qu'elle reçoit. ».*

### **2.3. Architecture interne d'un agent**

L'architecture interne d'un agent fait référence aux mécanismes responsables de sa dynamique. Dans le système multi agents on parle aussi du système décisionnel ou délibératif, un agent est considéré comme un processus cyclique s'il comporte les trois étapes successives : perception, délibération et action [Ferber, 1999, Michel, 2004]. Voir (figure (2.3)).

Selon [Zeigler and Ören,1986], un agent est vu comme un système dynamique particulier avec une structure interne variable, il peut donc ne pas réagir à la même façon au même stimuli.



**Figure 2.3.** L'agent, un processus cyclique à trois phases : perception, délibération et action  
[Michel, 2004]

D'une façon générale l'architecture interne d'un agent se caractérise par un processus décisionnel bien plus que par celles de perception et action. On distingue deux grandes architectures: l'architecture cognitive et l'architecture réactive.

### **2.3.1. Agent cognitive**

Ce type d'agents possèdent des butes et des plans pour leurs actions dans son environnement, ainsi qu'un ensemble d'informations nécessaire pour la réalisation d'une tâche bien précise, ce qui leur permet de gérer leurs interactions avec l'environnement et les autres agents. Dans cette architecture les agents peuvent être coordonnés et négociés entre eux pour résoudre les conflits [Ferber, 1995].

### **2.3.2. Agent réactive**

Contrairement aux agents cognitifs, l'agent réactive n'a pas un but explicite, ni un mécanisme de planification, il réagit seulement aux stimuli qu'il perçoit. Alors, dans ce cas le comportement complexe du système émerge de la coexistence et la coopération des comportements simples de chaque agent.

L'exemple de la fourmilière et termitières présentent ce système qualifiés et émergent : ils n'ont pas a priori de capacités cognitives mais elles peuvent accomplir des tâches complexes [Drogoul, 1993].

### **2.3.3. Agent hybride**

Un agent qui reprend les caractéristiques des agents réactifs et des agents cognitifs est dit agent hybride, il se comporte déferrement des deux autres. De même pour un grand nombre de problèmes, les solutions basées uniquement sur les agents cognitifs, ne convenaient point,

alors les chercheurs s'investissent dans la possibilité de combiner les deux types afin d'obtenir une architecture hybride [Ferguson, 1992, Müller 1996, Chaib-Draa and Levesque, 1996]. Nous évoquons juste son existence dans un souci d'exhaustivité mais nous n'en parlerons donc plus dans les prochains chapitres de ce document.

## **2.4. Le comportement**

Dans le contexte des agents, l'agent change son comportement en fonction des événements qu'il arrive. Lorsque un événement se produit l'agent analyse la situation avant d'actionner, l'action va modifier l'état du monde, n'est pas seulement envisagée comme le résultat de ce que font les agents mais comme le résultat des réactions du monde aux influences des agents.

Le comportement d'un agent peut être défini soit en extension, lorsque l'action qu'il effectue est donnée explicitement pour chaque situation qu'il peut rencontrer, soit en intention, s'il s'agit d'une fonction analytique qui fournit une action par un calcul sur l'état de l'agent et de son environnement [Ferber, 1995].

## **2.5. Interaction entre agents**

On appelle la relation dynamique qui existe entre deux ou plusieurs agents par le biais d'un ensemble d'actions réciproques une interaction. Cette interaction présente un élément indispensable à la constitution d'organisations sociales [Morin, 1977].

L'interaction entre les agents constitue à la fois la source de puissance et l'origine de problème car les agents peuvent gagner du temps par la coopération en même temps tomber dans le conflit qu'il doit résoudre avec leurs multitudes.

Un agent sans interaction avec d'autres agents n'est plus qu'un corps isolé dépourvu de caractéristiques adaptatives, cette multitude d'agents nécessite la coordination et doivent résoudre les conflits qui nécessitent un protocole de communication [Esmahi, 2001].

L'interaction permet alors le transfert de l'information entre agents ou agents et son environnement soit par perception ou par communication, cette notion suppose :

- A.** Un ensemble d'agents capables d'agir et/ou de communiquer ;
- B.** Des situations susceptibles de servir de point de rencontre entre agents : collaboration, utilisation de ressources limitées, régulation de la cohésion du groupe ;
- C.** Des éléments dynamiques permettant des relations locales et temporaires entre agents : communication, choc, champ attractif ou répulsif... ;
- D.** Une certaine flexibilité dans les relations entre agents leur permettant à la fois de créer, de maintenir ou de briser les relations, c'est-à-dire de disposer d'une certaine autonomie.

Les différentes formes d'interactions sont la collaboration et la coordination d'actions. La première consiste à répartir le travail entre plusieurs agents. La seconde s'intéresse à la manière dont les actions des agents sont organisées dans le temps et dans l'espace pour accomplir les buts. Si on prend l'exemple d'un agent réactif, il poursuit un objectif individuel au sein du système et ne focalise pas son énergie pour interagir avec les autres même s'il y est amené. Par contre, un agent qui participe à la satisfaction du but global du système tout en poursuivant un objectif individuel, va passer une partie de son temps à coopérer ou à se coordonner avec les autres agents. Pour cela, il doit posséder des connaissances sociales qui modélisent ses croyances sur les autres agents.

Pour constituer un système collectif cohérent il faut savoir coordonner et contrôler les échanges entre l'ensemble d'agents, donc l'interaction se constitue à niveau d'abstraction supérieur de la communication et l'action, alors le problème est de savoir comment combiner ces éléments (communication et interaction).

Ferber dans son livre donne une classification des situations d'interaction selon un observateur extérieur avec compatibilités ou incompatibilités qui expliquent les objectifs des agents, des ressources dont ils disposent suffisantes ou insuffisantes et de leurs compétences pour la résolution d'un problème [Ferber, 1995].

**Tableau 2.1.** Classification des situations d'interaction [Ferber, 1995]

<b>Buts</b>	<b>Ressources</b>	<b>Compétences</b>	<b>Situation</b>
Compatibles	Suffisantes	Suffisantes	Indépendance
Compatibles	Suffisantes	Insuffisantes	Collaboration simple
Compatibles	Insuffisantes	Suffisantes	Encombrement
Compatibles	Insuffisantes	Insuffisantes	Collaboration coordonnée
Incompatibles	Suffisantes	Suffisantes	Compétition individuelle pure
Incompatibles	Suffisantes	Insuffisantes	Compétition collectif pure
Incompatibles	Insuffisantes	Suffisantes	Conflit individuelle pour des ressources
Incompatibles	Insuffisantes	Insuffisantes	Conflit collectif pour des ressources

Ce tableau présente l'influence des trois critères objectifs, compétences et ressources et aboutit à la typologie des situations d'interaction qui sont : l'indépendance, collaboration simple, encombrement, collaboration coordonnée, compétition individuelle pure, compétition collective pure, conflit individuel et conflits collectifs.

Les agents sont dans une situation d'indifférence ou de coopération si leurs buts sont compatibles et dans une situation d'antagonisme dans le cas contraire.

### **2.5.1. Situation d'indifférence**

La situation d'inférence ou indépendance se résume à une juxtaposition des actions des agents. Ces actions prises indépendamment par les agents sans aucune interaction directe entre eux pour atteindre leurs buts qui sont indépendants avec des ressources suffisantes

[Ferber, 1995], mais selon [Franklin and Graesser, 1997] les agents peuvent interagir entre eux pour atteindre leur objectifs même les agents peuvent avoir les même buts mais chacun travaille indépendamment et différent des autres.

### **2.5.2. Situation de coopération**

D'après Ferber [Ferber , 1995] « *la coopération est caractéristique d'une attitude (posture) des agents. On dit que des agents coopèrent s'ils s'engagent dans une action commune après avoir identifié et adopté un but commun* ». La différence par rapport à la situation d'indifférence c'est que les objectifs des agents sont renforcés par une satisfaction des autres agents. Selon les compétences et les ressources on peut distinguer trois situations de collaboration.

#### **2.5.2.1. Collaboration simple**

Les compétences sont insuffisantes, cette collaboration nécessite l'addition des compétences pour réaliser des tâches difficile à réaliser individuellement. Les ressources sont suffisantes et ça ne nécessite pas de faire des actions supplémentaires [Ferber, 1995].

#### **2.5.2.2. Encombrement**

Les ressources sont insuffisante qui gênent mutuellement les agents pour réaliser leur tâches. Les compétences suffisante et les objectifs sont compatibles qui nécessite une coordination des actions pour battre l'encombrement [Ferber, 1995].

#### **2.5.2.3. Collaboration coordonnées**

C'est la situation la plus complexe des situations de collaboration, les ressources sont insuffisantes qui poussent les agents à coordonner leurs action. Les compétences sont insuffisantes qui nécessite une collaboration pour combler ce manque [Ferber, 1995].

### **2.5.3. Situations d'antagonisme**

Selon [Ferber , 1995] « [...] *On dira que des agents sont dans une situation de coopération ou d'indifférence si leurs buts sont compatibles et dans une situation d'antagonisme dans le cas contraire.*», de cela la satisfaction de l'un des buts des agents est affectée négativement par celle de l'autre, on distingue quatre situations :

#### **2.5.3.1. Competition individuelle pure**

Dans cette situation le conflit est absent, les ressources suffisantes (ne sont pas limitées et disponibles) et les agents luttent pour atteindre leurs buts individuels [Ferber, 1995].

#### **2.5.3.2. Conflit individuelle pour les ressources**

Dans cette situation les agents négocient pour atteindre leurs objectifs, les ressources insuffisantes et les compétences suffisantes pour lutter individuellement et acquérir ces ressources [Ferber, 1995].

### **2.5.3.3. Competition collective pure**

Dans ce cas les ressources sont suffisantes mais les compétences sont insuffisantes pour atteindre les objectifs qui nécessitent une collaboration simple entre les agents du même groupe, et une compétition entre les collections d'agents (groupes) [Ferber, 1995].

### **2.5.3.4. Conflits collectifs pour des ressources**

Dans cette situation les ressources sont limitées, les compétences sont insuffisantes qui oblige a création des groupes entrant en compétition pour atteindre les objectifs [Ferber, 1995].

## **2.6. Communication entre agents**

Comme chez les êtres humains, la communication dans les systèmes multi agent constitue la base de toute interaction et organisation entre les agents d'une part et entre les agents et l'environnement d'autre part. Donc la communication est un outil qui permet l'interaction (coopération ou antagoniste) [Ferber, 1995]. Cette communication peut prendre plusieurs formes : directe ou indirecte.

### **2.6.1. Communication indirecte**

Dans ce mode les agents se communiquent par des signaux via l'environnement, ce type consiste à laisser des signaux dans l'environnement qui peuvent être perçus par les autres agents, ce type est spécifique aux agents réactifs, donc seulement l'agent émetteur soit conscient de cette communication.

### **2.6.2. Communication directe**

La communication directe est réalisée volontairement en l'envoi des messages pour échanger des informations, de déléguer des tâches à un ou plusieurs individus, elle repose sur l'intention de communiquer des agents, cet échange est réalisé soit par un partage d'information ou par l'envoi de messages.

#### **2.6.2.1. Par partage d'information**

Dans ce type les agents se communiquent en utilisant la technique tableau noir<sup>1</sup> (*Blackboard*) [Penny, 1986]; [Russell and Norvig, 1995] Cette technique se base sur un espace commun où les agents peuvent y accéder pour partager et échanger des informations pour résoudre un problème. Les travaux de F. Hayes-Roth, L. Erman et V. Lesser avec Hearsay II ont été à la base de toutes les recherches ultérieures portant sur ce type de système [Erman et al., 1980].

---

<sup>1</sup> Le tableau noir : Ce type de communication n'existe pas dans les SMA où l'on dispose que d'une vision partielle de systèmes alors que la communication par partage d'information suppose l'existence d'une base partagée sur laquelle les agents viennent lire et écrire.

### 2.6.2.2. Par envoi de messages

Les agents se communiquent par un échange des messages entre eux, cette technique est inspirée de l'interaction sociale telle que la communication entre les êtres humains [Hewitt, 1977].

## 2.7. Actes de langage

La communication directe consiste en l'envoi des messages ou d'actes de communication. La théorie des actes de langage est initiée dans les années 60 par [Austin, 1962], il a montré qu'une communication ne sert pas juste à décrire le monde mais aussi à agir dessus et que toute énonciation permet la réalisation d'un acte qui sert à accomplir quelque tâches, par exemple provoquer un comportement ou changer l'état du destinataire.

Il définit ainsi trois actes où à chaque fois qu'un locuteur produit un énoncé, il fait du même un acte illocutoire : c'est de dire quelque chose (le fait d'énoncer le message), la force illocutoire qui permet de faire la différence entre deux phrases de même contenu et de sens différent comme par exemple dans les deux phrases (elle neige) et (Est ce qu'elle neige ?).

Dans les ACL (*Agent communication Language*), un acte est modélisé par un acte illocutoire et il se forme d'une force illocutoire et d'un contenu.

Il existe plusieurs langages de communication entre agents qui se diffèrent par le type de forces illocutoires qui présente la manière de décrire exhaustivement des actes de communication d'un point de vue syntaxique et sémantique.

Deux langages de communication FIPA-ACL (*FIPA Agent Communication Language*), et KQML (*Knowledge Query and Manipulation*) sont proposés dans le cadre d'un travail de standardisation et portés par des besoins d'interopérabilité logicielle [Finin et al., 1997].

KQML est développé pour l'échange d'information et des connaissances dans les systèmes à base de connaissance, par la suite il est utilisé pour décrire les messages envoyés entre agents.

FIPA est le plus usité dans les SMA, il est fondé sur 21 actes communicatifs exprimé par des performatifs<sup>1</sup>, un message est constitué des champs suivants :

- A.** Adresse de l'émetteur ;
- B.** Adresse du destinataire ;
- C.** Le performatif (acte du message). C'est lui qui représente la force illocutoire du message ;
- D.** Contenu du message ;
- E.** Le langage et l'ontologie utilisés pour décrire le contenu ;

---

<sup>1</sup> Passage d'information : Inform, Inform-if, Inform-ref, Confirm, Disconfirm, Réquisition d'information : Query-if, Query-ref, Subscribe, Négociation : Accept-proposal, Cfp, Propose, Reject-proposal, Distribution de tâches (ou exécution d'une action) : Request, Request-when, Requestwhenever, Agree, Cancel, Refuse, Manipulation des erreurs : Failure, Not-understood.

F. L'identifiant du message ;

G. L'identifiant de la conversation.

## 2.8. L'auto-organisation dans les SMA

L'auto-organisation dans un système multi-agents c'est la capacité à modifier son organisation pour qu'il s'adapte à l'environnement où il évolue. Selon Ferber les agents dans les systèmes multi-agents classique où l'organisation est statique et prédéfinie définissent eux même leur rôle et leurs relations avec les autres agents. Cette auto-organisation rend le système flexible et plus apte aux changements de l'environnement [Ferber, 1995].

Ses propriétés selon [Di Marzo et al, 2005].

A. Robustesse : la défaillance d'un composant du système ne remet pas en cause l'auto-organisation ;

B. Adaptabilité : le système s'adapte aux changements de l'environnement ;

C. Dynamisme : le système n'est pas stable, mais il s'organise continuellement pour garder un comportement stable devant la dynamique de l'environnement.

## 2.9. Emergence

Contrairement à l'approche classique qui décompose la tâche globale en sous tâche pour résoudre le problème l'approche émergentiste ne fait que corder l'interaction entre les agents et l'environnement et lors de l'exécution que le processus construit un cheminement vers l'éventuelle solution.

Le terme émergence est un terme employé à l'origine de puis longtemps dans l'histoire des sciences. Par les philosophes, les premières idées énoncées dans la maxime « Le tout est plus que la somme des parties » aux grecs présocratiques Anaximandre et Thalès. Au milieu du XIX ème siècle Lewes<sup>1</sup> cité dans [Ali et al., 1998] explique bien la différence entre deux types de phénomènes : les émergents et les résultants. Lewes définit que le phénomène résultant est se réalise dans une séquence de pas (le phénomène est traçable) contrairement du phénomène émergent, avec cette première tentative épistémologique Lewes a présenté le rôle important d'un observateur dans un processus émergent.

Dans les années 20, les chercheurs proposent de nouvelles approches qui traitent l'émergence. Morgan dans son livre « *Emergent Evolution* » a introduit de la nouveauté sur le concept pour

---

<sup>1</sup> George Henry Lewes (18 Avril 1817-1830 Novembre 1878) était un philosophe et critique anglais de la littérature et du théâtre. Il est devenu une partie du ferment de la mi-victorienne des idées qui encouragent la discussion du darwinisme, le positivisme et le scepticisme religieux.

décrire un phénomène émergent [Morgan, 1923, Smuts, 1926] à ajouté au concept la notion des structures stables.

Dans son livre, Alexander a décrit ce phénomène comme l'état où les choses s'ordonnent dans des nouvelles formes qui sont elles même possèdent des nouvelles structures et propriétés. Elle devenu par la suite une théorie informationnelle et computationnelle du concept d'émergence [Alexander, 2004], cette vision éloigne les théories non matérialiste qui décrit le phénomène émergent comme « Une création spirituelle émanant d'un monde inconnu et mystérieux » cité par [Kenyon, 1941]

Malgré les tentatives d'Alexander ce concept reste flou a définir et discuter par les physiciens, les biologistes les informaticiens pour désigner le fait qu'une chose sorte d'une autre sans que celle-ci la produise, ensuite ce terme est repris de nouveau dans l'IAD (Intelligence Artificielle Distribuée) particulièrement dans les SMA. On peut dire c'est l'interaction et l'activité collective massive entre agents au niveau globale du système ne présente que le résultat des interactions simples aux niveaux d'agents constituant le système.

### **2.9.1. L'émergence au sein des SMA**

Dans son chapitre Drogoul définit le phénomène émergence et son nécessité dans un système multi-agent pour obtenir un phénomène globale émergent, il a définit comme production un processus, état stable ou invariant observé par un utilisateur extérieur ou par des entités constituant le système [Drogoul et al., 2004].

Si on prend un système multi agent réactif, l'observateur est généralement hors du système car les agents eux même n'ont pas une capacité pour identifier un comportement globale, contrairement à un agent cognitif qui peut identifier les phénomènes globaux et modifier son dynamique pour les arriver.

Avec une identification du niveau local, un système multi-agents réalise ses objectifs globaux, l'évolution dans le temps dépend de l'initialisation d'agent dans un état particulier, les données récupérées de l'environnement, l'ouverture du système. L'émergence est réalisée si le système multi-agents atteint un objectif ou un état non traçable dès le premier pour chaque évolution.

Ce phénomène reste imprévisible, car l'évolution du système provient du hasard qui caractérise le comportement des entités du système. Cette caractéristique inhérente par les agents est provienne de l'ouverture du système ou de l'influence de l'utilisateur. Cet utilisateur qui a un double rôle ou saine du système, agir sur le système et le dirigé vers l'émergence et au même temps identifier les résultats émergentes qui a un sens intéressant et qui ramène des solutions recherchés pour lui.

Pour reprendre à la question liée à l'émergence et le système multi agent, créer, organiser et faire l'interaction au niveau locale. Pour obtenir un niveau globale qui reprend aux attendu, la démarche consiste en une élaboration préalable<sup>1</sup> du système grâce à des considérations heuristiques suivi d'une phase expérimental et une succession des testes qui permet d'arriver à une configuration voulue. Donc il n'y a pas vraiment une vérification formelle des projets appart quelques protocoles expérimentaux qui couvrent largement les repenses et qui rend la spécification formelle du comportement émergent difficile.

### **2.9.2. Caractéristiques d'un phénomène émergent**

Une bibliographie sur le phénomène d'émergence est donnée par [Deguet, 2008], particulièrement sur les SMA, les caractéristiques et l'importance des interactions par rapport à un comportement individuel qui mène à l'apparition de tel phénomène.

- A.** Sens commun : l'apparition soudaine d'une chose cachée ou inexistante ;
- B.** Causalité descendante : le phénomène émergent est perçu indépendamment des entités qui composent le système même s'il survie d'elles ;
- C.** Observation et prédiction : L'émergence décrit l'état actuel d'un système et son évolution dans le future mais on ne peut pas déduire le niveau global à partir du niveau local. La simulation présente un moyen important pour percevoir l'évolution d'un système ;
- D.** Complexité à partir d'une simplicité : c'est la description simple d'un comportement complexe ;
- E.** Interprétation: l'émergence c'est le résultat d'une interaction entre les agents et l'environnement. L'observation et l'interprétation peuvent être faites par une personne de l'extérieur du système ou par les agents eux même ;
- F.** Emergence et auto-organisation : ces deux termes en les trouve souvent ensemble malgré l'indépendance la combinaison des deux fait leur force [De Wolf and Holvoet, 2005], pour [Shalizi, 2001] le processus auto-organisation<sup>2</sup> engendre une complexité au système qui incite à chercher des régularités émergentes pour la résoudre ;
- G.** Le tout est supérieur à la somme des parties: le gain apporté au système est par des interactions entre ses entités locales ;
- H.** Point de vue pragmatique: Ce phénomène est surprenant pour un utilisateur extérieur même s'il connaît les règles locales du comportement des entités qu'il compose.

---

<sup>1</sup> La phase d'élaboration préalable repose sur des principes d'ingénierie qui cherchent à reproduire des phénomènes naturels connus, en particulier le bio-mimétisme et le socio-mimétisme.

<sup>2</sup> Le processus d'auto-organisation présente une augmentation de quantité d'information (qui implique donc une complexification du point de vue de la théorie de la complexité) en l'absence d'action programmatrice de l'extérieur.

### 2.9.3. Emergence dans les phénomènes naturels

On observe un phénomène d'émergence dans le comportement des insectes sociaux, par exemple les éthologues ont rapporté que des insectes très simples parviennent à regrouper des objets comme des brindilles sur un ou plusieurs endroits d'un environnement. C'est un exemple de communication indirecte entre agents.

### 2.10. Agents d'inspiration éthologique

Nous tirons d'autres exemples d'auto-organisation dans la nature c'est l'exemple des fourmis qui cherchent un chemin minimum pour rapporter la nourriture vers leur nid avec une communication indirecte entre eux sans utiliser le sens de la vue [Deneubourg et al., 1990].

Elle s'adapte aussi au changement de l'environnement dans l'exemple d'un obstacle où le chemin est inaccessible. la figure suivante montre que les fourmis se déplacent linéairement dans un chemin court entre le nid et la nourriture.

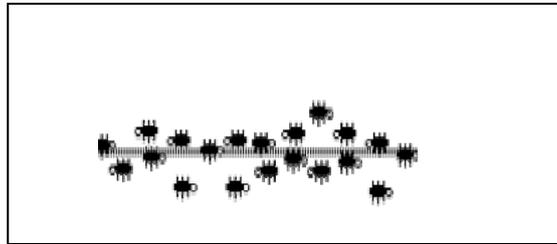


Figure 2.4. Déplacement linéaire des fourmis

Dans leur déplacement les fourmis déposent une trace chimique (des phéromones), elles préfèrent suivre un chemin riche en phéromones plutôt qu'un autre dont la concentration est faible.

Cette règle explique bien comment elles peuvent trouver le chemin le plus court pour arriver à leur nid. Dans le cas d'un obstacle inattendu la moitié des fourmis passent par la gauche et l'autre moitié passe par la droite dans le but de choisir un chemin plus court.

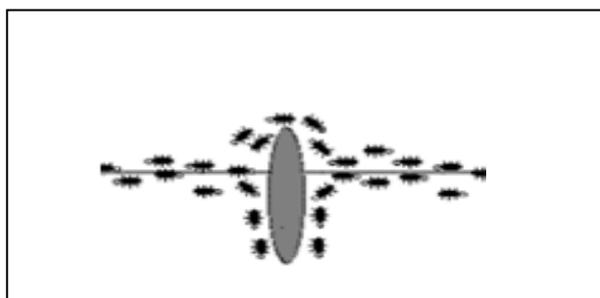
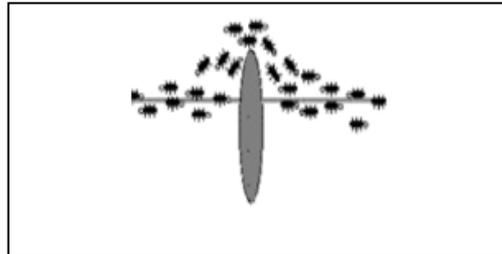


Figure 2.5. Déplacement à la recherche du chemin le plus court

Les fourmis qui par chance ont choisi le chemin le plus court, alors la concentration en phéromones est élevée sur ce chemin par rapport au chemin plus long.

Plus la concentration va augmenter, plus d'autres fourmis vont choisir le chemin riche en phéromones.



**Figure 2.6.** Les fourmis trouvent le chemin le plus court

Trouver le chemin le plus court est le résultat de l'émergence de l'interaction des agents autonomes. Le déplacement des fourmis est de la même vitesse qui implique une concentration de la même quantité de phéromone sur le passage. Le phéromone a une tendance à s'accumuler plus rapidement sur le chemin court que le chemin long et comme les fourmis ont une tendance de d'ériger vers les chemins où la concentration des phéromones est plus élevé il en résulte ce processus auto-catalytique [Deneubourg et al., 1992].

### **2.10.1. Algorithme d'optimisation**

Des travaux sur les colonies d'insectes ont toujours inspirés les informaticiens qui ont d'abord adapté les modèles afin de créer des algorithmes pour la résolution des problèmes. La découverte du chemin le plus court a été formalisé et exploité sous le nom d'une collection d'algorithme qui s'appelle ACO (*Ant Colony Optimisation*).

Une colonie de fourmis est représentée par un ensemble d'agents, ces agents se déplacent dans un espace plus précisément selon un graphe qui représente les coûts de chaque étape du problème. Son objectif est de trouver une solution optimale en terme du plus bas coût si ils utilisent l'approche gloutonne ou par le suivie du chemin le plus court marqué par les phéromones [Dorigo et al., 1996, Botte and Bonabeau, 1998].

## **2.11. Utilisation et application des SMA**

Les agents sont utilisés dans une grande variété d'applications allant de la comparaison des petits systèmes tels que des filtres de messagerie personnalisés aux grands, systèmes critiques complexes telles que le contrôle du trafic aérien, au cours des dernières années, le système

multi-agents est devenu de plus en plus une approche pour résoudre les limitations dans le domaine médicale.

Dans cette section, nous présentons quelques applications importantes dans le domaine médical qui ont été développés en utilisant le système multi-agents.

- **Dans le diagnostic**

Ils ont proposé un modèle de l'exploration de données de diagnostic pratique vise à soutenir le diagnostic médical réel par deux nouvelles technologies - exploration de données et système multi-agents [Klusch et al., 2003, Foster et al., 2005, Zhang et al., 2005]

Kazar et al introduisent le système multi-agents dans les domaines de la médecine, ce qui aide l'administration à prendre des décisions et des réactions, et assure également la communication et la coordination en réduisant les erreurs de diagnostic et de traitement, et en améliorant le temps nécessaire pour rechercher les ressources médicales, et d'autres services médicaux [Kazar et al].

Qiao Yang et Jhon proposent un modèle de diagnostic MADHS (*Multi-Agent Diagnosis Helping System*), qui se compose de quatre types d'agents: Coordonnateur, examinateurs, spécialistes et le constructeur de la décision commune [Qiao and Shieh, 2008].

Dans le même aspect, ils ont proposé un algorithme opérationnel pour décrire les opérations d'un système hybride multi agent à base intelligente du système de diagnostic médical appelé CDS (*Système de Diagnostic Clinique*) [Gupta and Pujari, 2009].

Gupta et al ont orienté le système multi-agents vers le diagnostic médical en ligne (sur le net), l'utilisateur peut avoir accès aux détails du système comme l'hôpital, clinique médical...etc. Le système fonctionne bien, mais les défis concernent principalement : la communication entre les services, la sécurité des données, la transparence des environnements de différents agents, la communication entre l'agent et l'utilisateur, la synchronisation entre les services distribués. L'objectif final du projet est de construire un système qui peut être adapté à n'importe quel environnement [Gupta et al., 2012].

- **Pour la simulation des phénomènes (protéines et des gènes)**

Un autre aspect de l'utilisation du système multi-agents dans le domaine médical était de construire la simulation des phénomènes d'interactions des protéines et des gènes, qui permet de réduire le nombre des patients exigent des traitements. Ce système principalement développé pour aider les médecins dans leur processus de diagnostic. Plusieurs perspectives sont possibles car le système est plus souple pour l'ajout ou la suppression des agents [Hayes-Roth, 1985].

### • Le génie logiciel et la technologie (e-Health)

Ces derniers temps, nous trouvons la technologie (*E-Health*) [Miller, 2007], l'interaction et la communication nécessitent des agents comme base du système, ses propriétés : nécessite quelques domaines du système multi-agents, augmenter la vitesse de parallélisme, la fiabilité, l'extensibilité et une programmation facile. Cette technologie été un moyen pour développer le télé-soins médicaux [Marlene et al., 2001], en fin dans ce travail l'auteur a tenté d'étudier le génie logiciel orienté agent et E-Health, en utilisant le système à base d'agents (propriétés et architecture) pour l'hôpital. Bien que des améliorations restent nécessaires dans cette technologie : services supplémentaires et la difficulté de communication.

Il ya des travaux qui traite un système multi-agents comme un système de soins de santé, qui est composé de modules de capteurs médicaux combiné avec la technologie de télécommunication sans fil. Le système comporte divers services comme la télémédecine, la surveillance de patient, la gestion et la communication entre les patients et le personnel médical [Nikolay et al., 2003, Byung-Mo et al., 2006].

Le système multi- agents est orienté aussi vers le génie logiciel, l'objectif principal était de construire un processus, des outils et des services pour le traitement et la maintenance du logiciel à base d'agents [Kalmarzi et al., 2013].

Aussi, dans le domaine des soins de santé le système multi-agent présente la technologie adaptée pour fournir des solutions aux problèmes déferents problèmes. Comme par exemple le projet ASainLog qui se base sur la coopération, la communication et la coordination. L'agent a été adapté pour le logiciel basé sur la réalisation de zones d'activité active, et contribue à la solution des problèmes d'informations logistique. Agents d'effectue des tâches d'interprétation et la concaténation des données, l'organisation de contrôle d'accès, contrôle des processus médicaux complexes, des rendez-vous de suivi, associant les types de maladies similaires, et de recommander des traitements. Les agents logiciels agissent au nom de leurs représentants dans le monde réel [Schweiger and Krcmar, 2004].

### • Data warehouse et le domaine médical

Dans data warehouse, les agents dans le domaine médicale sont utilisés dans une agence pour l'intégration, la diffusion et l'archivage de l'information médicale, se sont des agents pro-actives qui jouent le rôle des travailleurs électroniques intelligents. Cet agent pro- active interagit avec les différents systèmes de transfert de l'information médicale, ainsi classer et enregistrer les informations pour accéder aux ressources nécessaires et répondre avec les informations demandés dans un temps acceptable, comme exemple on peut citer : MIS (*Medical Information Systems*), HIS (*Healthcare Information System*), AIS (*Administrative*

*Information System* ), MSIS (*Medical Support Information System*), NIS (*Nursing Support Information System*), EMR (*Electronic Medical Record Information System*)... etc [Kimball, 1996, Chaudhuri and Dayal, 1997, Abelha et al., 2004]

- **Système expert et SMA**

Un autre travail utilise une architecture hybride d'un système de consultation pour l'obésité axée sur des problèmes de santé. Cette architecture hybride composé d'un système expert (ES) et un agent intelligent (IA). L'auteur a proposé qu'ils continuent à produire et collecter des connaissances biomédicales efficacement en utilisant le système multi-agents dans le domaine médical. L'idée principale derrière cette recherche était d'étudier la capacité de multi-agents hybride avec système expert qui capable de diagnostiquer les symptômes et fournir des consultations liées à des problèmes d'obésité [Mateo et al., 2007].

- **La télémédecine et la modélisation d'un système de classification du cancer**

Un autre travail qui utilise NBMS (*la machine d'apprentissage Naive Bayesian* ) qui se base sur une architecture de système multi-agents pour la télémédecine est proposé par [Croteau and Vieru., 2002, Garcia et al., 2003] et la simulation un système modèle pour la classification du cancer [Wooldridge, 2002].

- **SMA pour la surveillance des protocoles**

Une architecture de système multi-agents pour la surveillance des protocoles. Elle décrit la construction d'un système multi-agent qui permet de soutenir et de réglementer la mise en œuvre des soins de santé ou la pratique médicale en milieu hospitalier. Un protocole médical effectue une série de cas en face de systèmes de diagnostic médical [Alsinet et al., 2000].

- **Dans la segmentation des images médicales**

Dans la segmentation des images, l'idée est basée sur un système multi-agents pour segmenter une image IRM du cerveau, il fonctionne sans interaction de l'utilisateur dans la segmentation de l'image. L'avantage le plus important de cette méthode a été segmente l'image en plus de deux régions d'une manière parallèle [Liu et al., 1997].

Melouah propose une méthode de segmentation qui se base sur la coopération des agents, l'image est composée d'un ensemble de structures, pour réaliser sa segmentation, l'auteur fait intervenir un ensemble de techniques, de sorte que pour une zone donnée de l'image la technique offrant de meilleur résultat selon un critère donné [Melouah, 2010].

Mahsa et WooChaw ont été utilisé un cadre pour extraire simultanément plusieurs objets d'intérêt de la tomographie des images en utilisant une priori des connaissances. Leur méthode utilise les propriétés de l'agent dans un environnement multi-agents, les méthodes de

segmentation d'images médicales ont des limites parce que les images médicales sont d'un niveau de gris très similaire [Mahsa and WooChaw, 2013].

### • Le suivi des patients a distance

Les agents utilisent souvent les services d'autres agents pour effectuer une tâche souhaitée. Pour effectuer cela, les agents ont d'abord localisé l'agent spécifique de prestataire de services et communiquer avec lui [Terry et al., 2002]. Dans un autre travail, l'agent est utilisé pour la récupération de l'ontologie sur la base de la médecine à travers la base de données de la littérature Bio (MEDLINE), par exemple La tâche de système d'agent d'interface est d'aider les patients pour examiner de deux ontologie médicale existante et l'agent à générer des requêtes afin d'obtenir des informations utiles et adaptés [Jung-Jin, 2003]. Dans le même contexte et avec l'utilisation d'ontologie, appelée GHDO (*Generic Human Disease Ontology*) et une architecture de système multi-agent pour le traitement des requêtes, la récupération de données et l'incorporation des informations [Hadzic and Chang, 2005].

Pour attribution de données médicales entre les hôpitaux [Martinez-Garcia et al., 2009]. Dans la gestion d'urgence médicale mobile dans une zone urbaine. Il définit une architecture abstraite, ainsi, une plate-forme logicielle où les agents interagissent selon des normes et des règles mises par la structure organisationnelle. Leur système était basé sur des données réelles fournies par le service médical d'urgence de la région de Madrid [Centeno et al., 2009].

Dans le suivi des patients à distance, le patient à domicile ou par la route peut être traité par le personnel médical en utilisant la technologie mobile, les médecins peuvent être avertis de l'état du patient en même temps [Fiaz et al., 2011]. Dans la construction d'un service d'information pour les utilisateurs mobiles, les systèmes multi-agents ont été adaptés pour de nombreux domaines médicaux, comme la programmation d'une rencontre entre les patients et la réunion des médecins et pour les services de planification et d'urgence [Srividya et al., 2011].

Les SMA sont proposés pour le service médical sophistiqué dans les zones rurales. Le système multi-agents aide à la décision où le nombre d'agents logiciels prend une décision d'un problème donné collectivement [Dutta et al., 2012]. Dans la découverte de connaissances à partir de bases de données KDD (*Knowledge Discovery from Databases*) était un processus complexe (itératif et interactif) l'idée c'est d'inclure tous les modules à un agent [Benomrane et al., 2013].

## 2.12. Conclusion

Dans ce chapitre nous avons survolé le domaine système multi-agents, rappelé son robustesse dans la résolution des problèmes.

Nous avons vu qu'un agent est une entité autonome qu'est capable de percevoir, communiquer, actionner et prend des décisions. Un des aspects importants définissant le concept d'agent repose notamment sur son autonomie décisionnelle. Cette autonomie dépend du type d'agent (Agent cognitif, réactif et hybride) qui est détaillé dans la section (cf.sec. 2.2.5).

Elle dépend aussi du modèle interne sur lequel repose le fonctionnement des agents, c'est-à-dire du raisonnement et du comportement pour réaliser ses buts, interagir et communiquer avec les autres agents, la communication ne peut être réalisée sans parler de l'environnement avec laquelle l'agent susdit interagit.

Nous nous sommes intéressés des propriétés de souplesse et de robustesse qui sont assurées par un équilibre dynamique entre la spécialisation et la redondance des agents, on parle dans cette partie du phénomène d'émergence et d'auto-organisation d'un système.

D'un autre côté le système n'est pas optimum au niveau de son fonctionnement car la présence d'agents proposant les mêmes services est consommatrice de ressources et donc de temps pour l'exécution d'une opération. On parle alors des techniques qui résolvent les problèmes de conflits entre agents.

Nous avons détaillé le phénomène d'émergence dans le domaine naturel que l'étude des insectes influence et donne des idées sur le comportement et la conception des systèmes multi-agents, ça conduit à l'utilisation des algorithmes d'optimisations qui constitue un axe important dans notre thèse.

En fin, nous avons montré comment le système multi-agents a été intégré dans le domaine médical pour résoudre divers problèmes. Pourtant, il ya des avantages et des inconvénients de l'utilisation MAS dans le domaine médical, le système multi-agents est un domaine émergent dans le domaine médical.

## Chapitre III

---

# Indexation et Recherche d'images

---

### 3.1. Introduction

L'objectif de ce chapitre est de faire un horizon des principaux concepts de base relatifs à la recherche et l'indexation des images basée sur le contenu.

Faire de la recherche d'images par le contenu CBIR (*Content Based Image Retrieval*) revient à extraire les caractéristiques des images, les comparés afin de trouver une similarité entre ces images; la similarité est utilisée pour l'indexation (rangement et classement des images) pour la recherche des images en fonction d'une requête donné.

Dans ce chapitre nous introduisons les différentes méthodes d'indexation, nous présentons l'architecture mise en œuvre dans la recherche d'images par le contenu et nous détaillons les différentes parties qui le composent.

Comme, nous décrirons quelques systèmes de recherche d'images générales et médicales par le contenu, puis nous présenterons un état de l'art relative aux classifications supervisée et non-supervisée des images. Puis nous consacrerons une section à la présentation des méthodes d'évaluations d'un système CBIR.

### 3.2. Méthodes d'indexation

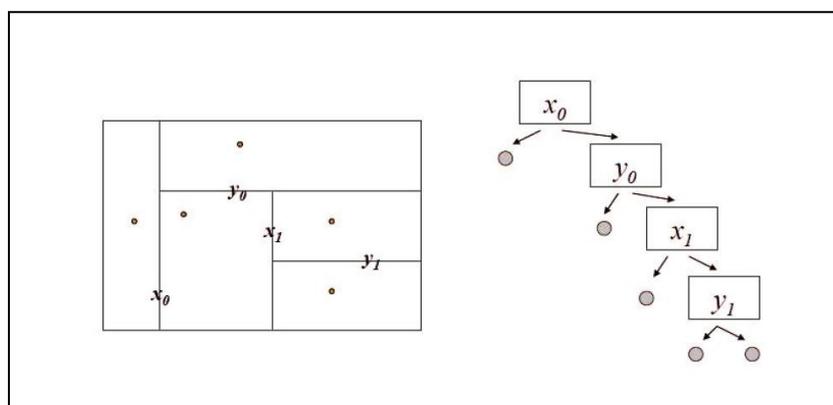
L'organisation et la structuration d'une collection d'images présente un grand défi pour les chercheurs de ce domaine. Les objets similaires devront être les uns près des autres afin de faciliter leur accessibilité. L'objectif principal des méthodes d'indexation d'images est de construire une structure adéquate qui permet une recherche rapide et efficace. Le voisinage pertinent permet aux utilisateurs de localiser la recherche, c'est-à-dire parcourir la plus petite partie dans l'espace de recherche, malheureusement la performance de ces techniques se dégrade lorsque le nombre des descripteurs dépasse un certain seuil. Il existe plusieurs

méthodes d'indexation, nous distinguerons trois catégories ; selon le type de partitionnement de l'espace, selon les caractères statiques et selon les caractères dynamiques.

### 3.2.1. Indexation selon le partitionnement de l'espace

#### • L'espace des caractéristiques

Ce partitionnement permet de diviser l'espace des caractéristiques indépendamment de la distribution des données, il présente une organisation physique. La méthode Kd-tree en est un exemple, son principe se base sur le partitionnement de l'espace des caractéristiques pour accélérer la recherche dans les espaces multidimensionnels, cette méthode fait partie de la famille BSP-trees (*Binary Space Partitioning Trees*) [Weber et al., 1998, Smeulders et al., 2000].



**Figure 3.1.** La structure Kd-tree, partitionnement d'un ensemble de cinq données représentées en deux dimensions [Gouet-Brunet, 2010]

L'espace est divisé en régions, si le nombre est élevé dans une région alors celle-ci va être divisé en sous régions (figure 3.1). Avec cette structure, la consommation de l'espace croît exponentiellement avec la dimension et la recherche des  $k$  plus proches voisins revient dans ce cas à explorer toutes les partitions [Weber et al., 1998].

#### • Partitionnement des données

Cette méthode associe à chaque point de l'espace des caractéristiques une région qui est représentative du voisinage de ce vecteur de points. Par exemple, la méthode SR-tree qui se base sur l'utilisation conjointe de volumes hyper-sphériques et hyper-rectangulaires, elle se base sur les principes des deux méthodes: R-tree et SS-tree [Katayama and Satoh, 1997].

#### • Partitionnement basé sur des calculs de distance

Ce type de structures est adapté à des systèmes de recherche d'images par le contenu lorsque l'espace des caractéristiques est métrique, comme les méthodes VP-tree ou MVP-tree. Le

principe de cette méthode est le suivant : on prend un point exemple parmi les données. Le reste est divisé en cercles concentriques autour de ce *point exemple* ce qui aboutit à une structure *basée distance* [Smeulders et al., 2000].

### 3.2.2. Indexation selon le caractère statique et dynamique

Les méthodes d'indexation dans ce type peuvent être également divisées en *trois catégories* : [Heesch, 2008]

- **Des méthodes statiques**

Sont notamment *les méthodes hiérarchiques* et *les réseaux*. Comme leur nom l'indique, les structures statiques sont des structures qui ne sont pas modifiées pendant la recherche tandis que les structures dynamiques sont des structures qui s'adaptent pendant la recherche.

- **Structure hiérarchiques statiques**

L'hiérarchie a une présence universelle dans notre vie quotidienne: comme exemple dans l'organisation des fichiers sur un ordinateur, l'agencement de livres dans une librairie physique sur le web et beaucoup plus.

Les structures hiérarchiques ont été étudiées depuis longtemps comme une solution qui traite la complexité en temps [Fukunaga and Narendra, 1975], nous trouvons l'application récente à la recherche d'images [Kurniawati et al., 1997, Milanese et al., 1996, Obdržálek and Matas, 2005, Roussopoulos et al., 1995, White and Jain, 1996].

Cette structure consiste à diviser l'ensemble des données en sous ensembles (clusters).

L'idée générale est de trouver les voisins les plus proches d'une image requête en explorant un arbre hiérarchiquement en centroïde (l'élément qui représente le mieux l'ensemble des données contenues dans un cluster) de clusters.

À chaque étape, une comparaison est effectuée entre la requête et la représentation du nœud relatif à l'étape. A chaque étape l'utilisateur compare l'image requête avec les centres de gravités classés à un niveau donné de la hiérarchie et décide alors du chemin à poursuivre.

Les méthodes de clustering, qui ont pour principe de regrouper les données en sous-ensembles, se divisent en deux catégories :

Les méthodes de clustering strict (*hard clustering*) et les méthodes de clustering flou (*fuzzy clustering*).

Dans les méthodes de la première catégorie, un élément est clairement attribué à un seul cluster, tandis que dans la deuxième catégorie un élément peut appartenir à plusieurs groupes.

### **-Les méthodes de clustering strict (Hard clustering)**

*Agglomerative methods* peuvent être obtenus soit par fusion itérative des clusters (*bottom-up*) c'est-à-dire des feuilles vers la racine, pour lesquelles on regroupe les données en petits ensembles *pyramide de similarité* [Chen et al., 2000] soit par division clustering *top-down* c'est-à-dire de la racine vers les feuilles, pour lesquelles l'ensemble entier des données est divisé en clusters. Les premières applications de la méthode agglomérative dans la navigation des images sont décrites dans [Yeung and Liu, 1995, Zhang and Zhong, 1995, Yeung and Yeo, 1997, Krishnamachari and Abdel-Mottaleb, 1999]. Les deux premiers articles portent sur la navigation des vidéos. Dans [Yeung and Yeo, 1997] les auteurs regroupent des séquences vidéo et construisent des (*scene transition graphs*) visualisant le développement des histoires dans la vidéo.

Dans [Zhang and Zhong, 1995] et [Yang, 2004] la carte d'auto-organisation (SOM) [Kohonen, 2001] est appliquée aux images sur une grille à deux dimensions. La grille obtenue est ensuite regroupée par la méthode agglomérative. Les plus récents algorithmes sont présentés par [Laaksonen et al., 2000] où les images sont organisées sur une carte SOM, hiérarchiquement organisées. Pour chaque image un ensemble de trois caractéristiques visuelles, basées sur le retour de pertinence, les images des différents SOM sont fusionnées. Dans [Chen et al., 1998] et [Chen et al., 2000], Chen et al proposent (*The similarity pyramid*) pour représenter la collection d'images. Chaque niveau est organisé de telle sorte que des images similaires sont organisées l'une près de l'autre sur une grille à deux dimensions.

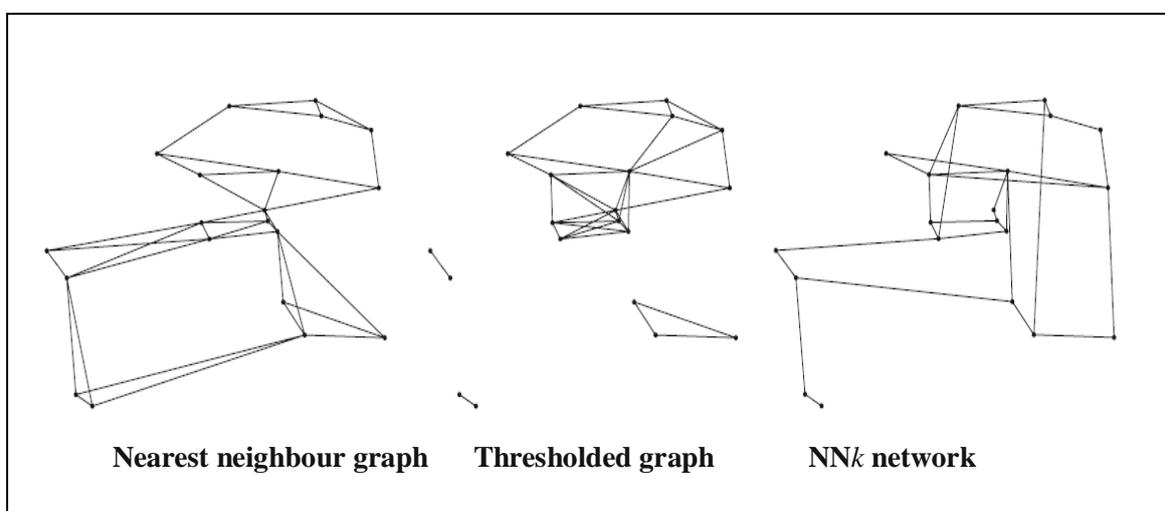
Les images sont d'abord organisées en un arbre binaire par une *agglomerative*, l'arbre binaire est par la suite transformé en un arbre où chaque nœud à quatre au lieu de deux enfants, c'est en fait, une structure de données qui est techniquement connu sous le nom d'arbre quaternaire. Un quadtree permet non seulement aux utilisateurs un grand choix à chaque niveau de la hiérarchie, mais il permet également une meilleure utilisation de l'espace d'affichage en deux dimensions.

### **-La méthode de clustering floue**

Dans cette méthode, la donnée peut appartenir à plusieurs clusters, la méthode fournit une mesure de confiance associée à l'assignement particulier d'un cluster à une donnée. Comme exemple Les méthodes probabilistes [Clough et al., 2005] démontre que les méthodes par division comme *K-means* [Pečenović et al., 2000] donnent des clusters moins cohérents que les méthodes *agglomeratives*, par ailleurs les centroïdes des clusters représentent rarement un bon résumé de l'ensemble des images présentes dans le cluster [Clough et al., 2005].

### -Les Réseaux statiques

La manière dont la connaissance est organisée dans le cerveau humain reste encore un point incompréhensible. Le stockage n'est pas hiérarchique mais prend la forme d'une structure associative dont les concepts sont reliés entre eux. Dans l'indexation les données ont des relations de similarité que l'utilisateur peut discerner parmi un ensemble d'images. L'inconvénient majeur, par rapport aux structures hiérarchiques, est de ne pas avoir une vue globale de la base de données mais elle permet une navigation moins contraignante dans la base. On peut mentionner : *Nearest neighbour networks*, *Thresholded graphs*, *NNk Network*, *Navigable hypercube*, *pathfinder networks* [Clough et al., 2005].



**Figure 3.2.** Comparaison de trois types différents de réseaux: nearest neighbour graph, thresholded graph et NNk network

### 3.2.3. Méthode hiérarchique dynamique

Selon le type de partitionnement à caractère dynamique, nous citons *La navigation ostensive* qui est une méthode dynamique.

#### -Navigation ostensive

La navigation ostensive est un exemple des méthodes dynamique. Elle se base sur le bouclage de pertinence pour raffiner la recherche, l'utilisateur doit sélectionner une image parmi les images résultantes pour formuler une nouvelle requête à partir d'une somme pondérée des caractéristiques de cette image et des précédentes [Clough et al., 2005].

### 3.3. Recherche d'images

En ce qui concerne la recherche d'images par le contenu, il existe trois paradigmes essentiels, selon l'objectif attendu par l'utilisateur : la recherche par *cible*, la recherche par *catégorie* et la recherche par *association* [Smeulders et al., 2000].

#### 3.3.1. Recherche par cible

Dans ce type de recherche l'utilisateur du système a une idée plus ou moins précise sur ce qu'il cherche. L'image recherchée doit contenir un objet particulier ou elle correspond à une forme mentale précise, par exemple l'image d'un paysage est représentée par une zone verte au tiers inférieur avec une zone bleu ciel au dessus [Smeulders et al., 2000].

#### 3.3.2. Recherche par catégorie

La recherche dans ce type est un peu plus large ou plus globale que la première, l'utilisateur cherche un ensemble d'images. Par exemple dans la recherche par cible l'utilisateur cherche l'objet *banane*, alors que dans la recherche par catégorie la requête serait de chercher des images de fruits.

#### 3.3.3. Recherche par association

Cette fois, l'utilisateur n'a pas une idée précise sur ce qu'il cherche. Si on prend l'exemple d'un journaliste qui écrit un article théorique et qui veut illustrer son travail par des images, beaucoup d'images peuvent convenir sa recherche et donnent alors une nouvelle portée à l'article. Dans ce cas la requête par exemple ne fonctionne plus, l'utilisateur a besoin de naviguer sur cette base et il peut trouver ce qu'il cherche dès la première fois, donc il faut des mécanismes d'interaction entre le système et l'utilisateur pour affiner le résultat.

### 3.4. Architecture du système

Avec la naissance du projet QBIC [Niblack et al., 1993], l'idée d'extraire une sémantique du contenu d'images est créée afin de les indexer. En premier temps les caractéristiques visuelles telles que la couleur, la texture et la forme sont extraites des images, ensuite la construction des signatures à partir des images afin de pouvoir mesurer la similarité entre signatures ou images est faite.

L'utilisateur exprime sa recherche par une requête, souvent à partir d'une image exemple ou par mots clés, à l'aide d'une mesure de similarité entre requête et la base d'images ou les images collectées un résultat est affiché à l'utilisateur. La figure (3.3) suivante explique le déroulement du mécanisme de la recherche d'images par le contenu.

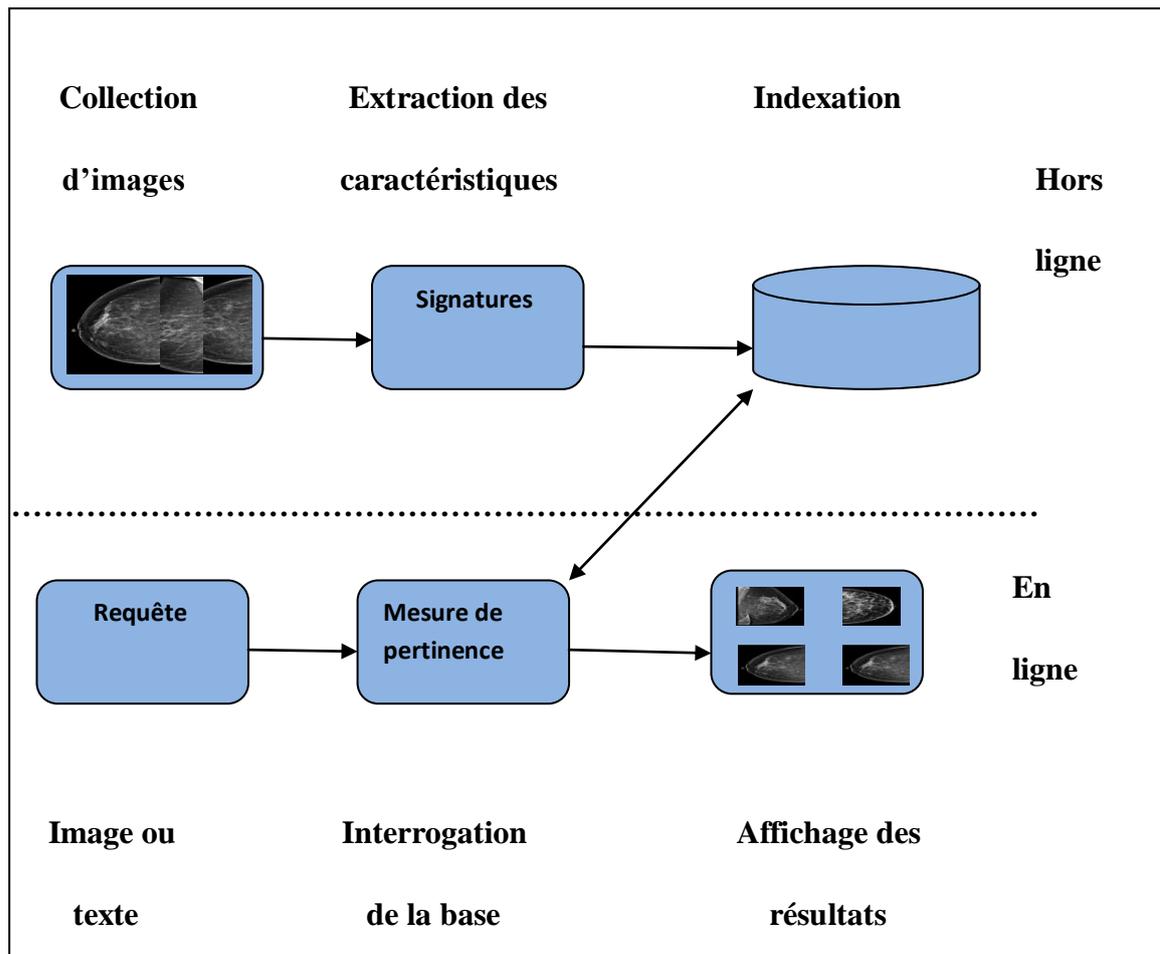


Figure 3.3. Déroulement de la recherche d'images par le contenu

### 3.4.1. Segmentation

La segmentation est une étape très importante dans les systèmes d'indexation et de recherche puisque les résultats ultérieurs *description, classification* sont bien liés de la segmentation. Une bonne détection de contour ou une région engendre une bonne description des caractéristiques.

Dans le domaine médical, l'étape de segmentation consiste à détecter la totalité du sein ou bien une partie qui présente les opacités ou bien les calcifications.

Il existe une multitude de méthodes de segmentation et elles sont classées en trois catégories: *les approches basées contour, les approches basées région et les approches basées pixels.*

- **Les méthodes de seuillage**

Ces méthodes consistent à segmenter l'image en deux ou plusieurs classes en utilisant l'histogramme. Ce dernier est partitionné en niveau de gris, mais cette méthode est utile seulement si l'histogramme contient des pics séparés. On peut citer le travail de Matsubara et

al qui ont utilisé différents seuils pour chaque type du tissu mammaire [Matsubara et al., 1997, Matsubara et al., 1998].

Mudigonda et al ont utilisé *le seuillage multiniveaux* pour détecter des contours fermés [Mudigonda et al., 2001], mais le seuillage est appliqué après l'utilisation d'un filtre, Kobatake et al ont été utilisé *un filtre d'iris* et dans le même contexte on peut citer le travail de Saha et al. qui ont utilisé une méthode de *connectivité floue* avant d'utiliser un seuillage pour détecter les masses [Saha et al., 2001]. Heath and Bowyer ont développé un algorithme de détection en utilisant *un filtre de la fraction moyenne sous le minimum*, connu en anglais AFUM (*Average Fraction Under Minimum*) [Heath and Bowyer, 2000].

#### • Les méthodes basées régions

Parmi les méthodes les plus connus basées régions sont :

**-Croissance de régions :** cette méthode est basée sur la proagation du point germe selon un critère d'homogénéité. Elle comporte deux étapes, *une étape d'initialisation* où l'on retrouve les germes des régions les plus représentatives de l'image et une *deuxième étape* où les germes s'accroissent en respectant des contraintes d'homogénéité, de forme géométrique ou de la taille [Huo et al., 1995, Cao et al., 2009].

Huo et al. ont introduit le contour comme information supplémentaire avec un critère d'homogénéité et ils ont développé une méthode semi automatique où le radiologue fixe le point initiale (point germe), ensuite l'étape *croissement de région* est alors calculée automatiquement [Huo et al., 1995]

Kupin et Giger ont amélioré la méthode et l'ont comparé avec celle de Huo et al, la première méthode fait intervenir *l'approche gradient d'indice radial* qui représente la mesure de la proportion moyenne du gradient radialement orienté vers l'extérieur. La deuxième méthode est basée sur la probabilité d'appartenance à une région par une distribution non gaussienne. L'étude comparative montre que la dernière version fournit de meilleurs résultats que la première. Le seul inconvénient réside dans le choix du point germe car si le point germe est situé dans une zone non homogène, la mesure de similarité va produire de fortes variations et la croissance va s'arrêter rapidement, ainsi il est nécessaire de choisir des régions homogène afin de lancer l'algorithme [Kupinski and Giger, 1998].

**-Division et fusion de régions :** Cette méthode est développée par Horowitz et Pavlidis, le processus est décomposé en deux étapes *l'étape division* où elle analyse les régions d'image si elles ne vérifient pas le critère d'homogénéité alors la région est divisée en blocs, on réitère l'algorithme sur chaque région. *L'étape fusion* étudie toutes les régions voisines, si l'union de ces deux régions vérifient le critère d'homogénéité alors elles seront fusionnées [Horowitz

and Pavlidis, 1974]. Rangayyan et al ont exploité cette méthode dans la segmentation des masses mammaires mais ils ont ajouté la mesure d'acutance [Rangayyan et al., 1997].

Egalement, Reyad et al ont présenté une technique pour la segmentation des tissus fibros glandulaires en se basant sur la technique *division et fusion* appliquée à l'histogramme de l'image mammographique et la distribution log-normal [Reyad et al., 2011].

#### • Les méthodes basées contours

Les approches les plus anciennes sont basées sur *la notion de dérivation* pour différencier entre l'objet et le fond de l'image qui correspondent à des zones de fortes variations d'intensité [Sobel, 1970, Prewitt, 1970, Canny, 1986]. Une autre méthode de détection de contour est la ligne de partage des eaux LPE (*Watershed*), elle décrit l'image à niveau de gris suivant des structures topographiques caractéristiques. Le principe de base de cette méthode est de remplir progressivement les bassins par une eau incessante. Alors la LPE s'établie à partir des débordements de chaque versant [Hamarnah and Li, 2009]. Une autre méthode est intitulée les contours actifs, le principe général se base sur les différents modèles de contour et différentes stratégies d'optimisation, comme exemple les modèles déformables connus aussi sous les noms (*snakes*) [Kass et al., 1987, Terzopoulos and Fleischer, 1988].

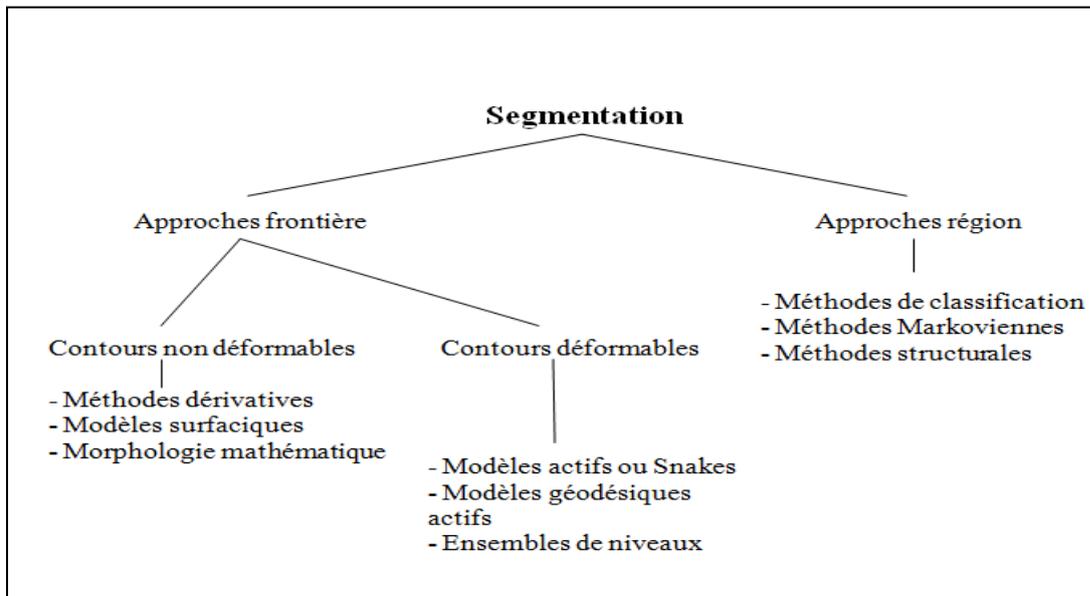
Il existe deux catégories majeures *les modèles basés contour* [Kass et al., 1987, Li et al., 2005] et *les modèles basés région* [Chan and Vese, 2001, Li et al., 2008].

Le contour actif évolue vers le bord de l'objet ayant un maximum de gradient d'intensité, ces méthodes ont trouvé de grands difficultés surtout dans le choix des paramètres d'initialisation et à la convergence dans les niveaux des bords concaves [Berger and Mohr, 1990].

Dans leur travail, Amini et al. ont déduit que la méthode peut s'avérer numériquement instable et que les points ont tendance à s'entasser sur certaines portions du contour [Amini et al., 1990]. D'autres travaux se basent sur le fait que les intensités sont statistiquement homogènes dans les régions à segmenter c'est ce qui n'existe pas dans les images médicales et particulièrement dans *les images de mammographie* [Chan and Vese, 2001, Paragios and Deriche, 2002]. Li et al. ont récemment présenté un modèle qui est capable ayant divers intensités inhomogènes, le modèle du contour actif basé région est capable de segmenter des images ayant diverses intensités inhomogènes, les résultats sont satisfaisants dans le cas où les objets sont à contours mal définis ou masqués, ce qui est souvent le cas des masses mammaires [Li et al., 2008].

Après un traitement et récupération des informations de bas niveau issues de la phase de prétraitement et de segmentation nous arrivons à la phase de description et extraction des caractéristiques, cette étape décrit d'une façon quantitative ou qualitative les objets dans une

image, c'est-à-dire elle les rend analytique. Les méthodes utilisées dépendent de l'objectif soit *une description globale*, soit *une description locale*, de type d'image à analyser, soit une *image binaire*, soit une *image en niveau de gris* soit *une image couleur*. En littérature, les chercheurs ont utilisé la couleur, la texture et la forme pour assurer une description d'images. Ces descripteurs sont détaillés dans la section (cf. sec.3.4.2).



**Figure 3.4.** Aperçu général des techniques de segmentation d'image

L'étape de la segmentation est très importante pour l'indexation et la recherche d'images par le contenu. En imagerie médicale on est obligés d'extraire les régions d'intérêts et de les caractériser (niveau de gris, forme, texture) afin de trouver des régions similaires dans la base.

### 3.4.2. Caractéristique visuelles

Principalement il existe deux approches pour les caractéristiques qui peuvent être extraites de l'image.

#### 3.4.2.1. Descripteur globale

Dans ce cas, il s'agit de fournir une observation sur toute l'image, le grand avantage c'est l'algorithme mis en œuvre et le nombre réduit d'observations obtenues qui influent sur les calculs effectués par la suite. Le grand inconvénient des descripteurs globaux c'est la perte de l'information de localisation d'objet ou d'élément dans l'image. Si on prend un petit objet dans une image, avec l'application de ce type des descripteurs on obtient une information qui se base principalement sur l'arrière plan qui peut être non pertinent.

### 3.4.2.2. Descripteur locale

Ce type de descripteur consiste à calculer quelques attributs sur des proportions dans l'image. Le grand avantage est de conserver une information localisée dans l'image évitant certains détails. L'inconvénient principal c'est la quantité importante d'informations qui rendent le traitement plus complexe, les différents types de descriptions et leur utilité sont détaillés dans les sections qui suivent.

- **Couleur**

La couleur est une caractéristique très utilisée dans les systèmes de recherche d'image, les images sont stockées dans les espaces de couleur *représentation mathématique d'un ensemble de couleurs*. Il en existe plusieurs espaces de couleur parmi lesquels : le RVB Rouge, Vert, Bleu mais son grand désavantage c'est qu'il ne correspond pas aux caractéristiques de la vision humaine, le TSL *Teinte, Saturation, Luminance*, le CMJ Cyan, Magenta, Jaune, le CMJN Cyan, Magenta, Jaune, Noir pur et le CIE *Codage de la couleur selon la Commission Internationale de l'Eclairage* [Pascale, 2003].

Dans le domaine du cancer du sein, la description des pathologies mammaires est un maillon très important. Le résultat de la classification est étroitement lié au résultat de cette étape qui fournit aux radiologues un ensemble important d'information sur les lésions mammaires. En effet, les images de mammographie sont des images en niveau de gris alors ce type de descripteur ne sera pas pris en compte dans notre étude.

- **Texture**

Il n'existe pas une définition conventionnelle de la texture, elle représente une zone de l'image qui n'ayant pas une couleur unie, les primitives formant la zone de l'image peuvent prendre un aspect répétitif ou aléatoire. Nous citons ici la définition de Haralick :

Selon Haralick « *Une image texturée est d'écrite par le nombre et les types de ses primitives tonales ainsi que leurs orientations spatiales. Elle ne peut pas être analysée sans une vue de la primitive tonale de référence. Au fur et à mesure que la résolution augmente, on observe une texture fine puis une texture grossière* » [Haralick, 1979].

Dans la littérature nous pouvons trouver plusieurs familles de texture : *les textures déterministes* qui se caractérise par la répétition spatiale du motif, *les textures stochastiques* sont irrégulières et obéisse à certaines lois de probabilité, *les texture observable* c'est un mélange des deux premiers *motifs similaire non identiques*.

L'existence de plusieurs familles de textures engendre la proposition de plusieurs descripteurs : les filtres de Gabor [Gabor, 1946], les matrices de cooccurrences [Haralick,

1979], la décomposition paramétrique Wold [Liu and Picard, 1996], dans ce qui suit, nous présentons les méthodes les plus utilisées dans le traitement des images de mammographies.

**-Les statistiques de premier ordre**

Les statistiques du premier ordre permettent de calculer la probabilité d'observer un niveau de gris à un emplacement aléatoire dans une image. Elles sont estimées sur les niveaux de gris sans la considération de leur distribution relative et sont représentées par la formule suivante:

$$Hist(i) = \frac{1}{ng} \sum_{p=1}^{ng} I(p) \equiv i \tag{3.1}$$

Avec : I représentant l'intensité des pixels, ng le nombre de niveaux de gris. On peut citer aussi, la moyenne (Moy), l'écart type ( $\sigma_{Hist}$ ) et la variance (Var).

**La moyenne :**  $Moy = \frac{1}{ng} \sum_{i=1}^{ng} i.Hist(i)$  (3.2)

**L'écart type :**  $\sigma_{Hist} = \frac{1}{ng} = \sqrt{\frac{1}{ng} \sum_{i=1}^{ng} (i - Moy)^2 . Hist(i)}$  (3.3)

**La variance :**  $Var = \frac{1}{N} \sum_{p=1}^N (I(p) - Moy)^2$  (3.4)

L'inconvénient des statistiques du premier ordre c'est qu'ils sont calculés à partir de l'histogramme des intensités des niveaux de gris et n'utilisent pas les relations spatiales entre les pixels.

**-La matrice de co-occurrence**

Cette approche classique, *statistiques du second ordre* consiste à calculer une matrice de co-occurrence sur les niveaux de gris des pixels, cette matrice décrit les régularités observables dans les niveaux de gris des pixels d'une région. Elle met en jeu deux pixels pour le calcul des caractéristiques, en conséquence, elle permet de déterminer la fréquence d'apparition d'un motif textural formé de deux pixels avec une certaine distance et direction. La distance permet d'avoir une description significative de la périodicité de la texture et l'angle permet d'évaluer la direction de texture, pour limiter le nombre de calculs on prend généralement les valeurs des angles : 0°, 45°, 90° et 135°.

Plusieurs statistiques peuvent alors être calculées sur la matrice de co-occurrence. Comme, l'énergie, l'entropie, le contraste et l'homogénéité, ces descripteurs sont utilisés et détaillés dans le chapitre V (cf.sec.5.2.3).

Dans le cas des images de mammographies, la matrice de co-occurrence a été exploitée pour la détection des masses, les microcalcifications et pour la classification des microcalcifications en type malin et bénin [Kim et al., 1997].

#### **-Description fréquentielle : transformée de Fourier**

La transformée de Fourier est l'une des méthodes les plus utilisées pour l'analyse de la texture d'une image. Comme la texture présente une répétition d'un motif elle sera caractérisée par des pics, les textures fines dans les fréquences hautes et les textures grossières dans les basses fréquences.

Soit  $i(x, y)$  une image, la transformée de Fourier est :

$$F(f_x, f_y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} i(x, y) e^{-2i\pi(f_x x + f_y y)} dx dy \quad (3.5)$$

Avec :  $f_x, f_y$  les fréquences spatiales de l'image  $i(x, y)$ .

D'un côté cette méthode est adaptée à l'analyse des comportements harmoniques *périodicité d'une texture*, d'un autre côté elle n'est pas appropriée à l'analyse des formes localisées spatialement et elle n'est pas beaucoup utilisée car ses valeurs sont difficilement à être exploiter et interpréter.

#### **-L'analyse fractale**

Helge Von Koch a inventé en 1906 *La courbe de Koch*<sup>1</sup> qui est une des premières courbes fractales. La méthode la plus répandue est celle de comptage de boîtes, connu en anglais par le nom (*Box counting*). La dimension fractale d'un ensemble  $A$  est défini par :

$$D = \log(N)/\log(1/r) \quad (3.6)$$

Avec:  $N$  est le nombre total des copies distinctes similaires à  $a$  et  $1/r$  correspond au facteur d'échelle avec lequel  $A$  est divisé.

Une autre méthode dite *Fractional Brownian Motion* est utilisée souvent pour décrire les phénomènes aléatoires, son point fort réside dans l'invariance aux transformations linéaires comme la translation. Par conséquence, cette méthode est utilisée pour gérer des textures, elle est utilisée en mammographie pour la détection des lésions en tant que descripteur de forme. L'inconvénient majeur est la complexité des calculs qui engendre un temps de calcul assez long [Sarkar and Chaudhuri, 1994].

---

<sup>1</sup> Le flocon de Koch est l'une des premières courbes fractales à avoir été décrite (bien avant l'invention du terme fractal. Elle a été inventée en 1906<sup>1</sup> par le mathématicien suédois Helge von Koch (1870 - 1924).

• **Forme**

De même que les notions de couleur et de la texture, la notion de forme est très importante puisqu'elle désigne l'aspect général d'un objet, son contour.

On dit que la forme est réussie si elle permet de décrire l'objet d'une manière proche à la perception humaine. La compréhension de la forme nécessite la compréhension de certaines notions tel que : la fermeture, l'ouverture... [Choquet, 1978].

**-Boule ouverte :** soit  $P$  un point de  $R^2$  et  $r$  un réel positif et  $d_E$  la distance. La boule ouverte de  $B^{d_E}(P, r)$  du centre  $P$  et de rayon  $r$  l'ensemble des point  $x$  tel que :

$$B^{d_E}(P, r) = \{x \in R^2 \mid d_E(P, x) < r\} \quad (3.7)$$

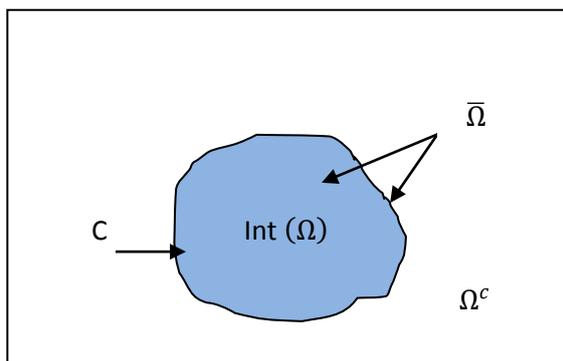
**-Ouvert :** un ensemble ouvert ou partie ouverte ou tout simplement un ouvert, tout ensemble  $\Omega$  de  $R^2$  qui ne contient aucun point de sa frontière.

**-Fermé :** soit  $\Omega^c$  l'ensemble des points  $P$  qui n'appartient pas à  $\Omega$ , un ensemble de  $R^2$  est dit fermé si son complémentaire  $\Omega^c$  est un ouvert.

**-Borné :** Un ensemble de points  $\Omega$  de  $R^2$  est dit borné s'il existe un point  $p$  de  $R^2$  et un réel positif  $r$  de façon que  $\Omega$  soit un sous-ensemble de la boule ouverte  $B^{d_E}(P, r)$ .

**-Intérieur :** On appelle intérieur  $int(\Omega)$  d'un ensemble de points de  $R^2$  le plus grand ouvert de  $\Omega$ , qui est en effet la réunion de tous les ouverts inclus dans  $\Omega$ .

**-Contour :** On considère un ensemble de points  $\Omega$  de  $R^2$ , le contour  $C$  de  $\Omega$  est défini par l'intersection entre la fermeture de  $\Omega$  et la fermeture de son complémentaire ( $C = \bar{\Omega} \cap \overline{\Omega^c}$ ), la compréhension de cette notion facilite la compréhension de la notion forme.



**Figure.3.5.** Schéma présentant les notions de fermeture, d'intérieur, de contour et de complémentarité

La figure 3.4 explique les notions : intérieur  $int(\Omega)$  de l'objet colorié avec le bleu, le contour  $C$  c'est la courbe noir fermé, la fermeture  $\bar{\Omega}$  est représentée par la partie colorié et le contour noir, le complémentaire de l'intérieur  $int(\Omega)^c$  est l'ensemble contour et le reste du plan, le complémentaire de la fermeture  $\bar{\Omega}^c$  est le reste du plan sans considéré le contour et le complémentaire du contour  $C^c$  est constitué par le reste du plan plus l'intérieur.

A partir de ces notions on peu définir la forme comme un ensemble de points  $\Omega$  de  $R^2$  qui est la fermeture d'un sous ensemble de  $R^2$  ouvert et borné, le contour de la forme est représenté par un nombre fini de courbes fermées.

#### -Les descripteurs de forme en mammographie

On a mentionné dans la section (cf. sec.4.7) les signes de malignité dans les deux cas de masses et de microcalcifications, pour les microcalcifications les signes dépendent du nombre de calcifications dans un cluster, en ce qui concerne les masses la malignité augmentent avec la complexité du contour.

#### • Les moments géométriques

Les moments géométriques permettent de décrire une forme à l'aide de propriétés statistiques.

Formule générale des moments :

$$m_{pq} = \sum_{p=0}^m \sum_{q=0}^n x^p y^q f(x, y) \quad (3.8)$$

L'ordre du moment est  $p + q$ . Le moment d'ordre 0  $m_{0,0}$  représente l'aire de la forme de l'objet. Les deux moments d'ordre 1  $m_{0,1}$  et  $m_{1,0}$  associés au moment d'ordre 0  $m_{0,0}$  permettent de calculer le centre de gravité de l'objet. Il est possible de calculer à partir de ces moments l'ellipse équivalente à l'objet.

À partir des moments géométriques, Hu a introduit sept invariants aux translations, rotations et changement d'échelle, appelés moments de Hu [Hu, 1962].

Les moments de Hu sont utilisés depuis longtemps pour calculer la position du centre d'une distribution mais aussi sa variance. Donc sur une image on peut calculer l'agencement des pixels d'une forme par exemple.

Nous avons calculé sept moments de Hu, ces moments sont invariants à la rotation, à la translation et le changement d'échelle. La formule du moment Hu  $m_{pq}$  d'ordre  $p + q$  d'une fonction de distribution  $f(x, y)$  est comme le suivant :

$$M_{p,q} = \iint (x^p y^q) F(x, y) dx dy \quad p, q = 0, 1, 2 \dots \quad (3.9)$$

Pour une image digitale  $g(x, y)$  de taille  $M \times N$ , la formule ci-dessus devient :

$$m_{pq} = \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} x^p y^q g(x, y) \quad (3.10)$$

Le moment central  $\mu_{pq}$  est défini par :

$$\mu_{pq} = \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} (x - \bar{x})^p (y - \bar{y})^q g(x, y) \quad (3.11)$$

Avec :  $\bar{x} = \frac{m_{10}}{m_{00}}$ ,  $\bar{y} = \frac{m_{01}}{m_{00}}$ , on définit le moment central normalisé :  $\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\frac{p+q+2}{2}}}$ ,  $p + q = 2, 3, \dots$

Basés sur les moments géométriques nous introduisons les moments de Hu cités dans le chapitre V (cf. sec.5.2.3).

#### **-Descripteurs géométriques**

Ce type de descripteurs caractérise l'aspect de la forme d'une façon générale indépendamment du domaine d'application, on peut par exemple mentionner l'aire, le centre de gravité, la circularité... qui vont être détaillés dans le chapitre V (cf. sec.5.2.3).

Dans ce qui suit, nous nous focalisons sur l'étude des descripteurs pour le cas des masses seulement. Nous expliquons quelques descripteurs et les descripteurs utilisés sont détaillés dans le chapitre V (cf. sec.5.2.3).

#### **-Descripteurs spécifiques**

Chen et al. ont proposé cinq descripteurs pour détecter les masses selon la régularité et l'allongement du contour dans les images échographiques [Chen et al., 2003]. Le premier descripteur est intitulé *Le nombre des protubérances et des dépressions importantes*, NSPD connu en anglais par (*Number of Substantial Protuberances and Depressions*).

Ce descripteur compte le nombre des protubérances et des dépressions les plus importantes dans le tissu, à partir du plus petit ensemble des points convexes, on construit une enveloppe convexe contenant la lésion mammaire. On calcule pour chaque point  $p_i$  du contour un angle polaire  $\theta_i$ . Le point  $p_i$  est considéré comme un point convexe si le point d'angle polaire  $\theta_i \geq \theta_p$  avec  $\theta_p \in \{20^\circ, 30^\circ, 40^\circ, 50^\circ, 60^\circ\}$ . La valeur du NSPD est élevée si le contour est irrégulier et spéculé, c'est-à-dire qu'il possède le critère de malignité.

### -L'indice de lobulation

L'indice de lobulation, LI connu en anglais par (*Lobulation Index*), est utilisé pour caractériser la distribution de la taille des lobes<sup>1</sup> dans une lésion.

$$LI = \frac{A_{max} - A_{min}}{\frac{1}{N_1} \sum_{i=1}^{N_1} A_i} \quad (3.12)$$

Avec :  $N_1$  est le nombre de lobes,  $A_i$  la dimension du ième lobe ( $i = 1, \dots, N_1$ ),  $A_{max}$  et  $A_{min}$  présentent les dimensions du plus grand et plus petit lobe respectivement.

Le point faible de cet indice c'est que LI peut prendre une valeur faible dans le cas d'une masse maligne avec plusieurs lobulations de dimensions similaires comparant avec un seul lobe important d'une masse bénigne. De ce fait cet indice ne reflète pas bien l'irrégularité de la masse.

### -Le squelette elliptique normalisé

Le squelette elliptique normalisé, en anglais ENS (*Elliptic Normalized Skeleton*) est utilisé pour détecter les lésions mammaires dans les images échographiques.

Le nombre de points constituant le squelette est la somme des points terminaux (PT), les points simples (PS) et des points multiples.

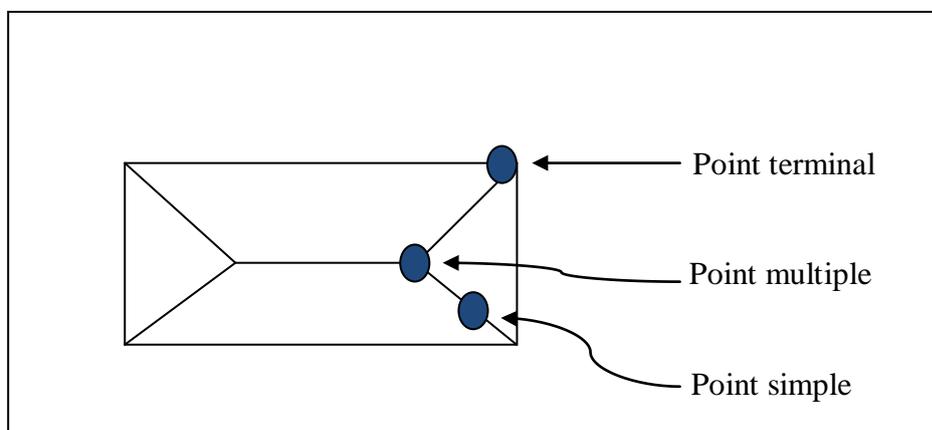


Figure 3.6. Les différents types de points dans un squelette

D'après Matheron, le squelette possède des propriétés importantes telles que : l'invariance, l'homotopie, la réversibilité, la description hiérarchique de la forme et l'épaisseur. Les auteurs ont suggéré de normaliser ce descripteur par *PEEq* le Périmètre de l'Ellipse Equivalente [Cheikhrouhou, 2012], la formule (3.13) suivante donne le squelette elliptique normalisé :

<sup>1</sup> Lobe est défini par la région délimitée par le contour de la lésion et la ligne reliant deux points adjacents concaves.

$$ENS = \frac{Pt+Ps+Pm}{PEEq} \quad (3.13)$$

### 3.4.3. Représentation des données

A partir des caractéristiques extraites des images on obtient les signatures qui peuvent être des graphes ou des vecteurs.

#### 3.4.3.1. La Signature

Le travail sur l'ensemble des descripteurs extraits est généralement difficile si on tenant compte de leurs nombre. La méthode la plus simple pour construire une signature relatif au contenu de l'image est de faire l'histogramme qui contient le nombre des occurrences calculées soit sur une partie ou toute l'image, en revanche la signature obtenue à une grande dimension.

Un dictionnaire visuel est constitué, on peut d'abord procéder par une quantification statique des descripteurs, par exemple en découpant de manière régulière l'espace des couleurs. De cela on peut dire que les signatures peuvent être plus complexe qu'un vecteur, par exemple un ensemble des vecteurs ou des graphes, dans ces conditions la fonction de similarité est plus simple dans le cas du vecteur que dans les autres cas [[Stricker and Orengo, 1995, Brunelli and Mich, 2001].

#### 3.4.3.2. La similarité

La comparaison entre deux signatures est défini par une mesure de similarité, la signature fournie par l'utilisateur est comparée avec celles de la base des images et le résultat sont les images jugées comme similaire à l'image requête.

Il existe un nombre important de distances qui dépend de la nature des signatures, par exemple la distance utilisée entre vecteurs c'est  $L_1$ ,  $L_2$  ou  $L_\infty$ .

Pour ce cas soit  $x_i, x_j$  deux vecteurs dans l'espace  $R^N$ , la fonction de similarité est un produit scalaire dans  $R^N$  :

$$f(x_i, x_j) = \langle x_i, x_j \rangle \quad (3.14)$$

Pour les histogrammes on peut utiliser deux types de distance, soit de considéré les histogrammes comme des vecteurs et ça permet d'utiliser les distances vectorielles ou d'utiliser des distance spécifiques comme *Earth mover distance* [Rubner et al., 1998]. Pour les distributions, on peut utiliser les distances statistiques.

Pour utiliser le produit scalaire dans le cas où les signatures ne sont pas des vecteurs, il faut projeter d'abord les signatures dans l'espace hilbertien à l'aide d'une fonction d'induction  $\emptyset$  et après utiliser le produit scalaire dans cet espace :

$$f(x_i, x_j) = \langle \emptyset(x_i), \emptyset(x_j) \rangle \quad (3.15)$$

D'une façon générale on peut dire que deux images sont jugées similaires si la distance entre leurs caractéristiques prend une valeur faible, mathématiquement la distance est définie comme une fonction qui varie dans  $[0 ; 1]$  et qui vérifie:

$$d(c_1, c_2) + d(c_2, c_3) \geq d(c_1, c_3) \quad (3.16)$$

$$d(c_1, c_2) = d(c_2, c_1) \quad (3.17)$$

$$d(c_1, c_2) = 0 \Leftrightarrow c_1 = c_2 \quad (3.18)$$

La majorité des systèmes existants utilisent la distance comme une mesure de similarité, notons que cette distance est définie selon les caractéristiques extraites. Nous présentons dans ce qui suit les distances les plus utilisées.

• **Les distances**

Comme exemple des fonctions de distance *entre vecteurs* et *entre fonctions accumulatives* on peut présenter :

**-La distance de Minkowski**

Soit  $C_1$  et  $C_2$  deux vecteurs, on définit par  $L_p$  la famille de distances de Minkowski et  $C_{k,i}$  désigne la  $i$ -ème composante du vecteur  $C_k$ . Avec une variation dans la valeur de  $p$ , on obtient d'autres fonctions de distance.

$$L_p(C_1, C_2) = \left( \sum_{i=0}^n |C_{1,i} - C_{2,i}|^p \right)^{\frac{1}{p}} \quad (3.19)$$

**-Distance de Manhattan**

City-block ou encore la distance Manhattan est obtenue avec :  $p = 1$ .

$$L_1(C_1, C_2) = \sum_{i=1}^n |C_{1,i} - C_{2,i}| \quad (3.20)$$

La distance Manhattan est recommandée par le comité MPEG-7 [Jeannin, 2001] pour comparer deux formes décrite avec la méthode ART de Kim et Kim (1999).

**-Distance Euclidienne**

C'est le cas le plus fréquemment rencontré, elle est obtenue avec  $p = 2$  :

$$L_2(C_1, C_2) = \sqrt{\sum_{i=0}^n |C_{1,i} - C_{2,i}|^2} \quad (3.21)$$

**-Distance uniforme**

Une distance  $L_\infty$  qu'est elle-même une mesure limite *au sens d'une fonction* est obtenue avec  $p$  qui tend vers l'infini.

$$L_\infty = \max_{i=0}^n |C_{1,i} - C_{2,i}| \quad (3.22)$$

**-Distance classique entre histogrammes**

Soit  $H_1$  et  $H_2$  deux histogrammes, on définit par  $d(H_1, H_2)$  la notion d'intersection [Swain and Ballard, 1991]

$$d(H_1, H_2) = \sum_{j=1}^n \min(H_{1,j}, H_{2,j}) \quad (3.23)$$

Où  $H_{k,j}$  désigne le j-ème composant de l'histogramme  $H_k$  et  $n$  le nombre de discrétisations de chaque histogramme. Si les images ont le même nombre de pixels alors cette mesure possède les mêmes propriétés ordinales qu'une mesure  $L_1$ .

**Distance par forme quadratique entre histogrammes**

Les auteurs ont proposé une mesure pondérée qui permet d'intégrer la notion de similarité entre deux couleurs [Hafner et al., 1995].

$$d(H_1, H_2) = \sqrt{\sum_{i=1}^n \sum_{j=1}^n a_{ij} (H_{1,i} - H_{2,j})^2} \quad (3.24)$$

Où  $H_{k,j}$  désigne le j-ème composant de l'histogramme  $H_k$  et  $a_{ij}$  la similarité entre les deux couleurs  $i$  et  $j$ ,  $a_{i,j}$  est calculé avec la distance Euclidienne  $d_{ij}$  entre les deux couleurs.

$$a_{ij} = 1 - \frac{d_{ij}}{\max_{i,j} d_{ij}} \quad (3.25)$$

### -Distance statistique entre les histogrammes

Soit  $H_1$  et  $H_2$  deux histogrammes, on considère l'histogramme comme une densité de probabilité et pour chaque canal qu'est représenté par les axes de l'espace de représentation nous calculons la moyenne, la variance le moment statistique d'ordre 3. La distance est calculée avec la formule suivante :

$$d(H_1, H_2) = \sum_i W_{i,1} |\mu_i^{H_1} - \mu_i^{H_2}| + W_{i,2} |\sigma_i^{H_1} - \sigma_i^{H_2}| + W_{i,3} |m_i^{H_1} - m_i^{H_2}| \quad (3.26)$$

Où  $\mu_i^{H_1}, \sigma_i^{H_1}$  et  $m_i^{H_1}$  représente respectivement la moyenne, la variance et le moment statistique d'ordre 3 du canal  $i$  et  $W_{i,j}$  la pondération associé.

#### 3.4.4. Requête

La requête c'est le point qui marque le commencement d'une recherche dans un système de recherche d'images par le contenu. La plupart des systèmes existants reposent sur la requête par l'exemple, même si sa pertinence en terme de souplesse d'utilisation et sa puissance de déduire à partir de l'exemple, ce que recherche réellement l'utilisateur, ne sont pas prouvées. Il existe un ensemble de moyens différents pour exprimer une requête, on peut citer :

##### 3.4.4.1. Requête textuelle

Le texte est la manière la plus simple pour exprimer une requête, lorsqu'on cherche un ensemble d'images on utilise les mots clés, si les images ne sont pas accompagnées d'annotations textuelles alors il va être difficile de traduire la requête dans l'espace de représentation pour indexer la base d'images. Il existe deux approches, dans la première approche l'image est associé à un texte comme les catalogues de vente, les encyclopédies, le web. Si on prend l'exemple des images de web, l'indexation peut être effectuée par un mot qui se situe à la tête de la page ou par les mots fréquents de cette page, d'un autre côté les images ne sont pas indexées avec le même mot sur cette page. La deuxième approche il y a auto-annotation par le contenu, cette méthode utilise l'apprentissage où les images sont associées à des classes de mots pour apprendre à prédire des mots sur de nouvelles images. [Wang and Ma, 2005] propose d'utiliser les réseaux de neurones afin d'apprendre l'association des mots clés avec les caractéristiques extraites des images. Une fois les images

sont indexées textuellement, une recherche va être effectuée par modèle booléen [Cooper, 1970], vectoriel [Salton, 1971] ou probabiliste [Robertson and Jones, 1976].

#### 3.4.4.2. Requête image

Plusieurs systèmes de recherche d'images par le contenu préfèrent l'utilisation d'une requête image exemple (*query by example*). Dans ce cas l'utilisateur pose sa requête *exemple*, le système calcule la signature avec la même méthode appliquée sur toute la base afin de l'indexer, puis recherche les images similaires. La requête exemple peut être toute une image ou une région dans laquelle le système recherche une région qui ressemble à la région sélectionnée par l'utilisateur du système. Elle peut être aussi une composition logique de catégories à partir des thésaurus photométrique des régions dans la base, l'utilisateur sélectionne la région qui cherche [Fauqueur, 2003]. Le croquis dessiné par l'utilisateur présente un cas particulier dans les requêtes par exemples car il utilise seulement le descripteur de la forme [Schomaker et al., 1999].

Il existe aussi les requêtes directes qui sont composées de descripteurs visuels, des requêtes avec une combinaison pondérée des caractéristiques visuelles. Des études montrent que l'utilisateur combine la requête par croquis, requête par l'exemple et une navigation dans une classification hiérarchique d'images pour construire sa requête, s'il n'a pas une chose précise à chercher il va utiliser la navigation dans la classification hiérarchique donc il utilise le moyen le plus simple pour formuler sa requête [McDonald et al., 2001, McDonald and Tait, 2003].

#### 3.4.5. Affichage des résultats

Après le traitement de la requête d'une manière à pouvoir calculer la similarité des images par rapport à la requête, un tri est effectué sur les images d'une façon à afficher les images les plus pertinentes aux images les moins pertinentes. Donc les images les plus similaires à l'image requête vont être affichées à l'utilisateur.

#### 3.4.6. Le bouclage de pertinence

Dans certains systèmes, l'utilisateur a la possibilité d'annoter quelques images résultats qui lui sont présentées, c'est ce qu'on appelle le bouclage de pertinence (*relevance feedback*). On peut décrire ce processus comme suit : Lorsqu'une requête est faite et qu'une réponse est donnée par le système, l'utilisateur donne plus d'indications au système comme des images positives et des images négatives à partir des premiers résultats. Le système recalcule et donne de nouveaux les deuxièmes résultats. L'utilisateur peut répéter ce processus plusieurs fois jusqu'à ce qu'il soit satisfait. Les images positives sont les images que l'utilisateur

considère comme les images semblables à la requête et inversement pour les images négatives. D'une façon générale ce processus peut être classifié en trois classes.

#### **3.4.6.1. Bouclage de pertinence par modèle d'apprentissage**

L'une des méthodes les plus utilisées dans le bouclage de pertinence par apprentissage est la méthode des séparateurs à vaste marge (*Support Vector Machine*) SVM [Liua et al., 2008], cette méthode offre une grande flexibilité dans le traitement des données, dans la rapidité d'apprentissage et dans le test, d'un autre coté le SVM ne fournit pas beaucoup des résultats pertinentes à cause du nombre limité des exemples positifs et négatifs sur lesquels il fonctionne. De nombreux travaux ont essayé d'améliorer cette méthode, soit par hybridation avec d'autres méthodes remédiant ainsi le nombre réduit des exemples [Liua et al., 2008]. Nous citons aussi la méthode d'apprentissage par amplification *Boosting*, cette méthode suit une démarche spécifique où en accordant à un classifieur un taux d'erreur élevé pour chaque descripteur de l'ensemble des descripteurs de la collection traitée, par la suite, le classifieur qui minimise le taux d'erreur relatif est choisi avec une certaine pondération. Finalement, le classifieur adopté est une combinaison d'un ensemble de faibles classifieurs avec différents poids accordés à chacun [Huang et al., 2010].

Le problème dans cette méthode c'est quelle fonctionne mieux sur les descripteurs qui ne sont pas fortement corrélés avec la perception humaine *exemple la couleur* [Deselaers et al., 2004], le deuxième problème posé est celui de la taille de l'information à traiter qui doit être importante car le nombre des descripteurs est aussi élevé afin de garantir un bon résultat.

Actuellement, les chercheurs utilisent pour réaliser cet apprentissage les arbres de décision [Vannoorenberghe, 2004], les algorithmes génétiques [Francisco da Silva et al., 2011], les colonies de fourmis [Piatrik and Izquierdo, 2011] et les colonies d'abeilles [Zhang et al., 2010].

#### **3.4.6.2. Bouclage de pertinence par modèle vectoriel**

Le principe de ce modèle se base sur la mise-à-jour successives des données relatives à la représentation des images. Il existe deux variantes : la première se base sur la pondération des descripteurs et la deuxième sur la pondération des images elles-mêmes.

Pour la première, le système propose l'intervention de l'utilisateur pour spécifier les descripteurs utilisés lors de la recherche *couleur, texture ou forme*. Chaque critère est associé à une pondération, après une première recherche l'utilisateur peut définir un nouvel ordre de pondération des descripteurs soit par la désélection de certains descripteurs ou par l'attribution d'un nouveau poids au critère.

Cette technique est aussi appelée (*feature re-weighting*), elle consiste à un déplacement vers la zone des exemples positifs et un éloignement de la zone des exemples négatifs dans l'espace de descripteurs des images [Kim and Yu, 2008].

Pour la deuxième variante, on attribut un score aux images qui sont pertinentes pour l'utilisateur. Initialement, ce score est identique pour toutes les images, par la suite cette valeur est mise à jour pour traduire le degré de similarité avec l'image requête ou exprimer l'ordre d'apparence lors de la recherche. Une image est sélectionnée comme pertinente quand elle gagne un poids additionnel favorisant son degré de similarité avec d'autres images et permettant ainsi de retrouver de nouvelles correspondances [Gallas et al., 2010]. Alors que les images notées comme non-pertinentes seront écartées des recherches postérieures avec une minimisation de score ou même une annulation [Kherfi and Ziou, 2006].

#### **3.4.6.3. Bouclage de pertinence par modèle probabilistique**

Dans ce cas le bouclage de pertinence se base sur les probabilités conditionnelles où à chaque étape il faut tenir compte des actions faites pendant les étapes précédentes, ces informations acquises ne servent pas seulement à éliminer les fausses correspondances mais aussi à définir de nouvelles métriques de calcul de similarité qui correspondent aux attentes de l'utilisateur.

Parmi les méthodes qui existent nous citons [Kherfi and Ziou, 2006], ou les auteurs proposent un modèle bayésien qui soutient les classes d'images qui assignent une probabilité élevée aux images définies comme exemples positifs et pénalise les classes qui accordent une probabilité élevée d'appartenance aux exemples négatifs. [Cox et al., 2000] utilisent aussi le modèle bayésien dans lequel ils essayent d'estimer la distribution des images pertinentes tout en réduisant au minimum la probabilité de rechercher des images non-pertinentes. Dans [Su et al., 2003], les auteurs présentent une classe bayésienne pour les exemples positifs. Ceci sert à estimer la distribution Gaussienne qui représente la classe des images cherchées, alors que les exemples négatifs sont employés pour modifier le rang des candidats recherchés [Paredes et al., 2008].

#### **3.4.7. Quelques systèmes de recherche existants**

Dans cette partie certains systèmes de recherche d'images générales par le contenu sont présentés, dans ce qui suit nous nous intéressons aux systèmes de fouilles d'images médicales, domaines des deux premiers systèmes IGDS (*Image Guided Decision Support*) et ASSERT sont restreint, celui du dernier est classé comme large.

Pour une référence presque complète mais peu détaillée sur des travaux réalisés sur la recherche d'images par le contenu pour des applications médicales le lecteur peut consulter l'article de Müller [Müller et al., 2004].

- **Le QBIC**

Le QBIC (*Query by Image Content*) est l'un des premiers systèmes de recherche d'images par le contenu développé par IBM au début des années 90. Il utilise des caractéristiques de couleur pris sous forme d'un histogramme dans l'espace RGB, la texture granularité, contraste et orientation. La forme se base sur l'axe principal et l'excentricité calculée sur la matrice de covariance des pixels de contour. Ces caractéristiques extraites sont mise sous forme d'histogramme et deux mesures déterminant la pertinence sont aussi utilisées. Une distance Euclidienne pondérée entre les vecteurs moyens de couleurs pour sélectionner un nombre réduit d'images après une distance  $L_2$  pondérée est appliquée. Il permet des requêtes par images exemple, croquis et directes, un texte est associé aux images est seulement utilisé pour le filtrage des résultats et non pas pour une amélioration dans la sémantique des résultats (Comme titre d'exemple : seules les images associées à tel écrivain sont affichées). Il permet aussi la navigation avec bouclage de pertinence [Niblack et al., 1993].

- **SIMPLIcity**

SIMPLIcity (*Semantics-Sensitive Integrated Matching for Picture Libraries*), un système développé par James Wang [Wang et al., 2001]. Il vise à réduire le fossé sémantique et reprend toutes les avancées faites dans la recherche d'images par le contenu dans les années 90. Il se base sur une segmentation d'images en région, chaque image est découpée en blocs de quatre pixels, le vecteur représentant se base sur les coefficients d'une transformée en ondelettes.

La signature de l'image est un sac de régions, et chaque région est représentée par un vecteur de six dimensions. Les images de la base sont ensuite classées en catégories. Lors d'une recherche, si l'image requête appartient à la base alors le système va utiliser les mesures de la catégorie détectées et envoi le résultat dans un ordre décroissant, dans l'autre cas le système va segmenter l'image requête, classée dans la catégorie qui lui convient et termine le processus comme si l'image appartenait à la base.

- **RETIN**

RETIN (*Recherche et traque Interactive*), ce système de recherche est développé au niveau de laboratoire ETIS. Il utilise l'attribut couleur dans l'espace LAB et les sorties de 12 filtres de Gabor *quatre orientations et trois échelles* comme caractéristiques de la texture.

Une normalisation de la dimension est appliquée, ensuite le système utilise la méthode précoce pour concaténer les vecteurs de la texture et de la couleur. La requête lancée par l'utilisateur est une image exemple, dans le bouclage de pertinence le système propose plusieurs classifieur SVM, réseaux bayesiens, etc.

- **Système IGDS (Image Guided Decision Support)**

Comanicu et al ont développé IGDS qui donne l'exemple d'un système de recherche par le contenu pour l'aide de diagnostic. L'utilisateur dépose sa requête, le processus de recherche se lance et le résultat est un affichage des cas morphologiques conformes à l'image requête, et suggère le diagnostic le plus pertinent en se basant sur la majorité des cas trouvés dans les images similaires. La base contient (261) images des spécimens microscopiques de cellules de sang, organisée en quatre classes dont trois sont des classes anormales et une classe saine.

Les auteurs utilisent le descripteur invariant de Fourier pour décrire la forme, une fois le vecteur est créé, une dissimilarité entre deux vecteurs de forme est mesurée par la distance Euclidienne. Ils utilisent aussi la couleur pour séparer le noyau du fond de l'image.

Comme l'attribut de couleur n'avait pas un pouvoir discriminant pour distinguer les trois classes, les auteurs ont utilisé l'information couleur pour séparer le noyau du fond de l'image, la surface du noyau est prise proportionnelle au nombre de pixels à l'intérieur de la région du noyau, car les images dans la base de données sont toutes prises à la même échelle.

L'expert, ici le médecin choisit une zone rectangulaire qui contient la région d'intérêt, le système utilise une méthode de segmentation et par défaut le système retourne les huit plus proche voisines images de l'image requête [Comanicu et al., 1999].

- **Système ASSERT**

ASSERT<sup>1</sup>, est un système de recherche basé sur le contenu, créé par Shyu et al, il est caractérisé par l'intégration du médecin dans la phase d'indexation.

Ce système est créé pour sauver le radiologue de la consultation d'un atlas des pathologies de poumon, le système recherche une image similaire à celle du patient.

Pour chercher une image le médecin entoure une ou plusieurs régions pathologiques dans l'image en question et le système recherche les images les plus semblables utilisant des descripteurs locaux des PBR (*Pathology Bearing regions*). Ensuite, chaque image est représentée par un ensemble de régions appelées LFS (*Lobular Feature Set*). LFS correspond à un lobe ou une combinaison de lobes du Poumon, le contour des lobes est délimité par l'utilisateur du système et une distance euclidienne est utilisée [Shyu et al., 1999].

- **Système à base de connaissance pour la recherche d'images médicales**

Ce système développé par University of California, Los Angeles (UCLA), se base dans son fonctionnement sur quatre couches :

---

<sup>1</sup> ASSERT: Travail de collaboration entre le département de radiologie de l'université d'Indiana, de l'école de médecine à l'université de Wisconsin, et des laboratoires de (Machine Learning) et (Robot Vision) à l'école de génie informatique et électrique à l'université de Purdue.

RDL (*Raw Data Layer*), la couche établit les liens avec la base d'images brutes et le système puis convertit l'image brute à un format utilisable par le système. FCL (*Feature and Content Layer*), cette couche fait l'extraction et la mémorisation des descripteurs comme les contours et les régions. SL (*Schema Layer*), elle établit les relations entre les entités médicales (lésions, ventricule droit,...) et les objets de la couche FCL. (*Knowledge Layer*), cette couche correspond à une représentation sous forme hiérarchique. La requête est lancée à l'aide d'un langage de requête visuel, il peut aussi délimiter des zones d'intérêts et les étiqueter avec des mots clés ensuite cette requête est interprétée vers une requête sous forme de prédicats [Chu et al., 1998].

### 3.3. Classification

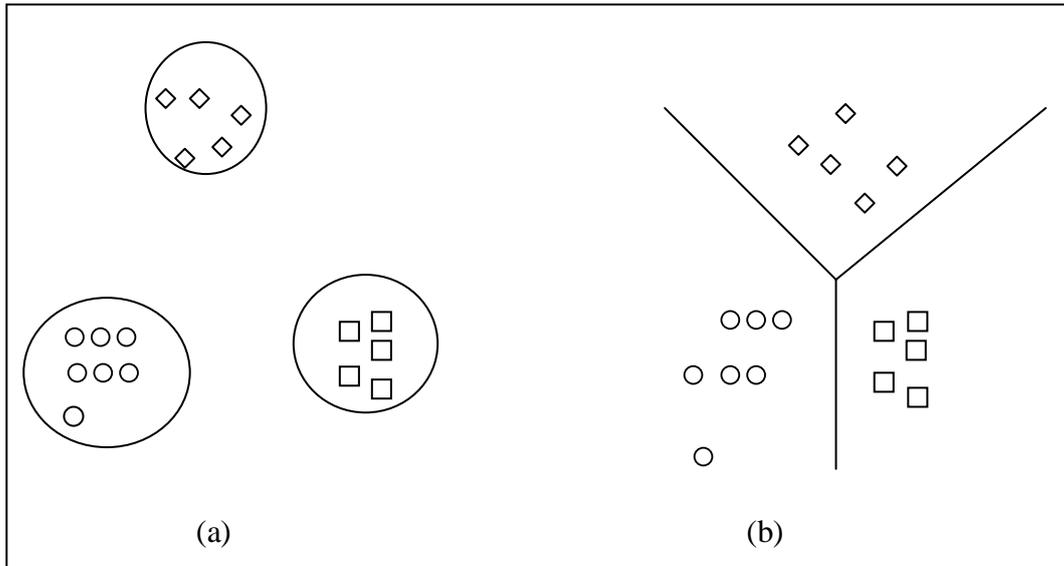
Il existe deux grandes catégories de classification, la classification supervisée et la classification non-supervisée.

#### 3.3.1. Methodes de classification supervisée

Selon [Govaert, 2003], on peut définir la classification supervisée par :

*« La classification supervisée (appelée aussi classement ou classification inductive) a pour objectif « d'apprendre » par l'exemple. Elle cherche à expliquer et à prédire l'appartenance des documents à des classes connues a priori. Ainsi c'est l'ensemble des techniques qui visent à deviner l'appartenance d'un individu à une classe en s'aidant uniquement des valeurs qu'il prend ».*

Si l'utilisateur a une connaissance suffisantes sur les données qui font la base de son étude *par exemple des images de type mammographie*, alors il peut effectuer une classification supervisée. Ce type suppose avoir un nombre d'images de chaque classe *le nombre des classes est fixe*, dont on connaît leur appartenance. On appelle cette phase *apprentissage*, elle est utilisée pour entraîner le classifieur, une deuxième phase qui s'appelle phase de *teste* sert à affecter de nouvelles images aux classes adéquates selon la modélisation en phase d'apprentissage. En générale il existe deux types de modélisation : la modélisation intrinsèque et la modélisation discriminante (Figure 3.7).



**Figure 3.7.** Types de modélisation : a) Intrinsèque et b) Discriminante

Plusieurs méthodes de classification supervisées dans la littérature se basent sur ces deux types de modélisations. Dans ce qui suit nous présentons quelques méthodes utilisées pour la classification des images.

Comme méthode intrinsèque, la classification bayésienne parfois appelé *Naïve* ou *simple* détermine la probabilité d'appartenance de chaque image aux différentes classes [Langley and Sage, 1999]. Dans le même contexte, les réseaux de neurones artificiels connus en anglais par ANN (*Artificial Neural Networks*) reposent sur la théorie des perceptrons [Maren, 1990], l'ANN est composé de plusieurs neurones répartis sur une couche d'entrée *descripteurs* et une couche de sortie *résultat de classification* et un nombre de couches cachées. L'inconvénient c'est qu'il n'existe pas de méthode pour résoudre le choix du nombre de couches du réseau. Pour le MLP (*Multi Layer Perceptron*) il est constitué par une ou plusieurs couches cachées [Freeman and Skapura, 1991]. Par contre les réseaux de neurones à base radiale RBF (*Radial Basis Functions*) sont constitués par une seule couche cachée [Oyang et al., 2005].

L'avantage de ce type de neurones est l'utilisation d'une seule couche cachée qui présente une structure moins complexe par rapport aux autres réseaux. De plus la complexité de calcul utilisé dans l'apprentissage est inférieure à celle des autres grâce aux algorithmes hybrides utilisés. Ainsi, la performance est bien liée aux nombres d'unités de la couche cachée et de l'estimation des paramètres du réseau créée. Généralement, les réseaux de neurones sont robustes aux erreurs et donc bien adaptés aux problèmes d'apprentissage d'images bruitées.

Une autre méthode de type discriminante, *l'analyse discriminante linéaire* qui se base sur la séparation entre les images qui appartiennent aux différentes classes en utilisant une analyse

linéaire. L'idée principale est de construire des limites de décision entre les classes qui minimise l'erreur, elle est performante dans le cas où les classes de la base sont linéairement séparables. Généralement elle ne permet pas des précisions congédiables [Kimura et al., 1997]. Les séparateurs à vaste marge, connus par SVMs (*Support Vector Machines*) sont orientés vers la classification binaires pour séparer linéairement les exemples positifs des exemples négatifs dans un ensemble d'images d'apprentissage, utilisant un hyper plan garantissant un maximum de marge [Vapnik, 1999].

Généralement l'efficacité de cette méthode est bien supérieure aux autres méthodes de classification supervisées, le plan peut effectuer une transformation des observations d'entrées non linéaire vers un cas linéaire, il s'adapte également pour la résolution des problèmes de la classification multi-classes [Weston and Watkins, 1999, Platt et al., 2000].

Face aux méthodes traditionnelles, il existe une méthode qui ne nécessite pas de modélisation (un modèle déduit des images d'apprentissage), c'est le cas du *raisonnement à partir d'exemples* comme la méthode Kppv *K-plus proches voisins* connue en anglais (*K-nearest neighbor*) [Weiss, 1991]. Dans cette méthode on attribue une image requête, en regardant quelle est la classe la plus proche dans son voisinage présenté par  $k$  dans la base d'apprentissage, il faut noter que la mesure de similarité et le  $k$  choisi joue un rôle important dans le déroulement de l'algorithme. Cependant, le grand problème c'est qu'il nécessite un espace mémoire important pour stocker toute la base d'apprentissage.

#### 3.3.2. Méthodes de classification non supervisées

Dans cette technique, les classes qui constituent la base ne sont pas identifiées. Ainsi, cette méthode appelée en anglais *clustering* consiste à déterminer les différentes classes sans aucune connaissance préalable, l'algorithme va créer des regroupements avec des caractéristiques similaires et identifie dans les images de la base en se basant sur leurs contenus. Les images attribuées aux différentes classes sont estimées selon une différenciation des images inter classes et la réduction dans la variation du contenu des images intra classes. Dans ce cadre, l'utilisation d'une distance reste une étape importante dans l'algorithme de classification, cette distance est spécifiée dans laquelle deux images sont considérées comme similaires, de ce fait certaines images vont être proches l'une de l'autre en fonction de cette distance et plus éloignées selon une autre [Dengsheng and Guojun, 2003].

Dans la littérature, il existe plusieurs approches pour la catégorisation de la méthode de classification non supervisée, ces deux catégories sont : les méthodes paramétriques et non paramétriques [Fukunaga, 1972].

Parmi les méthodes communément utilisées l'algorithme *K-moyenne*, en anglais connu sous le nom (*K-means*) [McQueen, 1967]. Dans cet algorithme, l'utilisateur attribue la valeur de *k* qui présente le nombre de classes, il calcule la distance dans la classe et refixe le centre selon les valeurs retenus. Les inconvénients majeurs de cette méthode c'est que le nombre de classes est fixé avant de commencer, et les classes suivent une loi de distribution normal réduite, c'est-à-dire la même importance dans toute les directions de l'espace ce qui n'est toujours vrai. Un algorithme itératif connu avec l'abréviation CAST (*Cluster Affinity Search Technique*) [Portnoy et al., 2002] à prouver sa puissance et son efficacité dans l'examen des données aberrantes [Shamir and Sharan., 2002].

Une autre structure incluse dans la classification non supervisée, c'est ce qu'on appelle *les méthodes de classification hiérarchique*, ce type génère les suites des partitions emboîtées, partant de la partition qui contient toutes les observations de début jusqu'à l'étape où chaque observation constitue une classe, elle contient cinq sous catégories : *Hiérarchique Ascendantes*, *Descendantes*, *Monothétiques* ou *Polythétiques*, *Déterministes* ou *Floues*, *Déterministes* ou *stochastiques* et *Incrémentales* ou *Non-Incrémentales* [Jain et al., 1999]. Il existe aussi *les méthodes par partitionnement* qui permettent d'organiser les images selon un paramètre *K*, et ça nécessite une connaissance du domaine, qui n'est pas toujours disponible [Ester et al., 1996]. EM (*Expectation- Maximisation*), présente aussi un modèle par partitionnement de Bayes. Dans la littérature, *Co-clustering* et *Bi-clusternig* présentent des algorithmes où non seulement les images sont classées mais aussi leurs caractéristiques [Kim and Cipolla, 2008].

La carte auto-organisatrice connue sous le nom anglais SOM (*Self Organizing Map*) présente une méthode non supervisée, cette carte est constituée d'un ensemble de réseaux de neurones et elle peut projeter un ensemble de données sur un espace à deux dimensions [Kohonen, 1984]. Généralement l'algorithme *K-means* reste le plus populaire mais la méthode SOM est plus robuste selon les conditions initiales des deux [Berry, 2003].

### 3.4. Evaluation des résultats

Il existe un ensemble de mesures pour une évaluation quantitative d'un système d'indexation, les plus simples pour *une évaluation globale* sont :

- **Le First Tier (FT)** : il est défini par la formule (3.27) suivante :

$$FT = \frac{N_{cor}(Q)}{Q} \quad (3.27)$$

Où  $N_{cor}(x)$  présente le nombre des résultats corrects parmi les  $x$  premiers résultats retrouvés,  $Q$  présente le nombre d'éléments de la catégorie qui correspond à la requête [Osada et al., 2001].

• **Le Score Bull-Eye (SBE)** : ou le second tiers, il est retenu pour l'évaluation de tous les descripteurs de forme dans GPEG-7, il est défini par :

$$SBE = \frac{N_{cor}(2Q)}{Q} \quad (3.28)$$

Où  $N_{cor}(x)$  présente le nombre des résultats corrects parmi les  $x$  premiers résultats retrouvés,  $Q$  présente le nombre d'éléments de la catégorie qui correspond à la requête.

Le SBE permet une marge d'erreur arrive à 50% dans le but de trouver l'ensemble des images recherchées, par contre le FT n'autorise aucune erreur, la marge d'erreur 0%.

Pour évaluer un système de recherche d'une manière plus fine *évaluation fine*, nous utilisons des mesures de la précision (*Precision*) et le rappel (*Recall*).

La précision  $p$ , cet indicateur précise combien l'ensemble des images résultats sont correctes (images pertinentes) sur l'ensemble des images résultats.

$$P = \frac{\text{Résultats pertinents}}{\text{Résultats}} \quad (3.29)$$

Le rappel, c'est la proportion d'images résultats trouvées par le système parmi l'ensemble d'images pertinentes contenues dans la base d'images. Ils sont donc complémentaires.

$$R = \frac{\text{Résultats pertinents}}{\text{images pertinentes}} \quad (3.30)$$

Il faut noter que la courbe de rappel, la courbe de précision et la courbe de rappel/précision donnent une représentation visuelle des résultats.

### 3.5. Conclusion

Dans ce chapitre nous avons détaillé les points clés du domaine d'indexation et recherche d'images par le contenu. Nous avons présenté quelques méthodes d'indexation. Comme la segmentation est une étape nécessaire dans un processus d'indexation et de recherche alors nous avons trouvé utile de présenter les approches basées contour, les approches basées région et les approches basées pixels, l'extraction des caractéristiques et les mesures de similarité. Nous avons aussi précisé qu'avec l'architecture du système et que n'importe quel système de recherches d'images par le contenu débute souvent par un type spécifique de requête ou différents types de requêtes sont présentés. Puisque les systèmes efficaces utilisent le bouclage de pertinence, alors celui est aussi détaillé. Le bouclage de pertinence permet d'obtenir des images et annotation associées, en utilisant différentes méthodes.

Ce chapitre prend fin avec une brève présentation des méthodes de classification qui existent, et nous donnons les formules les plus utilisées pour l'évaluation d'un système d'indexation et recherche.

## Chapitre IV

---

# L'image de Mammographie

---

### 4.1. Introduction

La détection précoce de cancer du sein est largement utilisée pour réduire la mortalité en permettant une intervention à un stade précoce de la progression de la maladie. Les données cliniques montrent que les femmes diagnostiquées avec un cancer du sein à un stade précoce sont moins susceptibles de mourir par rapport à celles qui sont diagnostiquées à un stade plus avancé. Un examen physique annuel et une auto-examinations mensuel approfondi du sein peuvent souvent détecter des tumeurs qui sont plus petites que celles trouvées avec d'autres examens, mais la performance de cet examen reste limitée.

En conséquence, avec ou sans examen médicale, l'approche de l'imagerie mammaire à montrer une détection précoce du cancer du sein et à réduire la mortalité de cette maladie.

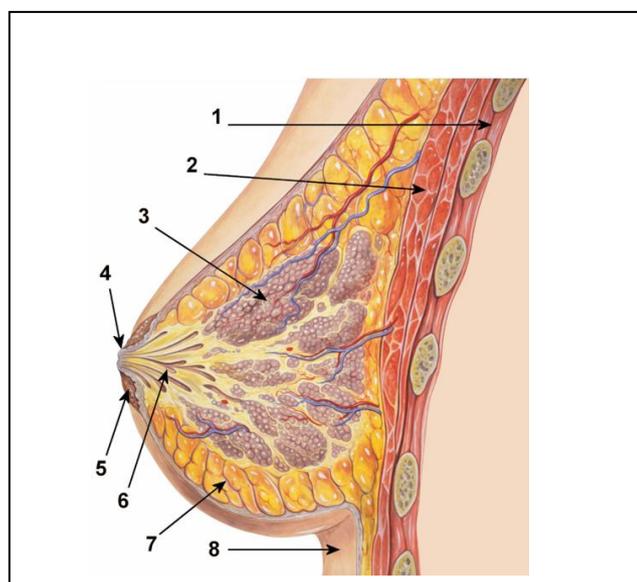
Cette approche englobe aussi l'expertise des radiologues afin de proposer des techniques performantes qui aide à dépister facilement le cancer du sein et de détecter les lésions mammaires avec une classification automatique [Feig et al., 1998, Smith, 2003, Belot et al., 2008].

Ce chapitre introduit une description de l'anatomie du sein. Ensuite, nous présentons quelques statistiques sur la mortalité et l'incidence dans les pays développés, les pays en voie de développement et en particulier le cas de l'Algérie. Nous détaillons aussi les pathologies mammaires ainsi que les classifications des lésions les plus utilisés par les radiologues. Enfin nous présentons quelques études sur la classification des images des seins normaux.

## 4.2. Anatomie du sein

Sur le plan interne, le sein est essentiellement constitué d'un tissu conjonctif adipeux graisseux qui donne sa forme et son volume. La glande mammaire est constituée d'une vingtaine de lobules, séparés par des cloisons conjonctives, qui sont des ensembles de bourgeons glandulaires, développés en période d'allaitement.

Le lait sécrété est déversé par des canaux galactophores séparés au niveau du mamelon. Le sein parcouru par une multitude de vaisseaux sanguins, est maintenu par la peau qui le recouvre et par des fibres liées au muscle pectoral. La figure (4.1) présente en détaille les composants du sein [Lynch, 2007].



**Figure 4.1.** Anatomie du sein :1) Cage thoracique, 2) Muscles pectoraux, 3) Lobules,4) Surface du mamelon, 5) Aréole, 6) Conduit galactophore, 7) Tissu adipeux, 8) Peau [Lynch, 2007]

## 4.3. Cancer du sein

Le cancer correspond à une multiplication incontrôlable de cellules anormales. Elles peuvent être situées dans n'importe quel organe et échappent à tout contrôle. Elles peuvent aussi migrer dans d'autres organes pour donner naissance à des métastases. Les mécanismes en cause sont très nombreux.

Le cancer se développe à partir d'une seule cellule anormale, qui se multiplie et prolifère de manière anarchique. Elles peuvent se situer dans n'importe quel organe et échappent à tout contrôle et même de migrer vers d'autres organes pour donner naissance à des métastases [Ferlay et al., 2013].

Le cancer du sein est la localisation cancéreuse la plus fréquente et la première cause de mortalité par cancer chez la femme dans le monde.

Depuis les dernières estimations de 2008, l'incidence a augmentée de plus de 20% et la mortalité de 14%. Le cancer du sein est la cause la plus fréquente de décès par cancer chez les femmes (522 000 décès) et le cancer le plus fréquemment diagnostiqué chez les femmes dans 140 des 184 pays couverts par GLOBOCAN<sup>1</sup> dans le monde. 1.7 millions de femmes ont un diagnostic de cancer du sein chaque année et en 2012, 6.3 millions de femmes vivaient avec un cancer du sein diagnostiqué au cours des cinq années précédentes, Il représente maintenant un cancer sur quatre chez les femmes [Ferlay et al., 2013].

Les tendances mondiales montrent plus généralement que dans les pays en développement en transition sociétale et économique rapide, la transition vers un monde de vie typique des pays industrialisés conduit à des cancers associés à des facteurs de risque génésiques, alimentaires et hormonaux [Bray et al., 2013].

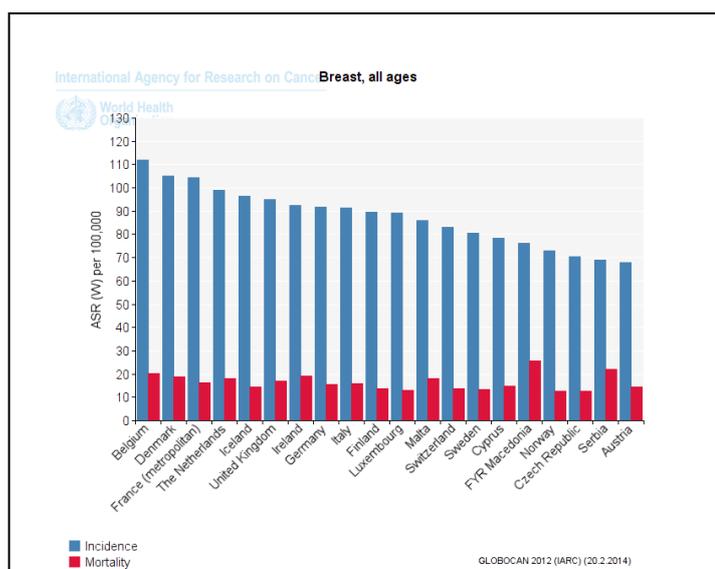
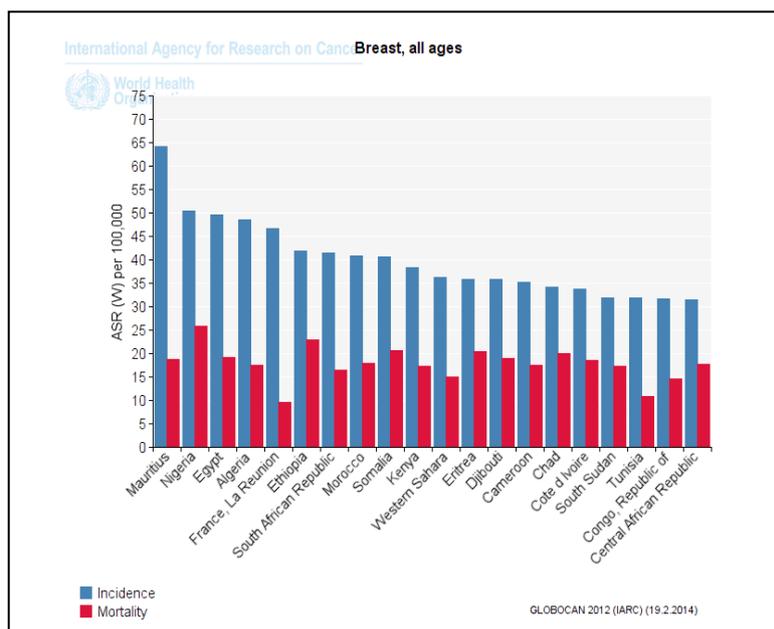


Figure 4.2. L'incidence et la mortalité dans les pays développés [Ferlay and al., 2013]

Bien que l'incidence soit en augmentation dans la plupart des régions du monde, il y a d'énormes inégalités entre les pays riches et les pays pauvres. Les taux d'incidence demeurent les plus élevés dans les régions les plus développées, mais la mortalité est beaucoup plus

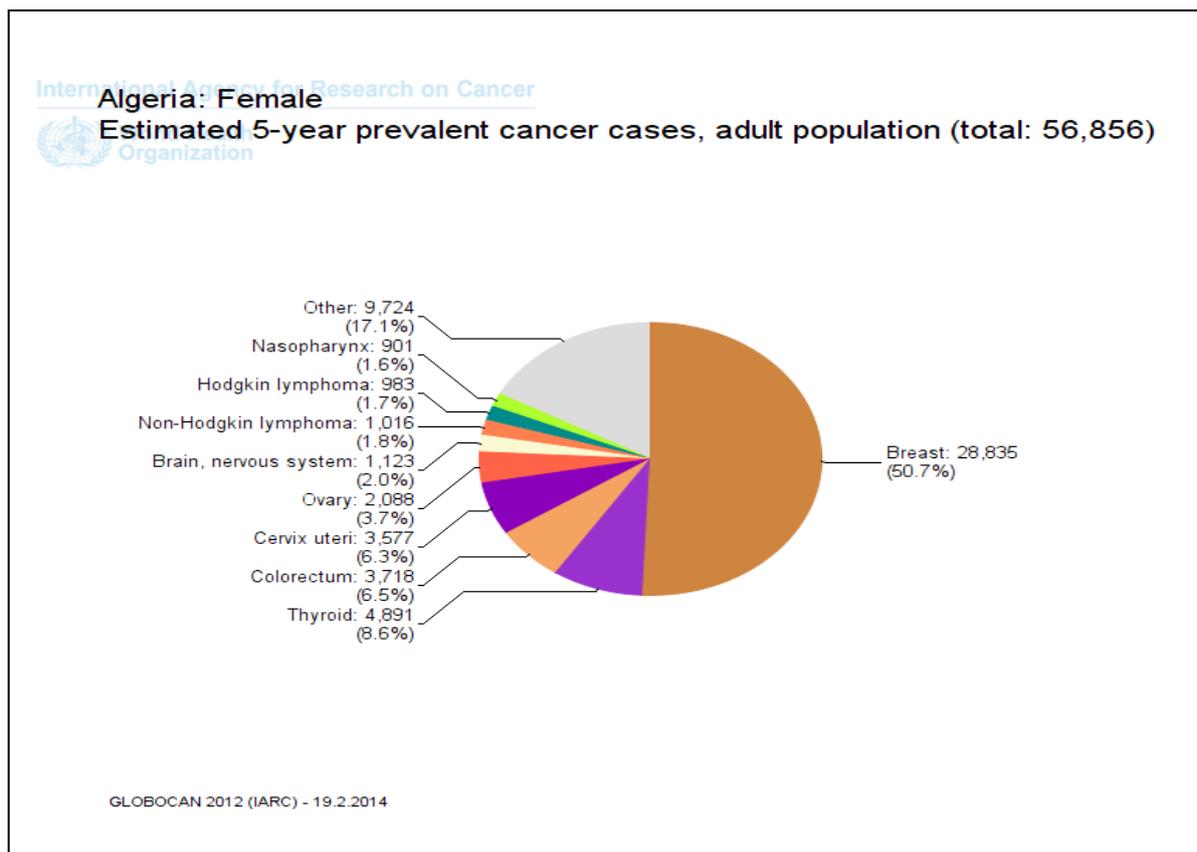
<sup>1</sup> GLOBOCAN project : L'objectif du projet est de fournir des estimations contemporaines de l'incidence de la mortalité et la prévalence des principaux types de cancer, au niveau national, pour 184 pays du monde.

élevée relativement dans les pays pauvres, faute de détection précoce et d'accès aux traitements [Bray et al., 2013].



**Figure 4.3.** L'incidence et la mortalité dans les pays en voie de développement [Ferlay and al., 2013]

Les taux d'incidence les plus élevés sont enregistrés dans les pays à haut niveau de vie, l'Algérie se situe parmi les pays du nord Afrique à plus forte incidence pour le cancer du sein et présente environ 50.70% du cancer comparativement avec 9 autres cancers les plus répandus, sur une période de cinq années.



**Figure 4.4.** Le pourcentage du cancer du sein par rapport à neuf types du cancer [Ferlay and al., 2013]

Alors le cancer du sein est la cause la plus fréquente de décès par cancer chez les femmes en Algérie.

Christopher Wild<sup>1</sup> le directeur du CIRC<sup>2</sup>, explique :

« *Il est aujourd'hui urgent, pour mieux lutter contre le cancer, de développer des approches efficaces et abordables pour la détection précoce, le diagnostic et le traitement du cancer du sein chez les femmes vivant dans les pays les moins développés du monde* » [Bray et al., 2013].

<sup>1</sup> Christopher Wild :Le Directeur est chargé de la direction générale du Centre : il donne le cadre de réalisation de la mission du Centre en conformité avec le Statut et les règlements du CIRC; il élabore une Stratégie à moyen terme et un Plan de mise en œuvre qui énoncent la vision globale, la direction et l'orientation du programme de recherche du Centre (actuellement 2010-2014).

<sup>2</sup> L'objectif du CIRC est de promouvoir la collaboration internationale dans la recherche sur le cancer. Le Centre est interdisciplinaire, et réunit des compétences dans les disciplines de laboratoire, en épidémiologie et en biostatistique pour identifier les causes du cancer, qui permettront d'adopter des mesures préventives afin de réduire le fardeau et les souffrances liés à la maladie.

#### 4.4. Facteur du risque

L'étiologie de la plus part du cancer de sein reste inconnu, ces facteurs peuvent être classés en deux grandes catégories :

##### 4.4.1. Facteurs endogènes

Cette maladie est considérée comme étant spécifiquement féminine avec un taux de 99% des cas.

- **L'âge** : le risque de développer cette pathologie augmente rapidement avec l'âge, il touche rarement les femmes de moins de 30 ans. Un taux de 85% survient à l'âge de 50 ans ou plus.
- **Hérédité et facteurs génétiques**: de 5 % à 10% des cancers du sein causés par une anomalie génétique transmise par l'hérédité (par exemple une forme altérée d'un gène). Ce risque concerne les femmes ayant un parent atteint (mère, sœur ou fille) cette anomalie.
- **Facteur hormonaux**: les hormones endogènes jouent un rôle important dans le développement de la glande mammaire, et de ce fait présentent un facteur de risque du cancer du sein. Ce risque se duplique chez la femme ménopausée, l'apparition précoce des règles et avec une ménopause tardive [Key and al., 2001, Gourinel and al., 1996].

##### 4.4.2. Facteurs exogènes

- **Grossesse** : Une grossesse tardive augmente le risque de cette maladie, selon des études une grossesse à un âge précoce va jouer le rôle d'un protecteur du sein avec une maturation de la glande mammaire et l'influence de la progestérone, également un nombre important de grossesse va réduire le risque [Gourinel and al., 1996].
- **Obésité et régime alimentaire** : l'obésité chez la femme après la ménopause duplique le risque du cancer du sein, cette hypothèse se base sur l'incidence du cancer du sein dans les pays comme le Japon et les pays sous- développés.

#### 4.5. L'imagerie médicale

L'imagerie médicale joue un rôle primordial dans le dépistage du cancer du sein. Il existe un certain nombre de techniques d'imagerie du corps humain, chacune d'elles est sensible à un type de contraste particulier et trouve ses applications pour des organes différents.

La mammographie est la technique la plus sensible pour la détection des lésions non palpables du sein, et donc la mammographie a une place importante dans le dépistage, bien qu'elle soit loin d'être parfaite, c'est pourquoi la mammographie peut réduire la mortalité par cancer du sein lorsqu'elle est combinée avec des interventions appropriées. Il existe aussi pour le dépistage, et pour le diagnostic du cancer du sein : autres types d'imagerie comme

l'échographie (*imagerie par ultrasons*), l'IRM (*Imagerie par Résonance Magnétique*) [Sharyl et al., 2001].

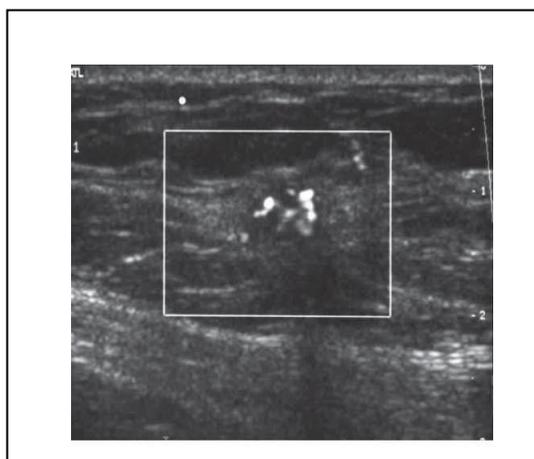
#### 4.5.1. L'échographie

Cette technique consiste à appliquer une sonde contre la peau explorant l'organe en question, la sonde émet des ultrasons qui traversent les tissus puis lui sont renvoyés sous la forme d'écho.

Un système informatique va recueillir, analyser le signal et le retransmettre directement comme une image sur un écran vidéo. Cette technique à un double avantage, d'une part elle n'est pas coûteuse, d'un autre côté elle ne présente aucun risque pour les patients.

Les inconvénients sont : présente un examen qui dépend de son examinateur et le positionnement de la sonde sur l'organe (possibilité de plusieurs interprétations pour le même patient), des fois il est difficile de s'assurer que le sein est diagnostiqué dans sa totalité ou non, en plus la détection des microcalcifications et les masses de petite taille est difficile. Elle est altérée par un bruit spécifique appelé la granularité qui est causé essentiellement par les interférences entre les ondes.

Généralement, il est conseillé d'utiliser l'échographie comme complément d'une mammographie pour distinguer entre les kystes et les lésions solides et une méthode pour la localisation de la tumeur avant une biopsie [Sharyl et al., 2001].



**Figure 4.5.**Un exemple d'une image du sein avec la technique d'échographie [Sharyl et al., 2001]

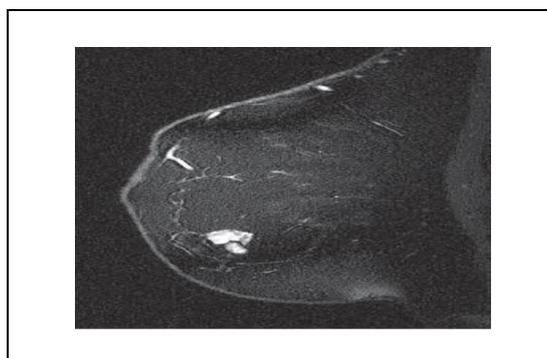
#### 4.5.2. L'imagerie par résonance magnétique

(IRM)(*Imagerie Résonance Magnétique*), en anglais connue par (*Magnetic Resonance Imaging*) est une technique d'imagerie médicale relativement récente (Les années 1980).

Des images de résonance magnétique sont créés en enregistrant les signaux générés après une excitation radiofréquence des noyaux d'hydrogène (ou d'autres éléments) dans le tissu est exposé à un champ magnétique statique.

Ces signaux ont des caractéristiques qui varient en fonction du type de tissu (par exemple, la graisse, le muscle, tissu fibrotique...).

Le but de l'IRM du sein est similaire à celle de la mammographie: pour identifier des anomalies structurelles dans le tissu. Elle est développée principalement comme un outil de diagnostic pour éviter des biopsies inutiles chez les femmes avec seins denses. La figure 4.6 présente une image IRM du sein.



**Figure 4.6.** Exemple d'une imagerie par résonance magnétique du sein [Sharyl et al., 2001]

L'IRM permet l'obtention d'images de grandes qualité en haute résolution qui permet une étude approfondie de la maladie, elle est coûteuse, destiner aux patients qui ne portent pas de prothèses métalliques. Ainsi, elle apporte beaucoup de précision à des anomalies déjà détectées par d'autres techniques [Sharyl et al., 2001].

En tenant compte de l'examen IRM qui coûte cher et la limitation de l'examen d'échographie, la technique de mammographie semble plus adaptée pour un dépistage ou un diagnostic du cancer du sein. Dans ce qui suit nous présentons cette technique.

### **4.5.3. La mammographie**

La mammographie est une technique utilisée pour obtenir des images (clichés) de l'intérieur des seins de la femme. C'est un examen radiologique qui utilise le principe des rayons X. La dose de rayons reçue de cette technique est très basse et sans danger.

L'appareil dédié à cet examen est le mammographe. Il existe deux types d'appareil, le mammographe dit analogique où les films mammographiques sont placés dans une cassette radiologique puis développés classiquement en chambre noire. L'autre type est dit numérique,

la réception des rayons X après la traversée du sein analysée soit sur plaques au phosphore lues par un système laser (numérisation indirecte) soit directement sur les capteurs plan (numérisation directe). Les images digitales obtenues transitent sur une console informatique avant leur reproduction sur film par procédé laser.

Cet examen n'est pas suffisant pour établir un diagnostic. Il permet de dépister des lésions suspectes qui seront ensuite investiguées pour poser un diagnostic, utilisée aussi pour le diagnostic et la localisation lors d'interventions chirurgicales (ponctions). Le grand avantage dans cet examen c'est qu'il permet d'examiner la totalité du sein avec une ou deux incidences seulement [Sharyl et al., 2001].

#### **4.6. Relation entre l'imagerie médicale et l'anatomie du sein**

La mammographie présente le résultat d'un faisceau de rayons X produit par un tube à rayons X est émis en direction de la zone du corps humain à examiner (tissus mammaire).

Les rayons vont rencontrer soit une composante adipeuse radiotransparente qui est la composante grasseuse, elle apparaît sombre sur le cliché à cause de sa densité, soit une composante qui est la fibroglandulaire radioopaque qui présente le composant d'une lésion mammaire. En conséquence, l'image de mammographie apparaît sombre alors que les composants d'une lésion (microcalcifications ou masses) sont plus clairs.

#### **4.7. Les pathologies mammaires**

On trouve les tumeurs et les calcifications, dans ce qui suit nous les détaillons.

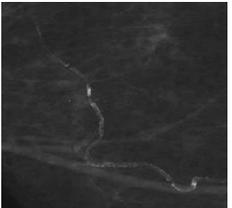
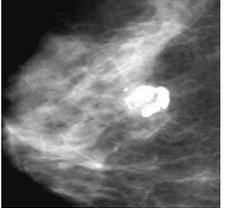
##### **4.7.1. Les microcalcifications (Mcs)**

Les microcalcifications mammaires sont des dépôts de calcium qui se forment dans le tissu du sein, il est composé des substances chimiques  $Ca_3(PO_4)_2$ ,  $CaCO_3$  et  $Mg_3(PO_4)_2$  qui se traduisent dans la mammographie par des points clairs. Elles se caractérisent par son contraste et la petite taille qu'est inférieur à 0.5 mm, si la taille dépasse 1mm alors on les appelle macrocalcification (elle est souvent bénignes).

Généralement elle n'a pas une taille minimale, et une confusion avec le bruit est probable se qui nécessite d'ajouter (pour éliminer le doute) d'autres critères comme la distribution dans le tissu, le nombre et la morphologie.

La morphologie des calcifications joue un rôle important dans la séparation entre les microcalcifications malignes qui sont irrégulières, hétérogènes et bénignes avec une forme et taille uniforme (ovales ou arrondies) [Lanyi, 1985, D'Orsi et al., 2003].

**Tableau 4.1.** Les différents types de calcifications [Frouge et al., 1994]

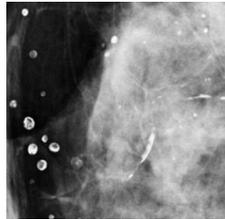
Types	Images	Explications
Calcifications cutanées (dermiques)		Dépôts calciques avec centre claire, pour les formes inhabituelles des clichés en incidence tangentielle permettent de confirmer la localisation cutanée de ces calcifications.
Calcifications vasculaires		Ces calcifications en rails ou linéaires clairement associées à des structures tubulaires.
Calcifications grossières ou coralliformes		Se sont des calcifications de grande taille (supérieures à 2/3 mm de diamètre) produites par un fibroadénome.
Calcifications en bâtonnets		Elles sont associées à une ectasie canalaire qui forment des bâtonnets à bords lisses, dirigées vers le mamelon avec une taille plus de 1 mm, elles sont souvent trouvez chez les patients qui dépassent 60 ans.

Calcifications  
rondes



Elles sont multiples, de taille variée, forme ronde et considérées bénignes si elles sont éparpillées. Lorsqu'elles sont inférieures à 0.5 mm, on utilise le terme punctiforme.

Calcifications à  
centres clairs



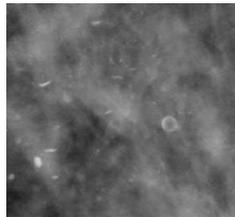
Elles sont rondes ou ovales avec une taille de 1 mm à plus de 1 cm. Une surface lisse et un centre clair avec une paroi plus épaisse que celle des calcifications en coquille d'œuf.

Calcifications en  
coquille d'œuf  
ou pariétales



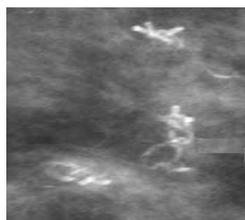
Se sont des calcifications très fines ayant l'apparence d'un dépôt calcique sur la surface d'une sphère, ces dépôts sont très fines, inférieur d'un 1 mm.

Calcifications à  
type de lait  
calciq



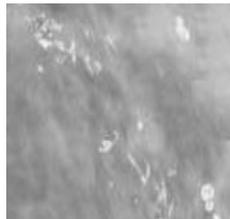
Elles constituent des dépôts amorphes avec des limites floues sur le cliché de mammographie. Elles sont souvent difficiles à discerner mais l'incidence de profil permet de démontrer leurs formes caractéristiques: semi lunaires, en croissants, curvilignes ou linéaires.

Calcifications de  
suture

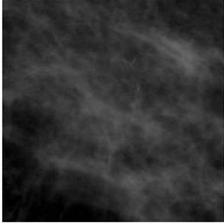
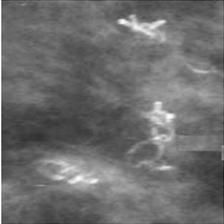
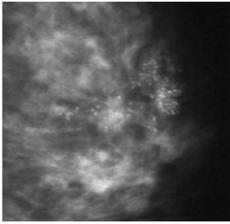


Présentent des dépôts calciques sur du matériel de suture. Sont typiquement linéaires suivant un trajet des fils de suture.

Calcifications  
dystrophiques

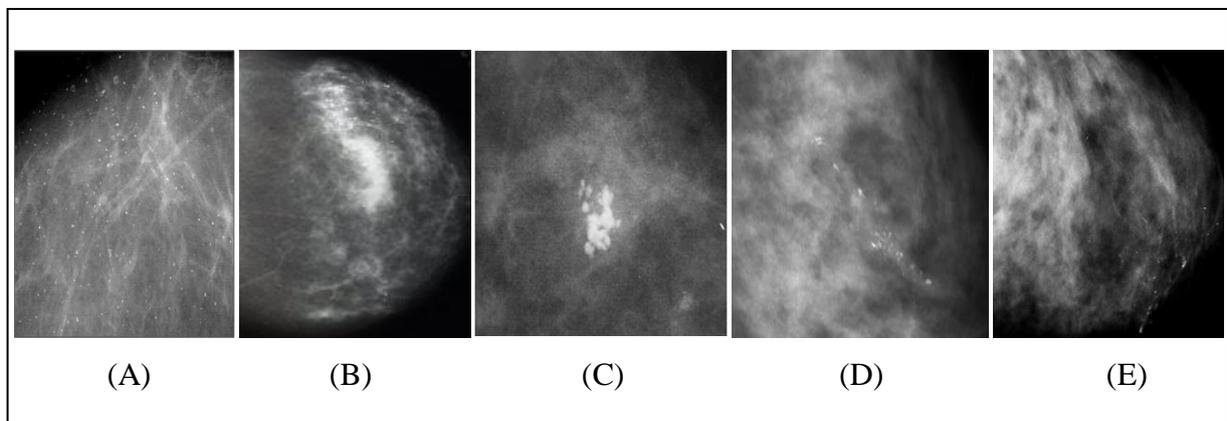


La forme est irrégulière, grossière et supramillimétrique (supérieure à 0.5 mm de diamètre). Elles représentent la majorité des cas retrouvés en pathologie mammaire.

Calcifications amorphes ou indistinctes		Elles sont souvent très fines et insuffisamment nettes pour les déterminer. Elles sont bénignes quand elles sont éparses.
Calcification grossières hétérogènes		La forme et la taille est variable, généralement supérieure à 0.5 mm. Elle se défère aux calcifications dystrophiques par sa petite taille, elles sont souvent bénignes.
Calcifications fines et polymorphes		Elles sont mieux visibles avec des contours bien distincts. De taille et de forme irrégulières généralement moins de 0.5 mm de diamètre.
Calcifications fines linéaires et ramifiées		Elles prennent une forme fine linéaire ou curviligne avec une taille moins de 0.5 mm d'épaisseur.

#### 4.7.1.1. La distribution

La distribution des calcifications présente un critère important pour une décision de la malignité. On peut distinguer comme type :



**Figure 4.7.** Les distributions des calcifications : A) calcification diffusées ou éparses, B) régionale, C) groupées, D) Linéaire et E) Segmentaire [Frouge et al., 1994]

**A.** Calcifications diffuses/éparses : les calcifications sont distribuées aléatoirement dans l'ensemble du sein et atteignent tous le quadrant mammaire.

**B.** Distribution régionale : les calcifications sont dispersées dans un large volume du tissu mammaire et ne présentent pas une distribution canalaire.

**C.** Calcifications groupées, en amas ou en foyer : correspond aux calcifications formant un groupe (au moins cinq calcification) occupent un petit volume tissulaire.

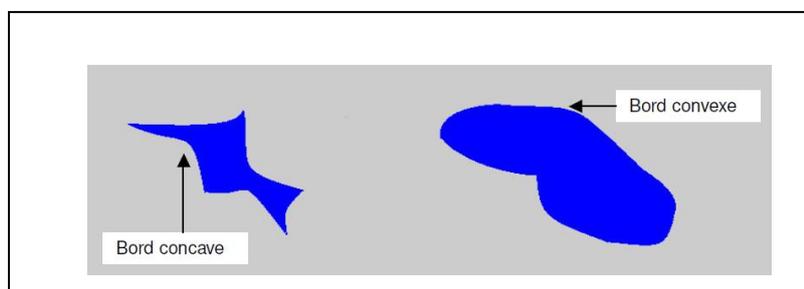
**D.** Distribution linéaire : les calcifications de disposition linéaire généralement se sont des dépôts calciques dans un galactophore.

**E.** Distribution segmentaire : elle suggère des dépôts calciques dans des canaux galactophores à une disposition triangulaire dirigée vers le mamelon.

#### 4.8. Les masses

Une masse ou une opacité est une lésion occupant un espace, vue sur deux incidences différentes.

Lorsque la lésion est visible sur deux incidences on parle d'opacité, l'expression asymétrie de densité est généralement utilisée jusqu'à confirmation du caractère tridimensionnel de la lésion, les limites d'une asymétrie de densité est décrites par concaves ou convexe [D'Orsi et al., 2003].



**Figure 4.8.** Asymétrie de densité (concave, convexe) [Murgo et al]

Les caractéristiques de ces masses sont décrites selon : la forme, le contour et la densité (cf. sec.4.9).

##### 4.8.1. La forme

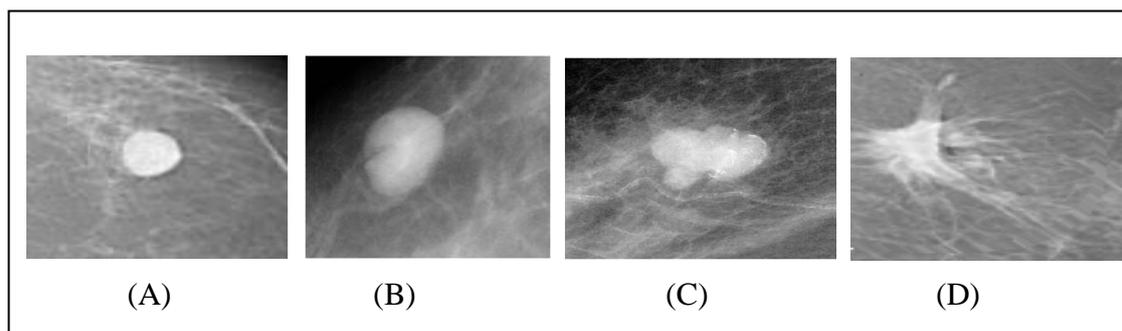
Selon BIRADS [D'Orsi et al., 2003], les masses peuvent prendre les formes suivante :

**A.** Ronde ou sphérique : la forme circulaire ou globuleuse.

**B.** Ovale : elle prend la forme d'une ellipse.

**C.** Lobulée : La forme présente une légère ondulation.

**D.** Irrégulière : caractérise les masses dont la forme est aléatoire.



**Figure 4.9.** Les formes d'une masse : A) Ronde, B) Ovale, C) Lobulée D) Irrégulière [Frouge et al., 1994]

#### 4.8.2. Le contour

Les masses peuvent prendre :

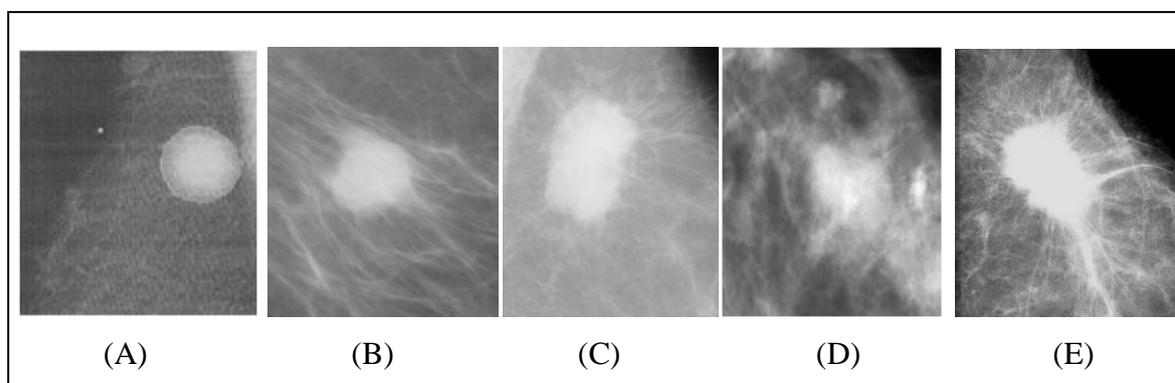
**A. Contour circonscrit :** bords nets et bien défini, la transition est brusque entre la lésion et le tissu environnant.

**B. Contour masqué :** bords à priori masqués par une superposition ou caché par le tissu normal adjacent. Ce terme est utilisé aussi pour caractériser une masse circonscrite dont une partie du contour est cachée.

**C. Contour micro-lobulé :** de courtes dentelures du contour créent de multiples ondulations (>3).

**D. Contour indistinct :** c'est un contour mal défini ou flou, peut correspondre à une infiltration.

**E. Contour spiculé :** elle est caractérisée par des lignes radiaires prenant naissance sur le contour [Lieberman and Menell., 2002].



**Figure 4.10.** Les contours d'une masse : A) Circonscrit, B) masqué, C) micro-lobulé, D) indistinct, E) spiculé [Frouge et al., 1994]

## 4.9. Classification des anomalies de mammographie

La mammographie a ses limites et ne permet en aucun cas à elle seule d'établir un diagnostic. Une double lecture, voire quelque fois une triple lecture lorsque les deux premiers radiologues ne portent pas des conclusions concordantes.

Le résultat de mammographie est classé selon la classification des radiologues américains (*Birads ACR*), car il est important de construire un lexique standard pour une description claire et précise des lésions mammaires.

Les études morphologique des lésions a fait l'objet de plusieurs classifications : [LeGal et al., 1984, Lanyi, 1985, D'Orsi et al., 2003] Les classifications les plus connues sont celles de Le Gal et de D'orsi and al.

### 4.9.1. La classification de Legal

En 1976, Le GAL avec sa classification à décrit cinq types morphologique qui ont une valeur prédictive de malignité croissante [LeGal et al., 1984] (tableau 4.2).

**Tableau 4.2.** Classification de Le Gal

<b>Types</b>	<b>Description</b>
<b>Type 1</b>	Msc annulaires, arciformes ou polyédriques. Risque de cancer du sein quasi nul.
<b>Type 2</b>	Mcs rondes et de tailles variables. Risque de carcinome : 22%
<b>Type 3</b>	Mcs poussiéreuses, pulvérulentes. Risque de cancer : 36%
<b>Type 4</b>	Mcs irrégulières associées à un risque de cancer : 56%
<b>Type 5</b>	Mcs vermiculaires ou branchées. Risque de carcinome : 90%

Le grand avantage dans cette classification est la simplicité, toutefois le défaut majeur c'est qu'elle se base sur les microcalcification seulement et elle n'intègre pas d'autre types ou

paramètres comme par exemple : l'étude morphologique des masses ou la disposition des microcalcifications dans le tissu mammaire.

#### 4.9.2. Classification de Birads

Pour palier les erreurs de la première classification, en 1986 l'ACR (*l'American college of Radiology*) a mis sur pied tout un programme de mammographie standardisé assure que les unités de mammographie répondent à des normes de qualité et que le personnel concerné a suivi une formation adéquate avec contrôles de qualité réguliers.

Le système *BI-RADS (Breast Imaging Reporting and Data System)* est le produit d'un comité d'experts réunis par l'ANAES, l'objectif principale est triple :

- A. standardiser les comptes-rendus de mammographie ;
- B. faciliter la comparaison entre les examens successifs, venant ou non de sources différentes;
- C. recueillir les données pour suivre les pathologies, évaluer les pratiques et les résultats du dépistage à plus grande échelle.

Dans ce qui suit, ce système comporte six catégories avec des anomalies très précisément décrites :

- A. **ACR0** : la mammographie est incomplète et des examens complémentaires sont nécessaires: agrandissement de la mammographie sur une zone donnée, ou échographie complémentaire, incidence particulière;
- B. **ACR1** : mammographie normale sans aucune anomalie même bénigne, surveillance standard tous les deux ans;
- C. **ACR2** : mammographie bénigne avec une ou des images typiquement classées bénignes, surveillance standard;
- D. **ACR3** : mammographie comportant une anomalie probablement bénigne, un contrôle à 4 ou 6 mois est alors recommandé, une biopsie est demandée dans des cas très particuliers (risque important de cancer du sein en raison d'antécédents particuliers);
- E. **ACR4** : mammographie comportant une lésion douteuse. Une biopsie est nécessaire. Cette lésion est maligne dans environ 40% des cas;
- F. **ACR5** : la mammographie comporte des anomalies très suspectes, l'anomalie est évocatrice d'un cancer [Bokobsa et al., 2005].

#### 4.10. Classifications des seins normaux

La densité mammaire a une grande importance dans l'interprétation des mammographies, un sein dense est un sein qui apparaît blanc dans un examen de mammographie car son contenu graisseux est peu important.

Par rapport aux autres organes, le sein présente une particularité, son développement lié aux changements hormonaux et son aspect varie tout au long de vie.

Pour les femmes jeunes le sein est plus dense si l'en compare avec les femmes âgées, cette densité se diminue dans le temps [Feig, 2000]. La plupart des cancers du sein présentent une densité supérieure ou égale au tissu conjonctivo-glandulaire. Il est rare (mais non impossible) qu'un cancer du sein soit de densité inférieure.

Wolf est le premier qui a établie la relation qui existe entre l'aspect du sein et le développement d'un cancer du sein [Wolfe, 1976]. D'autres travaux mettent ça en doute, une corrélation nette entre la densité mammaire et le risque de cancer apparaît dans le travail de [Heine and Malhotra, 2002].

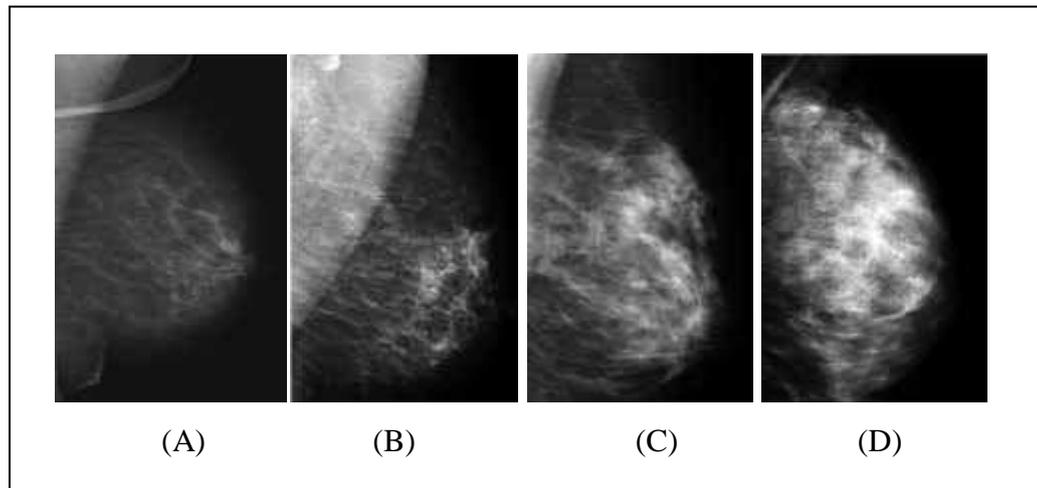
Donc, il est difficile de définir avec précision à partir de quand un sein est dense (normal) ou non, la classification de Wolfe a défini 4 classes de la composition du sein : (N1, P1, P2, DY) :

- A.** les seins (N1) sont presque entièrement graisseux ;
- B.** les seins (P1), contient des travées denses galactophoriques, elles occupent moins de 25 % de la surface du sein ;
- C.** les éléments opaques occupent plus de 25% de la surface, les seins (P2) à un risque élevé de cancer.
- D.** Les seins (DY) ont le risque le plus élevé avec une densité supérieure à celle de la graisse et ne comporte pas d'éléments canaux proéminents [Wolf, 1967a, Wolf, 1967b].

Des études à montrer que la classification de Wolfe est bien corrélée avec des classifications qui tiennent compte de la densité seulement [Heine and Malhotra, 2002].

La classification BI-RADS permet de classer les seins en quatre catégories :

- A.** (type1) le sein est presque complètement graisseux ;
- B.** (type 2) il y a des opacités fibroglandulaires éparses ;
- C.** (type 3) le tissu glandulaire est dense et hétérogène et cela peut diminuer la sensibilité de la mammographie ;
- D.** (type 4) le tissu mammaire est extrêmement dense ce qui peut masquer une lésion à la mammographie.



**Figure 4.11.** Densité mammaire, le lexique BI-RADS de l'American College of Radiology A) type 1, B) type 2, C) type 3, D) type 4 [Stinès, 2004]

Selon des études la classification de BI-RADS des seins en fonction de leur apparence sur la mammographie est satisfaisante [Leconte et al., 2003].

Tabar a proposé une classification en cinq catégories selon l'architecture tridimensionnelle de la glande mammaire [Tabar and Dean, 1985], Boyd a proposé une classification en six groupes selon la quantité de graisse [Boyd et al., 1998]. Des auteurs ont essayé d'améliorer les choses comme Harvey, il a démontré que le risque du cancer augmente selon la densité du tissu mammaire [Harvey and Bovbjerg, 2004]. La numérisation des films en utilisant la texture [Byng et al., 1997, Magnin et al., 1986] ou d'utiliser des mammographies d'emblée numériques [Heine and Malhotra, 2002].

Plus le critère de la densité du tissu mammaire, Hagay and Cherel présentent les principaux critères de bénignité et de malignité des opacités, le tableau (4.3) illustre cette étude [Hagay and Cherel, 1998].

**Tableau 4.3.** Critères de malignité et bénignité [Hagay and Cherel, 1998]

Critères de malignité	Critère de bénignité
Age > 50 ans	Age < 35 ans
Apparition récente	Stabilité
Augmentation de taille	Multiplicité
Haute densité	Faible densité et lacunes graisseuses
Micro calcifications	Macro calcifications
Contours partiellement effacés (flous ou irréguliers)	Contours nets
Spicules ou convergences	

Pour des besoins scientifiques, les expertes ont gommé les défauts de la classification BI-RADS par :

Une lecture multiple, l'utilisation d'un système d'aide informatisé, utiliser l'IRM pour quantifier quelques éléments du sein.

En conséquence, il est très difficile d'évaluer de la densité du tissu mammaire dans l'absence d'une standardisation

#### **4.11. Conclusion**

Dans ce chapitre, nous avons présenté l'anatomie du sein, la notion du cancer du sein ainsi que les outils d'imagerie utilisés pour le diagnostic.

Nous nous sommes intéressés aux pathologies mammaires à savoir les calcifications et les masses. En suite, nous avons présenté la classification de Le Gal, la classification de BIRADS qui est la plus adaptée par les experts pour classer les lésions mammaires.

Nous avons vu et étudié l'importance de la densité dans la classification des images normales. Dans ce chapitre nous déduisons qu'il faut tenir compte non seulement de la densité mammaire mais également de tous les autres facteurs de risques et de la probabilité d'avoir un cancer ou non.

# **Deuxième partie**

## **Travail réalisé**

<b>Chapitre V : Extraction des Caractéristiques.....</b>	<b>104</b>
<b>Chapitre VI : Classification Basée AC.....</b>	<b>125</b>
<b>Chapitre VII : Indexation et Recherche des Images par le Contenu Basée Agents.....</b>	<b>152</b>

# Chapitre V

---

## Extraction des Caractéristiques

---

### 5.1. Introduction

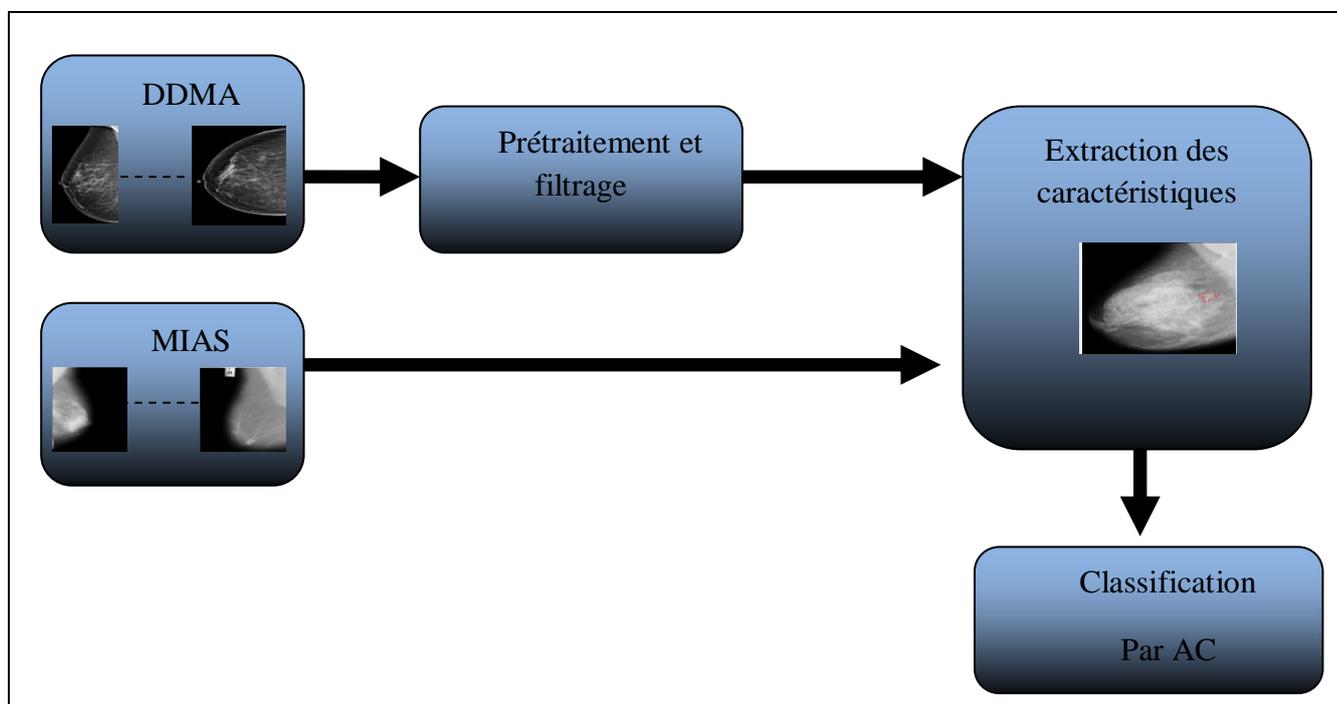
Dans ce chapitre nous présentons l'architecture d'un système de classification des images, cette classification est non supervisée, elle se base sur un système évolutionnaire Automate cellulaire (AC).

La première section est consacrée à une présentation de cette architecture, ensuite, nous détaillons la partie prétraitement et filtrage. La section qui suit explique la méthode d'extraction des caractéristiques se basant sur une étape de segmentation. Une méthode de segmentation basée agents et AC est développée, vue les résultats obtenus, l'avis du médecin et la complexité des images mammographies un algorithme *Scalable Fitting Energy* est utilisé.

Enfin, nous détaillons quelques descripteurs géométriques, spécifiques et les moments de Hu qui sont destinés pour décrire la forme ; les descripteurs Haralick pour la représentation de la texture.

## 5.2. Architecture du système

L'architecture suivante résume le fonctionnement du système.



**Figure 5.1.** Schéma du système de classification d'images de mammographie

Les étapes du système présenté en figure (5.1) seront détaillées dans les sous-sections suivantes :

### 5.2.1. Acquisition des images

Le choix d'une base sur laquelle repose les évaluations du travail est très important.

#### • Problèmes d'acquisition des images

- Images collectées sont des images sans interprétations ni diagnostics.
- Les images sont sous un format DICOM ( *Digital Imaging and Communication in Medicine*), donc il faut extraire les images de ce format et les coder afin de garder l'anonymat des patients.
- Le nombre des images reste insuffisant devant des bases de mammographies telle que MIAS.
- Les images brutes ne contenant aucune information sur les axes de la tumeur, ou de nous ce qu'on appelle (*truth*) qui présentent des marques sur l'emplacement des anomalies.
- La qualité des images prises par le médecin ou par les techniciens est parfois de mauvaises qualités.
- Le médecin n'enregistre ou n'archive pas les images de mammographie qui sont dans le dossier du malade ce qui implique la réduction du nombre d'images obtenues.

-La base de mammographies de la ville d'Annaba est un produit qui est encore en évolution et que nous devons l'enrichir et l'améliorer pour pouvoir avoir une base propre à la ville d'Annaba.

De cela :

-l'évaluation du système demande une base d'images avec une vérité-terrain.

-Notre base d'images est traitée comme un ensemble de deux classes

-La construction d'une telle base médicale est problématique car demande beaucoup de temps et de travail pour archiver et coder les images au fur et à mesure de leur disponibilité.

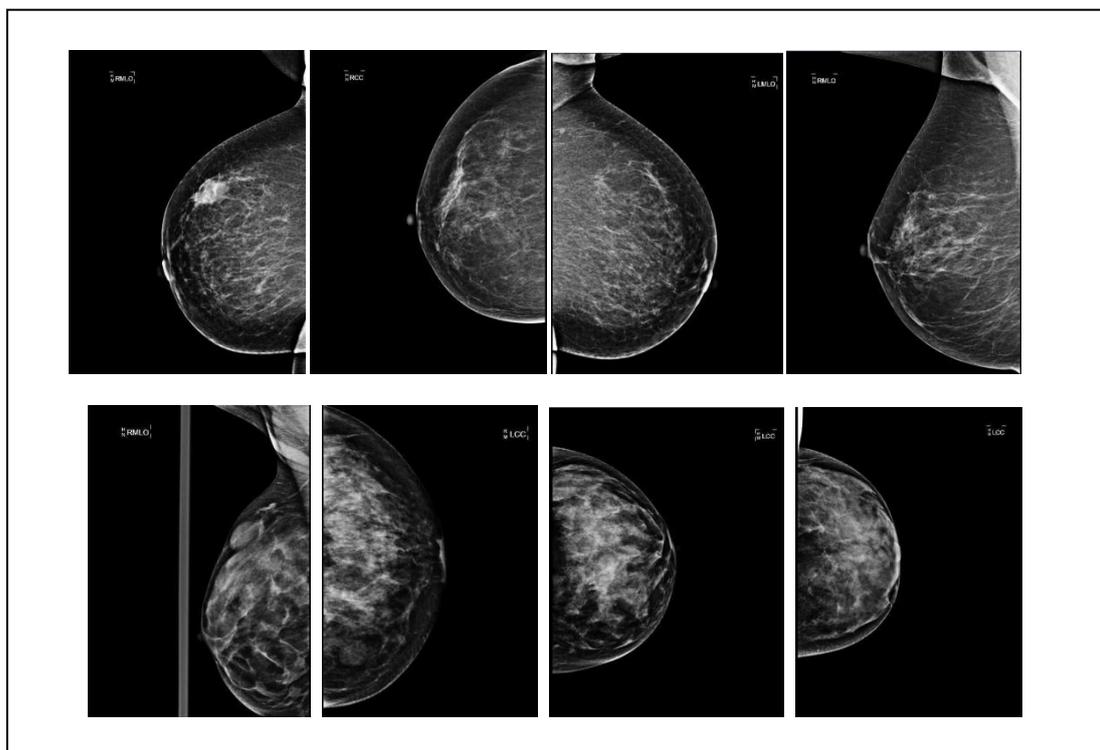
-Elle demande aussi la participation d'au moins un médecin spécialiste pour étiqueter ou diagnostiquer les images archivées

-Un échantillon de la base d'images obtenu dans ces conditions pour accomplir notre travail dans les délais prévus est illustré dans la (figure 5.2).

- **Base de données de mammographies**

- DDMA (Digital Database for Mammography of Annaba)**

Cette base est construite à partir d'une collaboration établie avec le service de radiologie Ibn Rochd d'Annaba.



**Figure 5.2.** Echantillons de la base DDM d'Annaba utilisés pour l'évaluation. La première ligne contient des images de mammographie à masse malignes et la deuxième ligne contient des masses bénignes

### 5.2.2. Prétraitement et filtrage

Comme nous l'avons déjà vu dans le chapitre IV précédant, le cancer du sein comme tous les cancers d'une façon générale, doit être détecté rapidement dans sa première phase pour maximiser les chances de survie.

Pour améliorer la qualité visuelle de l'image, on doit éliminer les effets des bruits en lui faisant subir un traitement appelé filtrage.

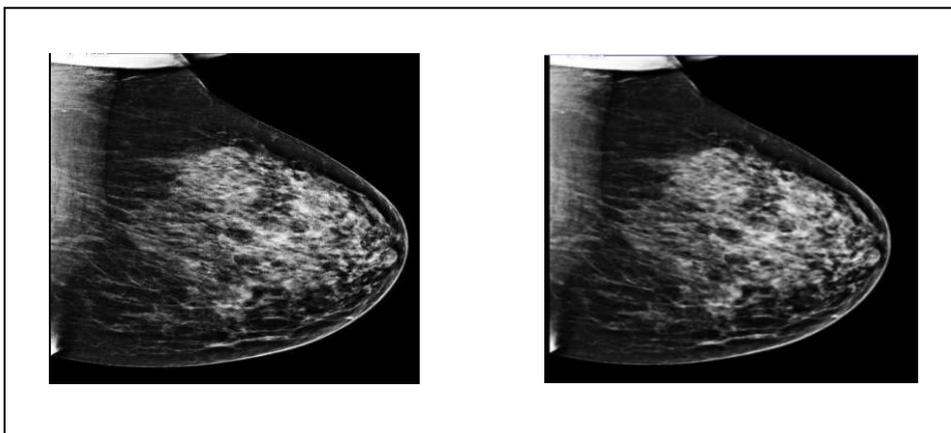
Il est nécessaire de faire un prétraitement qui présente une étape cruciale avant de passer à n'importe quelle étape du système, l'objectif principal est d'augmenter le contraste entre l'opacité et le reste de l'image pour faciliter le travail qui suit.

Les opérations de filtrage peuvent être utilisées pour supprimer sélectivement le bruit de l'image, pour détecter et améliorer les bords, ou pour modifier le contraste de l'image. Les filtres utilisent les variations statistiques locales dans une image pour révéler les éléments de la texture dans l'image.

Du fait du caractère non stationnaire du processus d'acquisition d'image, la qualité de cette dernière est généralement de faible contraste et pour cela nous avons besoin de filtrer l'image.

Pour résoudre ce problème nous proposons ici :

Le contraste des images médicales soit très faible, nous appliquons le filtre médian pour éliminer le bruit et d'améliorer les faibles limites d'images médicales ainsi ce filtre respecte les contours. Pour la suppression des bruits, le filtre médian, appliqué sur un voisinage de 3x3 permet d'éliminer les valeurs aberrantes de l'image (figure 5.3). Son fonctionnement consiste à remplacer la valeur d'un pixel par la valeur médiane de l'ensemble des pixels de son voisinage.



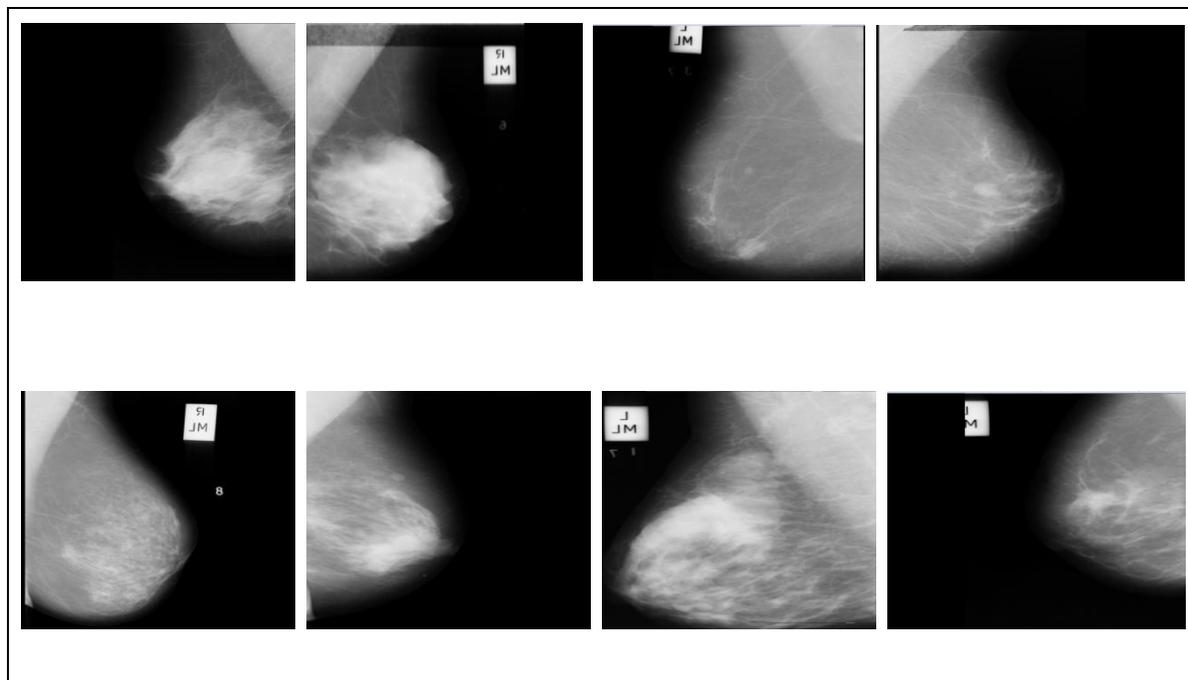
**Figure 5.3.** Application du filtre Médian

Pour une analyse comparative de l'algorithme, il est recommandé d'utiliser une base de données de test standard. La plupart des bases de données de mammographie ne sont pas accessibles au public. Parmi les bases des images les plus facilement accessibles et utilisées par le comité scientifique nous avons MIAS Mini Mammographic Database (i.e. mini-MIAS database of mammograms). Par ailleurs, il existe actuellement peu de projets de développement de nouvelles bases de données d'images de mammographie.

#### -MIAS

La société d'analyse des images de mammographie (MIAS) est une organisation de groupe de recherche du Royaume-uni qui s'intéresse à la compréhension et la génération des bases des images de mammographie numérique.

Les Films prises par (*UK National Breast Screening Programme*) ont été numérisés jusqu'à 50  $\mu\text{m}$  par pixel sur 8 bits. La base des images contient 322 films numérisés et sont disponibles sur 2,3 Go 8mm ( Exabyte ), elle comprend également (*truth*), qui représentent les marques de radiologue sur les emplacements de toutes les anomalies qui peuvent être présents et la résolution spatiales de toutes les images est de 1024x1024 pixels. Les 322 images sont réparties en 202 images normales et 120 images de type malin et bénin.



**Figure 5.4.** Echantillons de la base MIAS utilisés pour l'évaluation. La première ligne contient des images de mammographie à masse bénignes et la deuxième ligne contient des masses malignes

### 5.2.3. La segmentation

Dans un premier temps nous avons utilisé la méthode proposée par [Benmazou et al., 2012], cette méthode se base sur un critère d'homogénéité, un ensemble d'agents est utilisé dans le but de marquer les pixels qui répond au critère utilisé. La durée de vie de tels agents se base sur le principe d'un automate jeu de la vie [Benmazou et al., 2010, Benmazou et al., 2011a, Benmazou et al., 2011b, Benmazou et al., 2011c, Benmazou et al., 2011d].

Cette méthode est basée sur les systèmes multi-agents et l'automate cellulaire jeux de la vie. Au départ un nombre d'agents fixé est déployé aléatoirement sur l'image, ces entités ont une habilité d'estimer l'homogénéité de la région.

Avant de présenter le comportement des agents nous définissons l'environnement qui est une image (2D) avec :

$$A = (I_x, I_y)$$

Avec  $I_x$  et  $I_y$  représentant l'intensité relative à un pixel à  $(x, y)$ .

Mathématiquement nous avons choisie la distribution de densité pour définir un segment homogène.

$$D_{(i,j)_{r\acute{e}gion}} = \sum_{P(i,J,K,l)} P(i,j,k,l) \quad \|(i,j) - (k,l)\| \leq R_{(i,j)_{r\acute{e}gion}}$$

Avec :

(5.1)

$$P(i,j,k,l) = \begin{cases} 1 & \text{Si } \|I(i,j) - I(k,l)\| \leq \delta \\ 0 & \text{autrement} \end{cases}$$

Tel que :

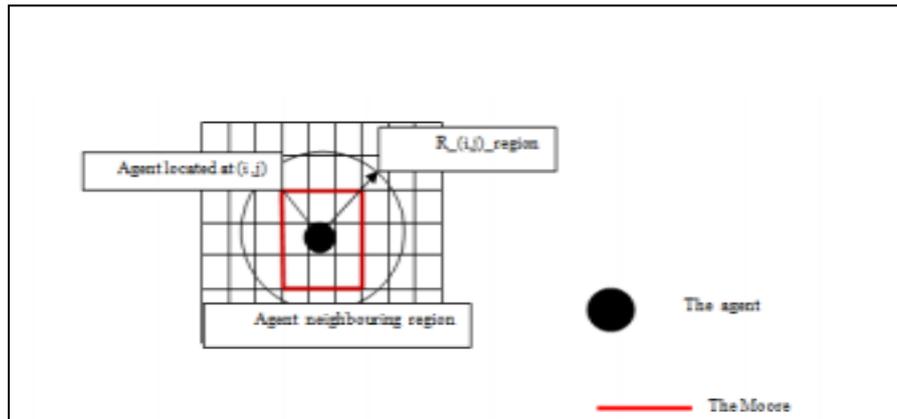
$(i, j)$ : Coordonnées du pixel sur le quel se trouve l'agent ;

$(K, l)$ : Coordonnées du pixel voisin de l'agent ;

$I(i, j)$ : La valeur d'intensité en niveau de gris du pixel  $(i, j)$  ;

$\delta$ : Seuil du contraste, il désigne la différence d'intensité admise entre le pixel  $(i, j)$  et les pixels voisins ;

$R_{(i,j)_{r\acute{e}gion}}$  : Le rayon de la région voisine fixé à 1 (utilisation du voisinage de Moore).



**Figure 5.5.** Voisinage d'un agent à une position (i,j)

Les agents déployés aléatoirement (figure 5.5) sur l'image se déplacent dans différentes directions dans le but de détecter le segment homogène selon le critère cité dans la formule (5.1). La diffusion se déclenche dès qu'un pixel ne répond pas au critère d'homogénéité dans le but d'exploiter au mieux l'image. Les agents ont un paramètre spécifique qui est défini par :

$$D_{(i,j) \text{ région}} \in n_1, n_2$$

$n_1$  et  $n_2$  sont prédéfinis, constants et reliés aux caractéristiques de l'image à segmenter.

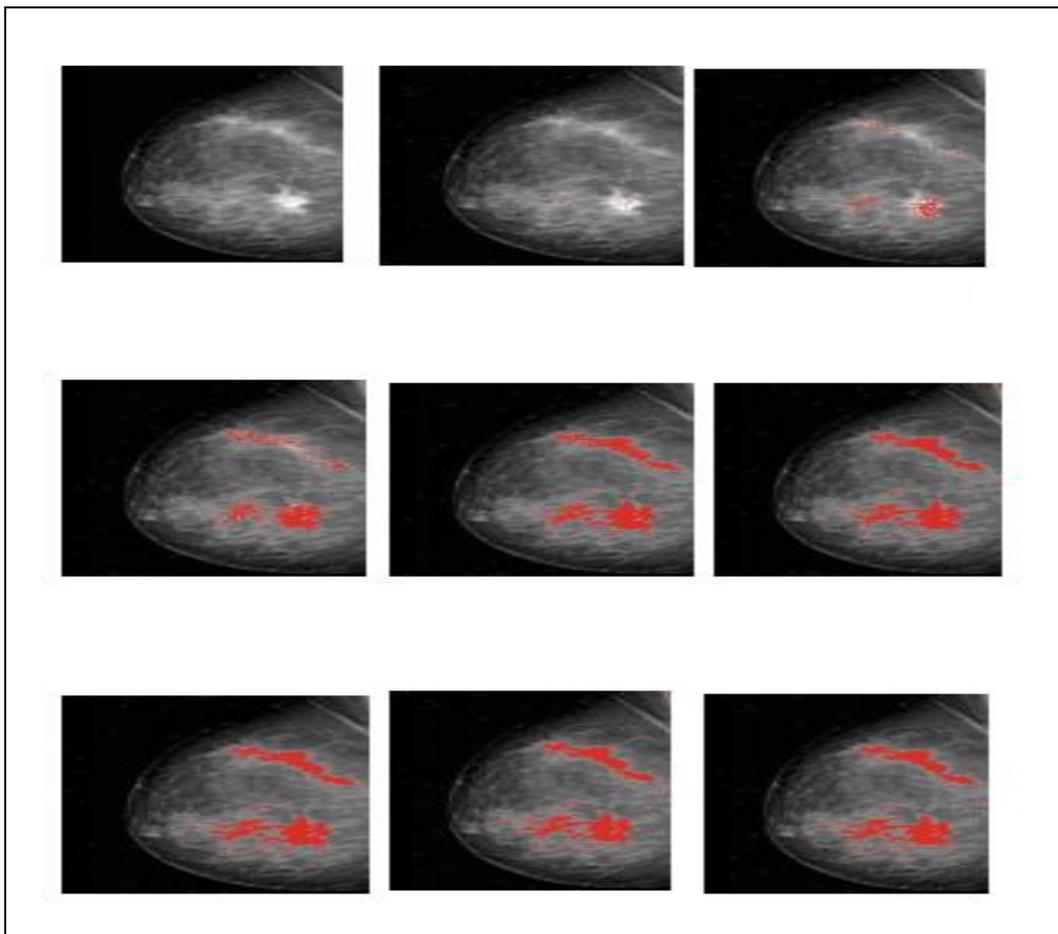
L'automate jeu de la vie est appelé pour explorer l'image rapidement, cet automate est défini par :

- Le domaine qui est composé d'un ensemble des cellules, chaque cellule présente le pixel de l'image à segmenté ;
- Le voisinage qui présente la relation avec un certain nombre de cellules (le voisinage de Moore est utilisé)
- L'ensemble des états de l'automate cellulaire est défini en fonction de l'application, dans notre cas les états sont : *life* et *dead*.

Règles de transition définissent le caractère dynamique de l'automate cellulaire et qui assure son évolution :

- Sur une cellule, si un agent a une couleur *white* entouré par 2 ou 3 agents *white* alors cet agent devient *gray* et prend l'état *dead* dans l'état qui suit.
- Si la cellule est vide est entourée par 3 agents qui ont une couleur *white* alors un nouveau agent est créé.

-Conditions aux bords, dans cette étude nous attribuons à l'espace une configuration torique, c'est-à-dire que l'on fait rejoindre les bords extrêmes de l'espace.



**Figure 5.6.** L'évolution des agents pour la recherche et l'étiquetage d'une région homogène

Nous avons discuté les résultats obtenus par cette méthode (figure 5.6) avec un spécialiste du domaine, l'utilisation de cette méthode rend le processus de segmentation un peu statique en plus la méthode segmente non seulement la région en question (la masse) mais aussi d'autres régions qui ont le même critère d'homogénéité de la masse.

Vue cette discussion, et le temps pour achever le travail, nous nous sommes orienté vers l'utilisation d'un autre algorithme.

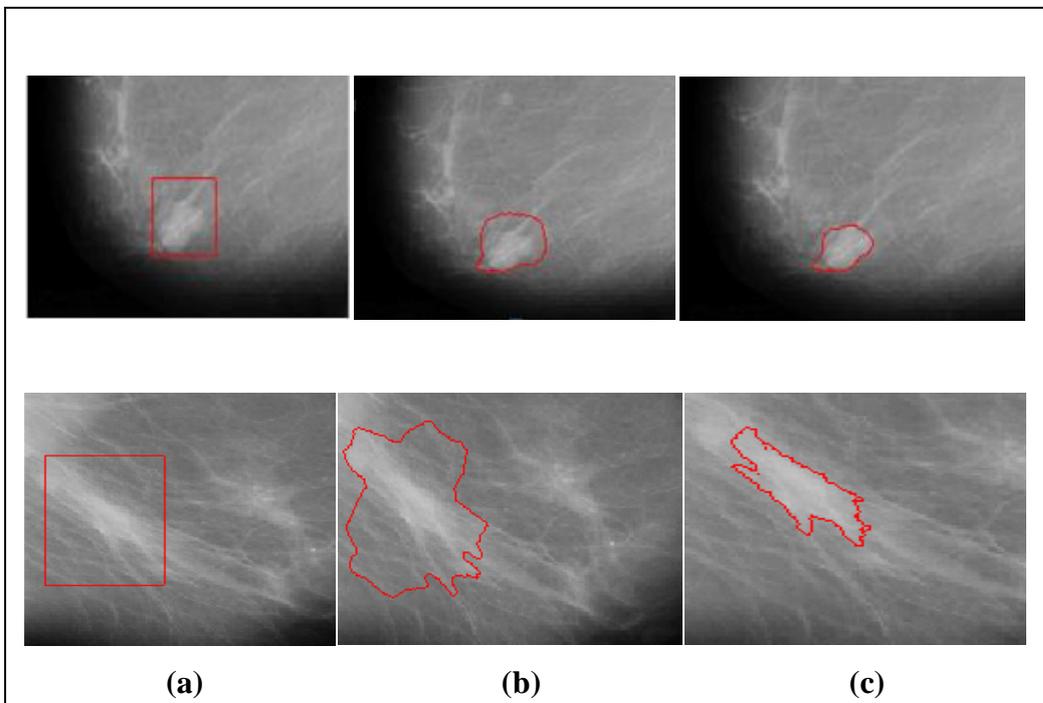
L'inhomogénéité de l'intensité pose un grand problème dans la segmentation des images d'une façon générale, pour les images médicales cette inhomogénéité présente le résultat des limitations techniques ou d'artefact introduit par l'objet étudiée, c'est le cas des images de mammographies.

[Li et al., 2008] ont proposé le modèle contour actif basé région (*Scalable Fitting Energy*) les résultats obtenus sont bonnes surtout pour le cas des objets mal définis ou masqués (les masses mammaires).

On peut présenter le modèle utilisé par ces points :

- La définition d'une région d'ajustement évolutif (RSF) d'énergie.
- La fonction d'énergie qui se base sur le contour et sur deux fonctions d'ajustement aidant à définir d'une façon approximative les intensités de l'image sur les deux bords du contour.
- L'ajustement de la région RSF (*Region Scalable Fitting Energy*) est dû la fonction Kernel.
- Un paramètre contrôlable de la fonction Kernel permet d'utiliser l'information de l'intensité dans les régions de l'image, allant d'un voisinage restreint jusqu'à couvrir tout le domaine.
- Le terme (*Level set regularization*) est introduit pour une précision des calcules et pour stabiliser l'évolution de *level set*. Ainsi, pour éviter les procédures coûteuses de réinitialisation de la fonction *level set*.

Dans notre travail, nous avons adopté cette méthode, la (figure 5.7) présente le résultat de la segmentation d'une masse circonscrite (la première ligne) et une masse spiculée (deuxième ligne). Les figures (5.7.a) présentent la superposition du contour initial sur une région d'intérêt. Les figures (5.7.b) présentent une propagation du contour sur la région d'intérêt et les figures 5.7.c) présentent la convergence vers le contour de la masse.



**Figure 5.7.** Segmentation d'une masse circonscrite et une masse spiculée utilisant l'algorithme proposé par (Li et al., 2008), a) Initialisation, b) Propagation et c) Convergence vers la limite de la masse

### 5.2.4. Extraction des caractéristiques

La comparaison directe des images entre elles n'est pas envisageable, il est donc nécessaire d'en extraire au préalable des informations représentatives :

Après une étape de segmentation, on utilise les moments de Hu, les descripteurs géométriques et spécifiques, dans ce qui suit nous nous intéressons à la texture, aux statistiques du deuxième ordre qui donnent une description de la structure locale dans l'image, la texture décrit habituellement ce qui est laissé après élimination de l'information de couleur et de la forme.

#### 5.2.4.1. Les Moments de Hu

Hu a défini sept moments qui caractérisent une image.

$$\begin{aligned}
 \varphi_1 &= \mu_{20} + \mu_{02} \\
 \varphi_2 &= (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2 \\
 \varphi_3 &= (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2 \\
 \varphi_4 &= (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \\
 \varphi_5 &= (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \\
 \varphi_6 &= (\mu_{20} - \mu_{02})[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03}) \\
 \varphi_7 &= (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] + (3\mu_{21} - \mu_{03})(\mu_{12} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{12} + \mu_{03})^2]
 \end{aligned} \tag{5.2}$$

Ces moments permettent de décrire la forme à l'aide de propriétés statistiques, dans notre travail, ces invariants sont extraits à partir d'images de mammographie. Le vecteur  $V$  est alors égale à  $(\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6, \varphi_7)$  [Benmazou et al., 2014 b].

#### 5.2.4.2. Les descripteurs géométriques

##### -L'aire

On notera par  $A$  l'aire de la masse, c'est le nombre de pixels qui compose une région *lésion mammaire*. L'aire est le descripteur de la forme le plus utilisé [Bottigli et al., 2006].

##### -Le périmètre

Le périmètre de la masse est noté par  $p$ , il présente le nombre des pixels du contour, ce descripteur est bien utilisé dans le domaine de diagnostic des lésions mammaires.

##### -La rectangularité

Ce descripteur décrit le taux de rectangularité et d'allongement d'une région (figure 5.8). Une boîte minimale  $B_E$  contient l'objet,  $R$  représente la région étudiée. La rectangularité peut être obtenue par la formule suivante :

$$\text{rectangularité} = \frac{\text{aire}(R)}{\text{aire}(B_E)} \tag{5.3}$$

La valeur obtenue est proche de 1 si l'objet décrit est rectangulaire, ce type ne tient pas compte de l'orientation de l'objet, il est sensible à la rotation [Zhang and Lu, 2004, Cheikhrouhou et al., 2007].

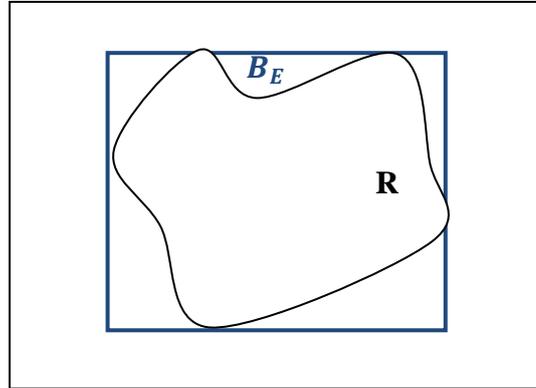


Figure 5.8. Calcul de la mesure de rectangularité

#### -La rectangularité modifiée

La mesure de rectangularité n'est pas invariante par rotation. Pour arriver à remédier à ce problème on considère une boîte englobante minimale selon la direction de l'objet, on définit et on calcule les moments centrés de l'objet d'ordre inférieur ou égale à 2.

Ces moments  $\mu_{02}, \mu_{20}$  et  $\mu_{11}$ ; sont alors calculés à partir du centre de gravité de l'objet défini par les coordonnées  $(x_i, y_i)$ .

$$\mu_{pq} = \sum_{i=0}^n \sum_{j=0}^m (i - x_g)^p (j - y_g)^q \quad (5.4)$$

Avec  $(p, q) = \{0, 1 \text{ ou } 2\}$ , et  $(n, m)$  sont les dimensions de la matrice.

Donc l'ellipse à les mêmes moments que l'objet définie par: Le grand axe  $\alpha_1$ , le petit axe  $\alpha_2$  et l'angle de rotation de l'objet par rapport à l'horizontal  $\alpha$ . Ces différents paramètres sont calculés de la façon suivante:

$$\alpha_1^2 = \frac{2(\mu_{02} + \mu_{20} + \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2})}{m_{00}} \quad (5.5)$$

$$\alpha_2^2 = \frac{2(\mu_{02} + \mu_{20} - \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2})}{m_{00}} \quad (5.6)$$

Avec  $m_{00}$  le moment d'ordre zéro qui représente l'aire de l'objet considéré.

$$t = \frac{\mu_{02} + \mu_{20} - \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{2\mu_{11}} \quad (5.7)$$

$$\alpha = \arctan(t) \quad (5.8)$$

**-La compacité**

C'est une mesure utilisée pour la complexité du contour (forme irrégulière ou simple), elle est calculée par :

$$com = \frac{P^2}{A} \quad (5.9)$$

Avec : P le périmètre de l'objet, A l'aire en pixels. Com prend une valeur élevée dans le cas d'une forme irrégulière, ce type est invariant à la translation, la rotation et le changement d'échelle [Peura and Iivarinen, 1997].

**-L'excentricité**

Elle représente le rapport entre la longueur du grand axe et celle du petit axe. L'excentricité peut être calculée par deux méthodes :

Les axes principaux de la forme sont définis par *deux segments orthogonaux* qui s'intersectent au niveau du centre de gravité de la forme et déterminés à partir de la matrice de covariance [Peura and Iivarinen, 1997]:

$$Cov = \frac{1}{N} \sum_{i=0}^{N-1} \begin{pmatrix} x_i & x_g \\ y_i & y_g \end{pmatrix} \begin{pmatrix} x_i & x_g \\ y_i & y_g \end{pmatrix}^T = \begin{pmatrix} Cov_{xx} & Cov_{xy} \\ Cov_{xy} & Cov_{yy} \end{pmatrix} \quad (5.10)$$

Avec :

$$Cov_{xx} = \frac{1}{N} \sum_{i=0}^N (x_i - x_g)^2 \quad (5.11)$$

$$Cov_{xy} = \frac{1}{N} \sum_{i=0}^N (x_i - x_g)(y_i - y_g) \quad (5.12)$$

$$Cov_{yx} = \frac{1}{N} \sum_{i=0}^N (y_i - y_g)(x_i - x_g) \quad (5.13)$$

$$Cov_{yy} = \frac{1}{N} \sum_{i=0}^N (y_i - y_g)^2 \quad (5.14)$$

Avec :  $(x_g, y_g)$  les coordonnées du centre de gravité de la forme. Le rectangle minimum englobant une forme, appelée aussi la boîte minimale englobante, est, en effet, le plus petit rectangle contenant tous les points de cette forme, la formule (5.15) présente l'excentricité qui

est le rapport entre la largeur  $l$  et la longueur  $L$  de la boîte minimale contenant la forme citée dans [Cheikhrouhou, 2012].

$$E_x = \frac{l}{L} \quad (5.15)$$

### 5.2.4.3. Les descripteurs spécifiques

L'utilisation individuelle de ce type de descripteurs présenté si dessus est insuffisants ce qui nécessite l'utilisation des descripteurs spécifiques surtout pour décrire une forme assez complexe comme les masses malignes de formes spéculées.

Les spécialistes de la reconnaissance de forme se sont penchés vers les descripteurs spécifiques qui sont d'après eux robustes et plus informatifs sur les détails de la forme que les descripteurs géométriques.

#### -La longueur radiale normalisée

Les six descripteurs nommés longueur radiale normalisée connu en anglais par LRN (*Normalized Radial Length*) ont été développés par Kilday et al, ils sont calculés à partir du centre du gravité de l'objet et les points du contour *ième pixel du contour* avec une distance Euclidienne [Kilday et al., 1993], donc c'est le résultat d'une normalisation par le maximum de la distance trouvée selon la formule (5.16) suivante :

$$d(i) = \frac{\sqrt{x(x(i)-x_g)^2 + (y(i)-y_g)^2}}{\max(d(i))}, i = \{1, 2, \dots, N\} \quad (5.16)$$

Avec :  $(x(i), y(i))$  et  $(x_g, y_g)$  les coordonnées du *ième pixel* et du centre de gravité respectivement,  $N$  est le périmètre de la masse.

-La moyenne de la longueur radiale normalisée  $d_{avg}$  : elle représente la façon dont le contour varie d'une manière macroscopique similaire à circularité, sa formule est :

$$d_{avg} = \frac{1}{N} \sum_{i=1}^N d(i) \quad (5.17)$$

-La déviation standard de la longueur radial normalisée ( $\sigma$ ) : c'est un testeur d'irrégularité du contour, la valeur est élevé tant que le contour est irrégulier, elle mesuré avec :

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (d(i) - d_{avg})^2} \quad (5.18)$$

**-L'entropie (E) :** cette mesure est calculée à partir de l'histogramme de la longueur radiale, elle intègre à la fois la circularité et l'irrégularité du contour.

$$E = \sum_{k=1}^{100} p_k \log(p_k) \quad (5.19)$$

Avec :  $p_k$  est la probabilité que la LRN soit entre  $d(i)$  et  $d(i) + 1/N_{bins}$ ,  $N_{bins}$  présente le nombre de bins de l'histogramme normalisé qui varié entre 0 et 1 qui a été divisé en  $N_{bins} = 100$ . Le rapport de surface ( $A_1$ ) : c'est le pourcentage de l'objet hors la région circulaire définie par la moyenne du LRN.

$$A_1 = \frac{1}{d_{avg} \cdot N} \sum_{i=1}^N (d(i) - d_{avg}) \quad (5.20)$$

Où :  $d(i) - d_{avg} = 0, \forall d(i) \leq d_{avg}$

**-La rugosité:** son objectif est d'isolé la forme de l'objet à partir de la structure du contour, le contour irrégulier fournisse une valeur élevé.

$$R = \frac{1}{N} \sum_{i=1}^N (d(i) - d(i + 1)) \quad (5.21)$$

**-Le taux de croisement en zéro ( $ZC_1$ ) :** ce taux est utilisé pour extraire les informations qui concernent les variations du contour (degré de spiculation). Il calcule le nombre de fois que la droite définie par la moyenne de LRN intercepte le contour de l'objet.

Dans le domaine du diagnostic assisté par ordinateur, ces descripteurs ont donné des résultats satisfaisants avec des bords ronds, pour les contours de bords complexes ces descripteurs sont moins appropriés. Les mêmes descripteurs ont été modifié, nommés la longueur radiale normalisée modifiée (LRNM) et ont donné des résultats satisfaisants dans la classification des opacités [Cheikhrouhou, 2012].

#### 5.2.4.4. La matrice GLCM (*Gray-Level Co-Occurrence Matrix*)

La matrice de co-occurrence GLCM, est une méthode statistique de la texture qui tient compte de la relation spatiale des pixels, également connu comme la matrice de la dépendance spatiale de niveaux de gris. Les fonctions de GLCM caractérisent la texture d'une image en calculant la fréquence de paires de pixels présentant des valeurs spécifiques et dans une relation spatiale spécifiée se produisent dans une image, la création d'un GLCM, puis à extraire les mesures statistiques de cette matrice.

##### **-La création du *Gray-Level Co-Occurrence Matrix***

Pour créer un GLCM on calcule la fréquence de la relation spatiale qui existe entre l'intensité (niveau de gris)  $i$  un pixel à un pixel à la valeur  $j$ . Par défaut, la relation spatiale est définie comme le pixel d'intérêt et le pixel immédiatement à sa droite (horizontalement adjacente), mais on peut spécifier d'autres relations spatiales entre les deux pixels. Chaque élément  $(i, j)$  dans la GLCM résultante est simplement la somme du nombre de fois de relation spatial entre deux pixels dans l'image d'entrée. Le nombre de niveaux de gris de l'image détermine la taille de la GLCM, on utilise une échelle pour réduire le nombre de valeurs d'intensité dans une image (8 bits), l'orientation  $0^\circ$ , la taille de la fenêtre, on a testé 8.8, 16.16 et 32.32 et la distance entre pixel qu'est égale à 1.

Pour l'illustration, la figure suivante (figure 5.9) montre comment on calcule les trois premières valeurs dans un GLCM.

Dans le GLCM de sortie, l'élément  $(1,1)$  contient la valeur 1 car il n'y a qu'une seule instance de l'image d'entrée lorsque deux pixels horizontalement adjacents ont les valeurs 1 et 1, respectivement. GLCM  $(1,2)$  contient la valeur 2 car il existe deux cas où deux pixels adjacents horizontalement ont les valeurs 1 et 2. Élément  $(1,3)$  dans le GLCM à la valeur 0, car il n'y a pas un cas de deux pixels adjacents horizontalement avec les valeurs 1 et 3. La matrice traite l'image d'entrée et balaie l'image pour d'autres paires de pixels  $(i, j)$  et en enregistrant les sommes dans les éléments correspondants de la GLCM.

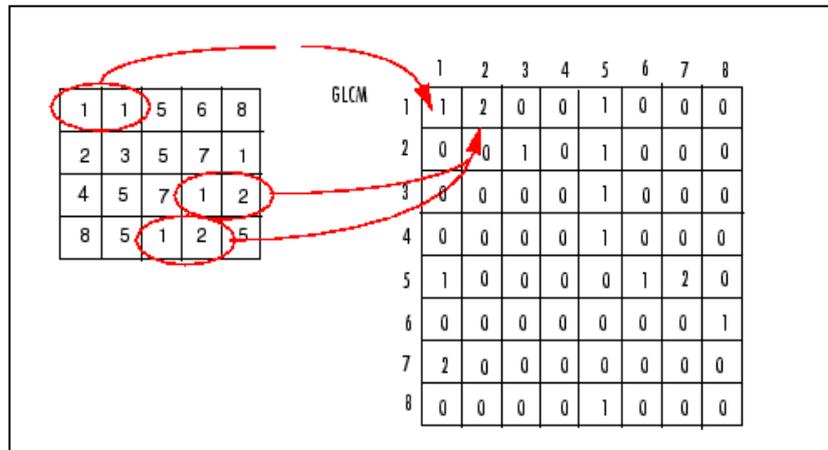


Figure 5.9. Processus utilisé pour la création d'une matrice de co-occurrence

### -Extraction statistique de la matrice GLCM

Après avoir créé la matrice GLCM, nous pouvons dériver plusieurs statistiques. Les statistiques fournissent des informations sur la texture d'une image. Les formules qui suivent représentent les statistiques utilisées.

Cet exemple (figure 5.10) montre comment créer un ensemble de GLCM et tirer des statistiques, et montre aussi comment les statistiques retournées ont un lien direct avec l'image originale d'entrée. L'exemple suivant calcule le contraste et la corrélation pour deux images de mammographie de textures différentes. Le schéma présente la corrélation des deux paramètres en fonction d'*offset*.

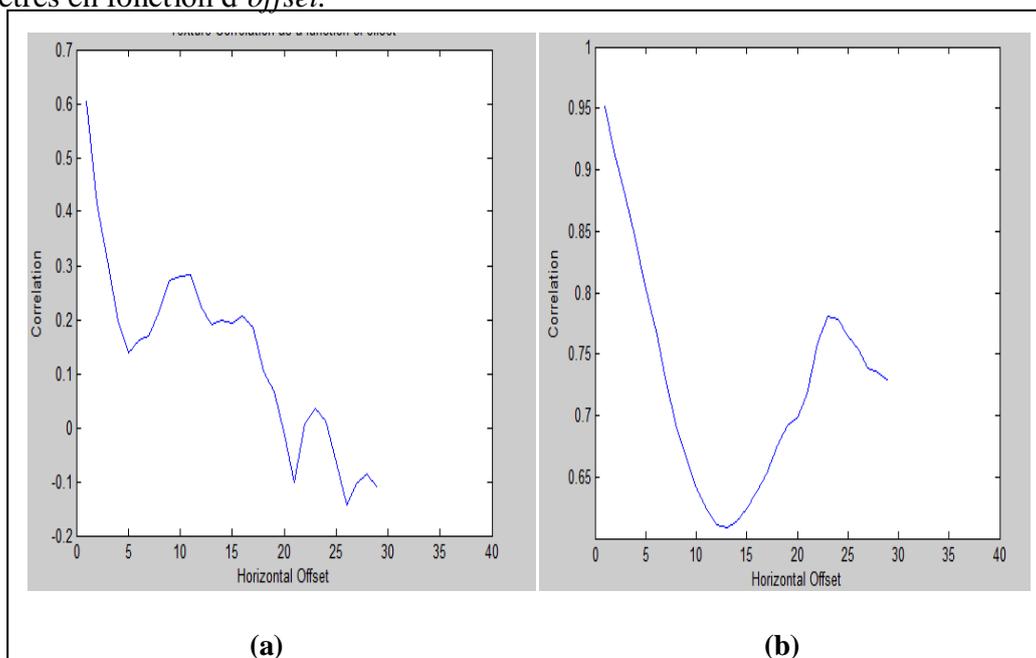


Figure 5.10. Exemple d'application de la matrice GLCM sur une image de masse maligne et une image de masse bénigne

Le schéma de l'image à masse bénigne présente des pics 8, 12, 14, 16, 18, 23, 24, 27. Si on examine l'image nous pouvons voir qu'il ya des éléments qui se répète chaque 2 et chaque 4 pixels. Dans l'image à masse maligne (figure 5.10. (b)) nous pouvons voir un seul pic qui annonce la variation dans la texture.

**-Les descripteurs utilisés**

Comme nous l'avons déjà vu au chapitre III, la comparaison directe des images entre elles n'est pas envisageable. Il est donc nécessaire d'en extraire au préalable des informations représentatives a savoir les descripteurs, utilisant la matrice de co-occurrence pour extraire les indices de textures.

Haralick a proposé 14 caractéristiques statistiques extraites à partir de cette matrice, actuellement, seulement les cinq caractéristiques les plus appropriées sont largement utilisées: l'énergie (*Angular Second Moment*), Contraste, la corrélation, la variance, le moment inverse de différence et l'entropie.

**-L'énergie (*Energy/Angular Second Moment*)**

Mesure le nombre de paires répétées. L'énergie est élevé si l'apparition de pixels paires répétée est aussi élevé. C'est à dire elle atteint de fortes valeurs lorsque la distribution des niveaux de gris est constante ou de forme périodique.

$$f_1 = \sum_i \sum_j \{p(i, j)\}^2 \tag{5.22}$$

**-Contrast**

Mesure le contraste local d'une image. Le contraste est faible si les niveaux de gris de chaque paire de pixels sont similaire.

$$f_2 = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{\substack{j=1 \\ |i-j|=n}}^{N_g} p(i, j) \right\} \tag{5.23}$$

**-Correlation**

Fournit une corrélation entre les deux pixels de la paire de pixels. La corrélation est élevée si les niveaux de gris des paires de pixels sont fortement corrélés (la texture est régulière), plus la coorelation est grande plus les valeurs sont uniformément distribuées dans la matrice de cooccurrence et une faible valeur dans le cas contraire.

$$f_3 = \frac{\sum_i \sum_j (ij)p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{5.24}$$

**-La variance (*sum of squares/Variance*)**

Cet indice fait également ressortir des différences entre les images, Ce paramètre mesure la distribution des niveaux de gris autour de la moyenne. Elle prend une grande valeur si les niveaux de gris de l'image ont grande dispersion.

$$f_4 = \sum_i \sum_j (i - \mu)^2 p(i, j) \quad (5.25)$$

**-Le moment inverse de différence IDM (*Inverse Difference Moment*)**

Inverse différence moment nous parle de *smoothness* de l'image, comme l'homogénéité. L'IDM est prévu à être élevée si les niveaux de gris des paires de pixels sont similaires.

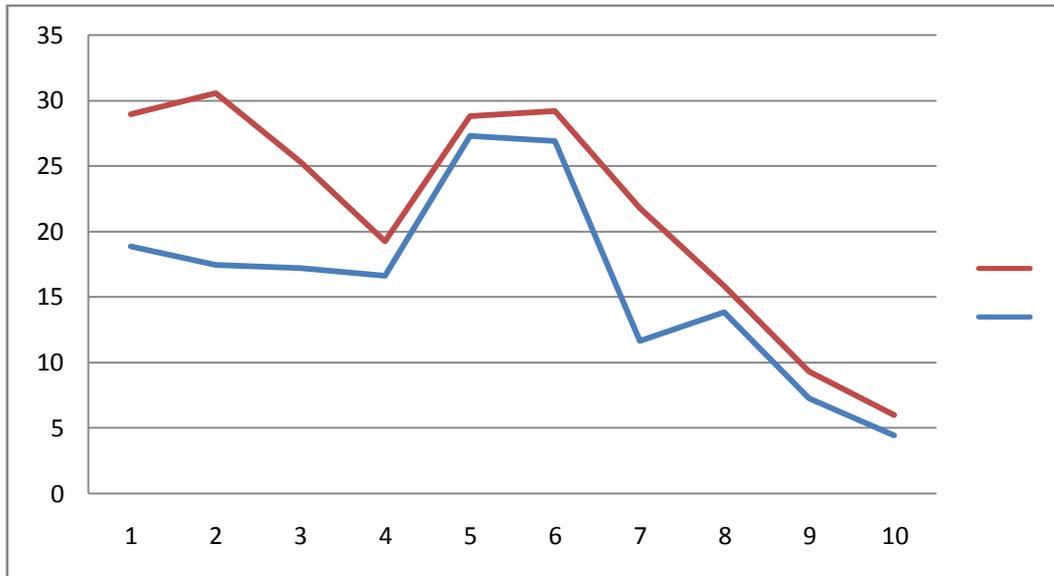
$$f_5 = \sum_i \sum_j \frac{1}{1+(i-j)^2} p(i, j) \quad (5.26)$$

Pour voir l'importance des caractéristiques, 20 régions d'images malignes et bénignes ont été choisis au hasard pour l'extraction de texture. La variance, IDM, Contraste, l'énergie et la corrélation ont été calculé sur les tissus mammaires avec des fenêtres 8x8, 16x16 et 32x32. Il se trouve que les meilleures caractéristiques pour discriminer les masses sont les caractéristiques de GLCM construits à direction de 0 °.

Il est également observé que les valeurs moment, variance et contraste pour les régions des masses et des zones non- masses sont très discriminantes. Ceci prouve l'utilité des descripteurs de texture pour la différenciation des masses.

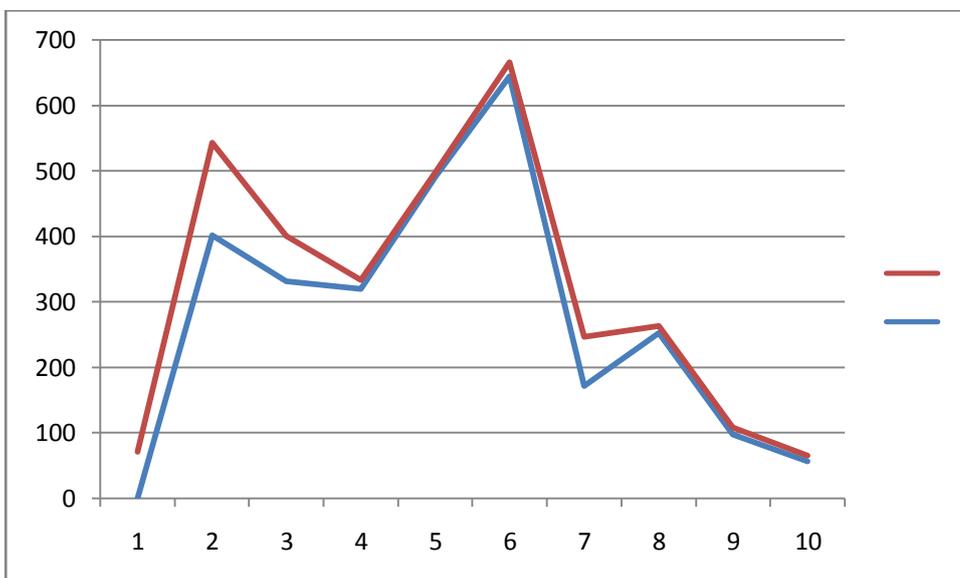
On observe également que le contraste et moment sont deux descripteurs de textures significatives, mais le moment est représenté pour être le plus efficace discriminateur décrit dans la figure (figures 5.11; 5.14).

Nous concluons que les cinq descripteurs peuvent être utilisés pour différencier les régions qui contiennent des masses malignes et bénignes.



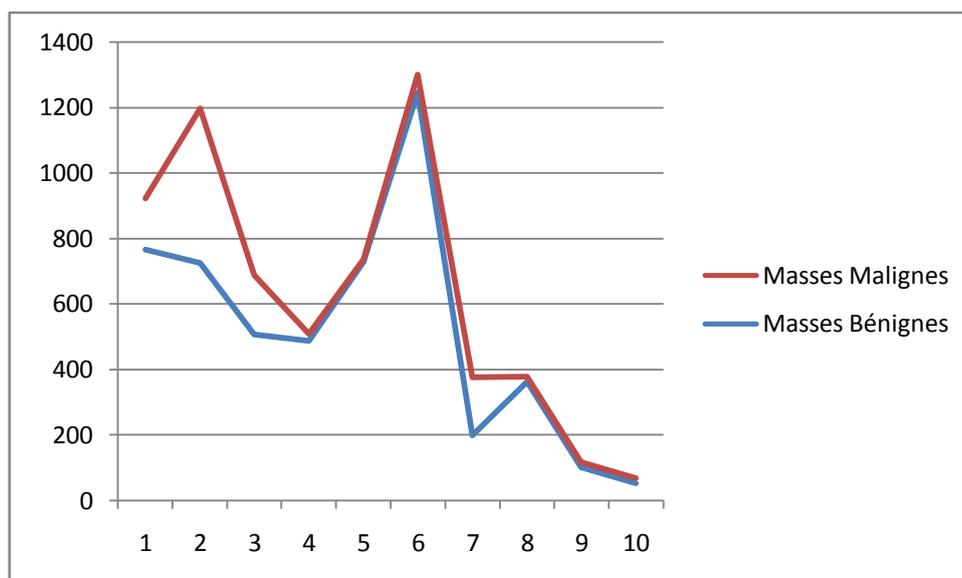
**Figure 5.11.** La valeur du contraste avec  $\theta=0^\circ$  utilisant une fenêtre 32x32 pixels.

On constate que la valeur du contraste pour les dix images malignes et bénignes joue un rôle important et peut différencier deux masses maligne et bénigne.



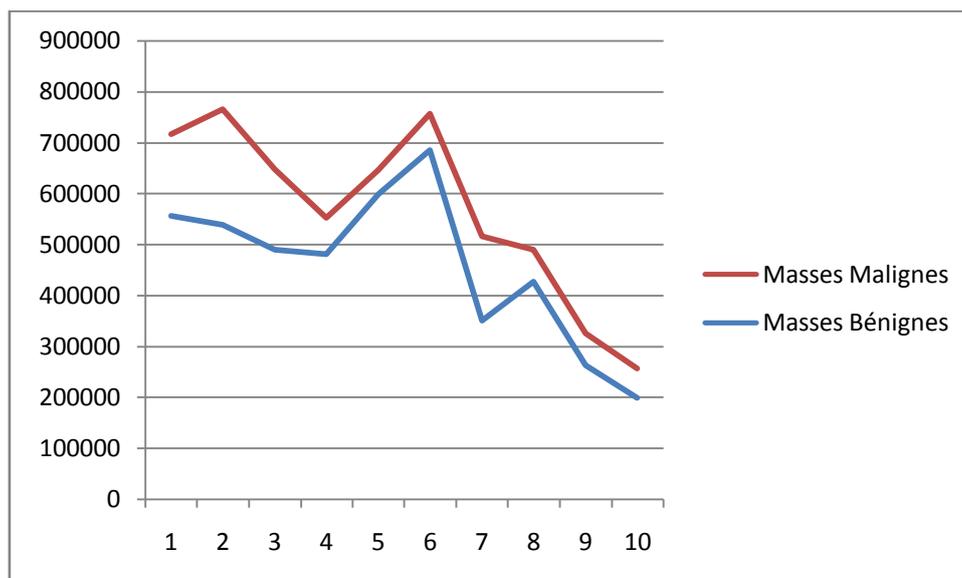
**Figure 5.12.** La valeur de la corrélation avec  $\theta=0^\circ$  utilisant une fenêtre 32x32 pixels.

Selon le graphe (figure 5.15), on peut voir que la valeur de la corrélation n'a pas vraiment un rôle important pour faire la différence entre masses malignes et bénignes.



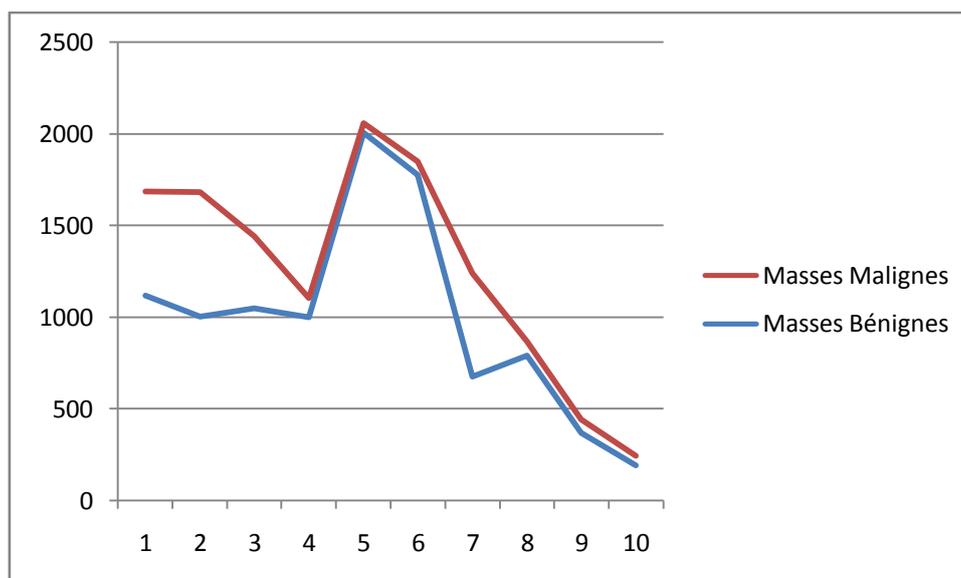
**Figure 5.13.** La valeur de la variance avec  $\theta=0^\circ$  utilisant une fenêtre 32x32 pixels

Le rôle de valeur (variance) reste important pour faire la différence entre les deux types.



**Figure 5.14.** La valeur du moment avec  $\theta=0^\circ$  utilisant une fenêtre 32x32 pixels

Avec l'utilisation du moment on peut voir la différence pour la distinction entre image maligne et bénigne.



**Figure 5.15.** La valeur de l'énergie avec  $\theta=0^\circ$  utilisant une fenêtre 32x32 pixels

Nous pouvons voir aussi que l'énergie a un rapport pour différencier les masses d'images (malignes, bénignes).

### 5.3. Conclusion

Dans ce chapitre nous avons présenté l'architecture générale d'un système de classification qui est destiné pour la classification des images malignes et bénignes.

Les éléments fondamentaux d'un tel système a été présenté, une grande attention a été allouée à l'étape de segmentation et extraction des caractéristiques, une méthode de segmentation est expliquée, vue les résultats et les discussions avec un spécialiste de domaine nous nous sommes orienté vers une autre méthode celle du Li et al. Ensuite, un ensemble de caractéristiques ont été décrits. Nous nous sommes focalisé sur les paramètres d'Haralik qui sont évalués avec quelques comparaisons, les moments de Hu, les descripteurs géométriques et descripteurs spécifiques sont présentés.

Dans le chapitre suivant, nous présentons la dernière partie de cette architecture, qui est une nouvelle méthode de classification se base sur les automates cellulaires.

## Chapitre VI

---

# Classification Basée AC

---

### 6.1. Introduction

Le travail effectué dans ce chapitre porte sur la classification des masses malignes et bénignes des images de mammographies.

Le paradigme construit à partir des éléments simples connectés entre eux, des règles simples sont appliquées dans le but de créer des ensembles d'images organisées sur un réseau de cellules. Les images similaires sont organisées l'une près de l'autre, dans ce travail nous choisissons la classification en deux classes pathologiques malignes et bénignes.

Une évaluation de cette méthode de classification en se basant sur la sélection des caractéristiques présenté dans le chapitre V (cf.sec. 5.2.4), la méthode ROC (*Receiver Operating Characteristic*).

Des comparaisons des deux méthodes de classification classique, la méthode K-means et la méthode Hiérarchique sont faites.

## 6.2. Principe de l'automate cellulaire

L'automate cellulaire est considéré comme un modèle basic pour générer les systèmes complexes par des règles simples et locales.

On a spécifié les termes suivants: la dimension de l'automate cellulaire, le voisinage, l'ensemble d'états et les règles locales.

L'automate cellulaire est un quadruple  $(Z^d, Q, V, \delta)$ , soit  $d$  un entier positif, dans notre travail nous nous intéressons à la structure  $2D$  où les cellules sont représentées par une matrice  $M \times M$ .

Nous avons fixé la taille de cette grille empiriquement selon la formule suivante :

$$M = A(\sqrt{2X_n}) + 1 \quad (6.1)$$

Avec :  $A$  représentant la partie entière d'un nombre réel,  $\sqrt{2X_n}$  calculé empiriquement afin de trouver une organisation spatiale de l'ensemble des vecteurs dans l'AC.

Soit  $Q = \{Q_1, Q_2, \dots, Q_n\}$  l'ensemble d'états finit, on appelle  $Q$  un état, dans notre travail nous utilisons quatre états:

*The alive* et *dead* prennent  $-1$  et  $0$  respectivement, *isolated state* prend une valeur égale à  $-2$ .

La cellule avec *dynamic state* va prendre les descripteurs de l'image.

L'état initial des cellules est égal à  $0$ .

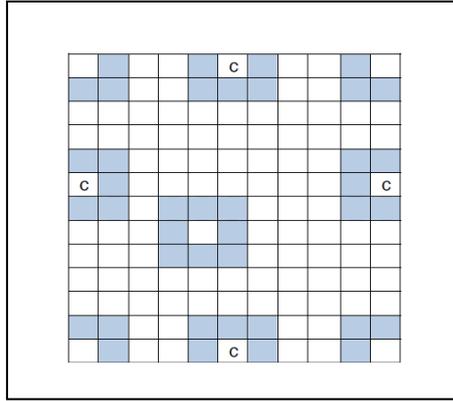
Théoriquement, la configuration du système est désignée par :

$$C: Z^d \rightarrow Q \quad (6.2)$$

Où les éléments de  $Z^d$  sont les cellules, chaque cellule prend un état, l'état de la cellule  $\vec{n} \in Z^d$  est  $Q(\vec{n})$ .

Dans un espace à deux dimensions le voisinage de Von Neumann et Moore est souvent utilisé. Dans notre travail nous utilisons une grille planaire avec un voisinage hybride.

Les quatre cellules dans les coins de la grille ont trois cellules comme voisinage, tandis que les cellules situées au niveau du périmètre ont un voisinage de cinq cellules (figure 6.1).



**Figure 6.1.** Voisinage hybride de l'AC

Soit  $V = \{\vec{n}_1, \vec{n}_2, \dots, \vec{n}_m\}$  un vecteur de voisinage pour n'importe quelle cellule  $\vec{n}$  qui appartient à  $Z^d$ .

Soit  $\delta$  la règle locale de l'automate cellulaire avec un ensemble des états et un voisinage  $n$  :

$$f: Q^n \rightarrow Q \tag{6.3}$$

Le nouveau état de chaque cellule est basé sur l'état précédent de son voisinage, si les voisinages de la cellule ont les états  $Q_1, Q_2, \dots, Q_n$  alors les nouveaux états seront:  $f(Q_1, Q_2, \dots, Q_n)$ .

### 6.2.1. Fonctionnement de l'automate cellulaire

Les règles locales pour l'évolution de l'automate cellulaire sont représentées ci-dessous :

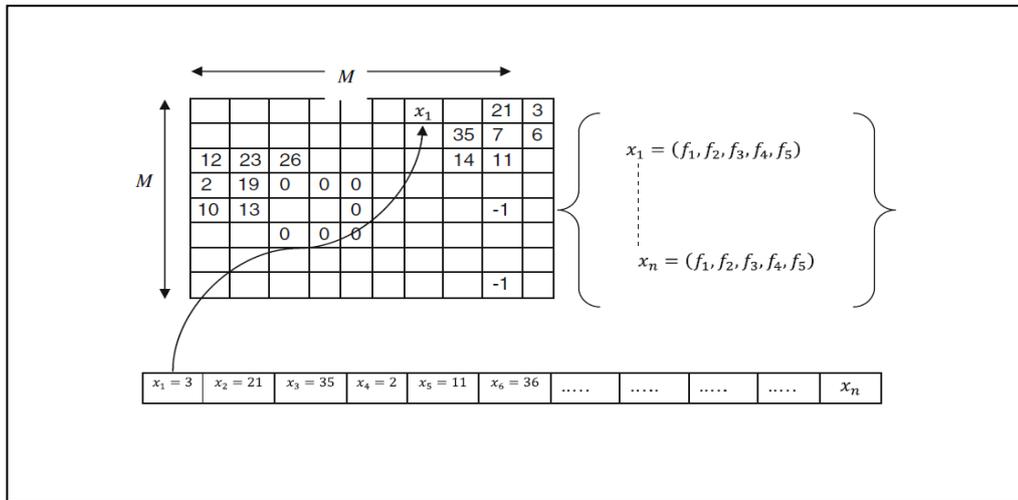
-Règle d'AC N° 1 : Si la cellule  $C_{ij}$  à l'instant  $t$  dans l'état *Dead* c'est-à-dire contient une valeur égale à 0 alors : dans l'état qui suit  $t + 1$  la cellule  $C_{ij}$  va recevoir un vecteur de liste où  $liste = [X_1, X_2, \dots, X_n]$  et le voisinage du voisinage de  $C_{ij}$  devient *Alive* c'est-à-dire qu'il va prendre la valeur -1.

-Règle d'AC N° 2 : Si la cellule  $C_{ij}$  à l'instant  $t$  est à l'état *Alive*, c'est-à-dire contient une valeur qu'est égale à -1 alors le voisinage doit être vérifié :

Si le voisinage de la cellule  $C_{ij}$  contient au moins une cellule avec l'état *dynamic* c'est-à-dire contient le vecteur représentant de l'image alors la cellule  $C_{ij}$  dans le temps  $t + 1$  va recevoir de la liste le vecteur le plus similaire de celle du voisinage et le voisinage devient *Alive* et reçoit une valeur -1.

-Dans les autres cas, l'état du voisinage devient *isolated*, c'est-à-dire que la cellule  $C_{ij}$  va recevoir la valeur -2 dans le temps  $t+1$ .

Pour appliquer ces règles à l'ensemble des cellules, nous avons testé différents parcours afin de choisir la première cellule. Nous avons choisi expérimentalement un seuil qui représente la similarité maximale entre les images de la base, la figure (6.2) explique la méthode.



**Figure 6.2.** Exemple de l'automate cellulaire. L'état *Alive* (-1), l'état *Dead* (0), l'état *Isolated* (2) et l'état *Dynamic* (contient le vecteur de l'image)

Avant d'utiliser les règles de la méthode AC pour une classification d'images de mammographie, il est nécessaire de choisir une mesure de similarité.

La similarité entre deux images peut alors être mesurée à l'aide d'une métrique définie sur l'espace vectoriel ainsi, les images sont considérées similaires si la distance entre eux est faible.

Notre choix est porté sur la distance de Manhattan. Cette norme est aussi connue sous le nom *city-block* et la plus simple à calculer.

Considérons deux images  $I_1$  et  $I_2$  indexées par des vecteurs dans  $R^N$ :  $I_1 = \{I_1(i), 1 \leq i \leq N\}$  et  $I_2 = \{I_2(i), 1 \leq i \leq N\}$ . Le calcul de similarité entre  $I_1$  et  $I_2$  est donné par la formule suivante :

$$Dist(I_1, I_2) = \sum_{i=1}^N (| I_1(i) - I_2(i) |) \tag{6.4}$$

Où  $i$  représente le nombre total de paramètres .

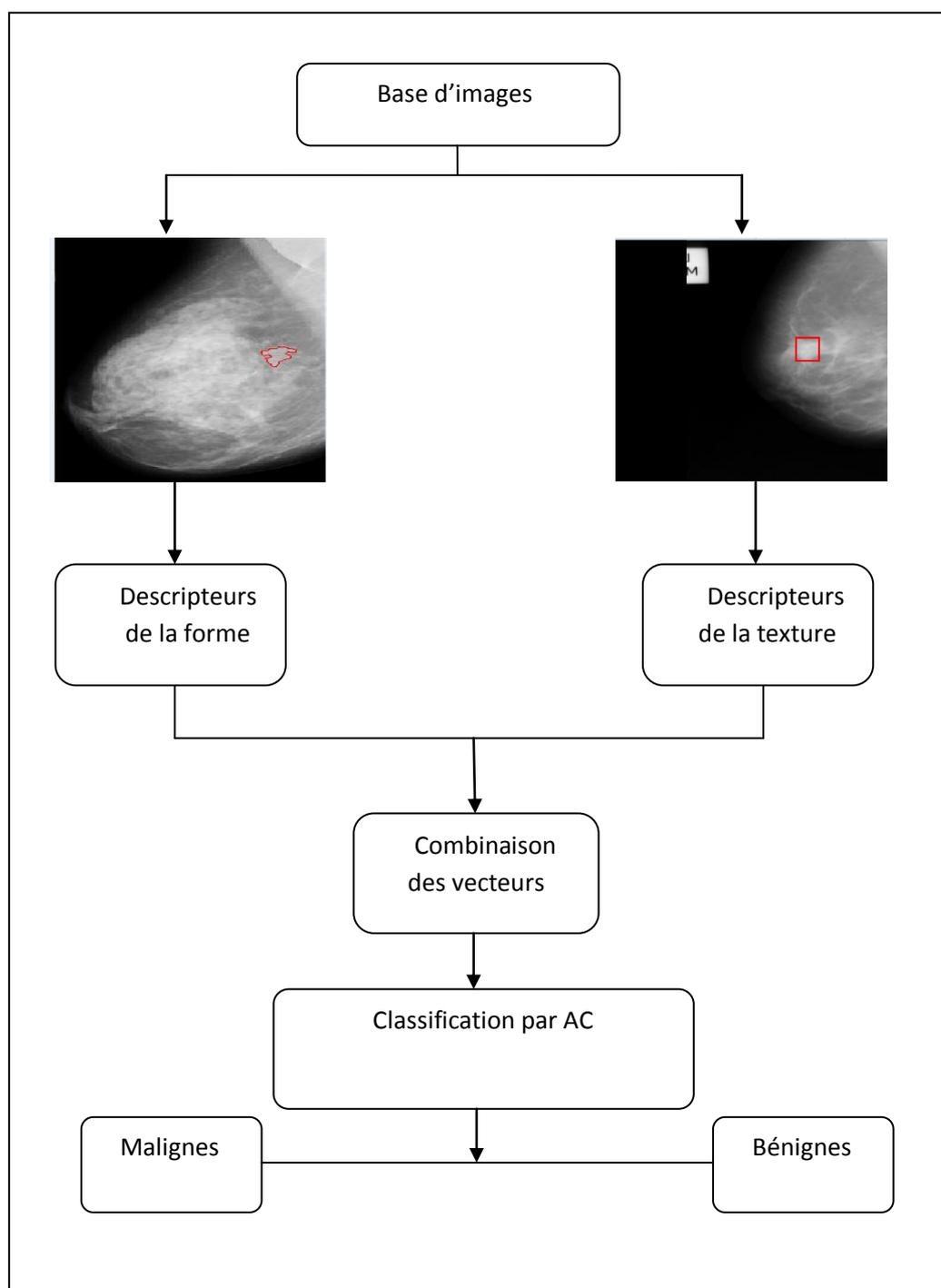


Figure 6.3. Schéma illustrative des étapes de la classification

Dans ce travail, nous utilisons deux bases DDM of Annaba et la base MIAS.

Pour la description de la texture un grand nombre de chercheurs utilisent des fenêtres de taille 8x8, 16x16 et 32x32, la fenêtre utilisée pour le traitement est importante car elle va déterminer l'aptitude du descripteur de texture à faire la différence entre les masses.

Dans notre travail la sélection de zone doit être fait comme suit:

Sur cette base les zones qui contiennent des masses se trouvent dans une fenêtre de 32x32, avant une classification la détection de la masse est importante dans le choix de la région candidate pour l'extraction de la texture.

Calculant la matrice GLCM et appliquant les paramètres d'Haralick sur ces fenêtres, ils donnent un ensemble des vecteurs, où chaque vecteur contient cinq valeurs.

Pour la description de la forme le processus se fait comme suit:

Une méthode de segmentation est appliquée (cf.sec. 5.2.3) sur l'ensemble des images pour extraire la région suspecte. Ensuite, un ensemble des descripteurs géométrique (cf.sec. 5.2.4.2) et spécifiques (cf.sec. 5.2.4.3) sont appliqués.

Ces attributs de bas niveau décrivent les images par le contenu visuel, ces attributs peuvent être combinés pour caractériser au mieux et d'une façon performante et efficace le contenu (figure 6.3). Alors, des descripteurs de forme robustes et adaptés au caractère pathologique des masses mammaires doivent respecter ces critères : Invariance, Tolérance par rapport au bruit, Unicité, Conservation de l'information et l'utilisation d'une métrique.

### **6.3. Résultats et évaluations de la performance de la classification proposée**

Pour une telle évaluation nous avons besoin de la méthode ROC (*Receiver Operating Characteristic*) qui est détaillée en Annexe (A.1).

#### **6.3.1. Aire sous la courbe ROC**

A partir d'une description de la performance du classifieur sous une forme d'une courbe ROC, nous sommes obligés de calculer l'aire sous cette courbe qui est noté par AUC.

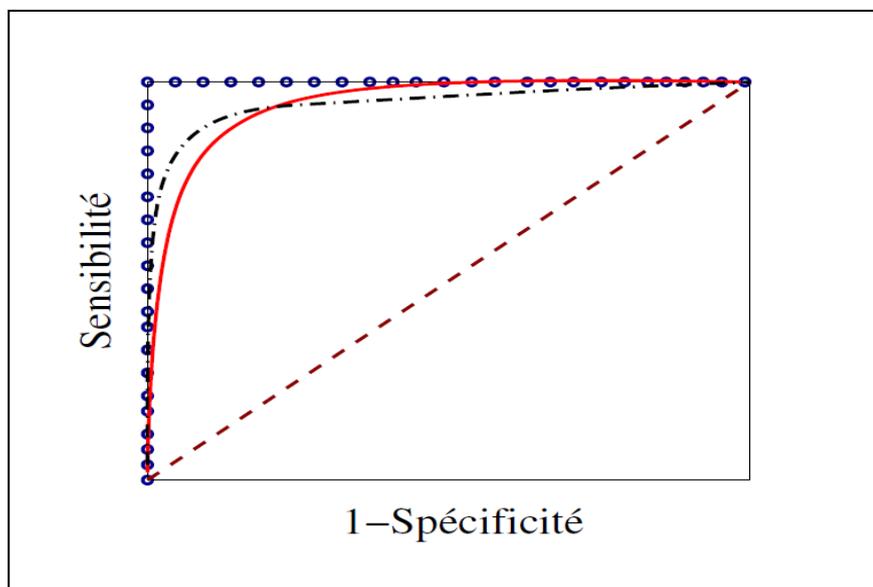
L'aire sous la courbe est une portion de l'aire d'un carré unitaire avec une valeur qui est toujours entre 0 et 1.

Dans le cas idéal cette valeur est égale à 1, c'est-à-dire plus la valeur de AUC est proche de 1, plus le résultat de classification est meilleure, dans le cas d'un système défaillant, l'aire est égale à 0.5.

Les courbes à trait mixte fin et à trait continu de la figure (6.4) ont presque la même valeur de AUC qui est plutôt proche du 1 que du 0.5. Ces courbes représentent de bons résultats de

classification, sachant que la courbe ROC d'un bon classifieur monte rapidement vers le coin supérieur gauche.

Enfin, nous pouvons dire que la courbe en trait mixte fin donne un meilleur résultat de classification que la courbe en trait continu même s'ils représentent la même valeur d'aire sous la courbe ROC.



**Figure 6.4.** Courbes ROC pour: trait en boules (un test diagnostique parfait assurant une séparation idéale entre les sujets malins et bénins), trait interrompu (un test diagnostique d'apport nul ne produisant aucune discrimination entre les individus), trait continu et trait mixte fin (des tests diagnostiques très informatifs)

D'une manière plus explicite, la signification de l'aire sous la courbe AUC (*Area Under the Curve*) est évaluée de la façon suivante:

- Si  $AUC = 0.5$  : le résultat du diagnostic est d'apport nul ;
- Si  $0.5 < AUC < 0.7$  : le résultat est peu informatif ;
- Si  $0.7 \leq AUC < 0.9$  : le résultat est moyennement informatif ;
- Si  $0.9 \leq AUC < 1$  : le résultat est très informatif ;
- Si  $AUC = 1$  : le résultat est parfait.

### 6.3.2. Résultats et évaluations

Nous avons travaillé principalement sur deux bases dans le but d'évaluer et de valider la classification proposée.

Nous présentons dans cette partie les résultats expérimentaux obtenus à partir de la méthode proposée, les résultats sont présentés selon deux formes :

- L'affichage des automates cellulaires obtenus avec l'utilisation des descripteurs.

-l'évaluation avec la courbe ROC, afin d'avoir une représentation visuelle globale de la performance de la classification (figure 6.28) et enfin une comparaison avec deux méthodes de la littérature les plus utilisés.

Dans notre travail nous avons utilisé les deux bases médicales DDMA et MIAS avec 90 et 120 images malignes et bénignes respectivement.

La matrice obtenue représente l'AC, chaque cellule contient le numéro qui donne le vecteur représentant l'image, les images similaires sont représentées les unes à côté de l'autres.

Nous tenons compte des descripteurs (cf. sec.5.2.3).

Nous appliquons l'AC avec l'utilisation de la base d'Annaba, vu la qualité des images, la segmentation n'était pas assez bonne et ceci à influencer sur les résultats de la classification c'est-à-dire que la segmentation est aussi primordiale dans notre travail (cf. sec.3.4.1), plus la détection de masse est nette plus la classification est réussie.

Pour tester la signification statistique de l'aire sous la courbe AUC, nous avons utilisé le teste de la courbe ROC sous les hypothèses suivantes :

$$H_0: AUC = 0.5$$

$$H_1: AUC > 0.5$$

Pour pouvoir tester la performance nous comparons l'aire sous la courbe ROC des résultats de classification :

Nous avons testé la classification basée sur les descripteurs géométriques, les descripteurs spécifiques et les moments de Hu, des combinaisons entres descripteurs. La figure (6.5) présente l'AC, l'utilisateur peut naviguer sur cette matrice ou cette classification visuelle. Dans ce qui suit nous présentons les résultats de la classification et l'évaluation avec la courbe Roc. Les premiers résultats concernent la base DDMA [Benmazou et al., 2014 a].

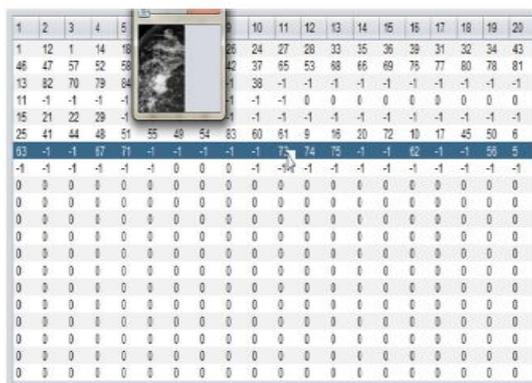


Figure 6.5. Carte navigable de l'AC





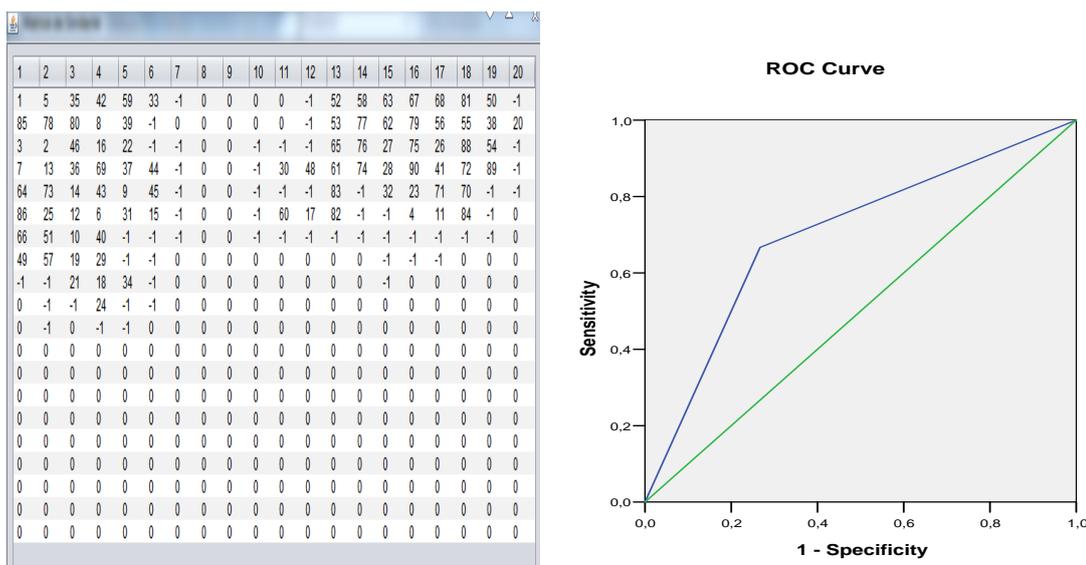






Etant donné que la valeur de sig (Annexe A.2.23) est de (0.035) qui est inférieur au niveau de signification ( $\alpha=0.05$ ), on doit rejeter l'hypothèse nulle ( $H_0$ ) et retenir l'hypothèse alternative  $H_1$ .

D'après la figure (6.13), le descripteur de LRN, qui décrit la forme annonce une valeur de l'aire sous la courbe avec AUC égale à 0.59 et ça aussi traduit par Std. Error (a) (Annexe A.2.23) faible qui égale à 0.053.



**Figure 6.14.** Courbe ROC obtenue avec classification AC et les descripteurs LRN/Hu

Etant donné que la valeur de sig (Annexe A.2.26) est de (0.011) qui est inférieur au niveau de signification ( $\alpha=0.05$ ), on doit rejeter l'hypothèse nulle ( $H_0$ ) et retenir l'hypothèse alternative  $H_1$ .

D'après la figure (6.14), la combinaison des descripteurs LRN et Hu, qui décrivent la forme annonce une valeur de l'aire sous la courbe avec AUC égale à 0.70 et ça aussi traduit par Std. Error (a) (Annexe A.2.26) faible qui égale à 0.044.

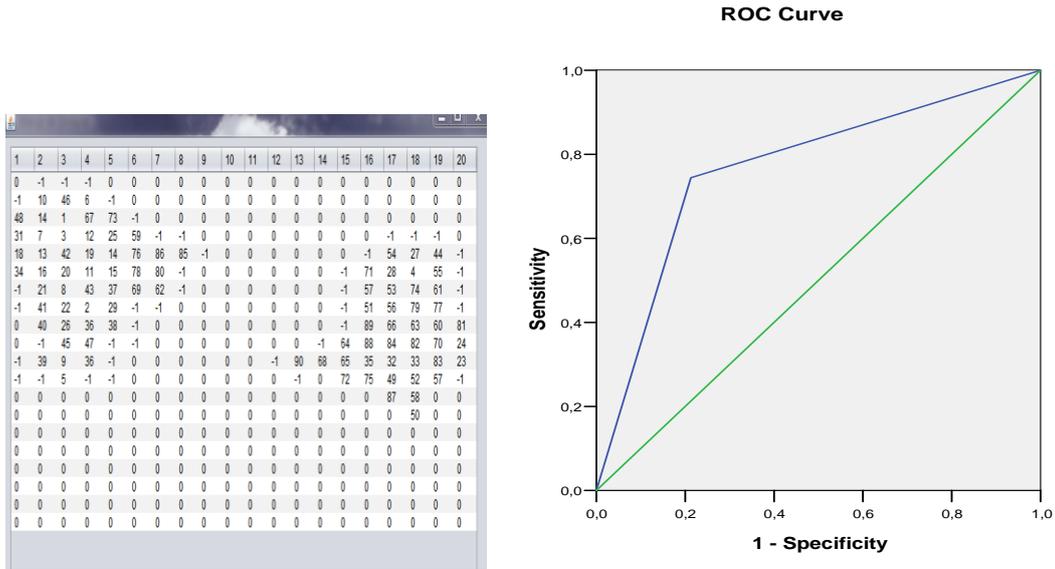


Figure 6.15. Courbe ROC obtenue avec classification AC et le descripteur Haralick/LRN

Etant donné que la valeur de sig (Annexe A.2.29) est de (0.011) qui est inférieur au niveau de signification ( $\alpha=0.05$ ), on doit rejeter l’hypothèse nulle ( $H_0$ ) et retenir l’hypothèse alternative  $H_1$ .

D’après la figure (6.15), le descripteur de Haralick/LRN, qui décrit la forme annonce une valeur de l’aire sous la courbe avec AUC égale à 0.76 et ça aussi traduit par Std. Error (a) (Annexe A.2.29) faible qui égale à 0.042, la figure (6.16) montre les courbes Roc pour les descripteurs Haralick, LRN et la combinaison des deux.

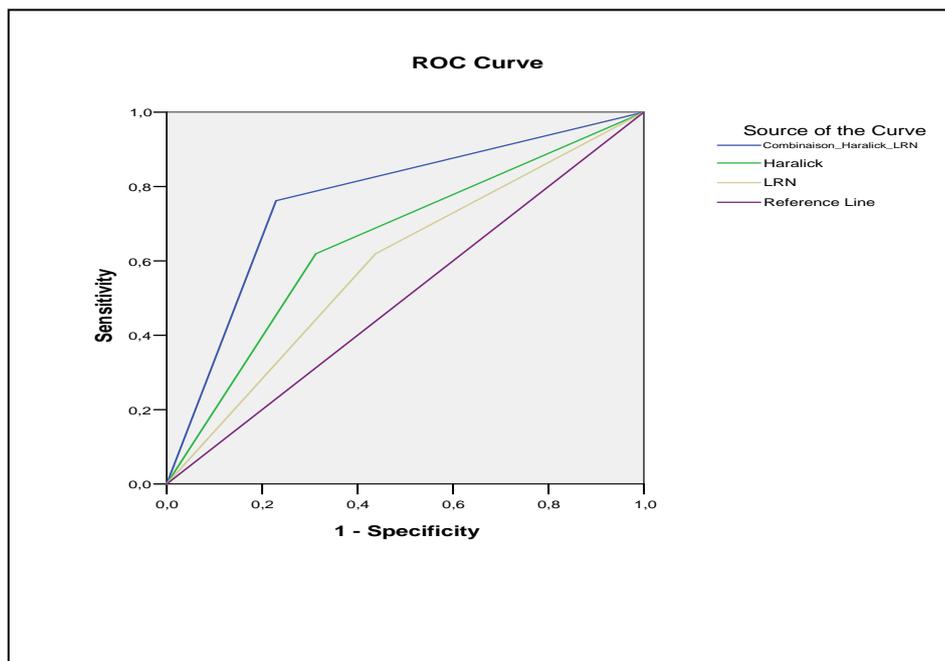


Figure 6.16. Courbes ROC obtenue avec classification AC et les descripteurs Haralick/LRN et la combinaison Haralick\LRN



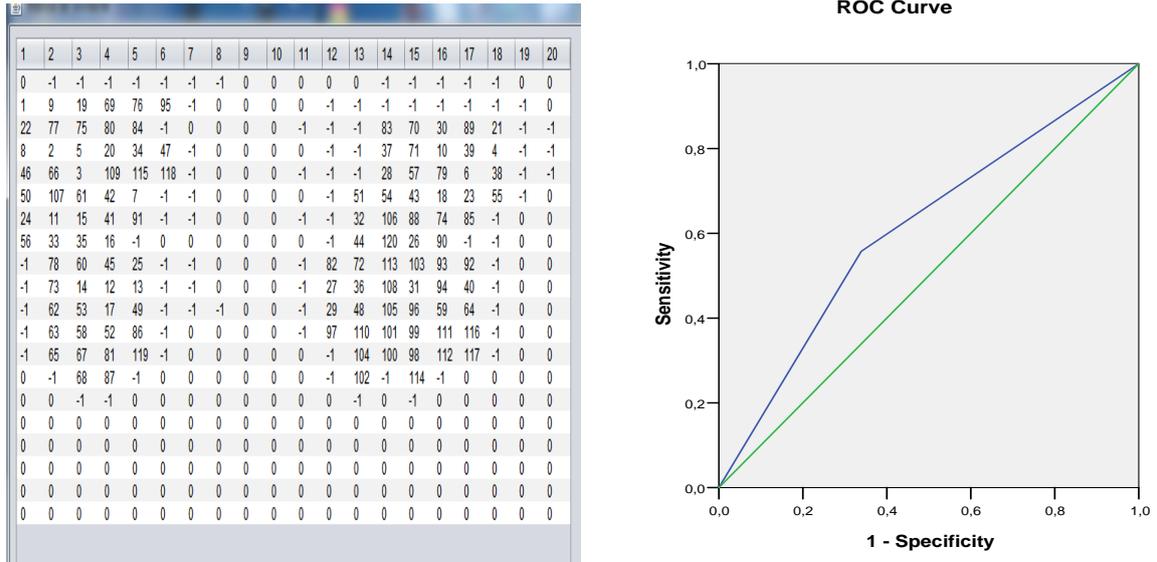


Figure 6.18. Courbe ROC obtenue avec classification AC et le descripteur Périmètre

Etant donné que la valeur de sig (b) (Annexe A.2.38) est de (0.031) qui est inférieur au niveau de signification ( $\alpha=0.05$ ), on doit rejeter l'hypothèse nulle ( $H_0$ ) et retenir l'hypothèse alternative  $H_1$ .

D'après la figure (6.18), le descripteur périmètre qui décrit la forme annonce une valeur de l'aire sous la courbe avec AUC égale à 0.61 et ça aussi traduit par Std. Error (a) (Annexe A.2.38) moyennement faible qui égale à 0.052.

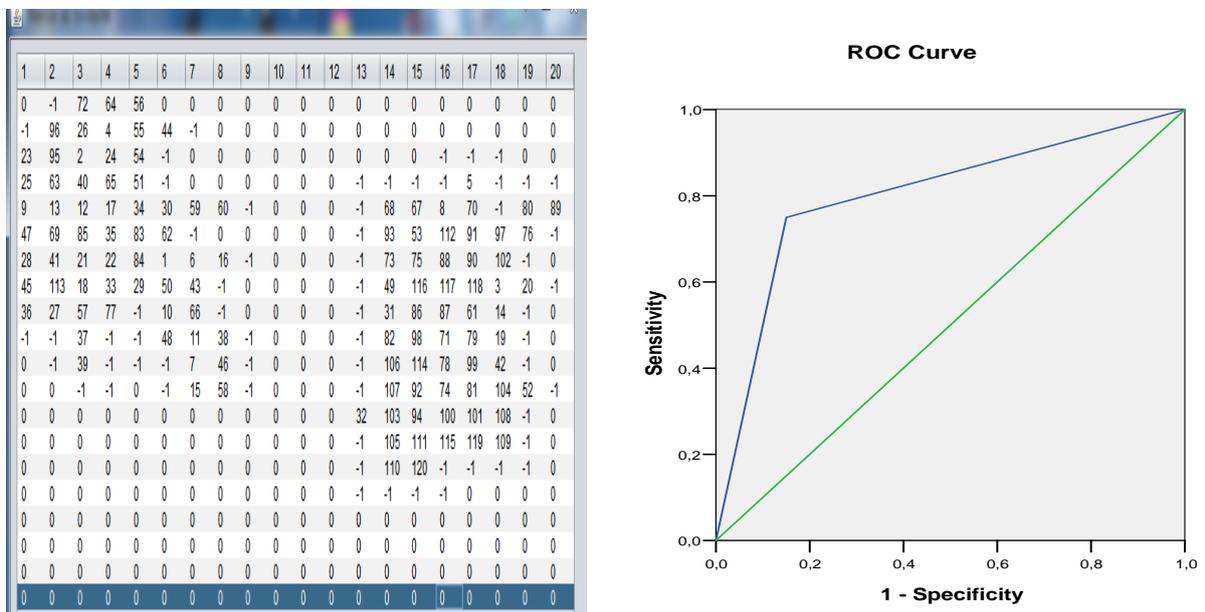
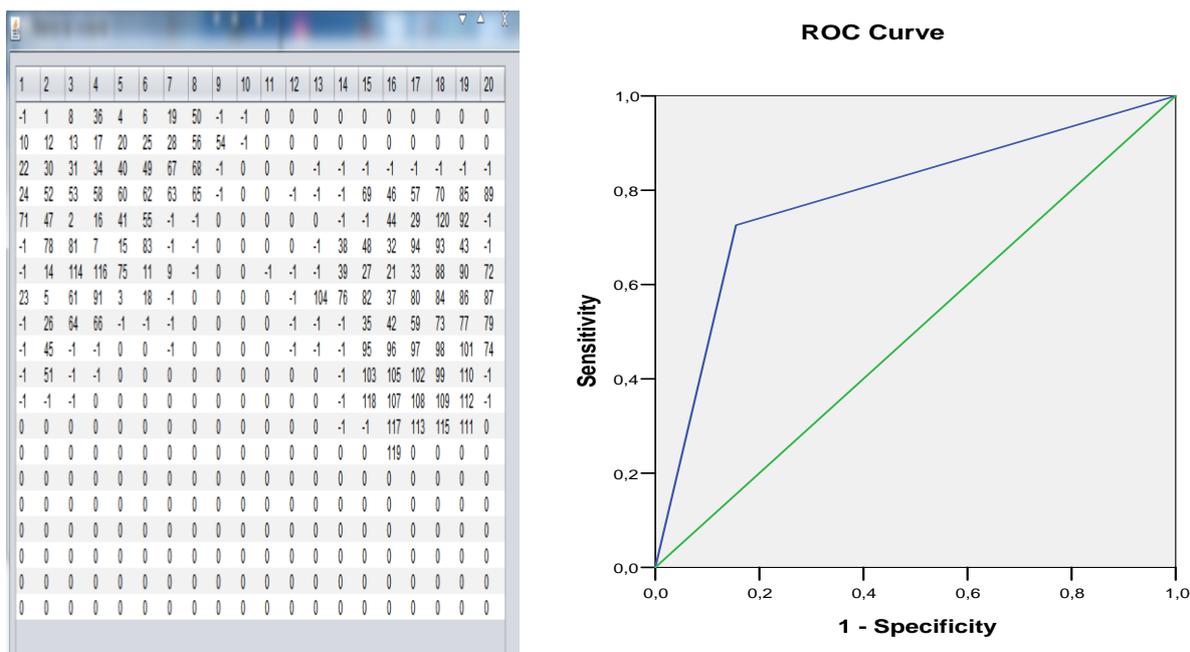


Figure 6.19. Courbe ROC obtenue avec classification AC et le descripteur Excentricité

Etant donné que la valeur de sig (Annexe A.2.41) est de (0.07) qui est inférieur au niveau de signification ( $\alpha=0.05$ ), on doit rejeter l'hypothèse nulle ( $H_0$ ) et retenir l'hypothèse alternative  $H_1$ .

D'après la figure (6.19), le descripteur excentricité qui décrit la forme annonce une valeur de l'aire sous la courbe avec AUC égale à 0.80 et ça aussi traduit par Std. Error(a) (Annexe A.2.41) faible qui égale à 0.42.



**Figure 6.20.** Courbe ROC obtenue avec classification AC et le descripteur Compacité

Etant donné que la valeur de sig (Annexe A.2.44) est de (0.019) qui est inférieur au niveau de signification ( $\alpha=0.05$ ), on doit rejeter l'hypothèse nulle ( $H_0$ ) et retenir l'hypothèse alternative  $H_1$ .

D'après la figure (6.20), le descripteur de la compacité qui décrit la forme annonce une valeur de l'aire sous la courbe avec AUC égale à 0.78 et ça aussi traduit par Std. Error (a) (Annexe A.2.44) faible qui égale à 0.043.



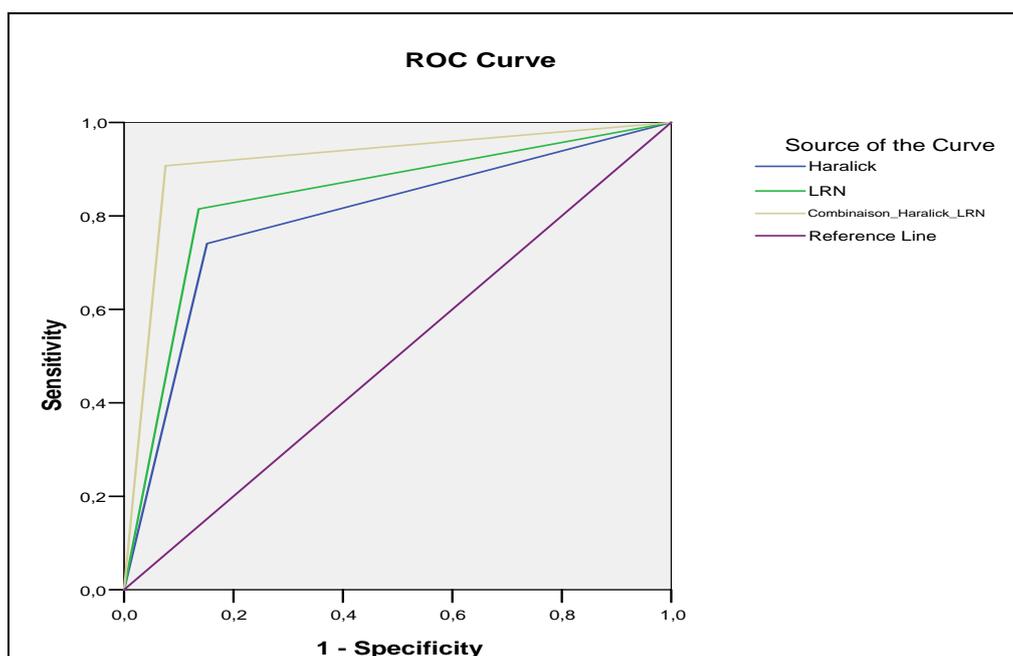






Etant donné que la valeur de sig (Annexe A.2.62) est de (0.007) qui est inférieur au niveau de signification ( $\alpha=0.05$ ), on doit rejeter l'hypothèse nulle ( $H_0$ ) et retenir l'hypothèse alternative  $H_1$ .

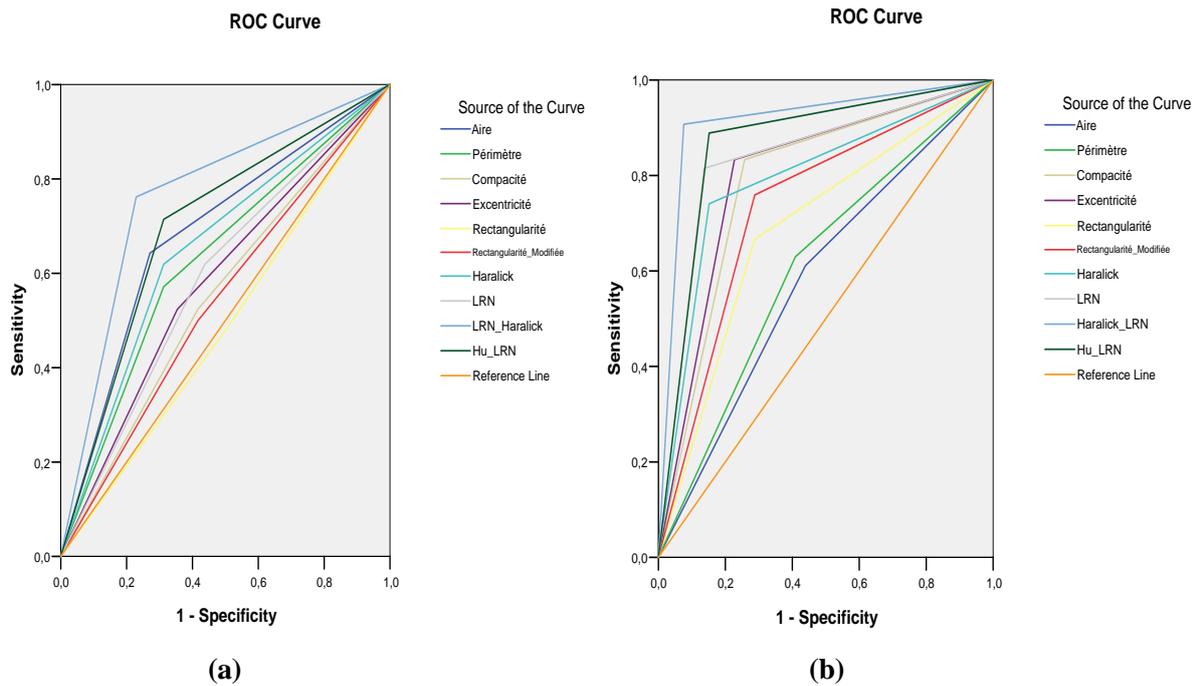
D'après la figure (6.26), la combinaison de la texture (les descripteurs d'Haralick) et la forme (les descripteurs spécifiques) annoncent une valeur de l'aire sous la courbe avec AUC égale à 0.91 et ça aussi traduit par Std. Error (a) (Annexe A.2.62) qui dépasse la moyenne avec une valeur égale à 0.020. Pour une explication plus détaillée, la figure (5.39) présente les trois courbes ROC.



**Figure 6.27.** Courbes ROC obtenues avec classification AC et les descripteurs Haralick/LRN et la combinaison Haralick\LRN

Afin d'évaluer l'apport de combiner les descripteurs spécifiques qui décrivent la forme aux descripteurs de la texture, nous avons tracé simultanément dans la figure (6.27), les courbes ROC correspondant aux descripteurs d'Haralick (Haralick), descripteurs de la longueur radiale normalisée (LRN) et à la combinaison des deux (Combinaison\_Haralick\_LRN).

On remarque bien que la courbe en couleur gris représentant la combinaison (AUC = 0.91) est nettement au dessus des deux courbes, une courbe bleu avec (AUC=79) et une courbe vert avec (AUC=83), Annexe (A.2.65).



**Figure 6.28.** Courbes ROC obtenues avec classification AC et les descripteurs et tous les descripteurs utilisés sur la base DDMA (a) et MIAS (b)

Les valeurs de sig (Annexe A.2.71) de tous les descripteurs sont inférieure au niveau de signification ( $\alpha=0.05$ ), on doit rejeter l'hypothèse nulle ( $H_0$ ) et retenir l'hypothèse alternative  $H_1$  sauf la valeur de la rectangularité calculé avec la base DDMA (figure 6.28.a).

D'après la figure (6.28), la combinaison de la texture (les descripteurs d'Haralick) et la forme (les descripteurs spécifiques) annonce la valeur la plus grande de l'aire sous la courbe avec AUC égale à 0.91 pour la base MIAS et 0.76 pour la base DDMA. Ça aussi traduit par Std. Error (a) (Annexe A.2.71) qui dépasse la moyenne avec une valeur égale à (0.020) et (0.010) pour la base MIAS et la base DDMA respectivement.

L'utilisation de l'AC avec le descripteur de la forme aire fournit le résultat le moins performant (figure 6.28.b). Le calcul d'AUC donne une valeur qui est égale à 0.58, cette valeur traduit que le descripteur aire est insuffisant pour décrire une masse complexe.

La figure (6.28.b) représentant la courbe qui décrit le périmètre, cette courbe s'éloigne du coté gauche ce qui se traduit par une sensibilité moyennement égale à 0.55 et une surface sous la courbe ROC égale à 0.61. Ce résultat traduit que le descripteur périmètre reste assez significative pour décrire le contour d'une masse, la combinaison avec d'autres descripteurs peuvent améliorer les résultats. Nous pouvons conclure que ces deux mesures restent proportionnelles pour décrire la complexité d'un contour et ça peut altérer le résultat de la classification.

La compacité considère le périmètre et l'aire à la fois, elle permet de reconnaître la régularité entre deux formes. Ces descripteurs reflètent les résultats qui sont obtenus avec la classification qui confirment que cette entité est plus discriminante que les descripteurs, la surface avec un AUC égale à 0.78, (figure 6.28.b).

L'excentricité (figure 6.28.b) est proportionnelle à la complexité de la forme, sa valeur varie entre 1 et 0 si la forme régulière ou allongée. Cette mesure reflète bien le résultat de la classification avec un AUC égale à 0.80.

En ce qui concerne la mesure de rectangularité (figure 6.28.b), l'aire sous la courbe ROC correspondante est de  $AUC = 0.68$ . Le résultat de cette mesure est moyennement informatif puisque elle est très sensible à la rotation.

L'apport de l'ajout d'un critère d'invariance par rotation au descripteur de la rectangularité, qui est nommée ensuite la rectangularité modifiée (figure 6.28.b) à donné un résultat meilleur que la première, l'aire sous la courbe est égale à 0.73.

Pour moments de Hu, les résultats ne sont pas satisfaisant, car ces mesures sont très sensible au bruit, une combinaison avec les descripteurs spécifiques influence le résultat avec une valeur d'AUC égale à 0.86.

Vue l'importance de la texture pour faire la différence entre une zone d'une image maligne et une image bénigne (cf. figure 5.10), les statistiques du deuxième ordre annoncent un résultat meilleur avec un AUC égale à 0.79.

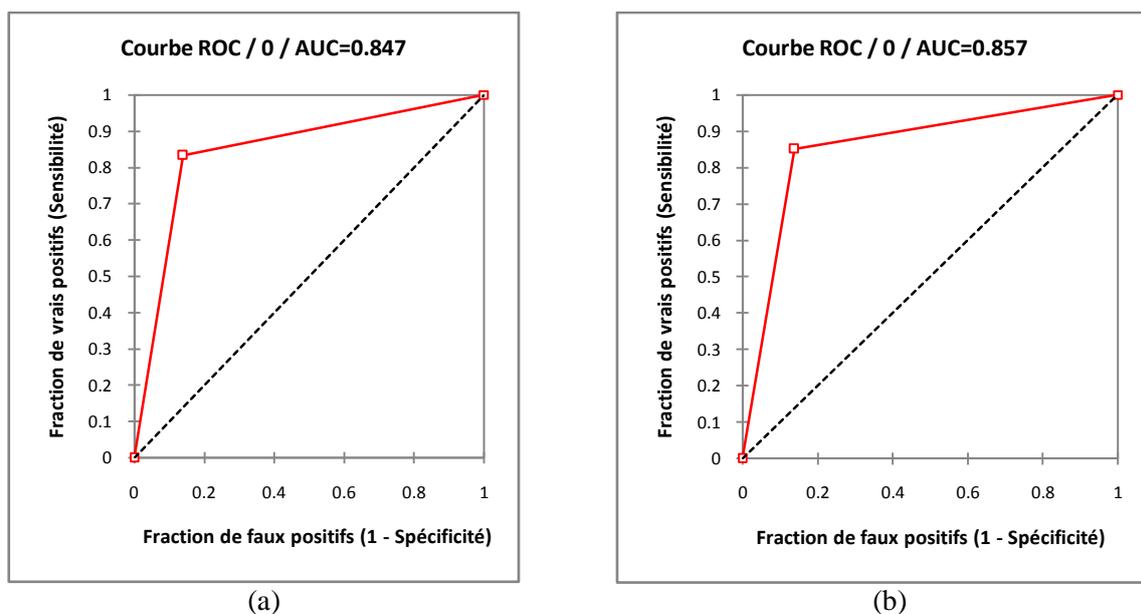
Pour décrire une forme assez complexe comme les tumeurs avec des formes irrégulières.

Les spécialistes se sont penchés vers les descripteurs spécifiques, l'utilisation de la longueur radiale normalisée a donnée un résultat encourageant, AUC égale à 0.83. Une combinaison avec les descripteurs de la texture améliore la courbe ROC avec une aire sous la courbe égale à 0.91.

### • Comparaison avec deux méthodes de la littérature

Nous comparons l'AC en utilisant LRN combiné avec les paramètres d'Haralick avec des méthodes largement utilisés dans la littérature qui sont Les méthodes K-means et hiérarchique, les algorithmes sont appliqués sur la base MIAS.

D'après les différentes courbes (figure 6.28) nous remarquons que la surface sous la courbe (figure 6.28.a) désigne le résultat de la classification hiérarchique qui représente la plus petite surface ce qui revient à la plus petite valeur de AUC.



**Figure 6.29.** Comparaison avec quelques méthodes de la littérature

La classification hiérarchique fournit des résultats de classification satisfaisants mais ils sont les moins performants dans ce teste. L'algorithme K-means fournit des résultats meilleurs par rapport à la première classification et ça se traduit par des valeurs plus importantes de l'aire sous la courbe ROC.

En comparant la classification évaluée avec la courbe ROC (figure 6.28) par la figure (6.29), la figure montre clairement que la courbe correspondant au classifieur AC se situe au dessus des autres courbes (correspondant aux K-means et hiérarchique), et ça revient aux valeurs d'AUC,  $AUC_{Hiérarchique} = 0.84 < AUC_{K-Means} = 0.85 < AUC_{AC} = 0.91$ .

#### 6.4. Conclusion

Dans ce travail nous avons proposé un algorithme de classification visuelle utilisant les automates cellulaires, nous avons montré expérimentalement que cet algorithme est capable de visualiser et présenter l'ensemble des images d'une manière informative et avec une interaction des utilisateurs avec les images.

Il peut représenter les deux groupes criés, il peut fournir une vue d'ensemble. Ainsi les résultats sont directement visualisés et organisés dans une grille de deux dimensions 2D. Les données qui sont proches les uns des autres sur la carte devraient être similaires dans le contenu et le meilleur résultat est obtenu avec la combinaison d'Haralik et LRN. Ce résultat est traduit par une valeur d'AUC qui est égale à 0.91.

Nous pensons pouvoir encore améliorer les performances de l'algorithme, du point de vue des erreurs de classification avec par exemple l'utilisation d'autres descripteurs et faire la

combinaison de ces descripteurs. Un autre point c'est l'automatisation du seuil par exemple changer le seuil selon l'écart type observé dans la similarité des images classées. Dans ce chapitre nous avons présenté une étape importante de notre travail, à l'aide de l'automate cellulaire, nous obtenons une classification des images, nous appliquons la recherche sur les images afin de trouver l'ensemble des images similaires à une image requête introduite par l'utilisateur dans le chapitre qui suit.

## Chapitre VII

---

### Indexation et Recherche des Images Par le Contenu Basée Agents

---

#### 7.1. Introduction

L'objectif de ce chapitre est de présenter le fonctionnement des systèmes de recherche interactive d'images par le contenu, ce travail est appliqué sur une base d'images médicale réelle.

Dans le troisième chapitre nous avons montré que la recherche d'images par le contenu exploitait des méthodes et des techniques du traitement des images et de la reconnaissance des formes.

Dans un premier temps, le traitement d'image offre les moyens d'extraire des attributs pertinents pour élaborer une représentation de l'image sous forme de signature (index) ou de structure (graphe). Ces attributs sont identifiés avec l'aide de l'utilisateur expert.

Une fois la base d'images est indexée, le problème est de trouver un jeu d'images semblables à l'image requête. A ce niveau là nous avons utilisé les méthodes issues de la reconnaissance des formes (mesures de similarité), nous quittons les approches classiques des systèmes de recherche car on veut prendre l'avis de l'utilisateur sur les images renvoyées. Cet avis peut alors être utilisé par le système pour effectuer une nouvelle recherche. On est donc en présence d'un système bouclé : appelé système avec boucle de pertinence.

Dans cette partie du travail, nous présentons la méthode d'indexation et recherche d'images par le contenu, contrairement aux autres méthodes la notre consiste à utiliser les automates cellulaires pour un but d'indexation et organisation des données et faire une recherche rapide et pertinente en utilisant les SMA.

Nous présentons l'aspect système de notre travail, nous introduisons dans un premiers temps, l'architecture du système qui a été développé en décrivant ses composants ainsi que le fonctionnement de chaque partie. Ensuite, nous détaillons le fonctionnement général et nous terminons par la validation, et la performance du système proposé.

## **7.2. Application de la recherche d'images en médecine**

Ce contexte concerne essentiellement la radiologie. En se limitant, dans un premier temps aux images malignes ou bénignes, il convient de construire une base d'images examinées par le médecin.

Le médecin recherche dans une base des images qui sont semblables à l'image requête et il complète sa connaissance d'une pathologie.

### **7.2.1. Description du contenu des images**

L'étape d'indexation ne peut pas se faire sans le médecin, le problème c'est de trouver une représentation pertinente de l'image et de la coder sous la forme d'un index.

Après un traitement de l'image, des attributs sont calculés sur une partie de l'image obtenue par une sélection de la zone suspecte.

### **7.2.2. Indexation**

Nous nous intéressons maintenant aux méthodes de structuration de ces informations pour la formation de la signature appelée aussi index. L'index doit résulter un compromis entre la pertinence de l'information pour chaque image et le temps de recherche dans cette base.

L'index est choisi selon la fonction visée, il existe un nombre de primitives pour faire une telle indexation, on peut mentionner : une région, des points d'intérêt, les contours...

Deux approches de structurations sont alors possibles : l'approche globale et l'approche partielle.

La première consiste à une accumulation de l'information sur toute l'image pour former la signature dite globale, deux structures de l'index global sont alors possibles: (une structure à plat, et une structure hiérarchique).

La première structure consiste à concaténer les vecteurs d'attributs de couleur, de forme, de texture, etc. pour construire le vecteur index. Dans la deuxième structure, les vecteurs d'attributs ne sont pas concaténés dans le but de définir une mesure de similarité qui correspond à chacun de ces vecteurs.

Dans l'approche partielle, l'image est représentée par un ensemble de régions obtenues après segmentation où chaque ensemble est caractérisé par un vecteur d'attributs. Dans ce cas, il

s'agit de rechercher une région ou un sous-ensemble de régions similaires dans la base d'images.

Pour représenter la topologie de l'image, il s'agit d'indexer l'image par son graphe d'adjacence. C'est un outil d'analyse complet mais complexe à gérer.

### **7.2.3. Similarité**

Nous nous intéressons dans cette partie à la question de la recherche par similarité dans une base d'images. Généralement la comparaison est effectuée entre les signatures images précédemment calculées.

Nous avons vu avant qu'un indexe présente une structure compacte des attributs extraits des images. La définition d'une mesure similarité implique le choix d'une mesure de similarité entre attributs. Nous citons les distances les plus couramment utilisées : distance Euclidienne, distance de Mahalanobis...

### **7.2.4. Interactivité**

Les index calculés hors ligne ne permettent pas toujours de caractériser efficacement la sémantique des images. Les images renvoyées par le système sont similaires à la requête du point de vue de ces index objectifs, mais leur sémantique n'a bien souvent rien à voir avec celle de l'image requête. Cette différence, dénommée fossé sémantique, limite grandement les performances des systèmes sans interaction.

L'intervention de l'utilisateur dans la recherche est proposée pour réduire voire combler ce fossé sémantique. De plus, l'Homme dans la boucle constitue un changement majeur dans l'approche de reconnaissance, dans la mesure de similarité notamment. Il doit guider le système, et les solutions sont dépendantes de son intervention.

## **7.3. Notre système et nos choix**

Nous avons choisi de concevoir un système de recherche d'images basé sur le contenu, nous utilisons les systèmes multi-agents selon l'architecture de la figure (7.4).

Nous avons utilisé une méthode de recherche par boucle de pertinence ou filtrage qui intègre l'avis de l'utilisateur dans le but d'affiner la recherche des images similaires pathologiques. Dans ce cas le système s'adresse aux spécialistes.

La méthode de recherche avec la phase bouclage de pertinence ne peut être que spécifique pour chaque application, dans notre cas le système proposé est destiné pour la recherche des images similaires à une image requête qui contient une pathologie dans une image de mammographie.

Il est difficile d'extraire des attributs de bas niveau de l'image sans avoir une connaissance sur le domaine ou l'interprétation du médecin qui se base dans son diagnostic sur des indices visuels.

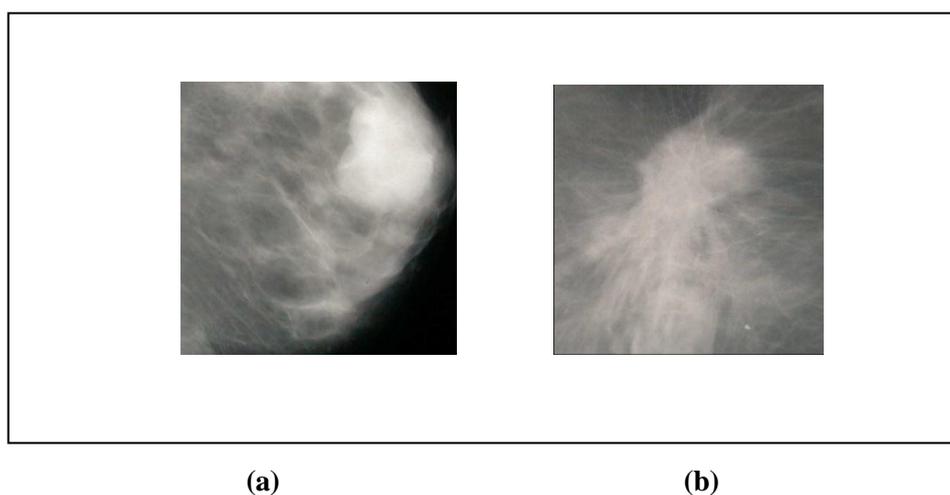
### 7.3.1. Contenu de la base

Pour la base d'Annaba, il s'agit des images mammographies, les examens radiologiques ont été fait sur des seins gauche ou droite (le sein contient une pathologie). Durant l'acquisition des images d'un patient, une série de ces coupes est enregistrée, avec le médecin nous avons seulement choisie les images qui contiennent une opacité (cf.sec. 5.2.1). La base MIAS est utilisée en parallèle pour l'évaluation de notre travail.

### 7.3.2. Indices visuels

Une phase essentielle de l'indexation des images de la base est la compréhension des indices visuels ou encore les éléments sémiologiques d'une pathologie qui sont des clés du diagnostic d'une pathologie. Ainsi la signature *indexe* extraite devra caractériser l'image en codant les informations importante et combler en partie la lacune sémantique en combinant l'interprétation humaine de haut niveau et le calcul des attributs des images de bas niveau.

Les symptômes important pour différencier les deux pathologies sont la texture de la tumeur, le tissu mammaire et les limites des opacités.



**Figure 7.1.** Pathologies mammaires : a) Kystes mammaires bénins, b) Cancer

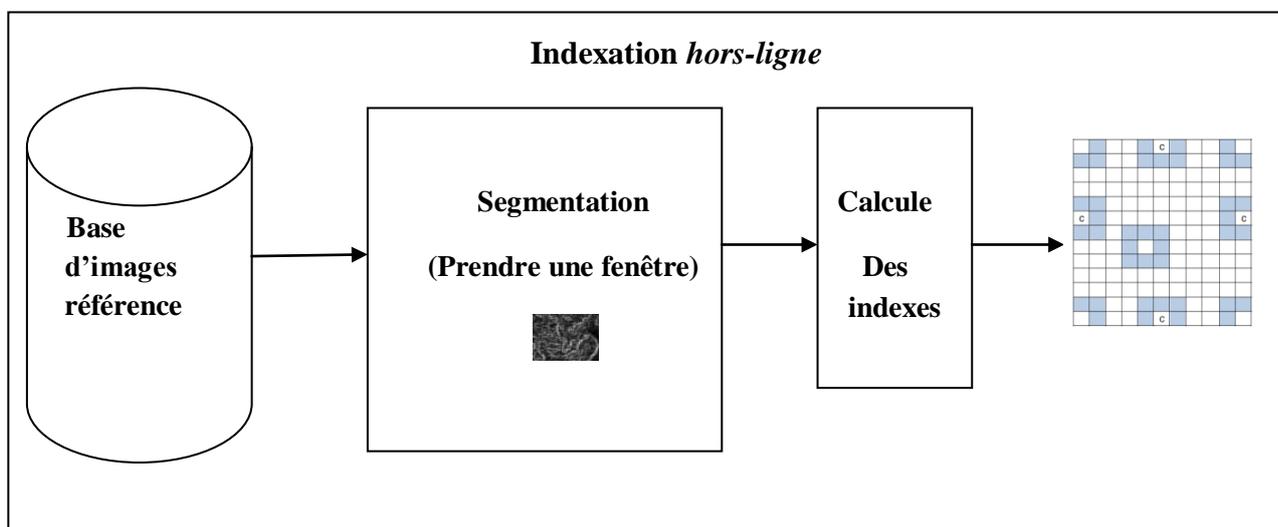
### 7.3.3. Présentation du système

Comme nous l'avons mentionné dans le chapitre *indexation et recherche d'images* (cf. Chapitre III) que le mot *indexation* recouvre les techniques qui permettent de retrouver des images dans des banques d'images.

En effet, il existe des techniques opérationnelles pour rechercher des informations à l'aide de mots-clés mais cela nécessite des informations codées à l'avance. Dans le cas des images médicales type mammographie, en accord avec l'analyse sémiologique effectuée par le médecin, les informations obtenues entreront dans la constitution d'un index qui sera utilisé pour classer les images et les retrouver efficacement.

Nous avons réalisé un système qui comprend classiquement une phase hors ligne d'indexation pour construire la base d'index.

Le fonctionnement en ligne comprend le calcul d'index sur l'image requête, et la phase de bouclage de pertinence (filtrage) pour annoter les images qui sont apparues pertinentes pour lancer une recherche qui affine les résultats de la recherche.



**Figure 7.2.** Schéma descriptif de la méthode d'indexation

L'index doit être fortement lié aux caractéristiques des images malignes et bénignes, pour cette raison la méthode d'indexation, comme illustré dans la figure (7.2) doit comporter une étape de segmentation d'images.

Une fois la phase de segmentation est établie, une phase de classification automatique (cf. chapitre VI) est proposée. Cette étape est délicate car il s'agit de faire une recherche en ligne simple et rapide.

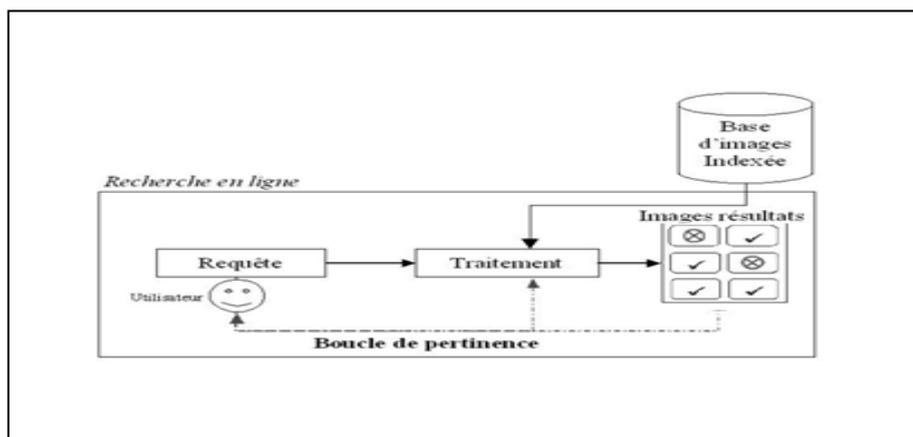
Après le prétraitement et l'indexation, nous voulons trouver, une méthode flexible de mesure de similarité et de recherche qui permet une économie de temps de recherche et un bouclage qui permet d'aider le médecin.

L'index étant choisi, et la base d'images étant indexée, le problème est de trouver un jeu d'images semblables à l'image requête. A ce niveau une mesure de similarité est introduite, notre choix c'est porté sur la distance Manhattan, vu qu'elle est largement utilisée dans les systèmes d'indexations et de recherches.

Cependant le contexte d'utilisation d'un système CBIR fait que l'on quitte les approches classiques car on veut prendre l'avis du médecin sur les images renvoyées, cet avis peut alors être utilisé par le système pour effectuer une nouvelle recherche.

Cette phase joue un rôle important, elle présente l'interface entre un système informatique de grande capacité calculatoire mais de faible capacité d'interprétation et un opérateur humain présentant les caractéristiques inverses, l'étape se déroule comme suit:

Le système de recherche présente un ensemble de résultats à l'utilisateur, l'utilisateur expertise ces résultats, ensuite le système intègre l'avis de l'utilisateur pour mettre à jour ses critères de recherche, enfin le processus est itéré si l'utilisateur est insatisfait.



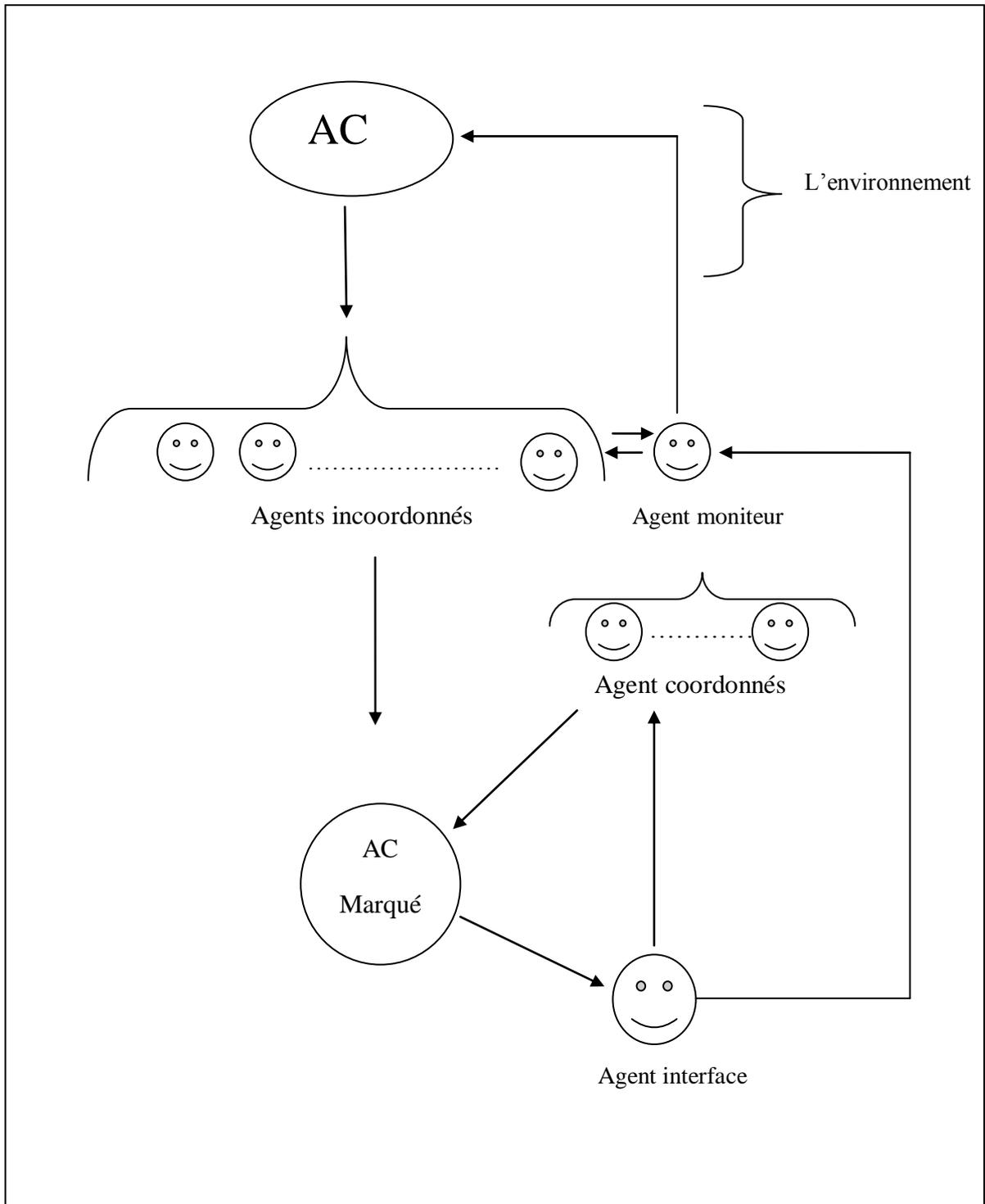
**Figure 7.3.** Schéma illustrant l'étape de raffinage

Dans ce cadre, la recherche repose sur deux entités, *la requête* et *la fonction de similarité*. La requête est un point d'ancrage dans l'espace de recherche auquel la fonction de similarité permet de comparer n'importe quelle image.

On distingue deux méthodes géométriques de bouclage de pertinence. La première c'est la mise à jour de la requête et la deuxième correspond à la modification de la mesure de similarité annotée par l'utilisateur. En fait, dans ce travail nous utilisons les deux approches.

### 7.3.4. Schéma général de la recherche

Le schéma suivant (figure 7.4) illustre le scénario du système, il est caractérisé par un environnement partiellement statique et trois catégories d'agents.



**Figure 7.4.** Architecture du système proposé

Dans cette partie nous détaillons le choix que nous avons effectué pour l'architecture de notre système, enfin nous expliquons les mesures de similarités utilisés ainsi que les caractéristiques visuelles.

L'environnement AC présente la partie statique du système, elle contient les vecteurs représentant la base d'images de mammographie. Nous proposons d'utiliser une population d'agents afin d'explorer l'automate cellulaire à la recherche des images similaires et pertinentes.

Chaque agent alors se déplace d'une cellule à une autre afin de trouver des images pertinentes, puis retourne le résultat une fois celle-ci trouvées.

Les Agents marquent leurs environnement, dans ce cas ces agents augmentent les chances pour les agents suivants ce déplace vers ces sites. Le marquage est effectué dans le but d'obtenir des chemins optimaux qui mènent à une recherche des images rapide et pertinente.

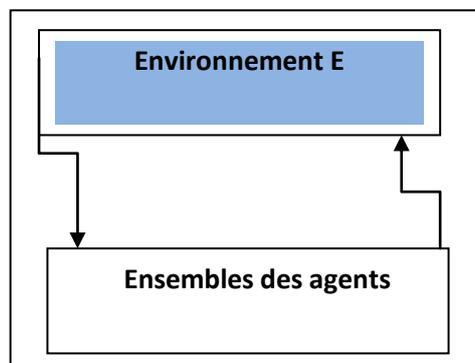
Les images de la base sont représentées par des vecteurs correspond à des caractéristiques de texture et de forme, donc la distance joue un rôle pour déterminer la pertinence de telle image et avec une interaction utilisateur le bouclage de pertinence va améliorer la recherche.

L'utilisateur commence par donner une image requête. A l'aide de l'interface agent le système construit la signature de cette requête. Le déplacement de ces agents est régi par des règles et une fonction de similarité est entraînée. Dès qu'un ensemble des images est retrouvé l'agent envoie un message contenant les vecteurs les plus similaires à celui de la requête plus une stratégie de marquage des cellules utilisées.

L'agent interface affiche l'ensemble des images similaires, l'utilisateur avec des annotations peut répéter le processus jusqu'à l'obtention d'un résultat le satisfaisant.

#### • Environnement

Un agent ne peut exister sans environnement, de même qu'on considère qu'un agent n'existe pas sans système multi-agents. L'environnement est une structure dans laquelle l'agent évolue.



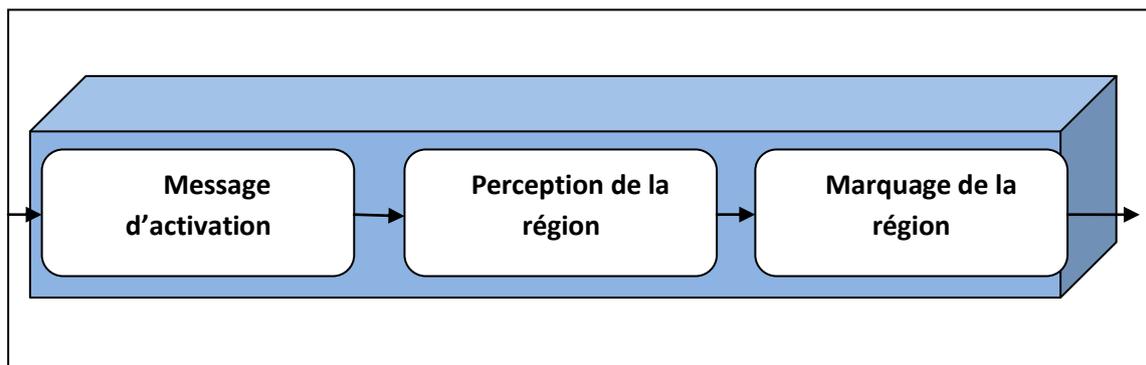
**Figure 7.5.** Environnement du système

Cet environnement (figure 7.5) présente l'automate cellulaire, chaque cellule contient un vecteur représentant l'image.

• **Agents Incoordonnés**

Les agents *Incoordonnés* se déplacent sur l'automate cellulaire dans le but de marquer les cellules.

Chaque agent de ce groupe à une vue partielle, il peut percevoir sa région et travail sur l'AC pour son propre but, les agents ne communiquent pas entre eux, ils fournissent des résultats à l'utilisateur, c'est-à-dire leurs actions dépendent uniquement de leur perception local.



**Figure 7.6.** Architecture d'un agent Incoordonné

Les agents dans le système suivent un cycle de vie classique, Perception-Décision-Action, l'architecture générale dans la figure (7.6) de ce type est proche de celle d'un agent réactif, ce type est déclenché par une requête envoyé par l'utilisateur qui présente un message d'activation.

• **Agent interface**

L'agent interface assure deux rôles importants : le premier il consiste à initialisé l'agent moniteur.

Le deuxième est de communiquer avec l'utilisateur en lui donnant la possibilité de faire une recherche pour trouver les images similaires à une requête donnée. Les résultats permettant à l'utilisateur de faire des annotations pour les images qui lui apparaissent pertinentes.

• **Agent moniteur**

Cet agent représente le centre de contrôle du système, il garde une copie de L'AC et il permet de lancer les agents incoordonnés pour le marquage des cellules. C'est-à-dire, il s'occupe d'organiser le pointage des agents incoordonnés dans l'automate cellulaire, en indiquant à ces derniers si la case (x,y) a été visitée ou non. Si l'agent de communication indique que la case est visitée, l'agent doit passer à une autre case, sinon il se pointe dans cette case.

• **Agents coordonnés**

Cet agent à le rôle de marquer les cellules lors de la phase (*feedback*), l'annotation des images par l'agent interface va lancer le processus de marquage en parallèle.

Ce type d'agent va suivre la stratégie envoyé lors de la première phase *recherche*, on doit noter que :

M est l'ensemble des cellules marquées dans la première phase sur lesquelles les agents incoordonnés vont faire la recherche. La requête Q avec K images obtenus peuvent être défini comme suit:

$$Q = \{ \langle q_1, w_1 \rangle, \langle q_t, w_t \rangle \dots \langle q_k, w_k \rangle \}$$

Avec  $w_i$  est entre 0 et 1, ce paramètre est employé pour indiquer l'importance de l'image  $q_1$  dans la requête Q lors de la notation.

Dans l'étape de bouclage chaque agent prend une image requête, ces agents coordonnés se lancent en parallèle et suivent une stratégie pour raffiner le résultat obtenu de la recherche dans la première phase.

Avec chaque image requête, nous prenons les  $n$  images les plus semblables.

$$R = \{ \langle i_1, dist_1 \rangle, \dots, \langle i_j, dist_j \rangle, \dots, \langle i_n, dist_n \rangle \}$$

La distance entre images doit satisfaire :  $dist_1 \leq \dots \leq dist_j \dots \leq dist_n$

Nous avons le résultat  $R_1$  :

$$R_1 = \{ \langle i_1, w_1 \rangle \} \text{ où } 1 \leq 1 \leq K.$$

On peut présenter le résultat obtenu de chaque sous requête après une annotation de l'utilisateur avec :

$$R_1 = \{ \langle i_{i1}, dist_1 \rangle, \dots, \langle i_{in}, dist_{in} \rangle \}$$

Où  $i_n \leq n$  et  $dist_{in}$  est la distance entre les images  $i_{in}$  obtenu de l'AC avec l'image requête  $R_1$ .

Afin d'obtenir le résultat final, nous fusionnons les  $R_i$  sous-résultat en utilisant la distance minimale.

• **Communication entre agents**

**-Communication directe**

Le moyen utilisé pour la communication entre agents est la communication directe, l'agent moniteur envoie le message aux agents coordonnés pour indiquer que les cellules sont libres ou non.

La communication est assurée par un envoi de message, ce message présente un objet qui est véhiculé entre agents, c'est l'unité de base pour l'échange des données. Essentiellement composé d'une enveloppe et un corps.

L'enveloppe se compose de :

- Un en-tête,
- Informations sur la source ;
- Information sur la destination.

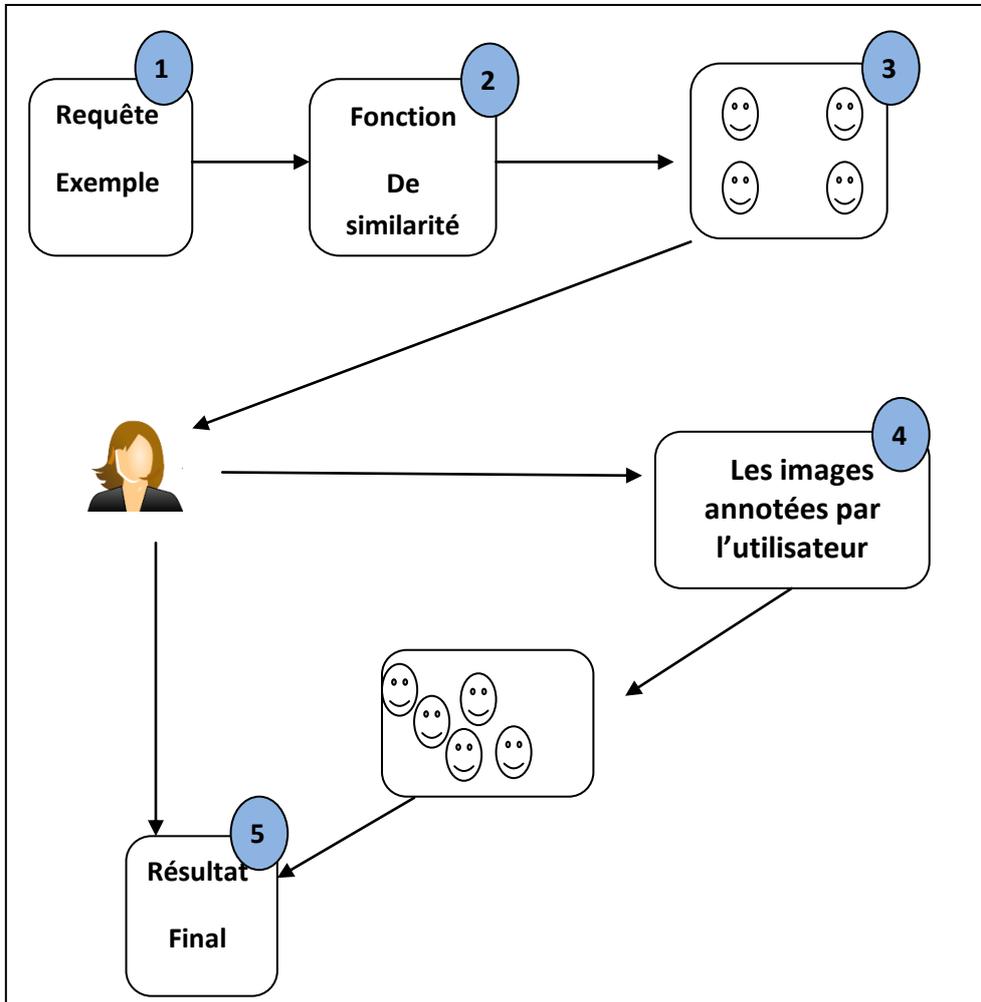
Le corps est variable, selon l'information du contenu.

**-Communication indirecte**

Les agents coordonnés et incoordonnés marquent l'AC qui sera par la suite contrôlé par l'agent moniteur, cela reflète une communication indirecte par l'utilisation de l'environnement.

### 7.3.5. Fonctionnement de l'ensemble du système

Nous détaillons le fonctionnement du système comme suit :



**Figure 7.7.** Fonctionnement du système

La figure (7.7) montre le déroulement du processus recherche d'images: l'utilisateur présente l'image requête, étant l'exemple de la catégorie recherchée à l'aide de l'agent interface (1), le système construit la signature de cette image de la même façon que cela à été fait pour les images de la base (2), et un signe d'activation est passé aux agents incoordonnés (3), ces agents sont lancés sur l'AC pour la recherche des images correspondantes à cette requête selon la première stratégie. Dès que les images similaires sont trouvées un message est envoyé à l'agent interface qui contient les images similaires ainsi que la stratégie utilisée pour le marquage des zones.

L'agent interface sélectionne les images annotées par l'utilisateur parmi l'ensemble des images retournées (4).

L'utilisateur annoté les images présentées, et ces annotations sont utilisées à la fois pour améliorer la mesure de similarité ainsi que l'amélioration du chemin utilisé par les agents coordonnés, de manière à se déplacer vers les cellules qui contiennent vraiment les signatures intéressantes du point de vue de l'utilisateur, les agents coordonnés sont relancés par la mesure de similarité et le chemin est mis à jour.

Cette boucle, l'annotation de l'utilisateur et le parcours des agents coordonnés va se répéter plusieurs fois jusqu'à ce que l'utilisateur soit satisfait du résultat obtenu. Les résultats finaux sont affichés à l'utilisateur (5).

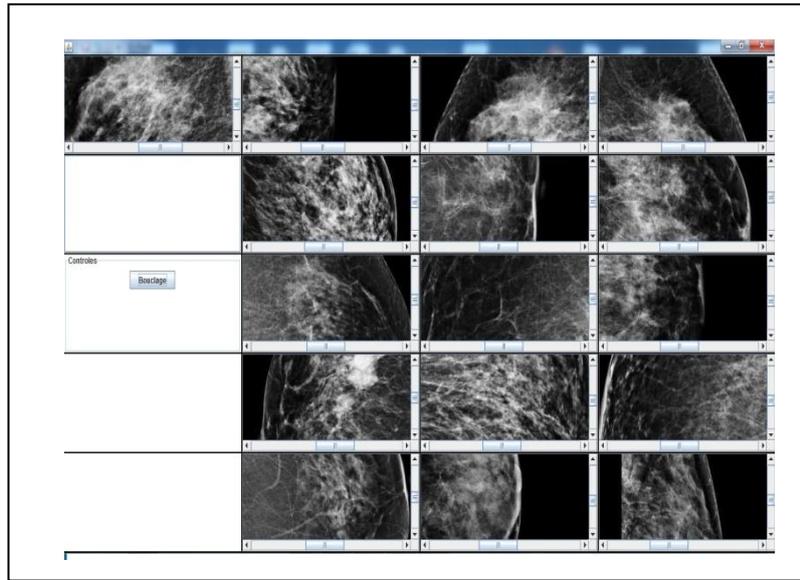
#### **7.4. Résultats, discussion et évaluation**

Dans cette partie nous présentons les expériences menées sur la stratégie proposée précédemment. Nous présentons dans un premier temps quelques résultats obtenus sur les deux bases DDMA et MIAS, ensuite nous évaluons les résultats obtenus.

L'évaluation de la performance du système se fait avec les outils classiques de la recherche d'images par le contenu, la méthode MAP (*Mean Average Precision*).

Pour présenter les résultats, nous avons développé un système de recherche d'images par le contenu basé SMA. La figure (7.8) décrit l'interface de notre système. Cette interface se divise en zone de requête et une zone de résultats.

Avec les deux bases DDMA et MIAS nous avons utilisé pour l'évaluation 20 requêtes, 10 images mammographies contiennent des masses malignes et 10 bénignes le reste des deux bases présentent les images sur lesquelles nous faisons la recherche (100 images et 70 images entre malignes et bénignes pour les deux bases MIAS et DDMA respectivement).



**Figure 7.8.** Résultat d’une recherche d’images mammographiques malignes dans la base DDMA par la texture

Lors d’une évaluation de système toute image retrouvée appartenant à la classe de l’image requête est comptée comme correcte, par exemple une basse précision pour un rappel élevé indique que le système capture difficilement des images de la classe de l’image requête.

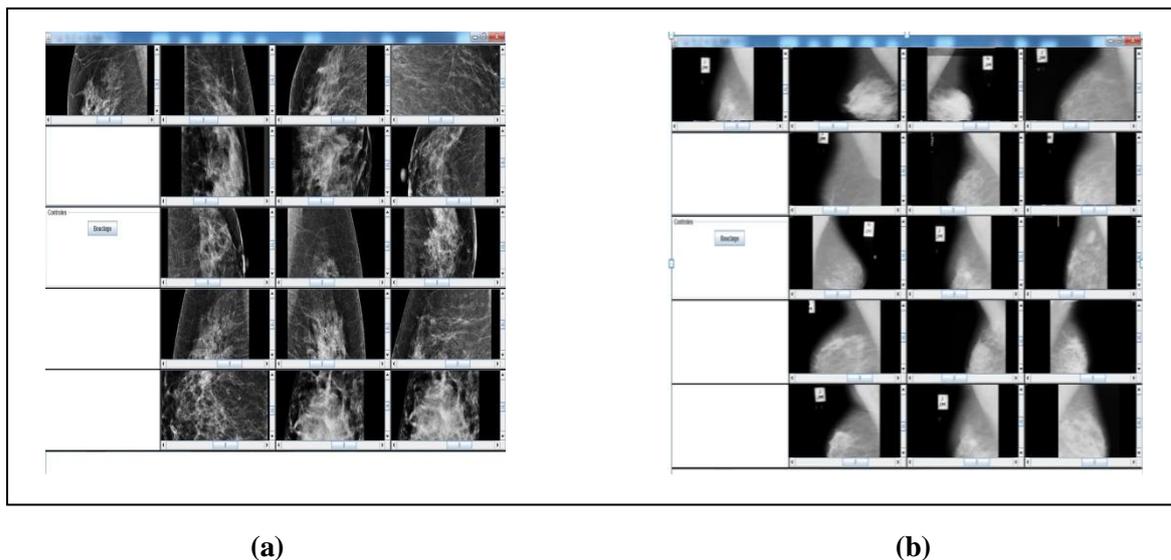
Dans le tableau (7.1) nous expliquons la méthode d’évaluation que nous utilisons pour évaluer nos travaux.

Les expériences ont permis de valider la stratégie proposée, plus les images recherchées sont précisément localisées, plus notre stratégie offre un gain dans la qualité des résultats retrouvés.

**Tableau 7.1.** MAP pour une requête maligne de la base DDMA

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
<b>Rappel</b>	$\frac{1}{44}$	$\frac{1}{44}$	$\frac{2}{44}$	$\frac{2}{44}$	$\frac{1}{44}$	$\frac{3}{44}$	$\frac{1}{44}$	$\frac{1}{44}$	$\frac{1}{44}$	$\frac{1}{44}$	$\frac{1}{44}$	$\frac{1}{44}$	$\frac{9}{44}$	$\frac{1}{44}$	$\frac{1}{44}$
<b>Précision</b>	1	1	$\frac{2}{3}$	$\frac{2}{4}$	1	$\frac{3}{6}$	1	1	1	1	1	1	$\frac{9}{13}$	1	1

La méthode d'évaluation est appliquée sur les requêtes des deux bases utilisant les descripteurs texture, forme et la combinaison des deux. La moyenne des moyennes est calculée et les résultats sont présentés ultérieurement. L'ensemble des courbes, une pour chaque requête sont combinées pour donner une courbe moyenne interpolée qui permet de mesurer les performances du système.



**Figure 7.9.** Résultat d'une recherche d'images mammographies malignes dans DDMA et MIAS (combinaison Haralick et LRN)

La figure (7.9) montre les résultats d'un exemple d'une recherche par le contenu dans la base DDMA. La mammographie requête présente un cas pathologique.

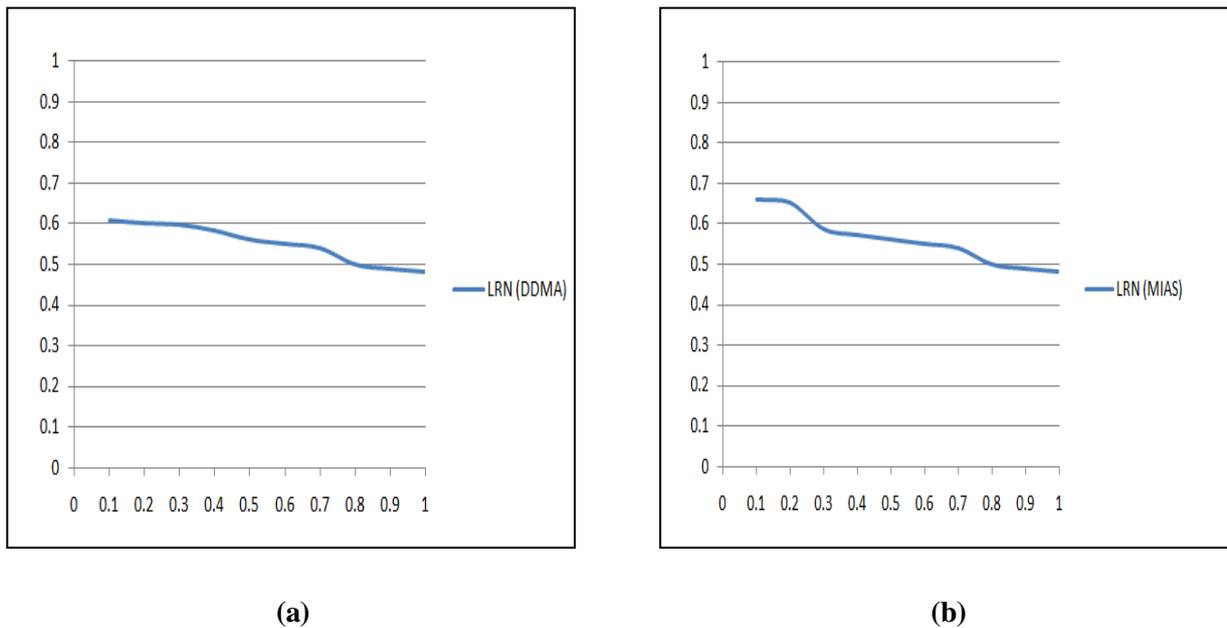
Les premières images données dans cette figure sont représentées dans l'ordre décroissant (de gauche à droite) en matière de similitude, un graphe illustrative est donné ultérieurement pour une explication des résultats.

**Tableau 7.2.** MAP des descripteurs utilisés sur les deux bases MIAS et DDMA

	Base MIAS			Base DDMA		
	Haralick	LRN	Haralick/LRN	Haralick	LRN	Haralick/LRN
<b>Rappel</b>	0.19	0.22	0.25	0.15	0.33	0.20
<b>Précision</b>	0.73	0.66	0.88	0.70	0.60	0.78

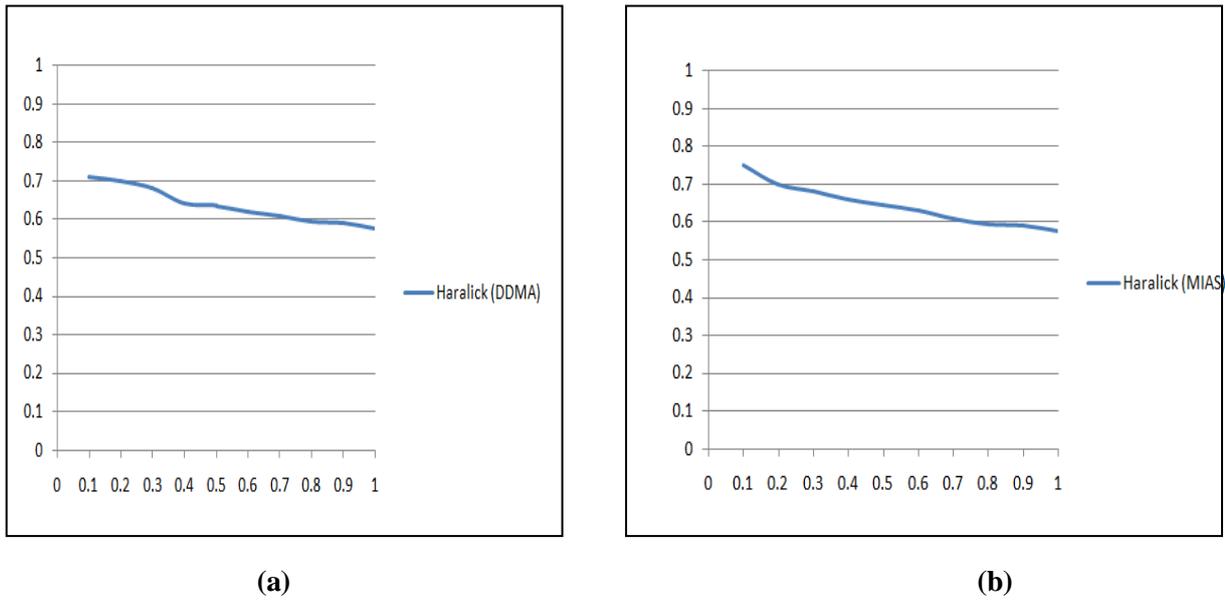
Nous trouvons qu’avec l’ensemble des images requêtes, la combinaison des descripteurs d’Haralick et les descripteurs spécifiques LRN sont les meilleurs descripteurs (tableau 7.2), L’LRN est assez efficace et la texture présente un descripteur fort.

Les figures (7.10; 7.11 et 7.12) donnent une vue globale de la performance des résultats avec les deux bases. Nous pouvons trier les caractéristiques en ordre décroissant de leurs efficacités comme suit: les paramètres LRN, Haralick et la combinaison des deux caractéristiques (Haralick/LRN).



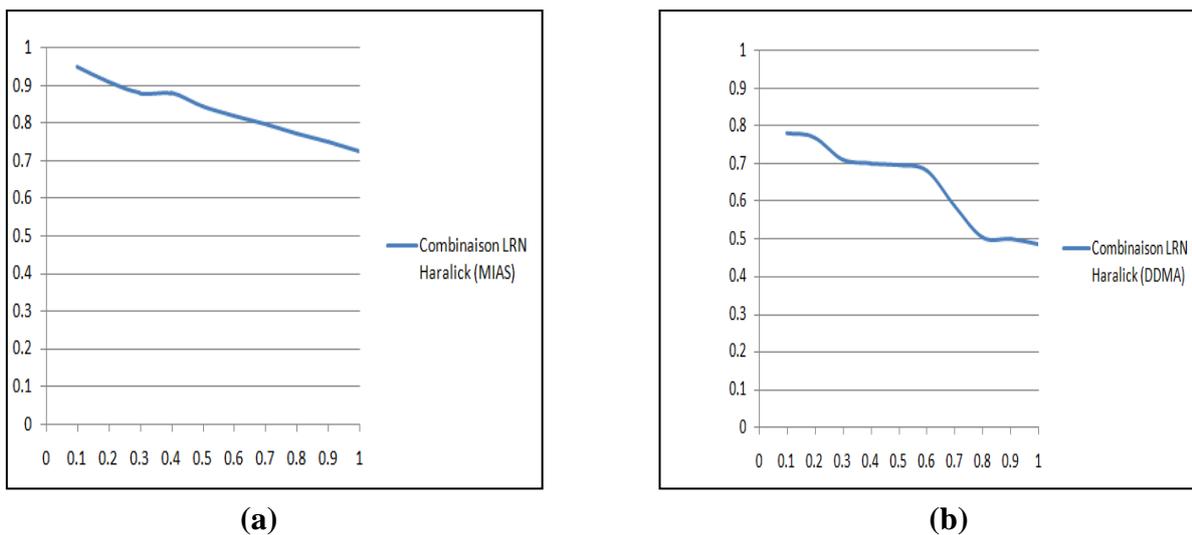
**Figure 7.10.** Courbe moyenne de rappel et précision du descripteur LRN, a) la base DDMA, b) la base MIAS

La figure de rappel et précision (figure 7.10.b), utilisant les descripteurs spécifiques LRN sur la base MIAS montre une précision qui arrive à 0.66 comparant à une précision des mêmes descripteurs sur la base DDMA la valeur été 0.61 (figure 7.10.a). A partir de ces résultats nous pouvons constater que les descripteurs spécifiques sont moyennement performants pour les deux bases, ainsi les résultats sur la base DDMA sont faible par rapport celles de MIAS vue la base qui influence sur la segmentation et bien sur les résultats de la recherche.



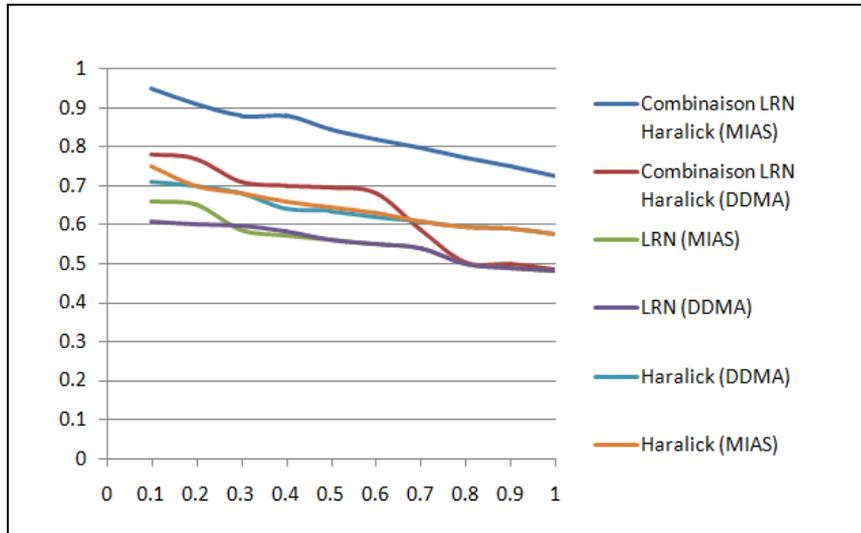
**Figure 7.11.** Courbe moyenne de rappel et précision du descripteur Haralick, a) la base DDMA, b) la base MIAS

Pour la texture, la précision est égale à 0.75 (7.11.b), les descripteurs d’Haralick sont performants dans l’ensemble sur les deux bases vue que nous nous sommes intéressé à l’organisation spatiale des pixels des masses, une combinaison avec d’autres descripteurs peut améliorer les résultats.



**Figure 7.12.** Courbe moyenne de rappel et précision des descripteurs combinés Haralick et LRN, a) la base MIAS, b) la base DDMA

Si nous comparons la combinaison des descripteurs avec une utilisation seul des descripteurs LRN et Haralick, nous trouvons que la combinaison donne de meilleure résultats (figure 7.12).



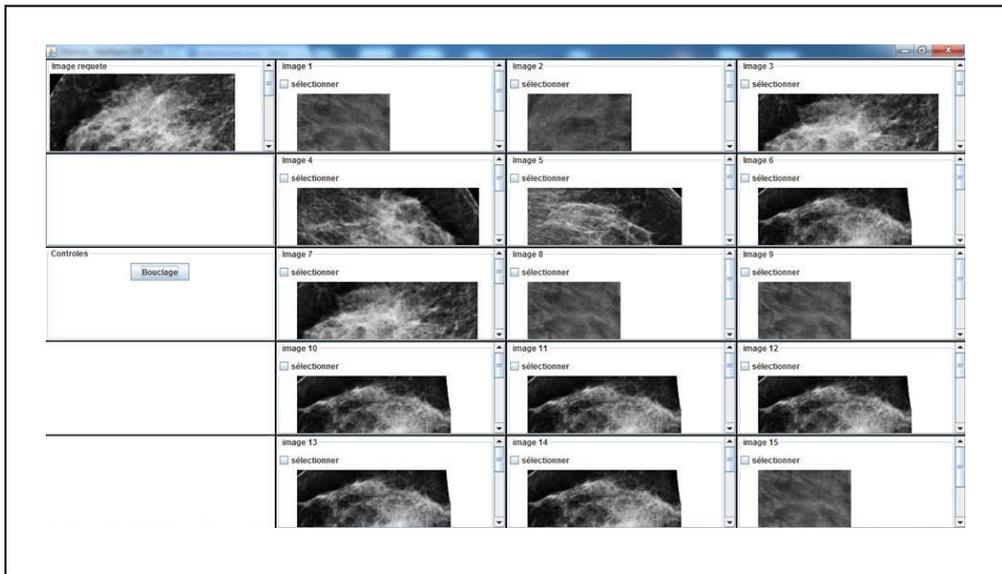
**Figure 7.13.** Courbe moyenne de rappel et précision des descripteurs sur DDMA et MIAS

Donc, dans cette d'évaluation du système nous avons utilisé plusieurs requêtes. Il y a lieu donc de calculer la précision moyenne PM pour toutes les requêtes correspondant à chaque niveau de rappel.

Le tableau (7.2) précédent résume les valeurs moyennes obtenues par notre système après plusieurs tests. La pertinence dans notre cas est définie sur la base des deux caractéristiques : la texture et la forme des masses.

La figure (7.13) résume les valeurs totales calculées dans nos testes. Les résultats obtenus sur les deux bases sont satisfaisants, l'utilisation de la texture seule ne donne pas de bonne résultats et reste insuffisante, la combinaison avec des descripteurs spécifiques surtout le taux de croisement en zéro donne de bons résultats avec une précision moyenne égale à 0.95.

Dans la phase de bouclage de pertinence ou raffinage des résultats obtenus qui est une étape optionnelle dans le système de recherche.



**Figure 7.14.** Exemple des images à annoter

Nous appliquons la stratégie que nous avons présenté dans la section (cf.sec. 7.3), l'utilisateur va annoter des images (figure 7.14) pour faire une autre recherche dans le même ensemble obtenus. Nous avons testé avec les deux bases, les résultats obtenus présentés à un spécialiste du domaine montrent que les images sont satisfaisantes et peut l'aider à préciser son diagnostic à l'image originale ou la requête initiale.

## 7.5. Conclusion

Dans ce travail, nous nous sommes intéressés à la conception d'un système de recherche par le contenu, en appliquant notre proposition sur des images de mammographies.

L'architecture de notre système est classique. Elle est formée de deux parties : l'indexation qui s'effectue *hors ligne* et la recherche qui s'effectue *en ligne*.

Le fossé sémantique, qui sépare les attributs bas-niveaux calculés hors ligne dans les images et le sens que l'utilisateur place dans sa requête, conduit à des résultats de recherche erronés.

Dans la phase en ligne nous expliquons l'utilisation des AC comme un environnement pour le développement des SMA. Ensuite, dans la phase hors ligne l'architecture du système multi-agents est détaillée.

Dans la deuxième phase nous intégrons l'expertise de l'utilisateur dans la recherche pour réduire le fossé sémantique et reprendre aux besoins grâce à ses annotations.

---

## Conclusion générale

---

Notre travail porte sur l'indexation et la recherche d'images par le contenu sur des bases médicales. Dans ce cadre, nous n'avons pas seulement travaillé sur une méthode qui organise les images d'une façon à faciliter la recherche *Automates cellulaires*, mais aussi de construire une mesure de similarité entre les images basées sur le contenu comme dans le cadre classique de la recherche d'images CBIR, ensuite trouver les cellules de cet automate hébergeant les images pertinentes répondant au souhait de l'utilisateur.

Notre architecture se base sur un système multi-agents, les agents se comportent de manière à ce que la construction soit faite avec la stratégie de recherche et le raffinement des résultats obtenus.

Quatre types d'agents ont été utilisés : agents interface qui est un intermédiaire entre l'utilisateur et le système, un agent moniteur qui lance le processus de recherche, des agents incoordonnées qui prennent le rôle de marquage des zones, et un type d'agents coordonnées qui suit la stratégie indiquée par le troisième type afin de raffiner le résultat.

Aussi, nous nous sommes intéressés à la segmentation. À la classification par AC, cette méthode est appliquée sur les deux bases de tests (DDMA et MIAS) avec différents types de descripteurs.

Les résultats obtenus sont satisfaisants, la combinaison des descripteurs LRN et Haralick ont donné une valeur d'AUC égale à 0.91. L'AC obtenu est utilisé dans le but d'organiser les données pour faire une recherche par le contenu, simple, rapide et appliqué aux images de mammographies.

Le système CBIR demeure un des problèmes des plus importants qui trouve des applications potentielles en médecine, étant donné la place que l'image numérique médicale occupe à présent dans le quotidien des hôpitaux.

L'architecture de notre système est classique, elle est formée de deux parties : une phase *hors ligne* qui présente l'indexation des images et une phase *en ligne* qui est une recherche par le contenu. Le *fossé sémantique*, qui sépare les attributs bas-niveaux sont calculés hors ligne dans les images et le sens que l'utilisateur place dans sa requête, conduit à des résultats de recherche erronés.

L'intégration de l'expertise de l'utilisateur *bouclage de pertinence* dans la recherche va réduire le fossé sémantique et améliorer les résultats grâce aux annotations effectuées.

Le système réalisé utilise le principe de recherche par boucle de pertinence pour intégrer les connaissances du médecin ou de l'utilisateur au sens large dans la procédure de recherche.

La nouvelle approche que nous avons introduite est d'utiliser les automates cellulaires, on a abordé le problème de la recherche et indexation sous l'angle de la séparation des images malignes et bénignes dans l'espace des index.

Il s'agissait de construire un classifieur permettant de grouper les images, une matrice de cooccurrence est appliqué avec des paramètres du deuxième ordre.

Pour l'indexation, en ce qui concerne la texture des fenêtres présent en se basant sur les indices visuels décrits par des médecins et une matrice de cooccurrence est appliquée avec des paramètres du deuxième ordre, pour la forme une étape de segmentation est faite et des descripteurs de forme sont appliqués afin de trouver le meilleur résultat.

Dans le but d'obtenir un index pour toutes les images afin d'accélérer la procédure de recherche en ligne, nous avons classé *classification non supervisée* les régions de la base par la carte auto-organisatrice *Automates cellulaires* (AC).

Ensuite l'index d'une image est calculé en fonction d'une distance calculée par rapport aux autres images de la base.

La plupart des systèmes CBIR orientés vers le domaine médicale opérant d'une façon semi automatique en indexation, prenant l'exemple de Assert, l'utilisateur doit délimiter manuellement la d'intérêt dans l'image, ce qui permettra par la suite le calcul d'un vecteur index traduisant au mieux les informations pertinentes de l'image.

L'évaluation a été réalisée sur la base DDMA, étiquetée par un médecin radiologue et nous même, on distingue deux classes, deux cas pathologique (malignes, bénignes).

La construction de cette base à nécessité beaucoup de temps pour trouver un nombre suffisant d'images pathologiques, la base MIAS est utilisée en parallèle pour l'évaluation du travail.

Selon l'avis du médecin de service les résultats de l'évaluation du système sont satisfaisants.

### **Perspectives de recherche**

- Le problème essentiel est la création de la base expertisée qui demande du temps et la connaissance et disponibilité du médecin.
- Notre méthode d'indexation est basée sur une sélection d'une fenêtre pour la texture et un seul algorithme pour la segmentation, Il serait intéressant d'appliquer différentes méthodes de segmentation automatique plus robustes aux bruits.
- L'extension du travail à un réseau.
- Nous pensons à long terme, d'utiliser des images de mammographie de contrôle, et de les comparer avec les résultats obtenus par notre système, avec un certain pourcentage pour que les résultats soient décisifs.
- La comparaison avec d'autres approches offre plusieurs sujets de réflexions.
- Développer un protocole d'évaluation qui permet d'évaluer et de comparé la phase interactive.
- Dans la classification, nous pouvons aussi utiliser d'autres types d'automates cellulaires, même d'intervenir sur ses paramètres.

---

# Annexes

---

## Annexe A

### (1) Analyse de résultats de la classification

La courbe ROC ( *Receiver Operating Characteristic*) est une méthode de représentation graphique des performances d'un classifieur à deux classes, qui est incontournable dans les méthodes d'évaluation et particulièrement dans les applications médicales.

Cette courbe représente la sensibilité en ordonnée en fonction de la quantité (1-spécificité) en abscisse.

En effet la sensibilité présente le taux des vrais positifs (TVP), c'est-à-dire elle représente la capacité d'un examen diagnostique à fournir un résultat positif en présence de la maladie. La spécificité (1-TFP) représente la capacité d'un examen à fournir un résultat négatif en absence de la maladie :

$$TVP = \text{Sensibilité} = \frac{VP}{VP + FN}$$

$$TFP = 1 - \text{Spécificité} = \frac{FP}{FP + VN}$$

Avec :

- VP (Vrai Positif) : est le nombre de lésions malignes qui sont classées malignes.
- FP (Faux Positif) : est le nombre de lésions bénignes qui sont classées malignes.
- VN (Vrai Négatif) : est le nombre de lésions bénignes qui sont classées bénignes.
- FN (Faux Négatif) : est le nombre de lésions malignes qui sont classées bénignes.

**(2) Analyses de la courbe ROC**

La base DDMA

L'Aire

**Tableau (1) Case Processing Summary**

Aire	Valid N (listwise)
Positive(a)	40
Negative	50
Missing	33

**Tableau (2) Area Under the Curve**

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.688	.045	.012	.575	.800

**Tableau (3) Coordinates of the Curve**

Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	.675	.300
2.0000	.000	.000

Le Périmètre

**Tableau (4) Case Processing Summary**

Périmètre	Valid N (listwise)
Positive(a)	39
Negative	51
Missing	33

**Tableau (5) Area Under the Curve**

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.621	.050	.031	.514	.748

**Tableau (6) Coordinates of the Curve**

Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
-1.0000	1.000	1.000

.5000	.615	.353
2.0000	.000	.000

**Compacité****Tableau (7) Case Processing Summary**

Compacité	Valid N (listwise)
Positive(a)	42
Negative	48
Missing	33

**Tableau (8) Area Under the Curve**

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.554	.060	.033	.434	.673

**Tableau (9) Coordinates of the Curve**

Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	.524	.417
2.0000	.000	.000

**Excentricité****Tableau (10) Case Processing Summary**

Excentricité	Valid N (listwise)
Positive(a)	39
Negative	51
Missing	33

**Tableau (11) Area Under the Curve**

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.586	.058	.032	.467	.705

**Tableau (12) Coordinates of the Curve**

Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
-1.0000	1.000	1.000

.5000	.564	.392
2.0000	.000	.000

**Rectangularité****Tableau (13) Case Processing Summary**

rectangularité	Valid N (listwise)
Positive(a)	46
Negative	44
Missing	33

**Tableau (14) Area Under the Curve**

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.490	.075	.085	.370	.610

**Tableau (15) Coordinates of the Curve**

Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	.457	.477
2.0000	.000	.000

**Rectangularité modifiée****Tableau (16) Case Processing Summary**

rectangularité_modifiée	Valid N (listwise)
Positive(a)	41
Negative	49
Missing	33

**Tableau (17) Area Under the Curve**

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.542	.061	.036	.422	.662

**Tableau (18) Coordinates of the Curve**

Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	.512	.429
2.0000	.000	.000

**Haralick****Tableau (19) Case Processing Summary**

Haralick	Valid N (listwise)
Positive(a)	41
Negative	49
Missing	33

**Tableau (20) Area Under the Curve**

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.654	.049	.030	.539	.769

**Tableau (21) Coordinates of the Curve**

Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	.634	.327
2.0000	.000	.000

**LRN****Tableau (22) Case Processing Summary**

LRN	Valid N (listwise)
Positive(a)	47
Negative	43
Missing	33

**Tableau (23) Area Under the Curve**

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.591	.053	.035	.473	.709

**Tableau (24) Coordinates of the Curve**

Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
---	-------------	-----------------

-1.0000	1.000	1.000
.5000	.553	.372
2.0000	.000	.000

**LRN/Hu****Tableau (25) Case Processing Summary**

Hu_LRN	Valid N (listwise)
Positive(a)	45
Negative	45
Missing	150

**Tableau (26) Area Under the Curve**

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.700	.044	.011	.590	.810

**Tableau (27) Coordinates of the Curve**

Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	.667	.267
2.0000	.000	.000

**Haralick/LRN****Tableau (28) Case Processing Summary**

Combinaison_Haralick_LRN	Valid N (listwise)
Positive(a)	43
Negative	47
Missing	33

**Tableau (29) Area Under the Curve**

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.766	.042	.010	.664	.868

**Tableau (30) Coordinates of the Curve**

Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	.744	.213
2.0000	.000	.000

**Haralick/LRN/La combinaison****Tableau (31) Case Processing Summary**

Diagnostic	Valid N (listwise)
Positive(a)	42
Negative	48
Missing	33

**Tableau (32) Area Under the Curve**

Test Result Variable(s)	Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Combinaison_Haralick_LRN	.766	.042	.010	.665	.868
Haralick	.653	.049	.030	.539	.768
LRN	.591	.053	.035	.473	.709

**Tableau (33) Coordinates of the Curve**

Test Result Variable(s)	Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
Combinaison_Haralick_LRN	-1.0000	1.000	1.000
	.5000	.762	.229
	2.0000	.000	.000
Haralick	-1.0000	1.000	1.000
	.5000	.619	.313
	2.0000	.000	.000
LRN	-1.0000	1.000	1.000
	.5000	.619	.438
	2.0000	.000	.000

**L'Aire****Tableau (34) Case Processing Summary**

Aire	Valid N (listwise)
Positive(a)	62
Negative	58

**Tableau (35) Area Under the Curve**

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.585	.052	.032	.483	.687

**Tableau (36) Coordinates of the Curve**

Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	.532	.362
2.0000	.000	.000

**Le périmètre****Tableau (37) Case Processing Summary**

Périmètre	Valid N (listwise)
Positive(a)	61
Negative	59

**Tableau (38) Area Under the Curve**

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.609	.052	.031	.508	.710

**Tableau (39) Coordinates of the Curve**

Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	.557	.339
2.0000	.000	.000

**L'excentricité****Tableau (40) Case Processing Summary**

Excentricité	Valid N (listwise)
Positive(a)	60
Negative	60

**Tableau (41) Area Under the Curve**

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval
------	---------------	--------------------	------------------------------------

			Lower Bound	Upper Bound
.800	.042	.07	.717	.883

**Tableau (42) Coordinates of the Curve**

Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	.750	.150
2.0000	.000	.000

**Compacité****Tableau (43) Case Processing Summary**

Compacité	Valid N (listwise)
Positive(a)	62
Negative	58

**Tableau (44) Area Under the Curve**

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.785	.043	.019	.700	.870

**Tableau (45) Coordinates of the Curve**

Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	.726	.155
2.0000	.000	.000

**La Rectangularité****Tableau (46) Case Processing Summary**

Rectangularité	Valid N (listwise)
Positive(a)	55
Negative	65
Missing	3

**Tableau (47) Area Under the Curve**

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.689	.050	.028	.592	.785

**Tableau (48) Coordinates of the Curve**

Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	.655	.277
2.0000	.000	.000

**La Rectangularité modifier****Tableau (49) Case Processing Summary**

Rectangularité	Valid N (listwise)
Positive(a)	59
Negative	58
Missing	6

**Tableau (50) Area Under the Curve**

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.735	.048	.025	.643	.828

**Tableau (51) Coordinates of the Curve**

Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	.695	.224
2.0000	.000	.000

**Haralick****Tableau (52) Case Processing Summary**

Haralick	Valid N (listwise)
Positive(a)	50
Negative	70

**Tableau (53) Area Under the Curve**

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.790	.044	.018	.716	.884

**Tableau (54) Coordinates of the Curve**

Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	.800	.200
2.0000	.000	.000

**LRN****Tableau (55) Case Processing Summary**

LRN	Valid N (listwise)
Positive(a)	53
Negative	67

**Tableau (56) Area Under the Curve**

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.830	.038	.015	.764	.917

**Tableau (57) Coordinates of the Curve**

Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	.830	.149
2.0000	.000	.000

**La base MIAS****LRN/Hu****Tableau (58) Case Processing Summary**

Hu_LRN	Valid N (listwise)
Positive(a)	58
Negative	62

**Tableau (59) Area Under the Curve**

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.865	.023	.012	.794	.937

**Tableau (60) Coordinates of the Curve**

Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
---	-------------	-----------------

-1.0000	1.000	1.000
.5000	.828	.097
2.0000	.000	.000

**LRN/Haralick****Tableau (61) Case Processing Summary**

Combinaison_Haralick_LRN	Valid N (listwise)
Positive(a)	54
Negative	66

**Tableau (62) Area Under the Curve**

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.916	.020	.007	.858	.974

**Tableau (63) Coordinates of the Curve**

Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
-1.0000	1.000	1.000
.5000	.907	.076
2.0000	.000	.000

**LRN/Haralick/La combinaison****Tableau (64) Case Processing Summary**

Diagnostic	Valid N (listwise)
Positive(a)	54
Negative	66
Missing	123

**Tableau (65) Area Under the Curve**

Test Result Variable(s)	Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Haralick	.795	.044	.018	.710	.880
LRN	.839	.038	.015	.762	.916
Combinaison_Haralick_LRN	.916	.020	.007	.858	.974

**Tableau (66) Coordinates of the Curve**

Test Result Variable(s)	Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
Haralick	-1.0000	1.000	1.000
	.5000	.741	.152
	2.0000	.000	.000
LRN	-1.0000	1.000	1.000
	.5000	.815	.136
	2.0000	.000	.000
Combinaison_Haralick_LRN	-1.0000	1.000	1.000
	.5000	.907	.076
	2.0000	.000	.000

**La base DDMA****La comparaison de tous les descripteurs****Tableau (67) Case Processing Summary**

Diagnostic	Valid N (listwise)
Positive(a)	42
Negative	48
Missing	174

**Tableau (68) Area Under the Curve**

Test Result Variable(s)	Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Aire	.686	.045	.012	.574	.798
Périmètre	.629	.050	.031	.513	.746
Compacité	.554	.060	.033	.434	.673
Excentricité	.585	.058	.032	.466	.703
Rectangularité	.490	.075	.085	.369	.610
Rectangularité_Modifiée	.542	.061	.036	.422	.662
Haralick	.653	.049	.030	.539	.768
LRN	.591	.053	.035	.473	.709
LRN_Haralick	.766	.042	.010	.665	.868
Hu_LRN	.701	.044	.011	.591	.811

**Tableau (69) Coordinates of the Curve**

Test Result Variable(s)	Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
Aire	-1.0000	1.000	1.000
	.5000	.643	.271
	2.0000	.000	.000

Périmètre	-1.0000	1.000	1.000
	.5000	.571	.313
	2.0000	.000	.000
Compacité	-1.0000	1.000	1.000
	.5000	.524	.417
	2.0000	.000	.000
Excentricité	-1.0000	1.000	1.000
	.5000	.524	.354
	2.0000	.000	.000
Rectangularité	-1.0000	1.000	1.000
	.5000	.500	.521
	2.0000	.000	.000
Rectangularité_Modifiée	-1.0000	1.000	1.000
	.5000	.500	.417
	2.0000	.000	.000
Haralick	-1.0000	1.000	1.000
	.5000	.619	.313
	2.0000	.000	.000
LRN	-1.0000	1.000	1.000
	.5000	.619	.438
	2.0000	.000	.000
LRN_Haralick	-1.0000	1.000	1.000
	.5000	.762	.229
	2.0000	.000	.000
Hu_LRN	-1.0000	1.000	1.000
	.5000	.714	.313
	2.0000	.000	.000

### La base MIAS

#### La comparaison de tous les descripteurs

Tableau (70) Case Processing Summary

Diagnostic(b)	Valid N (listwise)
Positive(a)	54
Negative	66
Missing	264

Tableau (71) Area Under the Curve

Test Result Variable(s)	Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Aire	.586	.058	.032	.483	.688
Périmètre	.610	.055	.031	.509	.712
Compacité	.788	.047	.019	.703	.872

Excentricité	.803	.041	.07	.721	.885
Rectangularité	.689	.050	.028	.593	.786
Rectangularité_Modifiée	.736	.048	.025	.644	.827
Haralick	.795	.044	.018	.710	.880
LRN	.839	.038	.015	.762	.916
Haralick_LRN	.916	.020	.007	.858	.974
Hu_LRN	.869	.023	.012	.799	.939

**Tableau (72) Coordinates of the Curve**

Test Result Variable(s)	Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
Aire	-1.0000	1.000	1.000
	.5000	.611	.439
	2.0000	.000	.000
Périmètre	-1.0000	1.000	1.000
	.5000	.630	.409
	2.0000	.000	.000
Compacité	-1.0000	1.000	1.000
	.5000	.833	.258
	2.0000	.000	.000
Excentricité	-1.0000	1.000	1.000
	.5000	.833	.227
	2.0000	.000	.000
Rectangularité	-1.0000	1.000	1.000
	.5000	.667	.288
	2.0000	.000	.000
Rectangularité_Modifiée	-1.0000	1.000	1.000
	.5000	.759	.288
	2.0000	.000	.000
Haralick	-1.0000	1.000	1.000
	.5000	.741	.152
	2.0000	.000	.000
LRN	-1.0000	1.000	1.000
	.5000	.815	.136
	2.0000	.000	.000
Haralick_LRN	-1.0000	1.000	1.000
	.5000	.907	.076
	2.0000	.000	.000
Hu_LRN	-1.0000	1.000	1.000
	.5000	.889	.152
	2.0000	.000	.000

## Annexe B

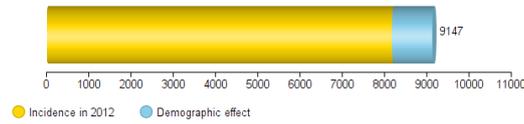
### (1) Des prédictions sur l'évolution du cancer du sein en Algérie (2015)



Algeria Breast cancer				
Year	Estimated number of new cancers (all ages)	Male	Female	Both sexes
2012	ages < 65	-	8177	-
	ages >= 65	-	995	-
2015	ages < 65	-	9147	-
	ages >= 65	-	1047	-
Demographic change				
	ages < 65	-	970	-
	ages >= 65	-	918	-
		-	52	-

Population forecasts were extracted from the United Nations, World Population prospects, the 2012 revision. Numbers are computed using age-specific rates and corresponding populations for 10 age-groups. GLOBOCAN 2012 (IARC) - 19.2.2014

International Agency for Research on Cancer  
**Algeria**  
**Breast cancer**  
 Number of new cancers in 2015 (all ages)



GLOBOCAN 2012 (IARC) (19.2.2014)

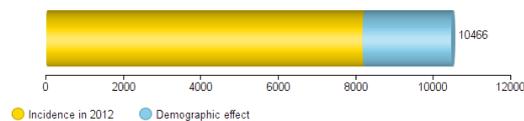
### (2) Des prédictions sur l'évolution du cancer du sein en Algérie (2020)



Algeria Breast				
Year	Estimated number of new cancers (all ages)	Male	Female	Both sexes
2012	ages < 65	-	8177	-
	ages >= 65	-	995	-
2020	ages < 65	-	10466	-
	ages >= 65	-	9084	-
			1382	-
Demographic change				
	ages < 65	-	2289	-
	ages >= 65	-	1902	-
		-	387	-

Population forecasts were extracted from the United Nations, World Population prospects, the 2012 revision. Numbers are computed using age-specific rates and corresponding populations for 10 age-groups. GLOBOCAN 2012 (IARC) - 14.4.2014

International Agency for Research on Cancer  
**Algeria**  
**Breast**  
 Number of new cancers in 2020 (all ages)



GLOBOCAN 2012 (IARC) (14.4.2014)

**Annexe C**

**(1) Compte rendu d'une mammographie gauche d'un malade réalisé le 16/06/2011**

**CABINET D'IMAGERIE MEDICALE  
ECHOGRAPHIE RADIOLOGIE SPECIALISEE MAMMOGRAPHIE  
Spécialiste en Radiologie et Imagerie Médicale**

**Le 16/06/2011**

Adressé par Dr:

Chère consoeur; Cher confrère;

Je vous remercie de votre confiance pour m'avoir adressée votre patiente

**âgée de ans.**

**COMPTE RENDU D'UN EXAMEN RADIOLOGIQUE**

**MAMMOGRAPHIE GAUCHE:**

Les incidences face et profil pratiquées objectivent un sein à composante graisseuse prédominante ne comportant pas d'image nodulaire ou stellaire.

Absence de foyer de micro-calcification pathologique.

La graisse sous cutanée est d'épaisseur correcte.

Le revêtement cutané est fin et régulier.

Le mamelon est en place, non rétracté.

**N.B :** Quelques coupes échotomographiques faites au décours de l'examen mammographique n'objectivent pas d'image nodulaire ou kystique

Absence d'adénopathie axillaire.

**AU TOTAL : Examen écho-mammographique ne montrant pas d'anomalie.**

**Confraternellement**

**Dr djilani**

**(2) Compte rendu d'une échographie mammaire gauche d'un malade**

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE**

**Wilaya de Guelma**

**Etablissement public hospitalier Ibn Zohr**

**Unité échographie**

*Patiente :*

*Age :*

**ECHOGRAPHIE MAMMAIRE GAUCHE**

**TECHNIQUE :** *Exploration ultrasonographique réalisée à l'aide d'une sonde linéaire multifréquence.*

**RESULTATS :**

- Sein de composante mixte.
- Glande mammaire de structure habituelle, sans images pathologiques ou lésions focales
- Revêtement cutané fin et régulier.
- Graisse sous cutanée intègre.
- Mamelon en place non rétracté.
- Creux axillaires libres.

**CONCLUSION :**

*EXAMEN NORMAL.*

*Salutations confraternelles et amicales*

*Dr DJILANI*

**Annexe D**

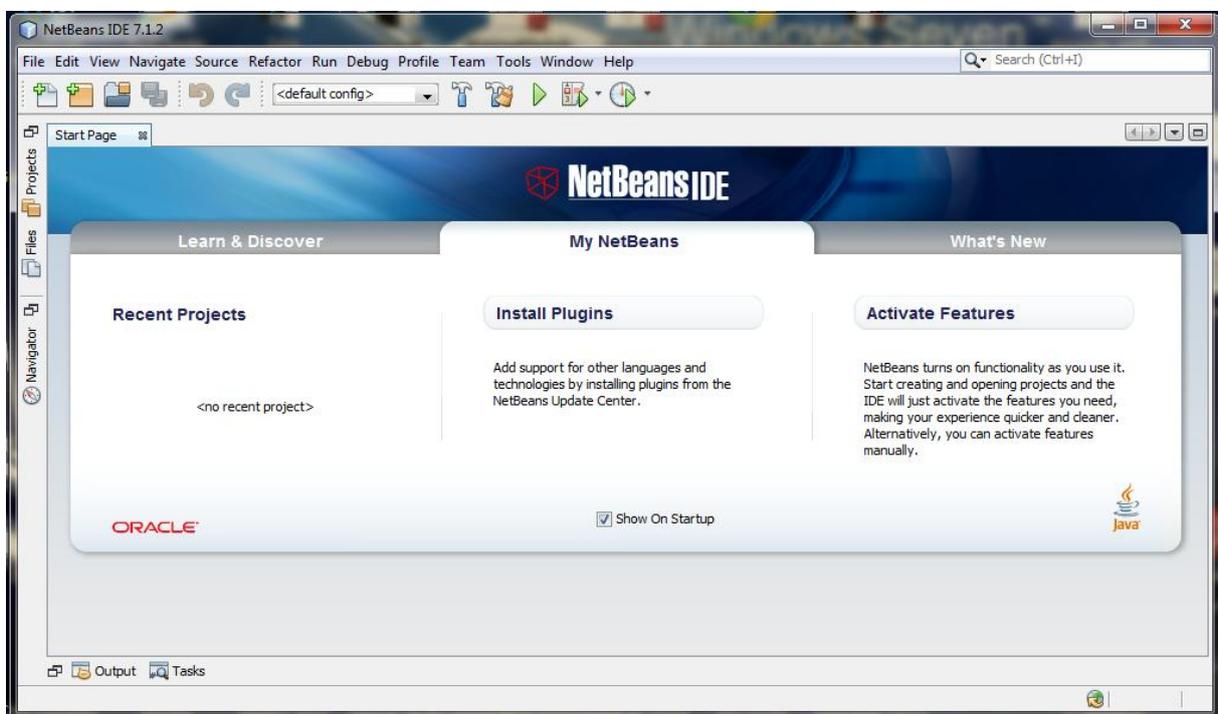
## Environnement du travail

### Le langage de programmation java

Le langage de programmation java est un langage orienté objet capable de s'exécuter sur n'importe quelle plate-forme. Il est semi compilé, semi interprété, le code source java est transformé en de simple instruction binaire (byte code = Instructions générées par le compilateur qu'un ordinateur abstrait peut exécuter) celui-ci constitue les instructions pour la machine virtuelle JVM (Java virtual Machine) pour la plus part des plates-formes (Linux, Windows, Mac, Solaris).

### L'environnement NetBeans

NetBeans est un environnement de développement intégrés pour java placé en open source par Sun en juin 2000 sous licence CDDL (Common Development and distribution licence). En plus de java, NetBeans permet également de supporter différents autres langages, comme python, C, C++, XML et HTML. Il comprend toutes les caractéristiques d'un IDE moderne (éditeur en couleur, projets multi-langage, refactoring, éditeur graphiques d'interfaces et de pages web). NetBeans est disponible sous Windows et Linux, Mac OS X et Open VMS, et qu'est lui-même développé en java, ce qui peut le rendre assez lent et gourmand en ressources mémoires.



## **Environnement de travail (SMA)**

Dans ce travail nous avons eu recours aux plates-formes qui existent pour le développement des systèmes multi-agents.

### **Plates-formes multi-agents**

Nous abordons dans ce qui suit quelques plates-formes, SeSAM, Jade, oRis, Madkit et NetLogo.

- **Shell for Simulated Agent Systems (SeSAM)**

La plateforme SeSAM2 présente un environnement de développement et de simulation écrit en Java [Klugl, 2003]. Le développement des agents, des ressources et de l'environnement ainsi que l'exécution des simulations se réalisent à partir d'une interface.

Le comportement des agents et de l'environnement est décrit sous forme de diagrammes d'activités. Il devient alors très simple de créer, modifier et visualiser des comportements.

A partir de l'interface l'utilisateur peut créer et simuler son système multi-agent, elle contient des éditeurs pour la conception des agents, des ressources, des environnements, des simulations et permet aussi de créer des analyseurs pour récupérer les résultats d'une simulation sous forme de graphes.

Cependant, l'interface pose un problème majeur. Codée en Java, son temps de réaction lors de l'exécution d'une simulation est assez élevé. La possibilité d'interagir avec la simulation s'en trouve alors diminuée

- **oRis**

oRis4 est un langage de programmation orienté agent associé à une plate-forme libre conçue pour le développement de systèmes multi-agents dans le domaine de la réalité virtuelle [Harrouet 2000]. Il permet le prototypage interactif" c'est-à-dire d'étudier des agents dans un environnement en modifiant leur comportement pendant leur simulation.

Il permet le contrôle de la simulation et la modification du comportement des agents sans avoir à relancer la simulation.

- **Madkit**

Madkit5 est un environnement de développement libre programmé en Java, il permet la création de systèmes multi-agents [Gutknecht 2000].

Sa force réside dans le fait qu'elle n'impose quasiment aucune contrainte sur l'architecture des agents ainsi que sur la forme des messages entre les agents. Cependant, Madkit n'apporte rien quant à la possibilité de modifier les comportements des agents pendant la simulation, fonctionnalité, en plus, l'organisation des systèmes multi-agents développés avec Madkit est statique.

- **NetLogo**

NetLogo6 est un environnement de simulation permettant de modéliser des phénomènes tels qu'une agrégation de cellules, la propagation d'un feu de forêt, un trafic de voitures ou encore la recherche de nourriture par des termites. Cet environnement manipule des agents, appelés tortues, et les objets de leur environnement, appelés patches.

Il n'y a pas vraiment d'agent dans NetLogo. Cette plate-forme permet seulement de créer un programme pour manipuler des objets situés dans un espace en deux dimensions. De plus, les sources de NetLogo ne sont pas disponibles gratuitement.

- **Cormas**

Cormas7 est un environnement de simulation de systèmes multi-agents dédiés aux problèmes de dynamique et d'usage de ressources. Programmée en SmallTalk.

Il propose une interface permettant un grand nombre de fonctionnalités. Il est possible de coder les agents à partir de l'interface, de configurer la simulation en détail et de visualiser les résultats de la simulation. Les agents de Cormas peuvent être situés ou communicants et l'interface permet de visualiser l'environnement graphique où sont situés les agents ou les relations sociales entre les agents. Cependant, l'interface de Cormas n'est pas aussi simple et intuitive, le langage SmallTalk est beaucoup moins pratique et utilisé de nos jours que Java.

Moduleco

- **Moduleco8**

Est un environnement développé par A. Beugnard et D. Phan de l'Ecole Supérieure des télécommunications de Bretagne. Conçu en Java, Moduleco permet la simulation de marchés, de phénomènes sociaux et de dynamiques de populations. Son interface permet seulement de lancer une simulation et d'en afficher les résultats.

Permet de créer des graphiques pour représenter les résultats, une plate-forme qui n'aide pas beaucoup le développeur. De plus, elle est restreinte aux domaines où chaque agent ne perçoit que ses voisins et modifie son propre état en fonction de l'état de ses voisins.

- **AgentTool**

AgentTool9 est une plateforme qui permet de travailler les premières phases du développement d'un système multi-agent. L'interface proposée par AgentTool permet de créer différents diagrammes définissant la communication entre les agents. A partir de ces diagrammes, AgentTool peut valider ou non les conversations qui peuvent exister entre les agents. Cet outil permet aussi de générer du code en Java, concernant les conversations entre

agents, à partir des diagrammes. Cependant, AgentTool ne permet pas la simulation [DeLoach 2001].

- **Zeus**

Zeus10 est une plate-forme de développement de systèmes multi-agents complète, elle permet de créer des agents BDI en suivant la méthode *role modeling*.

Cependant, la prise en main et la compréhension de Zeus nécessitent beaucoup de temps. [Nwana,1998].

Nous pensons qu'une plate-forme devrait être simple et intuitive de manière à ce qu'elle soit accessible à quiconque ayant quelques notions d'informatique.

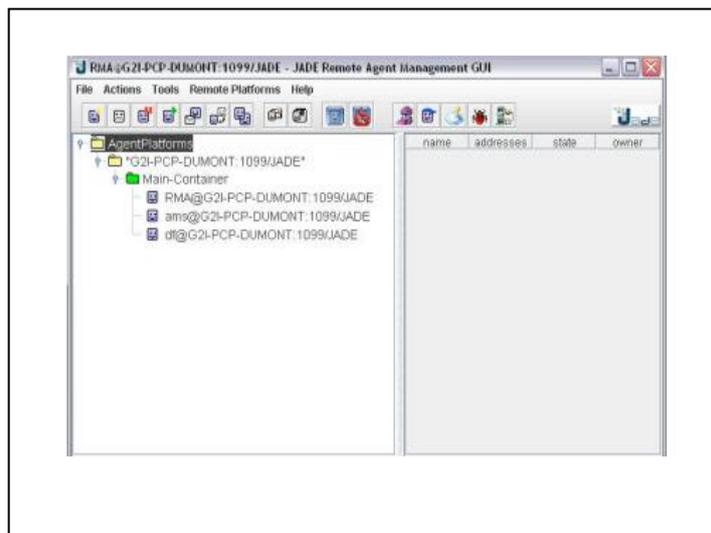
### Choix de la plate-forme

Parmi les plates-formes quand à vu, la plateforme Jade apparaît comme la plus adaptée à notre étude. En effet, contrairement à d'autres plates-formes

- **La plate forme multi-agent Jade**

La plate forme Jade3 ou Java Agent DEvelopment Framework est une plate-forme créée au niveau du laboratoire TILAB, elle est libre et programmée en Java [Bellifemine, 1999].

Cette plate forme est fondée sur FIPA qui permet la normalisation de la communication entre agents pour permettre aux différentes plates-formes de coopérer, donc Jade est particulièrement intéressante pour les agents communicants au sein d'une plate-forme ou dans un réseau de machines.



### -FIPA

FIPA définit trois agents : l'Agent Management System, l'Agent Communication Channel et le Directory Facilitator. Ces trois agents assurent le fonctionnement du système multi-agent et la communication entre agents.

FIPA assure aussi un langage de communication ACL qu'est nécessaire pour les échanges entre les agents. Les trois agents définis par FIPA ainsi que le langage ACL se retrouvent dans Jade pour assurer la communication entre les agents.

#### **-Agent Communication Language (ACL)**

L'ACL se base dans son fonctionnement sur des performatives telles que *request*, *reply* ou *proposition* informant l'agent du type de message qu'il a reçu. ACL permet aussi d'utiliser les ontologies et l'échange des informations entre agents dans des domaines particulier.

#### **-L'Agent Management System (AMS)**

Son rôle réside dans la gestion de cycle de vie des agents, il fait aussi le contrôle de l'accès et l'utilisation de l'Agent Communication Channel.

Il supervise et connue l'adresse de tous les agents situés sur la plate-forme et tous nouvel agent doit s'enregistrer auprès de l'AMS pour exister.

#### **-Le Directory Facilitator (DF)**

Le DF est un agent qui fait office de pages jaunes dans la plate-forme. Les agents doivent enregistrer les services qu'ils offrent auprès du DF. Ainsi, lorsqu'un agent a besoin d'un service, il lui suffit de consulter le DF pour obtenir la liste des agents qui peuvent le lui rendre.

#### **-L'Agent Communication Chanel (ACC)**

Ce type d'agent assure la livraison des messages à l'intérieur de la plateforme et la communication avec des agents des plateformes extérieures.

#### **-L'agentification**

Jade ne peut définir que des agents, la création des objets impossible si ces derniers ne soient pas des agents.

Les agents dans Jade est exécuté par un thread, ceci pose un problème lorsque le nombre d'agents ou d'objets est important car le passage d'un thread à un autre prend beaucoup plus de temps qu'un appel à une méthode.

#### **-Le comportement des agents**

Le comportement des agents créés est très simplement, il est défini en étendant la classe Behavior du noyau de Jade.

On peut créer des comportements simples, cycliques, s'exécutant un nombre de fois fini ou encore jusqu'à ce qu'une condition soit satisfaite.

Un agent peut prendre plusieurs comportements à la fois, ces comportements s'exécutent de manière séquentielle ou en parallèle. Cette flexibilité est intéressante, elle donne la possibilité à l'utilisateur de choisir le comportement que suivra un agent selon une situation donnée.

#### **-L'interface**

L'interface de Jade se focalise sur la communication entre les agents. Elle est directement réalisée à partir d'agents issus du noyau de Jade. Ces agents sont :

#### **-Jade GUI**

Permet de créer, de tuer ou de suspendre un agent et lancer autres agents fournis par Jade.

#### **Dummy Agent**

-Il teste le système multi-agent en envoyant des messages ACL aux agents.

#### **-Sniffer Agent**

Affiche les échanges de messages au sein d'une plate-forme.

#### **-Introspector Agent**

Affiche l'état d'un agent ainsi que les messages reçus et émis par cet agent. On doit noter que Jade ne permet pas de visualiser la simulation des agents, il revient à l'utilisateur de créer son propre agent graphique affichant l'état de la simulation.

Jade ne fournit pas une interface de développement d'agent, donc l'agent doit être codé à partir d'un éditeur de textes ou à partir d'un environnement de développement Java tel que NetBeans.

Jade est une plate forme très intéressante de développement des systèmes multi-agents. Elle permet de concevoir des comportements de toutes sortes pour les agents, il est possible de combiner les comportements, de les exécuter séquentiellement ou en parallèle.

Il est évident que la modularité des comportements qu'offre Jade nous est indispensable pour permettre à l'utilisateur d'associer un comportement à une situation non coopérative.

#### **• JADE**

Le choix de la plate-forme JADE s'avère raisonnable.

-Méthodologie associée à l'outil : La plate-forme doit associer une méthodologie couvrant les différentes étapes du cycle de vie du développement d'un SMA.

-Facilité d'implémentation et de déploiement : Pour réaliser un système de recherche d'information sur Internet, il faut utiliser un langage de programmation de haut-niveau supportant la programmation Orientée-Objet. Dans ce cas la programmation des *threads* et leur synchronisation où l'échange de message est aussi indispensable. D'autre part l'implémentation des communications doit être transparente.

-Interface graphique pratique et multifonction : La plate-forme doit posséder une interface utilisateur qui facilite le développement. La plate-forme doit permettre la visualisation des agents ainsi que leur gestion et celle des interactions avec le système.

-Connexion à d'autres composants : Notre système est connecté à une base de données enregistrant les requêtes déjà jouées par ce système afin de faciliter la gestion de l'historique et d'améliorer le temps de réponse de notre système.

-Disponibilité de documentation : La documentation est disponible dans JADE non seulement pour l'étape de développement mais aussi pour l'étape de déploiement, pour pouvoir remédier au problème de maintenance et d'exécution.

-Standard : La plate-forme JADE est conforme à un standard pour une interopérabilité avec des agents hétérogènes et interactifs et des systèmes multi-agents.

-Accès au code source : La plate-forme JADE est accessible ainsi que son code source. Ce qui lui permet une flexibilité d'extension.

-Portabilité : L'outil JADE est portable sur différents environnements et permet aussi une exécution simple du système indépendamment de l'environnement.

En effet, à partir du résultat de la comparaison, la plate-forme JADE a montré une nette supériorité pour l'ensemble de ces critères.

Par ailleurs, Jade est un outil qui répond aux normes FIPA 97. Il fournit des classes qui implémentent *JESS29* pour la définition du comportement des agents. L'outil possède trois modules principaux (nécessaire à la norme FIPA30). Le DF (Director Facilitator) fournit un service de pages jaunes à la plate-forme. Le ACC (Agent Communication Channel) gère la communication entre les agents. Le AMS (Agent Management System) supervise l'enregistrement de chaque agent, son authentification, son accès au système et son utilisation. Les agents communiquent par le langage FIPA ACL. Un éditeur est disponible pour l'enregistrement et la gestion des agents. Aucune autre interface n'est disponible pour le développement ou l'implémentation ce qui nécessite une bonne connaissance des classes et des différents services offerts [Bellifemine et al., 2003].

---

## Références Bibliographiques

---

[Abdellaoui, 2003] Abdellaoui, M. (2003). *Approche des systèmes distribués par automates cellulaires. Application en mécanique des milieux déformables*. Thèse doctorat, laboratoire de théorie des systèmes, université de Perpignan, France.

[Abelha et al., 2004] Abelha, A., machado, J., Alves, V., Neves, J. (2004). *Data warehousing through MAS in the medical arena*. WSEAS Transactions on Computers. ISSN 1109-2750. 3:6.

[Adamatzky, 1994] Adamatzky, A. (1994). *Identification of cellular automata*. Taylor and Francis (Ed). ISBN 0748401725, 9780748401727.

[Alexander, 2004] Alexander, S. (2004). *Space, Time And Deity: The Gifford Lectures At Glasgow*, Kessinger Publishing, (Ed). ISBN: 0766187012, 9780766187016.

[Ali et al., 1998] Ali, S.M., Zimmer, R.M., Elstob, C.M. (1998). *The question concerning emergence: Implications for artificiality*. Proc Computing Anticipatory Systems: Casys First International Conference, Dubois, D. M (Ed), Vol. 437, pp. 138-156, USA.

[Alsinet et al., 2000] Alsinet, R., Bejar, C., Fernandez., Manya, F. (2000). *A Multi-Agent System Architecture for Monitoring Medical Protocols*. In International Conference on Autonomous Agents, Proc The fourth international conference on Autonomous agents, Vol 3(7), pp. 499-505.

[Alt et al., 1997] Alt, W., Deutsch, A., Dunn, G. (1997). *Dynamics of Cell and Tissue Motion. Mathematics and Biosciences in Interaction*, Springer (Ed). Birkhauser, Basel.

[Amblard, 2003] Amblard, F. (2003). *Comprendre le fonctionnement de simulations sociales individu scentrees: application a des modeles de dynamiques d'opinions*, Thèse de doctorat en Informatique, Université Blaise Pascal – Clermont II, France.

[Amini et al., 1990] Amini, A. A., Weymouth, T., and Jain, R. (1990). *Using dynamic programming for solving variational problems in vision*. IEEE Transaction on Pattern Analysis and Machine Intelligence, 12(9), pp. 855-867.

[Austin, 1962] Austin, J. L. (1962). *How to do things with words*, Oxford. Clarendon Press (Ed).

[Axelrod, 1984] Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books (Ed), ISBN 0-465-02122-0, New York.

[Baptiste, 2006] Baptiste, A. (2006). *Les métaheuristiques en optimisation combinatoire*. Mémoire présenté en vu d'obtenir l'examen probatoire en informatique, Paris, France.

[Bellifemine et al., 1999] Bellifemine, F., Poggi, A., Rimassi, G. (1999). *JADE: A FIPA-compliant agent Framework*. Proc The Pratical Applications of Intelligent Agents and Multi-Agents, pp. 97-108.

[Bellifemine et al., 2003] Bellifemine, F., Caire, G., Trucco, T., Rimassa, G. (2003). *JADE, Programmer's Guide*.

- [Belot et al., 2008] Belot, A., Grosclaude, P., Bossard, N., Jouglu, E., Benhamou, E., Delafosse, P., Guizard, A., Molinié, F., Danzon, A., Bara, S., Bouvier, A., Trétarre, B., Binder-Foucard, F., Colonna, M., Daubisse, L., Hédelin, G., Launoy, G., Stang, N. L., Maynadié, M., Monnereau, A., Troussard, X., Faivre, J., Collignon, A., Janoray, I., Arveux, P., Buemi, A., Raverdy, N., Schvartz, C., Bovet, M., Chérié-Challine, L., Estève, J., Remontet, L., Velten, M. (2008). *Cancer incidence and mortality in France over the period 1980-2005*. *Epidemiol Sante Publique*, Vol 56(3), pp. 159-175.
- [Benmazou et al., 2010] Benmazou, S., Layachi, S., Merouani, H. F. (2010). *Proposition d'une approche de segmentation base AC & SMA*. Journées Ecole Doctorale & Réseaux de Recherche en sciences et technologies de l'information JED'10, pp. 46-50, Annaba, Algérie.
- [Benmazou et al., 2011 a] Benmazou, S., Layachi, S., Merouani, H. F. (2011). *Segmentation d'images basée sur un couplage AC-SMA*. Conférence International au systèmes d'information et technologie ICIST 2011, Tebessa, Algérie.
- [Benmazou et al., 2011 b] Benmazou, S., Layachi, S., Merouani, H. F. (2011). *Une approche AC-SMA pour la segmentation IRM cérébrale*. Journées Doctorales du Laboratoire D'Informatique d'Oran JDLIO'11, pp.327, Algérie.
- [Benmazou et al., 2011 c] Benmazou, S., Layachi, S., Merouani, H. F. (2011). *Approche automate Jeu de la vie-SMA pour la segmentation des images IRM cérébrales*. conférence EGC-M, Tanger, Maroc.
- [Benmazou et al., 2011 d] Benmazou, S., Layachi, S., Merouani, H. F. (2011). *Automate cellulaire SMA pour la segmentation des images IRM*. Conference on signal,image,vision et ces applications, pp.46, Guelma, Algérie.
- [Benmazou et al., 2012] Benmazou, S., Layachi, S., Merouani, H. F. (2012). *Segmentation of Images based Cellular Automata-Reactive Agent Implemented in Netlogo Platform*. *International Journal of Computer Applications*, Published by Foundation of Computer Science, New York, USA, Vol 45(24), pp. 40-43.
- [Benmazou et al., 2014 a] Benmazou, S., Merouani, H. F., Layachi, S. (2014). Classification of mammography images based cellular automata and Haralick parameters. *International journal Evolving Systems*, ISSN 1868-6478, DOI 10.1007/s12530-014-9105-1, Springer Berlin Heidelberg, Vol 5(3), pp. 209-216
- [Benmazou et al., 2014 b] Benmazou, S., Layachi, S., Merouani, H. F. (2014). Classification of Images by Cellular Automata in Medical Basis of Mammography. *International journal of Imaging*, ISSN 2231-525X, CESER Publications, Vol 14(3), pp. 28-34.
- [Benomrane et al., 2013] Benomrane, S., Ben Ayed, M., Alimi, A. M. (2013). *An Agent-Based Knowledge Discovery from Databases applied in Healthcare domain*. *IEEE 2013 - 978-1-4799- 0313-9/13*, pp. 176-180.
- [Berger and Mohr, 1990] Berger, M. O., Mohr, R. (1990). *Towards autonomy in active contour models*. In 10th International Conference on Pattern Recognition, Atlantic City, pp. 847-851.
- [Berry, 2003] Berry, M. (2003). *Survey of Text Mining: Clustering, Classification, and Retrieval*. 0387955631. Springer-Verlag, New York.
- [Boissier et al., 2004] Boissier, O., S. Gitton., P. Glize. (2004). *Caractéristiques des Systèmes et des Applications. Sy stèmes Multi-Agents*. Observatoire Français des Techniques Avancées, (ARAGO 29), Diffusion Editions TEC & DOC, pp. 25-54.
- [Bokobsa et al., 2005] Bokobsa, J., Levy, L., Suissa, M., Tristant, H., Benmussa, M. (2005). *Construction et adaptation française de la classification BI-RADS*, Construction and french adaptation of the BI-RADS classification, 27 journées de la SFSPM, pp. 226-234, Deauville.
- [Botee and Bonabeau, 1998] Botee, H. M., Bonabeau, E. (1998). *Evolving and colony optimization*. *Advanced Complex Systems*, Vol 1, pp. 149-159.

- [Bottigli et al., 2006] Bottigli, U., Cascio, D., Fauci, F., Golosio, B., Magro, R., Masala, G. L., Oliva, P., Raso, G., Stumbo, S. (2006). *Massive lesions classification using features based on morphological lesion differences*. Proc The World Academy of Science, Engineering and Technology, Vol 12, pp. 228-232.
- [Boyd et al., 1998] Boyd, N. F., Lockwood, G. A., Martin, L. J., Knight, J. A., Byng, J. W., Yaffe, M. J., Tritchler, D. L. (1998). *Mammographic densities and breast cancer risk*. Breast Disease 1998, Vol 10, pp. 113-26.
- [Boykov and Jolly, 2001] Boykov, Y., Jolly, M.P. (2001). *Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images*. Proc Of the International Conference on Computer Vision, Vol 1, pp. 105-112.
- [Bray et al., 2013] Bray, F., Ren, J. S., Masuyer, E., Ferlay, J. (2013). *Global estimates of cancer prevalence for 27 sites in the adult population in 2008*. Int J Cancer, Vol 132(5), pp. 1133-1145.
- [Brunelli and Mich, 2001] Brunelli, R., Mich, O. (2001). *Histograms analysis for image retrieval*. Pattern Recognition, Vol 34, pp. 1625-1637.
- [Byng et al., 1997] Byng, J. W., Yaffe, M. J., Lockwood, G. A., Little, L. E., Tritchler, D. L., Boyd, N. F. (1997). *Automated analysis of mammographic densities and breast carcinoma risk*. Cancer, Vol 80, pp. 66-74.
- [Byung-Mo et al., 2006] Byung-Mo, H., Seung-Jae, S., Kyu, M.L., Kyung-Soo, J., Dong-Ryeol, S. (2006). *Multi Agent System based Efficient Healthcare Service*, Proc Advanced Communication Technology ( ICACT 2006), The 8th International Conference, ISBN 89-551 9-1 29-4, Vol 1, Suwon.
- [Canny, 1986] Canny, J. (1986). *A computational approach to edge detection*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 8(6), pp. 679-698.
- [Cao et al., 2009] Cao, Y., Hao, X., Xia, S. (2009). *An improved region-growing algorithm for mammographic mass segmentation*. In Medical imaging, parallel processing of images, and optimization techniques, Vol 7497: 74971O-1–74971O-7.
- [Centeno et al., 2009] Centeno, R., Fagundes, M., Billhardt, H., Ossowski, S., Corchado, J.M., Vicente, J., Fernandes, A. (2009). *An Organisation-based Multiagent System for Medical Emergency Assistance. Bio-Inspired Systems: Computational and Ambient Intelligence*. Lecture Notes in Computer Science, Springer Berlin Heidelberg, Vol 5517, pp. 561-568.
- [Chaib-Draa and Levesque, 1996] Chaib-Draa, B., Levesque, P. (1996). *Heirarchical model and communication by signs, signals and symbols in multi-agent environments*, Journal of Experimental and Theoretical AI (JETAI), Vol 8, pp. 7-20.
- [Chan and Vese, 2001] Chan, T. F., Vese, L. A. (2001). *Active contours without edges*. IEEE Transaction on Image Processing, Vol 10(2), pp. 266-277.
- [Chaudhuri and Dayal, 1997] Chaudhuri, S., Dayal, U. (1997). *An Overview of Data Warehousing and OLAP Technology*. ACM SIGMOD record, Vol 26(1), pp. 65-74.
- [Cheikhrouhou et al., 2007] Cheikhrouhou, I., Djemal, K., Sellami, D., Derbel, N., Maaref, H. (2007). *Abnormalities description for breast cancer recognition*. Proc The First International Conference on E-Medical Systems, pp. 198-205, Morocco.
- [Cheikhrouhou et al., 2009] Cheikhrouhou, I., Djemal, K., Sellami, D., Maaref, H., Derbel, N. (2009). *Empirical descriptors evaluation for mass malignity recognition*. Proc The First International Workshop on Medical Image Analysis and Description for Diagnosis Systems (MIAD'09), pp. 91-100. Chen, C. M., Chou, Y. H., Han (Eds)
- [Cheikhrouhou, 2012] Cheikhrouhou, I. (2012). *Description et classification des masses mammaires pour le diagnostic du cancer du sein*, Thèse doctorat, l'Université d'Evry-Val d'Essonne, France.

- [Chen et al., 1998] Chen, J., Bouman, C., Dalton, J. (1998). *Similarity pyramids for browsing and organization of large image databases*. Proc conf human vision and electronic imaging (SPIE), Vol 3299, pp-563-575, Bellingham.
- [Chen et al., 2000] Chen, J., Bouman, C., Dalton, J. (2000). *Hierarchical browsing and search of large image databases*. Image Processing, IEEE Transactions, Vol 9(3), pp. 442-455.
- [Chen et al., 2003] Chen, C. M., Chou, Y. H., Han, K. C., Hung, G. S., Tiu, C. M., Chiou, H. J., Chiou, S. Y. (2003). *Breast lesions on sonograms: Computer-aided diagnosis with nearly setting-independent features and artificial neural networks*. Radiology, Vol 226(2), pp. 504-514.
- [Choquet, 1978] Choquet, G. (1978). *Cours de topologie*, Vol 54. Dunod, 2e édition.
- [Chu et al., 1998] Chu, W. W., Hsu, C.C., Cardenas, A. F., Taira, R.K. (1998). *A knowledge-based image retrieval with spatial and temporal constructs*. IEEE Transactions on Knowledge and Data Engineering, Vol 10(6), pp. 872-888.
- [Clough et al., 2005] Clough, P., Joho H., Sanderson, M. (2005). *Automatically organising images using concept hierarchies*. Proc ACM multimedia workshop (SIGIR), Singapore.
- [Comaniciu et al., 1999] Comaniciu, D., David, P.M., Foran, J. (1999). *Image guided decision support system for pathology*. Machine Vision and Applications, Vol 11(4), pp. 213-224.
- [Conway, 1976] Conway, J. H. (1976). *Game of life*. Academic press.
- [Cooper, 1970] Cooper, W. S. (1970). *The potential usefulness of catalog access points other than author, title and subject*. Journal of the American Society for Information Science, Vol 21(2), pp. 112-127.
- [Cox et al., 2000] Cox I.J., Minka T.P., Papatomas T.V., Yianilos P.N., (2000). *The Bayesian image retrieval system, PicHunter: Theory, implementation, and psychophysical experiments*. IEEE transactions on image processing, Vol. 9(1), pp. 20-37.
- [Croteau and Vieru., 2002] Croteau, A. M., Vieru, D. (2002). *Telemedicine Adoption by Different Groups of Physicians*, 35<sup>th</sup> Hawaii international Conference on System Sciences (HICSS'02), Big Island, Hawaii.
- [Culik et al., 1988] Culik II. K, Yu, S. (1988). *Undecidability of CA classification schemes*. Complex Systems 2, Vol 2, pp. 177-190.
- [Deguet, 2008] Deguet, J. (2008). *Integration de l'émergence au sein des systèmes multi-agents*. Une étude appliquée à la recherche heuristique. Thèse de Doctorat, l'Université Joseph Fourier, France.
- [DeLoach, 2001] DeLoach, S. (2001). *Analysis and Design Using MaSE and agentTool*, Proc The 12<sup>th</sup> Midwest A.I. and Cognitive Science Conference (MAICS01), Ohio.
- [Demazeau, 1995] Demazeau, Y. (1995). *From interactions to collective behaviour in agent-based systems*. Proc The 1st. European Conference on Cognitive Science (ECCS '95), pp. 117-132, Saint-Malo.
- [Demazeau, 2001] Demazeau, Y. (2001). *Voyelles*. Rapport technique, CNRS.
- [Deneubourg et al., 1990] Deneubourg, J. L., Aron S., Goss. S., Pasteels, J.M. (1990). *The self-organizing exploratory pattern of the argentine ant*, Journal of Insect Behavior, Vol 3, pp. 159-168.
- [Deneubourg et al., 1992] Deneubourg, J., Theraulaz, G., Beckers, R. (1992). *Swarm made architectures*. Varela, F. et Bourgine, P (eds), MIT Press, Proc The first European conference on artificial intelligence, pp. 123-133, Paris, France.
- [Dengsheng and Guojun, 2003] Dengsheng, Z., Guojun, L. (2003). *Evaluation of similarity measurement for image retrieval*. Neural Networks and Signal Processing, Vol 2, pp. 928-931.

[Deselaers et al., 2004] Deselaers, T., Keyserse, D., Ney, H. (2004). *Features for image retrieval: a quantitative comparison*. Notes in computer science, Vol 3175, pp. 228-236.

[De Wolf and Holvoet, 2005] De Wolf, T., Holvoet, T. (2005). *Emergence Versus Self-Organisation: Different Concepts but Promising When Combined*. Engineering Self Organising Systems: Methodologies and Applications (Brueckner, S. et Di Marzo Serugendo, G. et Karageorgos, A. et Nagpal, R., eds.), Lecture Notes in Computer Science, Vol 3464, pp. 1-15.

[Di Marzo et al, 2005] Di Marzo Serugendo, G., Gleizes, M. P., Karageorgos, A. (2005). *Self-Organization in Multi-Agent Systems*, The Knowledge Engineering Review, Cambridge University Press, Simon Parsons (Eds)Cambridge, Vol 20(2), pp. 165-189, UK.

[Ding et al., 2012] Ding, J., Cheng, H. D., Huang, J., Liu, J., Zhang, Y. (2012). *Breast ultrasound image classification based on multiple-instance learning*, Journal of Digital Imaging, Vol 25(5), pp. 620-725.

[Dorigo et al., 1996] Dorigo, M., Maniezzo, V., Colomi, A. (1996). *The ant system :Optimization by a colony of cooperating agents*. IEEE Transactions on Systems, Man, and Cybernetics- Part B, Vol 1(26), pp. 29-41.

[D'Orsi et al., 2003] D'Orsi, C. J., Bassett, L. W., Berg, W. A., Feig, S. A., Jackson, V. P., Kopans, D. B., Linver, M. N., Mendelson, E. B., Moss, L. J., Sickles., E. A. (2003). *breast imaging reporting and data system*. American college of radiology, Troisième édition française réalisée par SFR (Société Française de Radiologie).

[Douglas and Craig, 1994] Douglas Tougaw, P., Craig, S.L. (1994). *Logical devices implemented using quantum cellular automata*. Journal of Applied Physics. Vol 75(3), pp. 1818-1825.

[Drogoul, 1993] Drogoul, A. (1993). *De la simulation multi-agent à la résolution collective de problèmes*, Thèse doctorat, France.

[Drogoul et al., 2004] Drogoul A., Ferrand N., Muller J.P. (2004). *Emergence : L'articulation du local au global*. Observatoire français des techniques avancées (OFTA). Systèmes multi-agents, Paris, pp. 105-136.

[Dutta et al., 2012] Dutta, A., Sudipta, A., Aneesh, K., Swapan, B. (2012). *Virtual Medical Board: A Distributed Bayesian Agent Based Approach*. Proc International Conference on Parallel Distributed and Grid Computing (PDGC), 2012 2nd IEEE, India.

[Erman et al., 1980] Erman,L., Hayes-Roth, F., Lesser, V., Reddy, D. (1980). *The HEARSAY-II Speech Understanding System: Integrating Knowledge to Resolve Uncertainty*. ACM Computing Surveys, Vol 12(2), pp. 213-253.

[Erceau, 1993] Erceau. J. (1993). *Intelligence Artificielle Distribuée et Systèmes Multi-Agents*. Support de cours, extrait de la thèse doctorat partie état de l'art, Nadia Kabachi, Lyon, France.

[Esmahi, 2001] Esmahi, L. (2001). *Coopération dans les systèmes multi-agents : une analyse et un point de vue*. Proc JFIADSMA'01, pp. 31-53, Montréal, Canada.

[Ester et al., 1996] Ester, M., Kriegel, H., Sander, J., Xu, X. (1996). *A density-based algorithm for discovering clusters in large spatial databases with noise*. Proc Second International Conference on Knowledge Discovery and Data Mining, Han, J (edr), pp. 226-231, Portland.

[Fauqueur, 2003] Fauqueur, J. (2003). *Contributions pour la Recherche d'Images par Composantes Visuelles*. Thèse doctorat, Université de Versailles Saint-Quentin en Yvelines, France.

[Freeman and Skura, 1991] Freeman, J., Skura, D. (1991). *Neural networks algorithms, applications, and programming techniques*. Reading, MA, Addison-Wesley, New York.

[Feig et al., 1998] Feig, S. A., D'Orsi, C. J., Hendrick, R. E., Jackson, V. P., Kopans, D. B., Monsees, B., Sickles, E. A., Stelling, C. B., Zininger, M., Wilcox-Buchalla, P. (1998). *American college of radiology guidelines for breast cancer screening*. American Journal of Roentgenology, Vol 171(1), pp. 29-33.

- [Feig, 2000] Feig, S. A. (2000). *Age-related accuracy of screening mammography: how should it be measured* ?. Radiology, Vol 214, pp. 633-640.
- [Ferber, 1995] Ferber, J. (1995). *Les systèmes multi-agents. Vers une intelligence collective*. InterEditions, ISBN 2-72-96-0572-X, 1995, Paris.
- [Ferber, 1999] Ferber, J. (1999). *Multi-Agent Systems - An introduction to distributed artificial intelligence*, Addison- Wesley, ISBN 0-201-36048-9.
- [Ferguson, 1992] Ferguson, I. A. (1992). *Touring Machines: An Architecture for Dynamic, Rational, Mobile Agents*. Thèse de doctorat, Université de Cambridge, UK.
- [Ferlay et al., 2013] Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., Bray, F. (2013). *Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11* [Internet]. Lyon, France: GLOBOCAN 2012 v1.0, International Agency for Research on Cancer; 2013. Disponible on ligne: <http://globocan.iarc.fr>.
- [Fiaz et al., 2011] Fiaz, M., Aslam, M., Rizwan, S., Martinez-Enriquez A. M. (2011). *Towards Maturity in Multi-agent Based Remote Patient Monitoring*. International Journal of Multidisciplinary Sciences and Engineering, ISSN 2045-7057, Vol 2(1), pp. 11-16.
- [Finin et al., 1997] Finin T., Labrou, Y., Mayfield, J. (1997). *KQML as an agent communication language*. Jeff Bradshaw (Ed.), MIT Press, Cambridge.
- [Föllmer, 1994] Föllmer, H. (1994). *Stock price fluctuation as a diffusion in a random environment*. Philosophical Transactions of the Royal Society A, pp. 471-483.
- [Formenti, 1998] Formenti, E. (1998). *Cellular automata and chaos: from topology to Kolmogorov complexity*. Thèse doctorat, École Normale Supérieure de Lyon, France.
- [Foster et al., 2005] Foster, D., McGregor, C., El-Masri, S. (2005). *A Survey of Agent-Based Intelligent Decision Support Systems to Support Clinical Management and Research*. Proc The International Conference on autonomous Agent and Multiagent Systems, ISBN 1595930930, ACM, pp. 16-34, Netherlands.
- [Francisco da Silva et al., 2011] Francisco da Silva, S., Xavier Ribeiro, M., Batista Neto, J., Traina-Jr , C., Traina A., (2011). *Improving the ranking quality of medical image retrieval using a genetic feature selection method*. Decision support systems, Vol 50, pp. 1-11.
- [Franklin and Graesser, 1997] Franklin, S., Graesser, A. (1997). *Is it an Agent ,or just a program?:A Taxonomy For Autonomous Agents*. Proc The Third International Workshop on Agent Theories, Architecture, And Languages ISBN 3-540-62507-0, pp. 21-35, Springer-Verlag, London, UK.
- [Frouge et al., 1994] Frouge, C., Guinnebretière, J. M., Contessor. (1994). *Physiopathologie des microcalcifications mammaires*. Feuilletts de radiologie, Vol 34, pp. 370-380.
- [Frouge, 1995] Frouge, C. (1995). *Stratégie diagnostique en sénologie*. (Ed) Masson, Paris.
- [Fukunaga, 1972] Fukunaga, K. (1972). *Introduction to Statistical Pattern Recognition*. Vol COMPUTER Science and Scientific Computing, Editeur Werner Rheinboldt, Second Edition.
- [Fukunaga and Narendra, 1975] Fukunaga, K, Narendra, P. (1975). *A branch and bound algorithm for computing k-nearest neighbors*. IEEE Transactions Computers, Vol 24(7), pp. 750-753
- [Gabor, 1946] Gabor, D. (1946). *Theory of communication*. Journal of the Institution of Electrical Engineers, Vol 93(26), pp 429-441.
- [Gallas et al., 2010] Gallas A., Barhoumi W., Zagrouba E. (2010). *Une approche rapide basée sur les régions floues pour une évaluation sémantique de la similarité d'images*. WIA 2010 Workshop en informatique et applications, Tozeur, Tunisie.

- [Garcia et al., 2003] Garcia, A., Enrique, J., Gomez, M., Elena Hernando, F. Perdices, J., del Pozo, F., Torralba, V. (2003). *Multi-Agent Architecture for the Provision of Intelligent Telemedicine Services in Diabetes Management*. disponible on ligne: <http://cyber.felk.cvut.cz/EUNITE03-BIO/pdf/EHernando.pdf>.
- [Gardner, 1970] Gardner, M. (1970). *The fantastic combinations of John Conway's new solitaire game life*. Scientific American, Vol 220(4), pp. 120-123.
- [Gattaneo et al., 1996] Cattaneo, G., Formenti, E., Margara, L. (1996). *Topological definitions of deterministic chaos: applications to cellular automata dynamics cellular automata*. Theoretical Computer Science, Vol 217(1), pp. 31-51.
- [Gaud et al., 2008] Gaud, N., Galland, S., Gechter, F., Hilaire, V., Koukam, A. (2008). *Holonic multilevel simulation of complex systems: Application to real-time pedestrians simulation in virtual urban environment*. Simulation Modelling Practice and Theory, Vol 16(10), pp. 1659-1676.
- [Gilman, 1987] R. H. Gilman. (1987). *Classes of linear automata*. Ergodic Theory and Dynamical Systems, Vol 7(1), pp. 105-118.
- [Gonzales and Woods, 2002] Gonzales, R. C., Woods, R. E. (2002). *Digital Image Processing*. Origins of Digital Image Processing: Newspaper Industry, ISBN, 0201180758, 2nd Edition, Prentice Hall.
- [Gouet-Brunet, 2010] Gouet-Brunet, V. (2010). *Bases de données multimédia : Structuration de données pour la recherche d'images par contenu visuel*. Conservatoire National des Arts et Métiers de Paris.
- [Gourinel et al., 1996] Gouriel, C., Buxeraud, J., Coulon, J.P., Lagarde, A., Comby, F., Bernard, M., Dossier. (1996). *Le cancer du sein*, Actualité pharmaceutiques, Vol 347, pp. 37-47.
- [Govaert, 2003] Govaert, G. (2003). *Analyse de données*. Gérard (ed), IC2, Lavoisier.
- [Gupta and Pujari, 2009] Gupta, S., Pujari, S. (2009). *A Multi-Agent Based Scheme for Health Care And Clinical Diagnosis System*. Proc International Conference on Intelligent Agent & Multi-Agent Systems (IAMA 2009), IEEE (Ed) ISBN 978-1-4244-4710-7, pp. 1-3.
- [Gupta et al., 2012] Gupta, S., Sarkar, A., Pramanik, I., Mukherjee, B. (2012). *Implementation Scheme for Online Medical Diagnosis System Using Multi Agent System with JADE*. International Journal of Scientific and Research Publications, ISSN 2250-3153, Vol 2(6), pp. 1-6.
- [Gutknecht, 2000] Gutknecht, O., Ferber, J., Michel, F. (2000). *Madkit: une architecture de plateforme multi-agent generique*. Rapport de recherche 00061, Laboratoire d'Informatique de Robotique et de Micro-electronique de Montpellier.
- [Hadzic and Chang, 2005] Hadzic, M., Chang, E. (2005). *Use of ontology-based multi-agent systems in the biomedical domain*. Proc The18th International Conference on Computer Applications in Industry and Engineering (CAINE), Dascalu, S.(ed), pp. 141-145, Honolulu, Hawaii, USA.
- [Hafner et al., 1995] Hafner, J., Sawhney, H. S., Equitz, W., Flickner, M., Niblack, W. (1995). *Efficient Color Histogram Indexing for Quadratic Form Distance Functions*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 17(7), pp. 729-736.
- [Hagay and Cherel., 1998] Hagay, C., Cherel, P. (1998). *Les opacités rondes en mammographie*. Radiologie, Vol 18(4), pp. 263-275.
- [Hamarneh and Li, 2009] Hamarneh, G., Li, X. (2009). *Watershed segmentation using prior shape and appearance knowledge*. Image and Vision Computing, Vol 27, pp. 59-68.
- [Haralick, 1979] Haralick, R. M. (1979). *Statistical and structural approaches to textures*. Proc of the IEEE, ISSN 0018-9219, Vol 67(5), pp. 786-804.

- [Harrouet, 2000] Harrouet, H. (2000). *oRis : s'immerger par le langage pour le prototypage d'univers virtuels à base d'entités autonomes*. Thèse doctorat, Université de Bretagne Occidentale, Brest, France.
- [Harvey and Bovbjerg, 2004] Harvey, J. A., Bovbjerg, V. E. (2004). *Quantitative assessment of mammographic breast density: relationship with breast cancer risk*. Radiology, Vol 230(1), pp. 29-41.
- [Hayes-Roth, 1985] Hayes-Roth, B. (1985). *A Blackboard Architecture for Control*. Artificial Intelligence Journal, Vol 26, pp. 251-321
- [Heath and Bowyer, 2000] Heath, M. D., Bowyer, K. W. (2000). *Mass detection by relative image intensity*. Proc The 5<sup>th</sup> International Workshop on Digital Mammography, pp. 219-225, Toronto, Canada.
- [Heesch, 2008] D. Heesch. (2008). *A survey of browsing models for content based image retrieval*. Multimedia Tools and Applications, ISSN 1573-7721, Vol 40, pp. 261-284.
- [Heine and Malhotra, 2002] Heine, J. J., Malhotra, P. (2002). *Mammographic tissue, breast cancer risk, serial image analysis, and digital mammography. Part 2. Serial breast tissue change and related temporal influences*. Acad Radiol, Vol 9(3), 317-35.
- [Hernandez and Herrmann, 1996] Hernandez, G., Herrmann, H. (1996). *Cellular automata for elementary image enhancement*. Graphical Models and Image Processing (GMIP), Vol 58(1), pp. 1-3.
- [Heudin, 1994] Heudin, J. C. (1994). *La Vie Artificielle*. Editions Hermès Sciences, Paris.
- [Hewitt, 1977] Hewitt, C. (1977). *Viewing Control Structures as Patterns of Passing Messages*. Artificial Intelligence, Vol 8(3), pp. 323-364.
- [Horowitz and Pavlidis, 1974] Horowitz, S. L., Pavlidis, S. (1974). *Picture segmentation by a directed split and merge procedure*. Proc The Second International Joint Conference on Pattern Recognition, Vol 424, pp. 424-433, Copenhagen, Denmark.
- [Hu, 1962] Hu, M. K. (1962). *Visual pattern recognition by moments invariants, computer methods in image analysis*. Transactions on Information Theory, Vol 8, pp. 170-187.
- [Huang et al., 2008] Huang W., Gao Y., Chan K.L. (2008). *A review of region-based image retrieval*. Sign process system, Vol 59, pp. 143-161.
- [Huo et al., 1995] Huo, Z., Giger, M. L., Vyborny, C. J., Bick, U., Lu, P., Wolverson, D. E., Shmidt, R. A. (1995). *Analysis of spiculation in the computerized classification of mammographic masses*. Medical Physics, Vol 22(10), pp. 1569-1579.
- [Jain et al., 1999] Jain, A., Murty, M., Flynn, P. (1999). *Data clustering: a review*. ACM Computing Surveys, Vol 31(3), pp. 264-323.
- [Jeannin, 2001] Jeannin, S. (2001). *MPEG-7 Visual part of experimentation model, Version 9.0*. Dans ISO/IEC JTC1/SC29/WG11/N3914, 55th Mpeg Meeting, pp. 27-41, Pisa, Italy.
- [Jung-Jin, 2003] Jung-Jin, Yang. (2003). *An Ontology-Based Intelligent Agent System for Semantic Search in Medicine*. Intelligent Agents and Multi-Agent Systems, Lecture Notes in Computer Science, ISSN 0302-9743, Vol 2891, pp. 182-193, Springer-Verlag Berlin Heidelberg.
- [Kalmarzi et al., 2013] Kalmarzi Moghaddam, M., Shojafar, M., Reza Nami, M., Rashidi, H. (2013). *An Efficient Multi-agent System for E-health Functionalities*. International Journal of Software Engineering and Its Applications, Vol 7(3), pp. 23-34.
- [Kass et al., 1987] Kass, M., Witkin, A., et Terzopoulos, D. (1987). *Snakes: active contour models*. International Journal of Computer Vision, Vol 1(4), pp. 321-331.
- [Katayama and Satoh., 1997] Katayama, N., Satoh, S. *The SR-tree: an index structure for high-dimensional nearest neighbor queries*. Proc The ACM SIGMOD Conference on Management of Data (SIGMOD'97), ISBN 0-89791-911-4, Vol 26(2), pp. 369-380, New York, USA.

[Kazar et al ] Kazar, O., Sahnoun, Z., Frecon, L. (2006). *Multi-agents system for medical diagnosis*. disponible on ligne: <https://www.uop.edu.jo/csit2006/vol4%20pdf/pg118.pdf>.

[Keenan and O'Brien, 1993] Keenan, D. C., O'Brien, J. M. *Competition, Collusion and Chaos*. Journal of Economic Dynamics and Control, Vol 17, pp. 327-353.

[Kenyon, 1941] Kenyon, F. (1941). *The Myth of The Mind*. Thinker's Library, Watts & Co, London.

[Key and al., 2001] Key, T. J., Verkasalo, P. K., Banks, E. (2001). *Epidemiology of breast cancer*, The Lancet Oncology, Vol 2(3), pp. 133-140.

[Kherfi and Ziou, 2006] Kherfi, M.L., Ziou, D. (2006). *Relevance feedback for CBIR: a new approach based on probabilistic feature weighting with positive and negative examples*. IEEE transactions on image processing, Vol 15(4), pp.1017-1030.

[Kilday et al., 1993] Kilday, J., Palmieri F, Fox, M. (1993). *Classifying mammographic lesions using computerized image analysis*. IEEE Transactions on Medical Imaging, Vol 20(12), pp. 1275-1284.

[Kim et al., 1997] Kim, J. K., Park, J. M., Song, K. S., and Park, H. W. (1997). *Adaptive mammographic image enhancement using first derivative and local statistics*. IEEE Transactions on Medical Imaging, Vol 16(5), pp 495-502.

[Kim and Cipolla, 2008] Kim, T., Cipolla, R. (2008). *Mcboost: Multiple classifier boosting for perceptual co-clustering of images and visual features*. Proc Twenty-Second annual Conference on Neural Information processing Systems, (NIPS'08), pp. 841-856.

[Kimball, 1996] Kimball, R. (1996). *The Data Warehouse Toolkit: Practical technique for Building Dimensional Data Warehouses*, John Wiley and Sons, Inc, ISBN 0-471-15337-0, New York, USA.

[Kimura et al., 1997] Kimura, F., Wakabayashi, T., Tsuruoka, S., and Miyake, Y. (1997). *Improvement of handwritten japanese character recognition using weighted direction code histogram*. Journal of the Pattern Recognition Society, Vol 30(8), pp. 1329-1337.

[Kim and Yu, 2008] Kim, D.H., Yu S.H, (2008). *A new region filtering and region weighting approach to relevance feedback in content-based image retrieval*. The journal of systems and software, vol 81, pp.1525-1538.

[Klugl et al., 2003] Klügl, F., Herrler, R., Oechslein, C. (2003). *From Simulated to Real Environments: How to use SeSAM for software development*. Multiagent System Technologies, Lecture Notes in Computer Science, ISSN 0302-9743, Vol 2831, pp 13-24.

[Klusch et al., 2003] Klusch, M., Lodi, S., Moro. G. (2003). *The Role of Agents in Distributed Data Mining - Issues and Benefits*. Proc IEEE/WIC International Conference on Intelligent Agent Technology (IAT'03), pp. 211-217, USA.

[Kohonen, 1984] Kohonen, T. (1984). *Self-Organization and Associative Memory*. Berlin : Springer-Verlag.

[Kohonen, 2001] Kohonen, T. (2001). *Self-organizing maps*. Vol 30 of Springer series in information sciences. Springer, Berlin Heidelberg, New York.

[Kramm, 2005] Kramm, A. (2005). *Automates cellulaires*. Mémoire de maîtrise d'informatique, Université Paris VIII. France.

[Krishnamachari and Abdel-Mottaleb., 1999] Krishnamachari, S, Abdel-Mottaleb, M. (1999). *Image browsing using hierarchical clustering*. IEEE symp computers and communications. IEEE, pp 301-307, Piscataway.

[Kupinski and Giger, 1998] Kupinski, M. A., Giger, M. L. (1998). *Automated seeded lesion segmentation on digital mammograms*. IEEE Transactions on Medical Imaging, 17(4), pp. 510-517.

- [Kurka, 1997] Kurka, P. *Languages, equicontinuity and attractors in cellular automata*. (1997). *Ergodic Theory and Dynamical Systems*, Vol 17(2), pp. 417-433.
- [Kurniawati et al., 1997] Kurniawati, R., Jin, J., Shepherd, J. (1997). *Techniques for supporting efficient content-based retrieval in multimedia databases*. *Journal of Research and Practice in Information Technology*, Vol 29(4), pp. 122-130.
- [Laaksonen et al., 2000] Laaksonen, J., Oja, E., Koskela, M., Brandt, S. (2000). *Analyzing low-level visual features using content-based image retrieval*. *Proc int'l conf neural information processing*, pp.14-18, Taejon.
- [Langley and Sage, 1999] Langley, P., Sage, S. (1999). *Tractable average-case analysis of naïve Bayesian classification*. *Proc of the Sixteenth International Conference on Machine Learning*, Bled, Slovenia: Morgan Kaufman, pp. 220-228.
- [Langton, 1986] Langton, C. G. (1986). *Studying artificial life with cellular automata*. *Physica*, pp. 120-149, North-Holland, Amsterdam.
- [Lanyi, 1985] Lanyi, M. (1985). *Morphologic analysis of microcalcifications*. *Early Breast Cancer* (Ed) J. Zander and J. Baltzer, Springer-Verlag, Berlin.
- [Leconte et al., 2003] Leconte, I., Feger, C., Galant, C., Berlière, M., Berg, B. V., D'Hoore, W., Maldague, B. (2003). *Mammography and subsequent whole-breast sonography of nonpalpable breast cancers: the importance of radiologic breast density*. *AJR Am J Roentgenol*. Vol 180(6), 1675-9.
- [LeGal et al., 1984] LeGal, M., Chavanne, G., Pellier, D. (1984). *Valeur diagnostique des microcalcifications groupées d'écouvertes par mammographie*. *Bull Cancer*, Vol 71, pp. 57-64.
- [Li et al., 2005] Li, C., Xu, C., Gui, C., Fox, M. D. (2005). *Level set evolution without re-initialization: A new variational formulation*. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 430-436.
- [Li et al., 2008] Li, C., Kao, C. Y., Gore, J. C., Ding, Z. (2008). *Minimization of region-scalable fitting energy for image segmentation*. *IEEE Transactions on Image Processing*, Vol 17(10), pp. 1940-1949.
- [Lieberman and Menell., 2002] Lieberman, L., Menell, J. H. (2002). *Breast imaging reporting and data system (BI-RADS™)*. *Radiol Clin North Am*, Vol 40(3), pp. 409-30.
- [Liu et al., 1997] Liu, J., Tang, Y. Y., Cao, Y. C. (1997). *An Evolutionary Autonomous Agents approach to Image Feature Extraction*. *IEEE Transactions on Evolutionary Computation*, Vol 1(2), pp. 141-158.
- [Liu and Picard, 1996] Liu, F., Picard, R. W. (1996). *Periodicity, directionality, and randomness: World features for image modeling and retrieval*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 18(7), pp. 722-733.
- [Liua et al., 2008] Liua R., Wang Y., Babab T., Masumotob D., Nagatab S., (2008). *SVM-based active feedback in image retrieval using clustering and unlabeled data*. *International journal of Pattern recognition*, Vol 41, pp. 2645-2655.
- [Lynch, 2000] Patrick J. Lynch. (2000). *illustrator, Medical Illustrations*. Generated for multimedia teaching projects by the Yale University School of Medicine, Center for Advanced Instructional Media.
- [Maes, 1995] Maes, P. (1995). *Intelligent Software*. *Scientific American*, Vol 273(3), pp. 84-86.
- [Maes and Shlosman, 1993] Maes, C., Shlosman, S. B. (1993). *When is an interacting particle system ergodic ?*, *Communications in Mathematical Physics*, Vol 151(3), pp 447-466.
- [Magnin et al., 1986] Magnin, I. E., Cluzeau, F., Odet, C. L., Bremond, A. (1986). *Mammographic texture analysis: an evaluation of risk for developing breast cancer*. *Optical Engineering*, Vol 25, pp. 780-784
- [Mahsa and WooChaw, 2013] Mahsa, C., Woo Chaw, S. (2013). *Medical Image Segmentation using a Multi-Agent System Approach*, *The International Arab Journal of Information Technology*, Vol 10(3), pp. 222-229.

- [Maren, 1990] Maren, A. (1990). *Handbook of Neural Computing Applications*. Academic Pr; First Edition (Ed), ISBN 13 978-0124712607.
- [Marlene et al., 2001] Marlene, M., Maheu, M., Whiten, P., Allen, A. (2001). *E-Health, Tele-health, and telemedicine: a guide to startup and success*, New York, Jossey-Bass.
- [Martin et al., 1984] Martin, O., Odlyzko, A. M., Wolfram. S. (1984). *Algebraic Properties of Cellular Automata*. Communications in Mathematical Physics, Springer-Verlag, Vol 93, pp. 219-258.
- [Martin, 2001] Martin, B. II. (2001). *Automates cellulaires, information et chaos*, Thèse de doctorat, École Normale Supérieure de Lyon, France.
- [Martinez-Garcia et al., 2009] Martinez-Garcia, C., Navarro-Arribas, G., Borrell, J., Martin-Campillo, A. (2009). *An Access Control Scheme for Multi-agent Systems over Multi-Domain Environments*. Y. Demazeau et al (Eds.), 7<sup>th</sup> International Conference on PAAMS'09, AISC 55, Springer-Verlag Berlin Heidelberg pp. 401-410.
- [Mateo et al., 2007] Mateo, R. M. A., Gerardo, B. D., Lee, J. (2007). *Healthcare Expert System Based on the Group Cooperation Model*. International Conference on Intelligent Pervasive Computing, South Korea.
- [Matsubara et al., 1997] Matsubara, T., Fujita, H., Kasai, S., Goto, M., Tani, Y., Hara, T., Endo, T. (1997). *Development of new schemes for detection and analysis of mammographic masses*. Proc International Conference on Intelligent Information Systems, (IIS'97), pp. 63-66, Grand Bahama Island, Bahamas.
- [Matsubara et al., 1998] Matsubara, T., Fujita, H., Hara, T., Kasai, S., Otsuka, O., Hatanaka, Y., Endo, T. (1998). *Development of a new algorithm for detection of mammographic masses*. International journal of Computational Imaging and Vision, ISSN 1381-6446, Vol 13, pp. 139-142
- [Maxwell, 1890] Maxwell, J. C. (1890). *Scientific paper of James Clerk Maxwell*, W.D. Niven., M.A., F.R.S (Ed), Cambridge University Press.
- [McDonald et al., 2001] McDonald, S., Lai, T-S., Tait, J. (2001). *Evaluating a content based image retrieval system*. Proc of 24<sup>th</sup> ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2001), ISBN 1-58113-331-6, pp. 232-240, New Orleans..
- [McDonald and Tait, 2003] McDonald, S., Tait, J. (2003). *Search strategies in content-based image retrieval*. Proc of 26<sup>th</sup> ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2003), ISBN 1-58113-646-3, pp. 80-87, Toronto.
- [McQueen, 1967] McQueen, J. (1967). *Some methods for classification and analysis of multivariate observations*. Cam, L. and Neyman, J (eds), Proc of the Fifth Berkeley symposium on Mathematical Statics and Probability, Vol 1, pp. 281-297.
- [Melouah, 2010] Melouah, A. (2010). *Réalisation d'une plateforme de segmentation d'images à base d'Agents*. Thèse de doctorat. Université Badji Mokhtar, Annaba, Algérie.
- [Michel, 2004] Michel, F. (2004). *Formalisme, outils et éléments méthodologiques pour la modélisation et la simulation multi-agents*. Thèse de doctorat, Université Montpellier II. France.
- [Milanese et al., 1996] Milanese, R., Squire, D., Pun, T. (1996). *Correspondence analysis and hierarchical indexing for content-based image retrieval*. Proc IEEE international conference image processing, pp. 859-862, Piscataway.
- [Miller, 2007] Miller, E. A. (2007). *Solving the disjuncture between research and practice: Tele-health trends in the 21st century*, Health Policy, Vol 82 (2), pp. 133-141.
- [Morgan, 1923] Morgan, C. L. (1923). *Emergent Evolution*. London. Williams and Norgate, (originally presented as the 1922 Gifford Lectures at the university of St. Andrews, Scotland).

- [Morin, 1977] Morin, E. (1977). *La méthode (t.1) : la nature de la nature*, Paris, Le Seuil. Réédition, édition de poche, Coll. Points, 1981, Traduction : USA (Peter Lang).
- [Mudigonda et al., 2001] Mudigonda, N. R., Rangayyan, R. M., Desautels, J. E. L. (2001). *Detection of breast masses in mammograms by density slicing and texture flow-field analysis*. IEEE Transactions on Medical Imaging, Vol 20(12), pp. 1215-1227.
- [Müller, 1996] Müller, J. (1996). *An Architectur for Dynamically Interacting Agents*. Thèse de doctorat, DFKI, Université de Saarlandes, Saarbrücken.
- [Müller et al., 2004] Müller, N., Michoux, N., Bandon, D., Geissbuhler, A. (2004). *A review of content-based image retrieval systems in medical applications—clinical benefits and future directions*. International Journal of Medical Informatics, Vol 73(1), pp. 1-23.
- [Murgo et al ] Murgo, S., Blaude, M. A., De grave, A., Delleuze, j. M., Dejardin, H., Farkas, B., Goddart, D., Mendez-Mayorga, V., Schoonjans, J., Sergeant, L., Vandenbroucke, A. *Classification des anomalies mammographiques*, Programme de dépistage du cancer du sein en Communauté Française.
- [Muller, 2010] Muller, J.P, [www.Lirimm.fr/~muller/colline/SMAenvorg1.0.doc](http://www.Lirimm.fr/~muller/colline/SMAenvorg1.0.doc), dernier accès 16/08/2012.
- [Neumann , 1966] Neumann, J. V. (1966). *Theory of self-reproducing automata*. In: Burks, A. W (Ed), University of Illinois Press.
- [Niblack et al., 1993] Niblack, W., Barber, R., Equitz, W., Flickner, M., Glasman, E., Petkovic, D., Yanker, P., Faloutos, C., Taubin, G. (1993). *The QBIC project: Querying images by content, using color, texture, and shape*. Storage and Retrieval for Image and Video Databases (SPIE), Vol 1908, pp. 173-187.
- [Nicholas and Paulien, 1997] Nicholas, J. S., Paulien, H. (1992). *Modelling Morphogenesis: From Single Cells to Crawling Slugs*, Journal of Theoretical Biology, Vol 18(3), pp. 229-235.
- [Nikolay et al., 2003] Nikolay, D., Aart, V. H., Ing, W. (2003). *BANip: Enabling remote healthcare monitoring with Body Area Networks*, International Workshop on scientific engineering of distributed Java applications (FIJI), Luxembourg.
- [Nowak and May, 1992] Nowak, M.A., May, R. M. (1992). *Evolutionary Games and Spatial chaos*. Nature Publishing Group 1992, Vol 359, pp. 826-829.
- [Nwana, 1998] Nwana, H., Ndumu, D., Lee, L. (1998). *ZEUS: An Advanced Tool-Kit forEngineering Distributed Multi-Agent Systems*. In Proc of the Practical Application of Intelligent Agents and Multi-Agent Systems, pp. 377-392, London, UK.
- [Obdržálek and Matas, 2005] Obdržálek, S., Matas, J. (2005). *Sub-linear indexing for large scale object recognition*. Proc conf British machine vision, pp. 1-10, Versailles.
- [Odell et al., 2002] Odell, J., Parunak, H., Van Dyke, F. M., Breuckner, S. (2002). *Modeling Agents and their Environment*. Giunchiglia, Fausto, Odell, James and Weiss, Gerhard (eds), Agent-Oriented Software Engineering (AOSE) III. Lecture Notes on Computer Science, Springer, Berlin, Vol 2585, pp. 16-31.
- [Osada et al., 2001] Osada, R., Funkhouser, T., Chazelle, B., Dobkin. (2001). *Matching 3D Models with Shape Distributions*. International Conference on Shape Modeling and Applications (SMI-01), IEEE Computer Society, pp. 154-166, Washington, USA.
- [Oyang et al., 2005] Oyang, Y., Hwang, S., Ou, Y., Chen, C., Chen, Z. (2005). *Data classification with radial basis function networks based on a novel kernel density estimation algorithm*. IEEE Transactions on Neural Networks, Vol 16, pp. 225-236.
- [Packard, 1986] Packard, N. H. (1986). *Lattice Models for Solidication and Aggregation*. Science on forme: Proc of First International Symosium for Science on Form, pp. 95-101, KTK Scientific Publishers, Tokyo.

- [Paragios and Deric, 2002] Paragios, N., Deriche, R. (2002). *Geodesic active regions and level set methods for supervised texture segmentation*. International Journal of Computer Vision, ISSN 1573-1405, Kluwer Academic Publishers, Vol 46(3), pp. 223-247.
- [Paredes et al., 2008] Paredes R., Deselaers T., Vidal E., (2008). *A probabilistic model for user relevance feedback on image retrieval*. Popescu-Belis, A., Stiefelhagen, R. (eds.) MLMI 2008, Vol 5237, pp. 260-271.
- [Pascale, 2003] Pascale, D. (2003). *A review of rgb color spaces...from xyy to r'g'b*. Montreal: The Babel Color company.
- [Pečenović et al., 2000] Pečenović, Z., Do, M., Vetterli, M., Pu, P. (2000). *Integrated browsing and searching of large image collections*, Proc int'l conf advances in visual information systems, LNCS 1929. Springer, Berlin Heidelberg New York, pp. 279-289
- [Penny, 1986] Penny, H. (1986). *Blackboard Systems: The Blackboard Model of Problem-solving and the Evolution of Blackboard Architectures*, AI Magazine, Vol 7(2), pp. 38-53.
- [Peura and Livarinen, 1997] Peura, M., Iivarinen, J. (1997). *Efficiency of simple shape descriptors*. Third International Workshop on Visual Form, pp. 443-451, Capri, Italy.
- [Piatrik and Izquierdo, 2011] Piatrik T., Izquierdo E. (2011). *Multi-feature fusion in image clustering using ant-inspired methods*. Third world congress on nature and biologically inspired computing, pp. 377-382, Salamanca.
- [Platt et al., 2000] Platt, J., Cristianini, N., Shawe-Taylor, J. (2000). *Large margin dagsvm's for multi-class classification*. In Advances in Neural Information Processing System NIPS'00, S. A. Solla, T. K. Leen and K-R. Muler (eds), MIT press, Vol 12, pp. 547-553.
- [Popovici and Popovici, 1990] Popovici, A., Popovici, D. (1990). *Cellular Automata in Image Processing*. International symp. On the Mathematical Theory of Networks and Systems.
- [Portnoy et al., 2002] Portnoy, D., Bellaachia, A., Chen, Y., Elkhahloun, A. (2002). *E-cast: A data mining algorithm for gene expression data*. In Zaki, M., Wang, J., and Toivonen, H., editors, Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD'02, Workshop on Data Mining in Bioinformatics (BIOKDD), pp. 49-54.
- [Prewitt, 1970] Prewitt, J. M. S. (1970). *Object enhancement and extraction*. Picture Processing and Psychopictorics. Academic Press, New York.
- [Qiao and Shieh, 2008] Qiao, Y., Shieh, J. (2008). *A Multi-Agent Prototype System for Medical Diagnosis*. Proc of 3rd International Conference on Intelligent System and Knowledge Engineering, 2008 IEEE, Vol 1, pp. 1265-1270.
- [Rangayyan et al., 1997] Rangayyan, R. M., El-Faramawy, N. M., Desautels, J. E. L., Alim, O. A. (1997). *Measures of acutance and shape for classification of breast tumors*. IEEE Transactions on Medical Imaging, Vol 16(6), pp. 799-810.
- [Rapaport, 1998] Rapaport, I. (1998). *Inducing an order on cellular automata by a grouping operation*, Thèse de doctorat, École Normale Supérieure de Lyon, France.
- [Regnault, 2008] Regnault, D. (2008). *Sur les automates cellulaires probabilistes: comportements Asynchrones*, thèse de doctorat, l'Ecole Normale Supérieure de Lyon, France.
- [Reyad et al., 2011] Reyad, Y., El-Zaart, A., Mathkour, H. (2011). *Segmentation of fibro-glandular discs in digital mammograms using log-normal distribution*. IEEE, International Conference on Data Engineering and Internet Technology. Malaysia.
- [Robertson and Jones, 1976] Robertson, S. E., Jones, K.S. (1976). *Relevance weighting of search terms*. Journal of the American Society for Information Sciences, Vol 27(3), pp. 129-146.

- [Roussopoulos et al., 1995] Roussopoulos, N., Kelley, S., Vincent, F. (1995). *Nearest neighbor queries*. Proc ACM int'l conf management of data (SIGMOD), ACM, New York, USA.
- [Rubner et al., 1998] Rubner, Y., Tomasi, C., Leonidas J. G. (1998). *A Metric for Distributions with Applications to Image Databases*, Proc International Conference on Computer Vision, Bombay, India.
- [Russell and Norvig, 1995] Russell, S., Norvig, p. (1995). *Artificial intelligence: A modern approach*, second edition, ISBN 0-13-103805-2, Englewood Cliffs, NJ: Prentice Hall.
- [Saha et al., 2001] Saha, P. K., Udupa, J. K., Conant, E. F., Chakraporty, P., Sullivan, D. (2001). *Breast tissue density quantification via digitized mammograms*. IEEE Transactions on Medical Imaging, Vol 20(8), pp. 792-803.
- [Salton, 1971] Salton, G. (1971). *The SMART Retrieval System*, Experiments in Automatic Document Processing. Prentice-Hall, New Jersey, USA.
- [Sarkar and Chaudhuri, 1994] Sarkar, N., Chaudhuri, B.B. (1994). *An efficient differential box-counting approach to compute fractal dimension of image*, Vol 24(1), pp. 115-120.
- [Schomaker et al., 1999] Schomaker, L., de leau, R., Vuurpijl, L. (1999). *Using Pen-Based Outlines for Object-Based Annotation and Image-Based Queries*. Proc International Conference on Visual Information and Information Systems, UK: Springer-Verlag. pp. 585-592, London.
- [Shamir and Sharan, 2002] Shamir, R., Sharan, R. (2002). *Algorithmic approaches to clustering gene expression data*. Current Topics in Computational Biology, T. Jiang, T. Smith, Y. Xu, M.Q. Zhang editors, MIT Press, pp. 269-299.
- [Sharyl et al., 2001] Sharyl, J., Nass, I., Craig, H., Joyce, C. Lashof. (2001). *Mammography and Beyond: Developing Technologies for the Early Detection of Breast Cancer*. Committee on Technologies for the Early Detection of Breast Cancer, National Cancer Policy Board, NATIONAL ACADEMY PRESS. 2101 Constitution Avenue, N.W. Washington, DC 20418.
- [Shyu et al., 1999] Shyu, C., C. Brodley, A. Kak, A. Kosaka, A. Aisen., L. Broderick. (1999). *ASSERT: a physician-in-the-loop content-based retrieval system for HRCT image databases*. Computer Vision and Image Understanding, Vol 75, pp. 111-132.
- [Smith, 2003] Smith, R. A. (2003). *IARC handbooks of cancer prevention, volume 7, Breast cancer screening*, ISBN 92 832 3007 8, Vol 5(4), pp. 216-217.
- [Stinès, 2004] Stinès, J. (2004). *La densité mammaire : un concept radiologique*. 26emes journées de la SFSPM, Nancy, France.
- [Sobel, 1970] Sobel, I. E. (1970). *Camera models and machine perception*. Thèse de Doctorat, Stanford University, USA.
- [Smeulders et al., 2000] Smeulders, A., Worring, M., Santini, S., Gupta, A., Jain. R. (2000). *Content-based image retrieval at the end of the early years*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 22, 1349-1380.
- [Stricker and Orengo, 1995] Stricker, M., Orengo, M. (1995). *Similarity of color images*. In SPIE, Storage and Retrieval for Image Video Databases III, Vol 2420, pp. 381-392.
- [Su et al., 2003] Su, Z., Zhang, H.J., Li, S., Ma, S. (2003). *Relevance feedback in content-based image retrieval: Bayesian framework feature sub-spaces and progressive learning*. IEEE transactions on image processing, Vol 12(8), pp. 924-937.
- [Swain and Ballard, 1991] Swain, M., Ballard, D. (1991). *Color indexing*. International Journal of Computer Vision, Vol 7(1), pp. 22-39.

- [Sahoo et al., 2009] Sahoo, G., Kumar, T., Raina, B. L., Bhatia, C. M. (2009). *Text Extraction and Enhancement of Binary Images Using Cellular Automata*, International Journal of Automata and Computing, Vol.6(3), pp. 254-260.
- [Sakoda, 1971] Sakoda, J. M. (1971). *The Checkerboard Model of Social Interaction*. Journal of Mathematical Sociology, pp. 119-132.
- [Schelling, 1969] Schelling, T. C. (1969). *Models of Segregation (in Strategic Theory and Its Applications)*. American Economic Association, Vol. 59(2), pp. 488-493.
- [Schelling 1971] Schelling, T. C. (1971). *Dynamic Models of Segregation*. Journal Mathematical Sociology, Vol. 1(2), pp. 143-186.
- [Schonsch, 2002] Schonsch, B. (2002). *Propagation of Fronts in Cellular Automata*. Physica D, Vol. 80, pp. 433-450.
- [Schweiger and Krcmar, 2004] Schweiger, A., Krcmar, H. (2004). *Multi-Agent Systems for Active, Dynamic Activity Areas*. Proc 10th Americas Conference on Information Systems, AMCIS, New York, USA.
- [Shalizi, 2001] Shalizi, C. (2001). *Causal Architecture, Complexity and Self-Organization in Time Series and Cellular Automata*. Thèse de doctorat, University of Wisconsin, Madison, USA.
- [Sichman, 1995] Sichman, J. S. (1995). *Du raisonnement social chez les agents : une approche fondée sur la théorie de la dépendance*. Thèse de doctorat, Institut National Polytechnique de Grenoble. France.
- [Smuts, 1926] Smuts, J. C. (1926). *Holism and Evolution*. MacMillan. Compass/Viking Press 1961, Greenwood Press 1973 reprint, Sierra Sunrise.
- [Srividya et al., 2011] Srividya, B., Nandini, S. S., Ravi, S. M., Sunilkumar, S. M. (2011). *Intelligent Scheduling in Health Care Domain*. International Journal of Computer Science Issues, (IJCSI), Vol. 8(5), ISSN 1694-0814.
- [Tabar and Dean, 1985] Tabar, L., Dean, P.B. (1985). *Teaching atlas of mammography*. 2nd revised edition. Stuttgart. Georg Thieme Verlag.
- [Terry et al., 2002] Terry, R. P., Massimo, P., Singh, R., Sycara, K. (2002). *Communicating Agents in Open multi agent systems*. First GSFC/JPL Workshop on Radical Agent Concepts (WRAC), pp. 365-371. McLean, VA, USA.
- [Terzopoulos and Fleischer, 1988] Terzopoulos, D., Fleischer, K. (1988). *Deformable models*. The Visual Computer, ISSN 1432-2315 Vol 4(6), pp. 306-331.
- [Theyssier, 2005] Theyssier, G. (2005). *Automates cellulaires: un modèle de complexités*. Thèse de doctorat. Laboratoire de l'Informatique du Parallélisme, l'École Normale Supérieure de Lyon, France.
- [Thomas et al., 2002] Thomas, V., Bourjot, C., Chevrier, V., Desor, D. (2002). *MAS and RATS : Multi-agents simulation of social differentiation in rats groups*, Interest for the understanding of a complex biological phenomenon, International Workshop on Self-Organization and Evolution of Social Behaviour, Monte Verita, Suisse.
- [Vannoorenberghe, 2004] Vannoorenberghe, P. (2004). *Bouclage de pertinence par arbres de décision crédibilistes en indexation d'images par le contenu*. Compression et représentation des signaux audiovisuels (CORESA), Lille.
- [Vapnik, 1999] Vapnik, N. (1999). *An overview of statistical learning theory*. IEEE Transactions Neural Networks, Vol 10(5), pp. 988-999.
- [Vercouter, 2001] Vercouter, L. (2001). *Une Gestion Distribuée de l'Ouverture dans un Système Multi- Agent*. Journées Francophones d'Intelligence Artificielle Distribuée et des Systèmes Multi- Agents (JFIADSMA'01), pp. 177-188, Hermes, Montreal, Canada.

[Vercouter , 2004] Vercouter, L. (2004). *MAST : Un modèle de composants pour la conception de SMA*, Journée Multi-Agents et Composants (JMAC'04), Paris, France.

[Wang et al., 2001] Wang, J., Li, J., Wiederhold, G. (2001). *SIMPLIcity: Semantics-Sensitive Integrated Matching for picture Libraries*. IEEE Transactions on pattern analysis and machine intelligence, 23(9), pp. 947-963.

[Wang and Ma, 2005] Wang, D., Ma, X. (2005). *A hybrid image retrieval system with user's relevance feedback using neurocomputing*. Informatica, Vol 29(3), pp. 271-280, Slovenia.

[Weber et al., 1998] R. Weber, H. J. Schek., S. Blott. (1998). *A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces*. Proc of the 24 rd International Conference on Very Large Data Bases, VLDB '98, pp. 194-205. New York, USA.

[Weston and Watkins, 1999] Weston, J., Watkins, C. (1999). *Multi-class support vector machines*. Proc In European Symposium on Artificial Neural Networks ESANN'99, Brussels, Bzlgium.

[Weiss,1991] Weiss, S. M. (1991). *Small sample error rate estimation for K-NN classifiers*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13 (3), pp. 285-289.

[White and Jain, 1996] White, D., Jain, R. (1996). *Similarity indexing with the SS-tree*, Proc IEEE int'l conf data engineering. IEEE, pp. 516-523, Piscataway.

[Wolfe, 1967a] Wolfe, J. N. (1967). *A study of breast parenchyma by mammography in the normal woman and those with benign and malignant disease*. Radiology, Vol. 89, 201-5.

[Wolfe, 1967b] Wolfe, J. N. (1967). *Mammography: ducts as a sole indicator of breast carcinoma*. Radiology, Vol. 89, 206-10.

[Wolfe, 1976] Wolfe, J. N. (1976). *Breast patterns as an index of risk for developing breast cancer*. AJR Am J Roentgenol., Vol. 126(6), 1130-7.

[Wolfram, 1984] Wolfram, S. (1984). *Computation theory of cellular automata*, Communications in Mathematical Physics, Vol. 96(1), pp. 15-57.

[Wolfram, 2000] Wolfram, S. (2002). *A new kind of science*. ISBN 1-57955-008-8, Publié par Wolfram Media, chapitre 6, pp. 231-235

[Wooldridge and Ciancarini, 2001] Wooldridge, M., Ciancarini, P. (2001). *Agent-oriented software engineering: the state of the art*. Agent-Oriented Software Engineering, Lecture Notes in Computer Science, ISBN 978-3-540-44564-7, Vol. 1957, pp. 1-28

[Wooldridge, 2002] Wooldridge, M. (2002). *An Introduction to Multiagent Systems*. ISBN-13: 978-0470519462, Publié par John Wiley et Sons.

[Yan et al., 2012] Yan, L, Cheng, H. D., Jianhua, H., Yingtao, Z., Xianglong, T. (2012). *An Effective Approach of Lesion Segmentation Within the Breast Ultrasound Image Based on the Cellular Automata Principle*, Vol. 25(5), 580-590.

[Yang, 2004] Yang, C. (2004). *Content-based image retrieval: a comparison between query by example and image browsing map approaches*. Journal of Information Science, Vol. 30(3), pp. 254-267.

[Yeung and Liu, 1995] Yeung, M., Liu, B. (1995). *Efficient matching and clustering of video shots*. Proc IEEE int'l conf image processing, pp. 338-341, Piscataway.

[Yeung and Yeo, 1997] Yeung, M., Yeo, B. L. (1997). *Video visualization for compact presentation and fast browsing of pictorial content*. IEEE Transactions on Circuits Systems for Video Technology, Vol. 7 (5), 771-7.

[Zeigler and Ören,1986] Zeigler, B. P., Ören, T. I. (1986). *Multifaceted, multiparadigm modeling perspectives : tools for the 90's*. Proc. of the 18th conference on Winter simulation (WSC 86). ACM Press, pp. 708-712, New York.

[Zhang and Lu, 2004] Zhang, D., Lu, G. (2004). *Review of shape representation and description techniques*. Pattern Recognition, Vol. 37, pp. 1–19, North-Holland, Amsterdam.

[Zhang and Zhang, 1995] Zhang, H., Zhong, D. (1995). *A scheme for visual feature based image indexing*. Proc. of SPIE Conf. On Storage and Retrieval for image and video databases, pp. 36-46, San Jose.

[Zhang et al., 2005] Zhang, C., Zhang, Z., Cao, L. (2005). *Agents and Data Mining: Mutual Enhancement by Integration*. Autonomous Intelligent Systems: Agents and Data Mining, Lecture Notes in Computer Science, Springer-Verlag Berlin Heidelberg,ISBN 978-3-540-31932-0, Vol. 3505, pp. 50-61.

[Zhang et al., 2010] Zhang, C., Ouyang, D., Ning, J. (2010). *An artificial bee colony approach for clustering*. Expert systems with applications, ISBN 978-3-540-20950-8, Vol. 37 (7), pp. 4761-4767.