

الجمهورية الجزائرية الديمقراطية الشعبية

MA-004-83-1

وزارة التعليم العالي و البحث العلمي

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

جامعة سعد دحلب البليدة

UNIVERSITE SAAD DAHLEB DE BLIDA



FACULTE DES SCIENCES
DEPARTEMENT D'INFORMATIQUE

MEMOIRE DE PROJET DE FIN D'ETUDES

POUR L'OBTENTION DU DIPLÔME DE
MASTER EN INFORMATIQUE
OPTION : INGENIERIE DES LOGICIELS

Analyse texturale de cellules tumorales mammaires pour l'aide à
l'interprétation

Proposé par :

Mlle Reguieg F.Z
Mlle Benblidia Nadja

présenté par :

Benhenia Wassila
Djillali Fatima

Session : Juillet 2012

MA-004-83-1



Remerciment:

Nous voudrions avant tous remercier le bon Dieu le tous puissant qui nous a donné la force et le courage par réaliser ce modeste travail.

Et la prière sur notre guide le prophète "Mohamed" la plus chère personne à notre âme que la prière soit sur lui pour l'éternité.

Nous tenant à remercier nos promotrices « Melle Z.Rguieg » «et « Melle N.benblidia » qui nous bien superviser et guider pour bien mener notre projet.

Nous tenant à remercier tous nos enseignants ainsi que tout le personnel du département de l'informatique.

Dédicace:

J'ai l'immense honneur de dédier ce modeste travail :

A mes très chers parents qui ont toujours été là pour moi. Pour leurs encouragements, compréhensions et amour. Que Dieu les protèges.

A ma petite sœur adorée Sabrina que Dieu la garde.

A mes chères tentes et oncles, mes grandes mères et mon grand père.

A mes cousines: Asma, Amina, Sara, Ikram, imene et Samia.

A mes cousins : Amine, Hicham et Malek.

A la mémoire de mes chères tentes adorées victime du cancer: elles ont été la source de mon inspiration qu'elles reposent en paix.

A toutes mes copines : Nour el houda, Yasmina, Asma, Amina et Nerimene.

A mon binôme Fatima que Dieu la garde pour sa famille.

A tous les étudiants de master de la promotion LMD 2011 /2012.

A tous ceux qui nous ont aidés de près ou de loin.

Dédicace:

J'ai l'immense honneur de dédier ce modeste travail :

A mes très chers parents qui ont toujours été là pour moi. Pour leurs encouragements, compréhensions et amour. Que Dieu les protèges.

A ma chère sœur adorée Meriem, mes frères Brahim et Amine que Dieu les garde.

A mes chères tentes et oncles, mes grandes mères .

A Selma, Sihem et Youcef

A mon fiancé :Mahmoud et toute sa famille .

A mes très chères amies :Asma,Sara et Zineb.

A mon binôme Wassila que Dieu la garde pour sa famille.

A tous les étudiants de master de la promotion LMD 2011 /2012.

A tous ceux qui nous ont aidez de prés ou de loin.

Table des matières

Introduction générale	1
I. Chapitre I	
I.1.Introduction.....	3
I.2 Notions physiologiques et cytologiques du sein.....	4
I.2.1 Définition.....	4
I.2.2 Rappel anatomique.....	4
I.2.3 Structure de cancer du sein.....	5
I.2.4 Cytopathologie du sein.....	6
I.2.5 Pathologies mammaires.....	8
I.3 Les modèles de la vision humaine et les espaces couleurs.....	12
I.3.1 Les espaces couleurs.....	13
I.3.2 Les modèles de la vision humaine.....	17
I.4 Le screening.....	19
I.4.1 Le screening manuel.....	19
I.4.2 Le screening automatique.....	19
I.4.3 Structure d'un système informatique de rescreening.....	19
I.5 Conclusion.....	19
II. Chapitre II	
II. 1 Introduction.....	21
II.2 Segmentation par analyse des propriétés spatiales.....	22
II.2.1 Segmentation par division de régions.....	22
II.2.2 Segmentation par croissance de régions.....	23
II.2.3 Segmentation par fusion de régions.....	23
II.2. 4. Segmentation par division fusion.....	24
II.2. 5. Discussion.....	24
II. 3 Morphologie mathématique.....	24
II.3.1 Concepts de base de la morphologie mathématique.....	25
II.3.2 Ligne de partage des eaux (watershed).....	27
II. 4 Segmentation en contours.....	29
II.4.1 La Segmentation par contours actifs.....	29
II.4.2 La segmentation par les ensembles de niveaux (levels sets).....	31
II. 5 L'approche coopérative.....	33
II.6 Parcours de quelques travaux en imagerie cytologique.....	35
II. 7 Choix de l'espace et approches adoptées.....	47
II.8 Conclusion.....	48

III.	Chapitre III	
	III.1 Introduction.....	49
	III. 2 Qu'est ce qu'une texture.....	49
	III.2.1 Distribution aléatoire ou régulière.....	50
	III.2.2 Les échelles des textures.....	50
	III.2.3 Caractéristiques de la texture.....	50
	III.3 Paramètres texturaux.....	51
	III.4 Paramètres topologiques.....	55
	III. 5 Classification connexionniste.....	55
	III.5.1 Fonctionnement général d'un réseau de neurones.....	56
	III.5.2 Quelques réseaux célèbres.....	59
	III. 6 Conclusion.....	67
IV.	Chapitre IV	
	IV.1 Environnement de travail.....	68
	IV.1.1 Matériel utilisé.....	68
	IV.1.2 Qu'est ce que Matlab ?	
	IV.2 Présentation du système développé.....	69
	IV.2.1 Images étudiées.....	70
	IV.2.2 Description des modules du système 'cytocoop'.....	70
	IV.3 Présentation de l'application cytocoop.....	71
	IV.4 Présentation des résultats et discussion.....	77
	IV.4.1 Extraction des régions d'intérêt.....	77
	IV.4.2 Caractérisation texturale et topologique.....	79
	IV.4.3 La classification connexionniste.....	80
	IV.5 Conclusion.....	92
	Conclusion générale.....	93
	Références bibliographiques.....	95

Figure I.1 : chaîne de la V.A.O.....	3
Figure I.2 : division cellulaire anarchique.....	4
Figure I.3 : Structure du sein.....	5
Figure I.4 : cancer canalaire.....	6
Figure I.5 : étalement de la cytoponction.....	7
Figure I.6 : Les principaux éléments de la cellule (4).....	8
Figure I.7 : étapes de la division cellulaire cancéreuse.....	9
Figure I.8 : Différentes formes des calcifications.....	9
Figure I. 9 : images cytologiques des calcifications.....	10
Figure I. 10 : images cytologiques de carcinomes mammaires.....	11
Figure I.11 : le spectre des couleurs visibles.....	12
Figure I.12 : Espace RVB en 3D	13
Figure I.13 : la représentation spatiale du modèle xyz.....	14
Figure I.14 : le modèle $L^*a^*b^*$	15
Figure I.15 : schéma du système mixte de repérage des points de couleur.....	16
Figure I.16 : la représentation spatiale du modèle HSL.....	17
Figure I.17 : modèle du tristimulus.....	18
Figure I.18 : modèle de Faugeras.....	18
Figure II.1 : procédé du traitement par l'utilisation des espaces couleurs.....	21
Figure II.2 : Segmentation par division de régions.....	22
Figure II.3 : esquisse de la segmentation en 4 régions.....	23
Figure II.4 Erosion de (a) par un élément structurant circulaire.....	25
Figure II.5 . Dilatation de (a) par un élément structurant circulaire.....	26
Figure II.6 : Ouverture d'un ensemble d'objets X par un élément structurant circulaire.....	26

Figure II.7 : fermeture d'un ensemble d'objets X par un élément structurant.....	27
Figure II.8 : ligne de partage des eaux entre 2 bassins.....	27
Figure II.9 : algorithme de la ligne de partage des eaux	28
Figure II.10 : illustrations des grandes étapes du déroulement d'un contour actif.....	31
Figure II. 11 : schéma de la coopération séquentielle.....	34
Figure II. 12 : synoptique de la coopération des résultats.....	34
Figure II.13 : synoptique de la coopération mutuelle.....	35
Figure II. 14 : stratégie de segmentation d'images couleurs de cytologie des séreuses.....	36
Figure II. 15 : Résultats de la segmentation d'images couleurs de cytologie des séreuses.....	37
Figure II.16 : résultats de la segmentation par analyse d'histogramme.....	38
Figure II.17 : Segmentation des cytoplasmes.....	40
Figure II.18 : segmentation des noyaux.....	40
Figure II.19 :Segmentation cytoplasmique et nucléaire avec précision du centre de gravité.41	
Figure II.20 : la segmentation des noyaux basée sur la couleur.....	42
Figure II.21 :schéma de la stratégie appliquée aux noyaux issus de la segmentation couleur.42	
Figure II.22 : Résultats des traitements de séparation et de fusion.....	43
Figure II.23 : résultats de la transformation logarithmique.....	44
Figure II.24 : résultats de la binarisation.....	44
Figure II.25 : résultats de l'orientation.....	44
Figure II.26 : séparation de la glande mammaire du fond.....	45
Figure II.26 : résultat final du fond.....	45
Figure II. 27 : détection de la région d'intérêt: (a) image originale, (b) détection de la glande mammaire,(c) application des ensembles de niveaux , (d) Région détectée.....	46
Figure II.28 : Synoptique du choix de l'espace couleur.....	47
Figure III.1 : Quelques exemples de textures naturelles : écorce d'arbre, poil court, perles, tapis d'aiguilles de pin et de feuilles.....	49

Figure III.2 : Modèle d'un neurone biologique.....	56
Figure III.3 : Modèle d'un neurone artificiel.....	57
Figure III.4 : Représentation matricielle du modèle d'un neurone artificiel.....	58
Figure III.5 : Modèle des PMC.....	60
Figure III.6 : Deux cas de séparation.....	62
Figure III.7 : déroulement du réseau PMC	63
Figure III.8 : Représentation de la fonction sigmoïde.....	64
Figure III.9 : Etapes d'apprentissage du réseau PMC.....	66
Figure III.10 : principe de classification par le PMC.....	67
Figure IV.1 : synoptique général du système développé	69
Figure IV.2 : exemples d'images cytologiques mammaires.....	70
Figure IV.3 : fenêtre principale.....	72
Figure IV.5 : menu « changement d'espace couleur ».....	73
Figure IV.6 : menu « prétraitement ».....	73
Figure IV.7 : sous menu « gradient ».....	73
Figure IV.8 : sous menu « les filtres numériques ».....	73
Figure IV.9 : menu « segmentation ».....	74
Figure IV.10 : menu « caractérisation texturale et topologique ».....	74
Figure IV.11 : menu « l'apprentissage ».....	75
Figure IV.12 : structure du réseau.....	75
Figure IV.13 : menu « classification connexionniste ».....	75
Figure IV.14 : sous menu « classification ».....	76
Figure IV.15 : menu « outils ».....	76
Figure IV.16 : menu « A propos ».....	76
Figure IV.17 : segmentation d'une image présentant des carcinomes.....	77
Figure IV.18 : segmentation d'une image présentant des calcifications bénignes.....	78
Figure IV.19 : segmentation d'une image présentant des calcifications malignes.....	79

Figure IV.20 : principales étapes de l'apprentissage.....	81
Figure IV.21 : Convergence du réseau de neurones lors de l'étape d'apprentissage.....	82
Figure IV.22 : convergence du réseau de neurones lors de l'étape d'apprentissage.....	86
Figure IV.23 : convergence du réseau de neurones lors de l'étape d'apprentissage.....	89
Figure IV.24 : test de la classe des carcinomes.....	92
Figure IV.25 : test de la classe des cas sains.....	92
Figure IV.26 : test de la classe des calcifications bénignes.....	92
Figure IV.27 : test de la classe des calcifications malignes.....	93

Liste des tableaux :

Tableau VI-1 Test des vecteurs de la classe1	80
Tableau VI- 2 sortie de test de classe1	80
Tableau VI-3 Test des vecteurs de la classe 2.....	83
Tableau VI- 4 sortie de test de classe2.....	83
Tableau VI-5 Test des vecteurs de la classe3	84
Tableau VI- 6 sortie de test de classe3.....	84
Tableau VI-7 Test des vecteurs de la classe4.....	84
Tableau VI- 8 sortie de test de classe4.....	84
Tableau IV.9 : vecteurs de la classe 4.....	85
Tableau IV.10 : sortie de test de la classe 4.....	85
Tableau IV. 11 : vecteurs de la classe 1.....	86
Tableau IV.12 : test des vecteurs de la classe 1.....	86
Tableau IV.13: vecteurs de la classe 2.....	87
Tableau IV.14 : Test des vecteurs de la classe 2.....	87
Tableau IV.15 : vecteurs de la classe 3.....	87
Tableau IV.16 : Test des vecteurs de la classe 3.....	88
Tableau IV. 17 : vecteurs de la classe 4.....	88
Tableau IV. 18: test des vecteurs de la classe 4.....	88
Tableau IV.19 : vecteurs de la classe 1.....	89
Tableau IV.20 : test des vecteurs de la classe 1.....	89
Tableau IV.21 : vecteurs de la classe2.....	90
Tableau IV.22 : Test des vecteurs de la classe2.....	90
Tableau IV.23 : vecteurs de la classe3.....	90
Tableau IV.24 : test des vecteurs de la classe3.....	91
Tableau IV.25 : vecteurs de la classe 4.....	91
Tableau IV.26 : test des vecteurs de la classe 4.....	91

Résumé

Ce mémoire, appréhende l'analyse d'images cytologiques du sein pour l'aide au diagnostic. Le système mis en œuvre, permet la localisation des régions d'intérêt des différentes structures, suivant une segmentation en régions via la morphologie mathématique et une segmentation en contours, suivant les ensembles de niveaux. Les zones issues de ces deux approches, sont caractérisées par des descripteurs texturaux et topologiques, en vue d'une classification connexionniste des différentes pathologies étudiées.

Mots-clés: Image cytologique du sein, aide au diagnostic, morphologie mathématique, Ensembles de niveaux, Descripteurs texturaux et topologiques, classification connexionniste.

Summary

This thesis, demonstrates the analysis of cytological images of breast that helps diagnosis assistance. The system implemented, enables the localization of regions of interest for different structures, depending on regions segmentation using mathematical morphology and segmentation of edges, regarding the level sets. Areas from these two approaches are characterized by textural and topological descriptors, for classification of different pathologies studied.

Keywords: cytological Image of breast, mathematical morphology, level set, textural and topological descriptor, classification connectionist aid to interpretation

ملخص

هذه المذكرة تحلل مجموعة من الصور الخلوية للثدي للمساعدة في التشخيص. ويعتبر نظام تنفيذها، موقع المجالات التي تهتم مختلف الهياكل، بإتباع تقطيع للمناطق بطريقة المورفولوجيا الرياضية وتقطيع الحدود، عقب مجموعة من المستويات. المناطق الناتجة من هذين النهجين، تتميز بالواصفات الطوبوغرافية، والتكوينية لتصنيف الأمراض المختلفة المدروسة

الكلمات الرئيسية

من الصور الخلوية للثدي، المورفولوجيا الرياضية، مجموعة من المستويات، الواصفات الطوبوغرافية، والتكوينية، لتصنيف.

Introduction générale

La recherche en imagerie médicale, est une des disciplines les plus actives du traitement d'images, l'un des modules prépondérants, de la vision par ordinateur. L'interprétation d'images, dans un contexte clinique, reste un problème majeur de l'imagerie médicale. La multiplicité des modalités d'imagerie, ainsi que les fortes variabilités des structures et pathologies à analyser, rendent cette tâche fastidieuse. L'importante quantité de données, ainsi que les nombreuses applications liées à l'imagerie médicale, nécessitent des logiciels de très haut niveau combinant des interfaces graphiques complexes avec des algorithmes interactifs rapides, pour une analyse plus concise. Le traitement d'images, répond à cette problématique, pour l'aide au diagnostic clinique

Le travail appréhendé dans ce projet, suscite l'intérêt des laboratoires LATSI et LRDSI, pour la mise en œuvre d'un système de détection de pathologies mammaires; à savoir les calcifications, les masses et les carcinomes. En effet, le cancer du sein, constitue la cause de décès, la plus fréquente chez la femme, en particulier en Algérie, où chaque année 10000 nouveaux cas sont décelés [1]. Avec le vieillissement de la population, on constate même que la mortalité par cancer du sein pour la tranche d'âge 20-75 ans a augmenté de 10% en vingt années d'observation [1]. Différentes études, ont confirmé que c'est la détection à un stade précoce des cancers infracliniques, qui peut en améliorer le pronostic. L'examen cytologique, est un examen complémentaire à la mammographie.

Dans ce cadre, en cytologie, le cytopathologiste doit établir à l'issu d'examens minutieux de lames de cellules, un diagnostic qui doit être le plus fiable possible. Le pathologiste reconnaît les types cellulaires présents sur une lame et c'est cette étude de la lame, qui détermine le diagnostic. Cet examen se fait via le microscope. Il est donc visuel, manuel et par conséquent très fastidieux et pourrait engendrer des erreurs de lecture. Afin de faciliter la lecture des lames, des recherches ont permis de développer des systèmes semi-automatiques permettant un diagnostic plus précis.

Le pathologiste est alors assisté dans son travail. De tels systèmes appellent à des techniques de traitement et d'analyse d'images, d'intelligence artificielle et de biologie. Nous cherchons à réaliser un système informatique semi-automatique permettant de réaliser la détection de cellules cancéreuses mammaires sur une lame de prélèvement cytologique. L'objectif appréhendé par le présent travail, est d'analyser les cellules cancéreuses mammaires dans des images cytologiques couleurs, afin d'aider le cytopathologiste dans son interprétation. Dans ce cadre, les images sont en

premier lieu, segmentées suivant une technique d'analyse de données basée sur la morphologie mathématique et les ensembles de niveaux, considérant ainsi, aussi bien l'information globale que locale, de façon à discerner toutes les structures pertinentes. En second lieu, les régions obtenues, sont caractérisées suivant des critères de malignité établis par la texture et la topologie, en vue d'une classification supervisée de la cellule suivant les types de pathologies rencontrées.

L'élaboration d'un tel système; est basée sur trois étapes séquentielles;

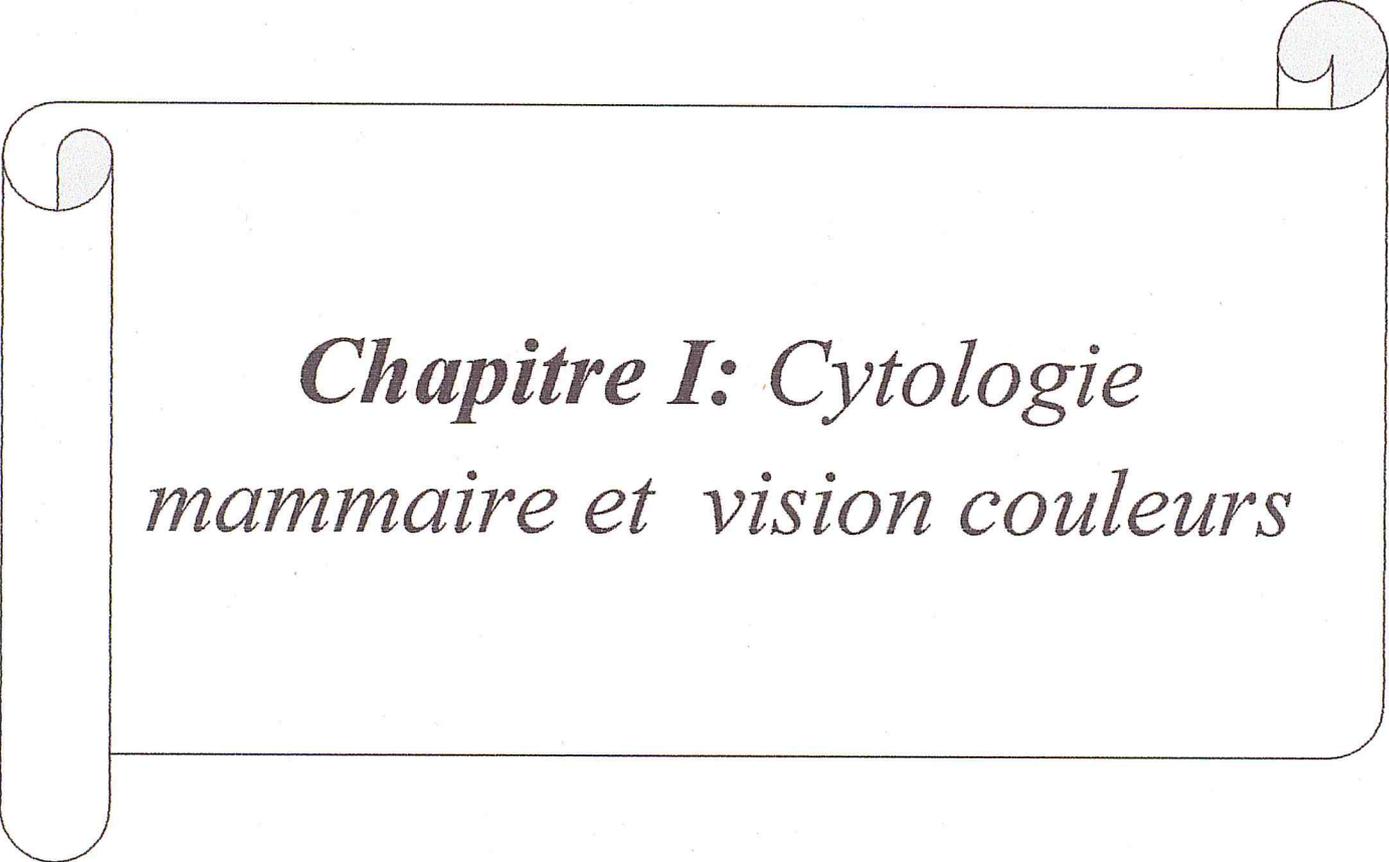
- ✓ **La segmentation** : elle consiste à créer des partitions dans l'image en identifiant des pixels similaires. Le critère de similarité permet de contrôler l'aspect final de la segmentation.
- ✓ **La caractérisation** : elle fournit des caractéristiques descriptives des régions d'intérêt.
- ✓ **La classification** : cette étape permet d'engendrer de classifier les différentes structures cytologiques caractérisées suivant un modèle connexionniste.

Afin de mieux comprendre notre approche, nous consacrons le premier chapitre à la vision couleurs ainsi qu'aux notions cytologiques;

Le deuxième chapitre, sera consacré aux techniques de segmentation en imagerie cytologique;

Le troisième chapitre, sera dédié à la caractérisation texturale et topologique et à la classification connexionniste;

Et enfin le quatrième et dernier chapitre, présentera l'application et les différents résultats obtenus sur des images cytologiques du sein.



*Chapitre I: Cytologie
mammaire et vision couleurs*

I.1 Introduction

La cytologie du sein, a une importance fondamentale pour l'étude des cancers où la cellule est l'unité élémentaire malade. En pratique, elle se consacre surtout à l'examen microscopique des cellules cancéreuses, pour reconnaître leur aspect, leur origine et leurs anomalies [2]. Dans le cadre de la détection du cancer du sein, le cytopathologiste, expertise alors les tumeurs dans les laboratoires cliniques et le méthodologiste, caractérise les données suivant, la vision par ordinateur pour l'aide au diagnostic.

La vision artificielle ou la vision assistée par ordinateur (V.A.O) symbolise toute la chaîne visuelle partant de l'image brute et allant jusqu'à l'interprétation de son contenu. Il existe donc un aspect décisionnel très important lors de la phase de compréhension, d'où l'interaction avec l'intelligence artificielle. Le traitement d'images s'intègre dans ce processus comme un outil, il n'a aucun pouvoir décisionnel car il se définit comme un ensemble de tâches destinées à extraire de l'image des informations qualitatives et quantitatives qui sont transmises au module de niveau supérieur la figure I.1 résume les composantes d'une chaîne de reconnaissance visuelle [3].

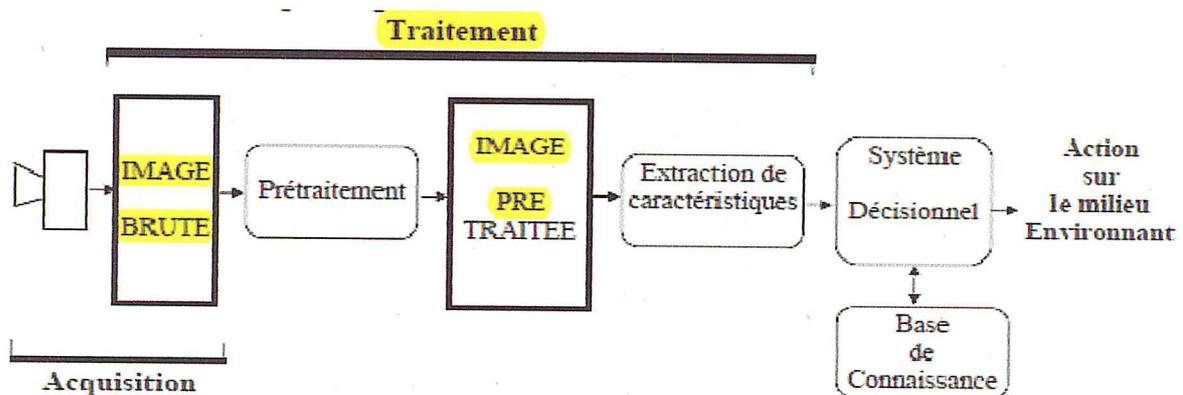


Figure I.1 : chaîne de la V.A.O

Le capteur dépend du domaine d'application où dans le cas de l'imagerie cytologique, il est représenté par un microscope électronique. Le prétraitement permet de corriger les artefacts de l'image; l'extraction des caractéristiques à pour rôle, de faire une description de l'image compatible avec les objets contenus dans la base de connaissance. Enfin, le système décisionnel, réalise la comparaison entre les caractéristiques qui lui sont fournies en amont et le contenu de la base de connaissance proprement dite.

Dans ce chapitre, nous introduisons la cytologie mammaire et la vision couleurs, afin d'appréhender le contexte médical et l'aspect qualitatif et quantitatif des données.

I.2 Notions physiologiques et cytologiques du sein [4] [5]

I.2.1 Définition

Une cellule cancéreuse est une cellule, qui s'est déformée, dont la structure du cytoplasme et du noyau est altérée. Elle se met alors à se multiplier anarchiquement et finit (cf. figure I.2), par former une masse qu'on appelle tumeur.

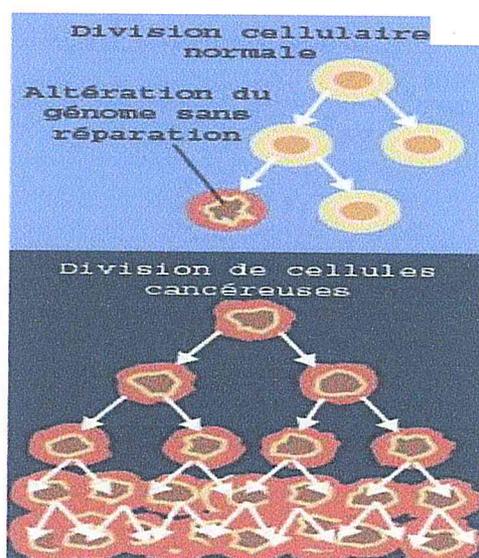


Figure I.2 : division cellulaire anarchique

I.2.2 Rappel anatomique

Les seins sont constitués de tissu graisseux contenant les glandes mammaires. Ils reposent sur une couche musculaire, les muscles pectoraux, recouvrant eux mêmes la paroi thoracique.

Ils sont soutenus par un système constitué par la peau et des fibres pénétrant à l'intérieur de la glande et participant ainsi à leur maintien (cf. Figure I.3).

C'est un tissu extrêmement hétérogène où coexistent des structures canalaire, glandulaires, fibro-conjonctives et adipeuses diversement enchevêtrées.

L'architecture de la glande mammaire comporte une vingtaine de lobes, eux-mêmes constitués de lobules au sein desquels, se base tout le raisonnement physiopathologique du sein. Tout ce système, est organisé autour d'un arbre galactophorique.

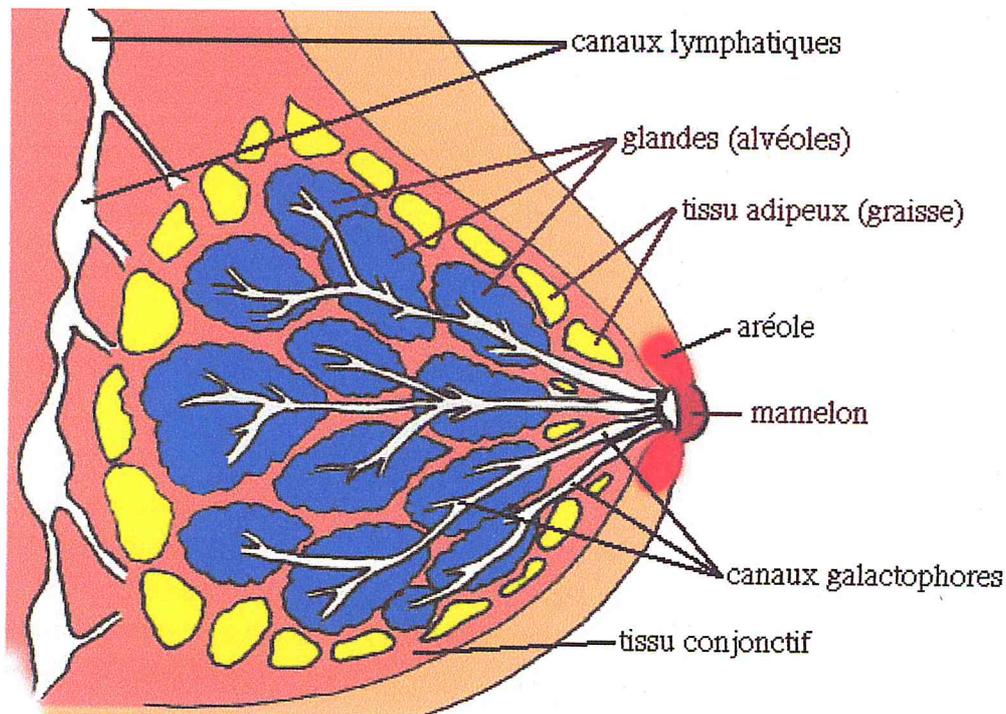


Figure I.3 : Structure du sein

I.2.3 Structure du cancer mammaire

Les cellules des tumeurs malignes ont tendance à quitter leur tissu d'origine et à envahir les tissus voisins: on parle de cancer infiltrant (cf. figure I.3).

Certaines tumeurs restent dans leur tissu d'origine sans infiltrer les tissus voisins: on parle alors d'un cancer in situ. Les cellules des cancers infiltrants, peuvent se détacher de la tumeur d'origine et se propager par les vaisseaux sanguins ou lymphatiques. Elles s'accumulent alors dans les ganglions lymphatiques voisins. On dit qu'il existe un envahissement ganglionnaire.

Les cellules cancéreuses ont tendance à migrer dans d'autres organes ou parties du corps, et à y développer de nouvelles tumeurs qu'on appelle métastases. On dit dans ce cas que le cancer est métastatique.

Tous les cancers ne se comportent pas de la même façon. C'est la raison pour laquelle les cancérologues, essaient d'adapter le traitement adapté à chaque cas. En l'absence de traitement adapté, le cancer risque d'évoluer plus vite.

On peut distinguer quatre niveaux structurant les cellules; les tissus, les organes et les appareils ou systèmes. A son niveau structural le plus simple, l'organisme est constitué de cellules: la cellule est l'unité fonctionnelle la plus petite de l'organisme. Les cellules peuvent être différentes entre elles selon leur fonction. Toutes les cellules vivantes possèdent des caractéristiques communes (cf. Figure I.5).

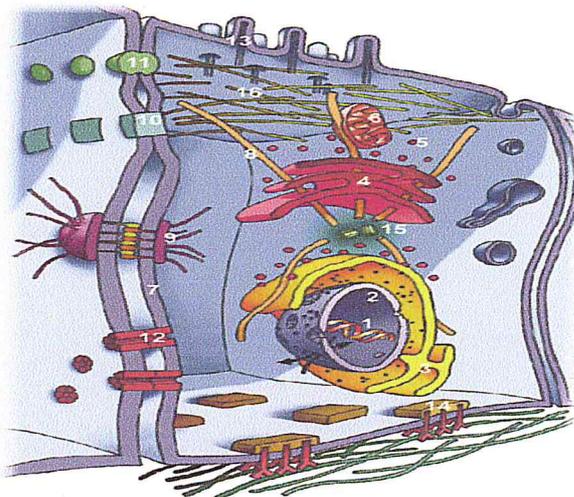


Figure I.4 : Les principaux éléments de la cellule (4).

1. nucléole, 2. Enveloppe nucléaire, 3. Réticulum endoplasmique rugueux, 4. Appareil de Golgi, 5. Vésicules, 6. Mitochondrie, 7. Membrane plasmique, 8. Microtubules, 9. Neurone, 10. Et 11. Jonctions adhérentes, 12. gap jonction, 13. microvilosité, 14. desmosomes, 15. Centriole, 16. microfilaments

L'image ci-dessus présente une structure détaillée d'une cellule. Elle est composée de petits éléments appelés structures subcellulaires. Un composant fondamental de la cellule, est la membrane qui la délimite. Cette membrane entoure également de petits éléments appelés organites. La substance fluide où baignent ces organites, est appelée cytosol ou cytoplasme. Le cytosol contient de nombreux enzymes essentiels et des métabolites. Le matériel génétique se situe dans le noyau. L'énergie nécessaire est fournie par la mitochondrie [2].

➤ **Caractères cytologiques**

La cellule pathologique, présente quelques critères pathologiques au niveau du cytoplasme et du noyau en particulier. Nous citons quelques unes de ces anomalies :

a. Anomalies du noyau

Des anomalies encore mal connues de la conformation de la chromatine, les anomalies de la chromatine, en particulier la ploïdie, sont à l'origine de multiples anomalies nucléaires:

- Les noyaux sont globalement augmentés de taille. Cette augmentation est, en partie, à l'origine d'une augmentation du rapport nucléo-cytoplasmique.
- Les noyaux sont de taille inégale d'une cellule à l'autre: anisocaryose.

- Les noyaux sont foncés, denses: hyper chromatisme.
- La chromatine est anormalement répartie, en mottes.
- La membrane nucléaire est épaissie, les contours nucléaires sont irréguliers.
- Une cellule tumorale peut avoir plusieurs noyaux : multinucléation.
- Les nucléoles sont multiples, volumineux, irréguliers.
- Le nombre de mitoses est augmenté. Ces mitoses sont fréquemment anormales (tri- ou tétrapolaires).

b. Anomalies du cytoplasme

- Le cytoplasme est moins abondant, ce qui contribue aussi à l'augmentation du rapport nucléo-cytoplasmique.
- La taille des cellules est variable (anisocytose).
- Le cytoplasme est basophile (par augmentation de son contenu en acides nucléiques).
- Il peut contenir des inclusions (par exemple des amas de filaments intermédiaires) et des vacuoles (de lipides ou de glycogène).

Dans certains cas, des cellules non cancéreuses peuvent présenter des anomalies identiques au cours des processus inflammatoires, des viroses, après irradiation... De plus, aucun de ces caractères morphologiques n'est constant; une cellule cancéreuse peut avoir une morphologie normale. Le cytodagnostic, cependant, n'a qu'une valeur d'orientation. Avant toute sanction thérapeutique, il doit être confirmé par l'étude histologique d'une biopsie.

Pour un même type de tumeur, l'intensité de ces anomalies est variable et permet de définir des grades de malignité des tumeurs, en fonction de l'importance des altérations morphologiques des cellules du clone tumoral, en particulier.

I.2.5 Pathologies mammaires

Le cancer du sein est le cancer le plus fréquent chez la femme dans le monde entier, en particulier en Algérie, où chaque, nous distinguons plus de 10000 cas [1]. Un quart des cancers de la femme, sont des cancers du sein. Le cancer du sein atteint des femmes à des âges différents, mais il se développe le plus souvent autour de 45 ans. Près de la moitié des Ce cancer peut aussi apparaître chez l'homme, mais c'est extrêmement rare (moins de 1 % des cancers du sein).

La division cellulaire se fait d'une manière anarchique suivant la figure ci-dessous;

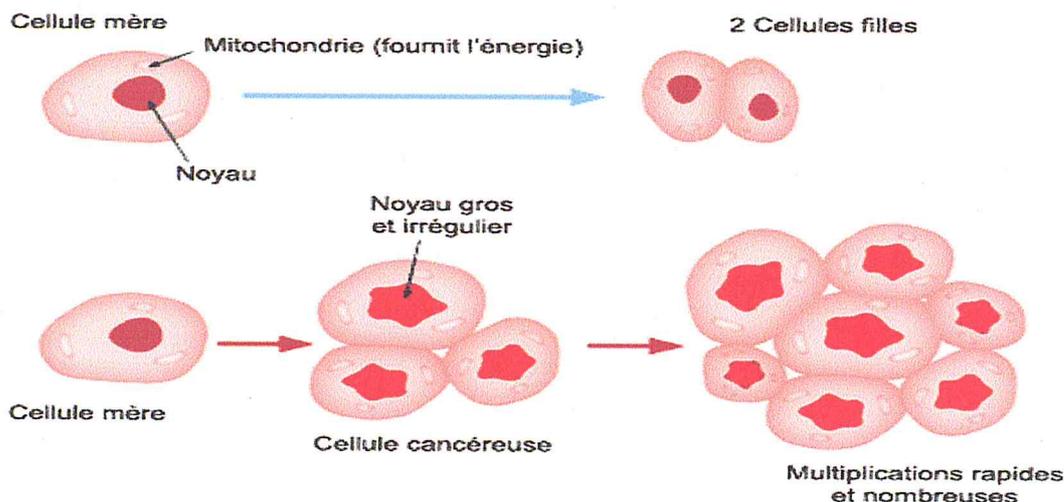


Figure I.6 : étapes de la division cellulaire cancéreuse

Nous remarquons parmi les types de pathologies mammaires, les calcifications, les carcinomes canaux et lobulaires.

➤ Les calcifications

Ce sont des dépôts de sels de calcium qui apparaissent sous forme de points "blancs et brillants" à la mammographie ; en fonction de leur taille, on parle de micro ou macro calcifications. Ces dernières, sont souvent bénignes alors que les microcalcifications demandent plus d'attention. Le diagnostic dépend de leur position dans le sein, de leur arrangement géométrique (linéaire, en forme d'agrégats, ...) et du nombre de microcalcifications dans un amas, et aussi de la comparaison avec une mammographie antérieure.

Selon Le Gal [2], on distingue 5 types de microcalcifications (cf. Figure I.6), les calcifications de type I (rondes et planes), celles de type II (images régulières), celles de type III (images en grains de sel), les calcifications de type IV (points irréguliers) et celles de type V (formes vermiculaires).

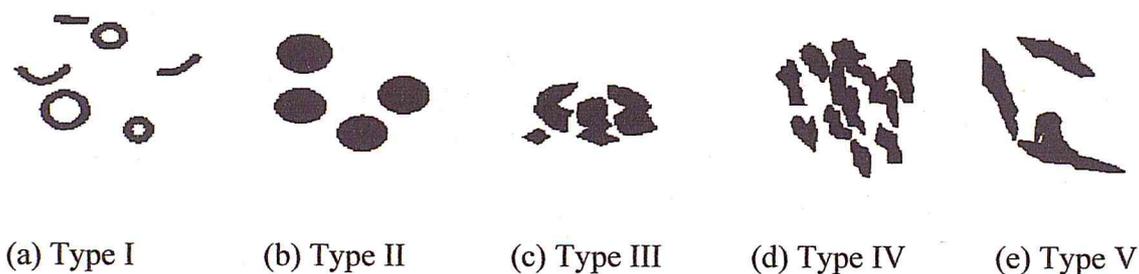
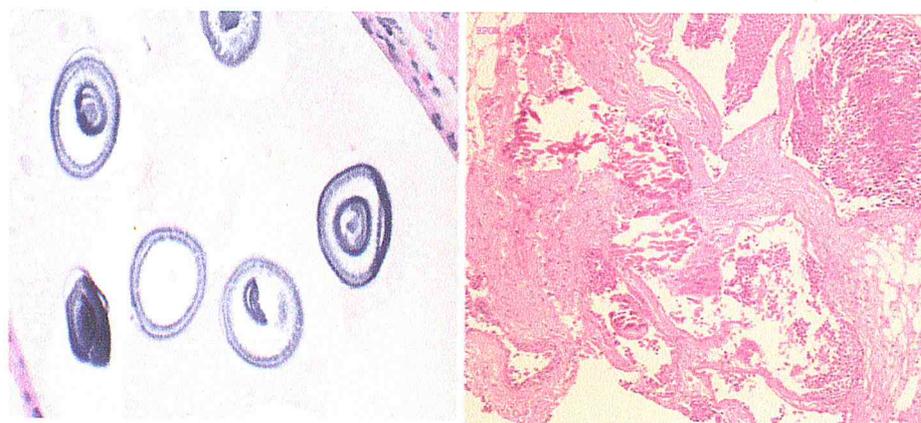


Figure 1. 7: différentes formes des calcifications

La figure I.8 illustre le cas de deux types de calcifications : le cas bénin et le cas malin.



A : calcifications bénignes

B : calcifications malignes

Figure 1. 8: images cytologiques des calcifications

➤ Les carcinomes

Lorsque les cellules cancéreuses ont infiltré le tissu qui entoure les canaux et les lobules, on parle de cancer ou carcinome infiltrant. Le cancer canalaire est le cancer infiltrant le plus fréquent. Huit cancers sur dix sont des cancers canaux infiltrants. Le cancer lobulaire infiltrant est plus rare. Les cancers infiltrants, peuvent se propager vers les ganglions ou vers d'autres parties du corps. Les ganglions le plus souvent, atteints par les cellules du cancer du sein se trouvent sous les bras, au niveau des aisselles. On les appelle les ganglions axillaires. Le traitement des cancers infiltrants repose sur un traitement local (chirurgie, radiothérapie), associé dans certains cas à un traitement général (chimiothérapie, hormonothérapie).

Lorsque les cellules cancéreuses se trouvent uniquement à l'intérieur des canaux ou des lobules, sans que la tumeur ait infiltré le tissu qui les entoure, on parle de cancer in situ.

Le cancer canalaire in situ ou carcinome canalaire in situ, est le plus fréquent: huit cancers in situ sur dix sont des cancers canaux in situ. Ils sont le plus souvent découverts lors d'un dépistage systématique (une mammographie ou un examen des seins par le médecin) et quelquefois révélés par un écoulement de sang du mamelon ou un eczéma du mamelon. Il arrive que les cancers canaux in situ, infiltrent les tissus voisins et deviennent des cancers infiltrants. Le traitement des cancers canaux in situ, est un traitement local comme la chirurgie ou la radiothérapie. Le cancer lobulaire in situ (ou néoplasie lobulaire in situ) est plus rare.

Le cancer du sein apparaît avant tout sous la forme d'adénocarcinomes. Ils sont les plus fréquents (95 %). Ils se développent à partir :

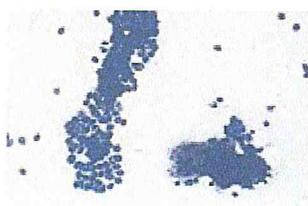
1. des cellules des canaux servant au transport de la prolactine (cancer canalaire): le carcinome canalaire in situ non-invasif (DCIS):

Prolifération cancéreuse qui se développe dans la lumière des canaux sans franchir la membrane basale et sans envahir le tissu conjonctif. Il constitue de 2-3% des cancers du sein. Il a un excellent pronostic s'il est découvert rapidement. Non traité, il peut devenir invasif; le risque de récurrence est de 40%.

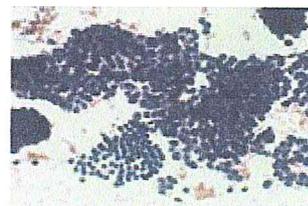
Le Carcinome canalaire infiltrant: Les cellules tumorales ont envahi le tissu conjonctif et peuvent donner des métastases. Il est le cancer infiltrant le plus fréquent. Huit cancers du sein sur dix sont des cancers canaux infiltrants. Ce cancer a le pire pronostic avec des métastases fréquentes aux ganglions, aux os, poumons, foie et cerveau.

2. Des lobules (cancer lobulaire) unités sécrétoires du sein : carcinome lobulaire in situ (LCIS) : Cancer qui se développe dans les lobules, sans franchir la membrane basale et sans envahir le tissu conjonctif 2-3% des cancers du sein. Il est habituellement découvert à l'occasion d'une biopsie du sein. Le cancer lobulaire infiltrant est plus rare. Ce cancer peut se retrouver à plusieurs endroits et sur les deux seins à la fois. Le carcinome lobulaire infiltrant, peut donner des métastases ganglionnaires axillaires. (5-10% des cancers du sein). Il se reconnaît par un épaissement du tissu mammaire.

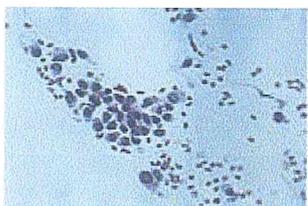
Nous distinguons d'autres types de cancers infiltrants tels que : le carcinome inflammatoire du mamelon ou la maladie de Paget.(1%-4%) des cancers du sein, le carcinome médullaire infiltrant (5-8% des cancers du sein) : le carcinome colloïde (mucineux): (2-4% des cancers du sein). Nous distinguons dans la figure I.9, quelques images cytologiques de carcinomes.



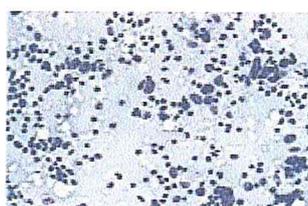
A, adénofibrome.



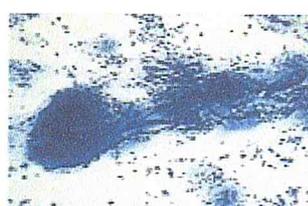
. B, carcinome canalaire invasif.



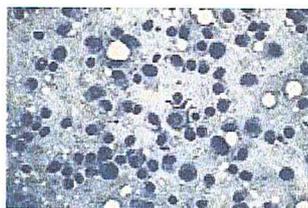
D, le carcinome canalaire invasif.



E, le carcinome médullaire.



, F carcinome mucineux.



G, carcinome lobulaire infiltrant.

Figure 1. 9: images cytologiques de carcinomes mammaires

I.3 Les modèles de la vision humaine et les espaces couleurs [6][7]

Un modèle de couleurs, décrit les couleurs que nous voyons et que nous utilisons dans les images numériques. Chaque modèle, RVB (Rouge, Vert, Bleu), CMJN (Cyan, Magenta, Jaune, Noir), (basé sur le, CMJN ou TSL), représente une méthode différente de description des couleurs.

Un espace colorimétrique, est une variante d'un modèle de couleur caractérisée par sa propre gamme des couleurs. Par exemple, le modèle de couleurs RVB, recouvre plusieurs espaces colorimétriques : Adobe RVB, sRVB, ProPhoto RGB, etc. Chaque périphérique, a son propre espace colorimétrique et ne peut reproduire que les couleurs situées dans la gamme correspondante. Lorsqu'une image passe d'un périphérique à un autre, ses couleurs peuvent varier, car chaque périphérique interprète les valeurs RVB ou CMJN selon son propre espace colorimétrique.

La couleur, est le résultat de la perception d'ondes de lumière sur la rétine dans une région visible du spectre, à savoir les longueurs d'onde comprises entre 380nm ($1\text{ nm} = 10^{-9}\text{m}$) pour le violet et 780nm pour le rouge (cf. Figure I.11).

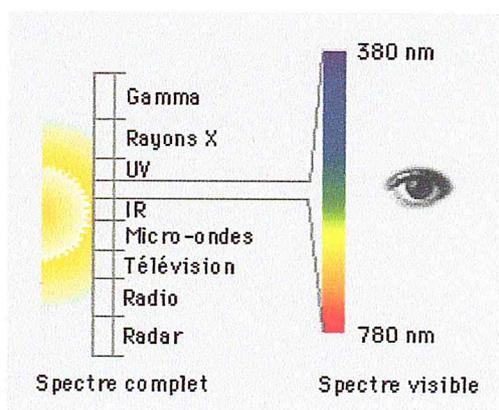


Figure I.10 : le spectre des couleurs visibles [6]

Le spectre des couleurs visibles est continu, il n'existe donc pas de couleur pouvant être véritablement appelée rouge, verte ou bleue car les transitions dans le spectre sont douces et produisent un effet de fondu (visible dans un arc-en ciel)

On peut représenter les différentes teintes du spectre lumineux par deux façons: soit par représentation fréquentielle ou par l'espace des couleurs (ce modèle a été étudié et modélisé par la Commission Internationale de l'Eclairage (CIE)).

I.3.1 Les espaces couleurs

Une couleur est généralement représentée par trois composantes. Ces composantes définissent un espace des couleurs. On peut citer l'espace RVB, l'espace CIE XYZ ou XYZ, ou encore l'espace Lab. Selon l'espace de couleurs choisi pour représenter une image couleur, le nuage des couleurs (c'est à dire l'ensemble des couleurs de l'image) n'aura pas la même répartition dans l'espace.

1. L'espace RVB

L'espace RVB est sans doute l'espace de couleurs le plus utilisé. Les systèmes de télévision, s'y appuient fortement. La représentation des couleurs dans cet espace (cf. Figure 2) donne un cube appelé **cube de Maxwell** [6].

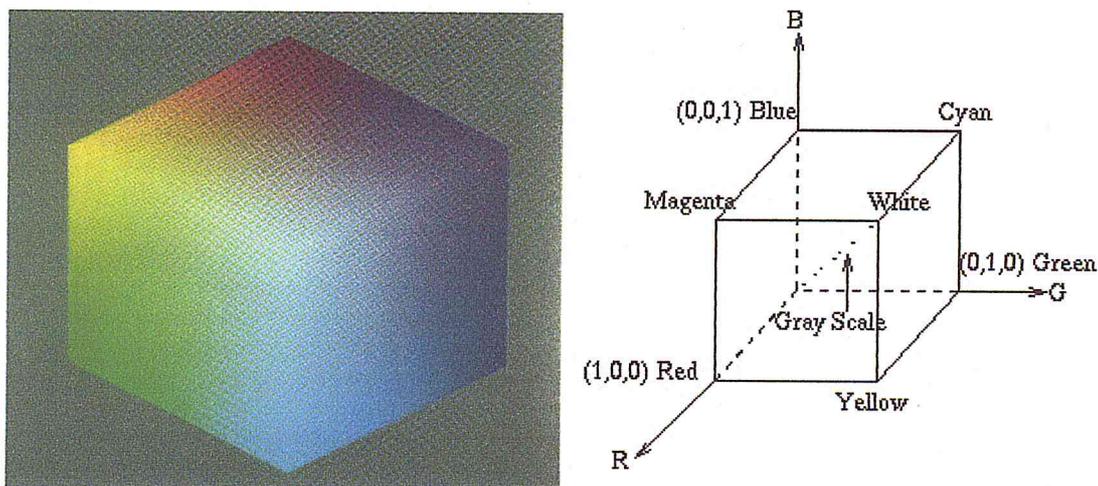


Figure I.11: Espace RVB en 3D [7]

Ce modèle utilise un système de coordonnées cartésiennes et définit un cube unité comme indiqué sur la figure I.11. Chaque axe correspond à une couleur primaire : rouge, verte et bleue. Une couleur est donc spécifiée, en indiquant les contributions de chaque couleur primaire. Ces couleurs sont additives, c'est-à-dire que les contributions de chaque primaire sont ajoutées pour créer le résultat.

2. L'espace CIE XYZ

L'espace XYZ (cf. Figure 1.12) dépend linéairement du système RVB. Il a été introduit par la CIE « Commission internationale de l'éclairage » vers une description des couleurs conforme à la vision humaine pour remédier aux inconvénients du système RVB ;

Il a introduit la notion de *luminance* Y , intensité lumineuse subjective indépendante de la couleur. Il utilise deux autres grandeurs choisies de telle manière qu'elles prennent toujours des valeurs positives pour décrire les couleurs visibles

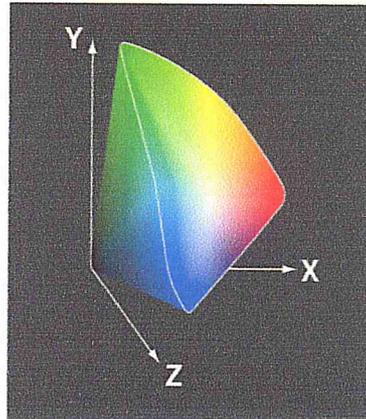


Figure I.12 : la représentation spatiale du modèle xyz

Le système XYZ pallie cet inconvénient et une couleur est représentée par une combinaison positive de X , Y et Z ; On a alors :

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} x_r & x_v & x_b \\ y_r & y_v & y_b \\ z_r & z_v & z_b \end{pmatrix} * \begin{pmatrix} R \\ V \\ B \end{pmatrix}$$

I.1

La matrice de changement de repère, est déterminée en mesurant les coordonnées XYZ correspondant aux trois couleurs primaires RVB suivantes :

$(R_{max}, 0, 0)$, $(0, V_{max}, 0)$, $(0, 0, B_{max})$ obtenues en appliquant le maximum des valeurs individuellement à chaque stimulus par rapport au blanc de référence utilisé.

Y est appelée la luminance (le degré de noir d'une couleur). En normalisant X , Y et Z , on obtient le système xyz défini par les relations suivantes :

$$\begin{aligned}
 x &= \left(\frac{X}{X+Y+Z} \right) \\
 y &= \left(\frac{Y}{X+Y+Z} \right) \\
 z &= \left(\frac{Z}{X+Y+Z} \right) \\
 x + y + z &= 1
 \end{aligned}
 \tag{I.2}$$

Comme $x + y + z = 1$, seules les variables x et y sont indépendantes, ces deux variables sont les composantes chromatiques de la couleur.

On peut disposer alors du diagramme de chromaticité de la CIE qui décrit le plan défini par les variables de chrominance x et y ; Le blanc est représenté par le point définissant un mélange en quantité égale des trois primaires XYZ .

En combinant ce diagramme avec la luminance on obtient l'espace de couleur xyY

3. L'espace CIE $L^*a^*b^*$

Le système $L^*a^*b^*$ est issu du CIE XYZ (cf. Figure I.13).. Il essaye de prendre en compte la réponse logarithmique de l'œil. Il possède le grand avantage d'être uniforme. Il est très utile dans le cas de mélanges de pigments, par exemple, pour l'industrie graphique ou du textile

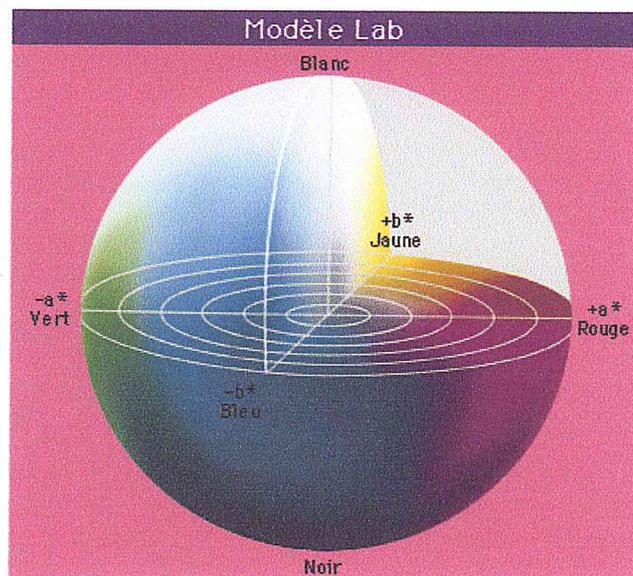


Figure I.13 : le modèle $L^*a^*b^*$ [7]

Une des difficultés majeures de ce système, est qu'il utilise un système mixte de repérage des points de couleur. La saturation est mesurée de manière cartésienne, alors que la teinte et la luminosité est mesurée de manière angulaire (cf. Figure I.14).

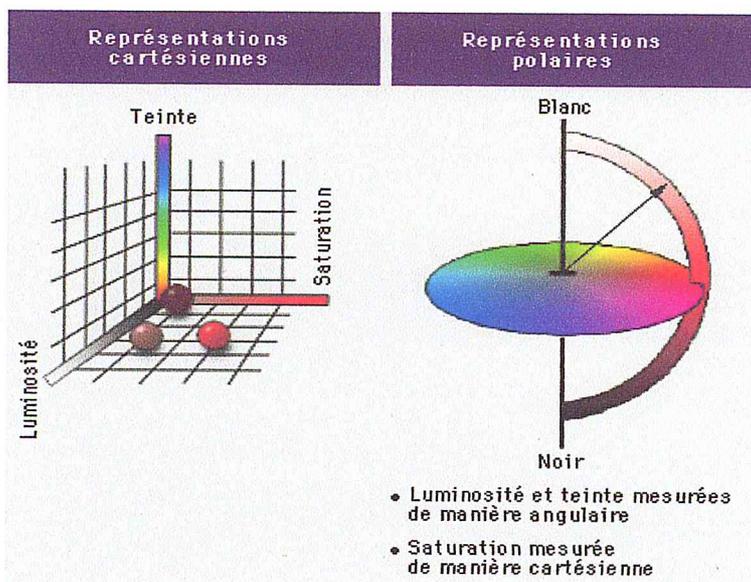


Figure I.14 : schéma du système mixte de repérage des points de couleur [7]

2. l'espace $L^*U^*V^*$

Le problème avec l'espace XYZ, est son manque d'uniformité dans la perception à savoir que deux couleurs à la même distance d'une autre, ne sont pas nécessairement perceptuellement équidistances [6]. Pour corriger cet inconvénient la CIE a proposé en 1976 le modèle $l^*u^*v^*$ qui est perceptuellement uniforme, c'est-à-dire que si deux couleurs A et B sont aussi proches entre elles que deux autres couleurs C et D, alors la différence est perçue entre C et D. Les équations permettant de passer du système XYZ au $l^*u^*v^*$ sont données ci-dessous :

$L^* = 116(Y/Y_0)^{1/3} - 16$	si $(Y/Y_0) > 0.008856$
$L^* = 9033(Y/Y_0)$	ailleurs
$U^* = 13L^*(U' - u')$	
$V^* = 13L^*(V' - v')$	
$U' = 4X/(X+15Y+3Z)$	$V' = 9X/(X+15Y+3Z)$
$u' = 4X_0/(X_0+15Y_0+3Z_0)$	$v' = 9Y_0/(X_0+15Y_0+3Z_0)$

4. L'espace de couleur HSL (Hue, Saturation, Luminance)

Cet espace est basé sur les trois grandeurs de Teinte, Saturation et Luminance. La Figure I.15 donne une représentation de cet espace (dans l'espace euclidien RVB). Cet espace est obtenu à partir de l'espace RVB et est décrit par les équations suivantes :

$$H = a \cos \left[\frac{((R - V) + (R - B))}{2\sqrt{(R - V)^2 + (R - B)(V - B)}} \right] \text{ et si } V < B : H = 2\pi - H$$

$$S = 1 - \frac{3 \min(R, V, B)}{L}$$

$$L = R + V + B$$

I.4

La non linéarité des deux équations permettant de calculer H et S entraîne des discontinuités dans leurs valeurs et deux couleurs perpétuellement proches peuvent se transformer en des composantes de teinte et de saturation très différentes.

Ceci est dû aux singularités de la teinte pour les points achromatiques ($R=V=B$) et de la saturation pour le noir ($R=V=B=0$).

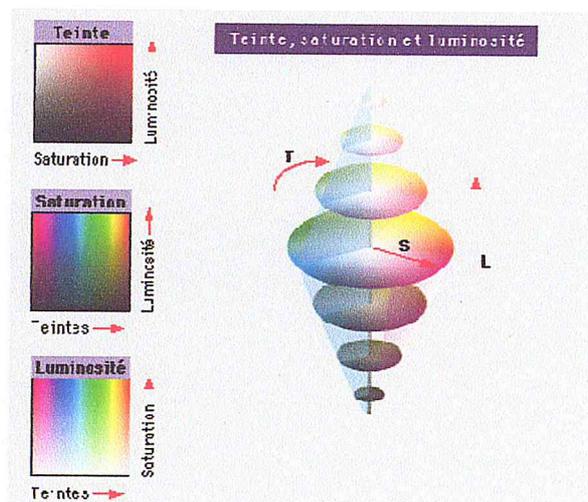


Figure I.15 : la représentation spatiale du modèle *HSL*

I.3.2 Les modèles de la vision humaine

1. Le modèle du tri-stimulus

La rétine humaine possède trois types de cônes photorécepteurs couleur qui répondent à la lumière incidente avec trois courbes de réponse. Ces réponses sont réparties suivant trois canaux qui sont le rouge, le jaune-vert et le bleu plutôt différenciés par leur largeur de bande (cf. Figure I.16).

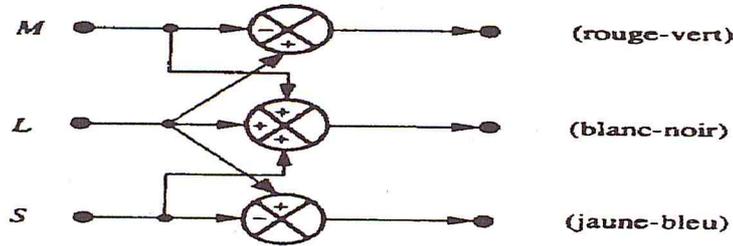


Figure I.16 : modèle du tri-stimulus [7]

2. Le modèle de Faugeras

Ce modèle combine le modèle du tri-stimulus et celui des couleurs opposées en s'appuyant sur le fait que chacun représente un niveau d'abstraction différent (cf. figure I.17).

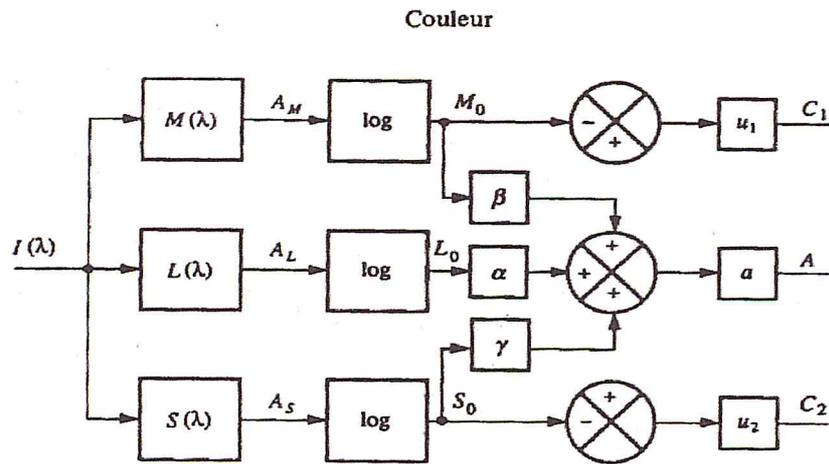


Figure I.17 : modèle de Faugeras [7]

3. Le modèle de Pratt

Bien que des modèles tels que le tri-stimulus soient communément utilisés, il est difficile pour une personne de faire le lien entre la perception de l'apparence visuelle d'une couleur et les composantes dans un espace de couleur [7].

Selon la CIE la teinte est la sensation selon laquelle, une zone apparaît similaire à une des couleurs perçues : rouge, jaune-vert, bleu ou une combinaison de deux d'entre elles.

La saturation est donnée par un adjectif comme vif, terne ou éclatant et est la proportion de couleur par rapport à la luminance.

Enfin la luminance est la sensation selon laquelle une zone émet plus ou moins de lumière et est donnée par un adjectif comme clair ou foncé.

4. La matrice de transformation RGB vers XYZ

Pour transformer les valeurs CIE XYZ en RGB₇₀₉ avec le point blanc D₆₅, on utilise la matrice suivante.

$$\begin{bmatrix} R_{709} \\ G_{709} \\ B_{709} \end{bmatrix} = \begin{bmatrix} 3.240479 & -1.537150 & -0.498535 \\ -0.969256 & 1.875992 & 0.041556 \\ 0.055648 & -0.204043 & 1.057311 \end{bmatrix} \times \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad \text{I.5}$$

Cette matrice a des coefficients négatifs ! En effet, les couleurs XYZ qui sont à l'extérieur du *gamut* correspondent par cette transformation à des valeurs RGB impropres, dont une composante au moins est négative ou supérieur à l'unité.

Voici la transformation inverse.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \times \begin{bmatrix} R_{709} \\ G_{709} \\ B_{709} \end{bmatrix} \quad \text{I.6}$$

Etant données les vecteurs RGB des primaires et du blanc, on peut, avec cette matrice, les transformer en coordonnées XYZ, puis calculer les composantes chromatiques x et y .

$$R_{RGB} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, G_{RGB} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, B_{RGB} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \text{ et } W_{RGB} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \text{I.7}$$

I.4 Le screening

L'analyse des prélèvements cytologiques se fait visuellement grâce à un cytopathologiste [6]. Nous distinguons ce screening dans ce qui suit :

I.4.1 Le screening manuel

Cette opération, consiste en une étape de lecture de la lame par un expert en cytologie (un cytotechnicien). Il repère les cellules intéressantes. Le screening est donc un repérage et une évaluation visuelle de toutes les cellules présentes sur une lame, afin de détecter et d'identifier d'éventuelles anomalies. Son but, est principalement la détection des cellules anormales ou suspectes afin d'établir un diagnostic fiable.

Un cytotechnicien observe entre 30 et 200 lames par jour. Sa culture médicale, est remise en cause chaque jour à cause de ce facteur subjectif inévitable. Effectivement, quelques erreurs peuvent apparaître et causer des faux négatifs. Les préoccupations principales en cytologie, sont celles qui permettent d'augmenter lepotène de vérité contenu dans le diagnostic. Un cytopathologiste s'attache donc, à réaliser les moins de faux positifs possibles mais surtout de faux négatifs.

I.4.2 Le screening automatique

Afin d'améliorer le contrôle de la qualité du screening, certaines recherches [6] se sont orientées vers une automatisation totale du screening. La mise en place d'un système informatique, peut apporter au cytopathologie un plus dans la précision de l'analyse qualitative. En effet, l'examen des lames par un cytotechnicien est une tâche subjective (celle-ci n'est pas infallible); un système informatique peut permettre de donner une analyse objective de lames.

I.4.3 structure d'un système informatique de rescreening

Les systèmes informatiques de rescreening, sont conçus pour opérer de façon postérieure au screening manuel. Cependant, ils peuvent éventuellement être utilisés pour fonctionner de manière antérieure. Effectivement, les pathologistes ont une préférence pour le rescreening informatique. L'architecture des systèmes utilisés pour le rescreening de lames de cytologie, reste identique quel que soit le système employé et est constituée de trois étapes séquentielles :

- **la segmentation** : elle consiste à extraire des les composantes cellulaires à partir d'images couleurs,
- **la caractérisation** : elle fournit des caractéristiques des objets précédemment extraits. Ces données doivent être suffisamment pertinentes et précises pour décrire les objets,
- **la classification** : les cellules sont classées dans un type cellulaire grâce à une méthode de classification (arbre de décision, réseaux de neurones, etc.), de façon à déterminer le type de la cellule à classer et plus précisément à détecter son caractère bénin ou malin.

I.5 Conclusion

Le traitement d'image est devenu un module considérable dans l'activité scientifique et technologique d'aujourd'hui. Ce chapitre, a été consacré aux notions générales de la vision couleurs; spécifiquement aux modèles et aux espaces couleurs. Une étude globale sur la cytopathologie du sein, a été appréhendée afin de mieux analyser l'information. Le chapitre suivant, est consacré aux méthodes de segmentation en imagerie cytologique.



*Chapitre II: Techniques de
segmentation en imagerie
cytologique*

II.1 Introduction

La segmentation est une tâche très importante et difficile, car elle dépend de la nature de l'image et de l'objectif de l'analyse. La littérature [8], propose de nombreuses techniques de segmentation, mais la plupart d'entre elles ont besoin de plusieurs paramètres dont le réglage exige souvent l'expertise humaine. Dans ce qui suit, nous présentons quelques méthodes de segmentation, en choisissant deux chemins de conception qui se basent soit sur une analyse des propriétés spatiales où l'on utilise les différents algorithmes présentés pour extraire une région ou un contour; ou sur la morphologie mathématique, qui utilise l'information couleur comme information a priori, en passant par quelques travaux réalisés dans le cadre de l'imagerie cytologique.

Avant de parcourir les différentes méthodes de segmentation couleur, il nous semble important de préciser en quoi celles-ci peuvent varier. Les méthodes de segmentation couleur présentent plusieurs particularités que l'on retrouve systématiquement [7] :

Les méthodes de segmentation couleur présentent plusieurs particularités [7] que l'on retrouve systématiquement suivant:

A) *La méthode utilisée* : elle caractérise la manière d'extraire les régions d'une image grâce à son information couleur [7].

B) *le choix de l'espace de travail* : les images couleurs sont généralement acquises et/ou stockées dans l'espace RVB [8]. Les traitements peuvent être effectués dans cet espace. Cette manière de procéder, présente l'avantage de garantir l'intégrité des données, mais elle entraîne un certain nombre d'inconvénients :

- *Les composantes y sont souvent fortement corrélées,*
- *Il est difficile d'y interpréter la notion de couleur telle qu'elle est habituellement perçue par l'œil humain,*
- *La différence perceptuelle entre deux couleurs ne coïncide pas avec leurs distances.*
- *Dans l'espace RVB, on peut chercher un espace plus approprié. Le traitement est alors effectué dans ce nouvel espace (après une transformation T), et l'application de la transformation inverse (T') permet si nécessaire de revenir dans l'espace RVB (cf. Figure II.1).*

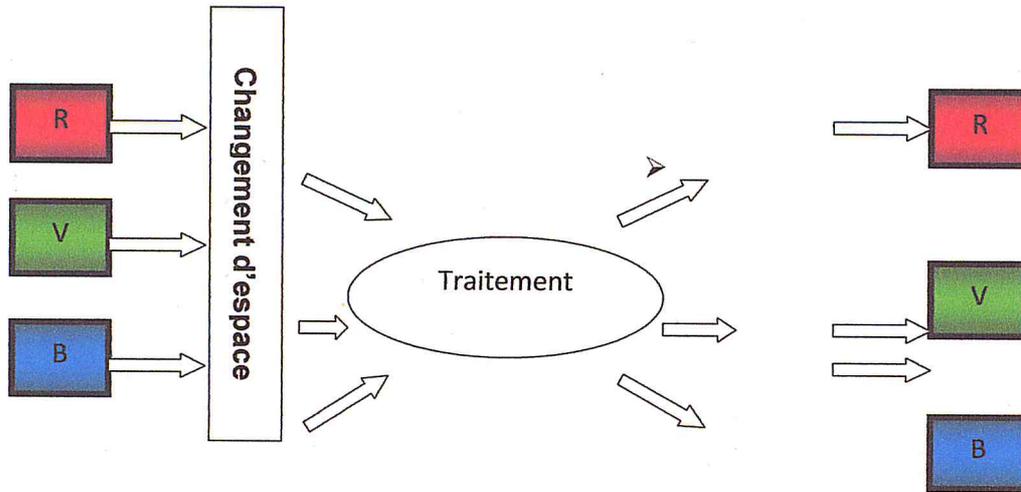


Figure II.1 : procédé du traitement par l'utilisation des espaces couleurs.

Nous présentons dans ce qui suit, les principales techniques de segmentation.

II.2 Segmentation par analyse des propriétés spatiales [7][8]

La segmentation des images, consiste à regrouper les pixels de ces images qui partagent une même propriété pour former des régions connexes. Les régions sont déterminées en fonction de leurs propriétés intrinsèques (agrégation). Contrairement à l'extraction des contours, qui s'intéresse aux bords des régions, la segmentation en régions homogènes vise à segmenter l'image en se basant sur les propriétés intrinsèques des régions. Nous distinguons ici quatre approches:

II.2.1 Segmentation par division de régions

Le principe de la segmentation par division (cf. Figure II.2) est réalisé suivant :

- ✓ la définition d'un critère d'homogénéité : forme a priori de l'histogramme, extrema de l'image (valeurs minimum et maximum), valeurs identiques, variance limitée, approximation par un polynôme de degré N, etc.
- ✓ Le test de la validité du critère sur l'image. Ainsi, si le critère est valide, l'image est segmentée (arrêt de la méthode), sinon, l'image est découpée en zones plus petites et la méthode est réappliquée sur chacune des zones.

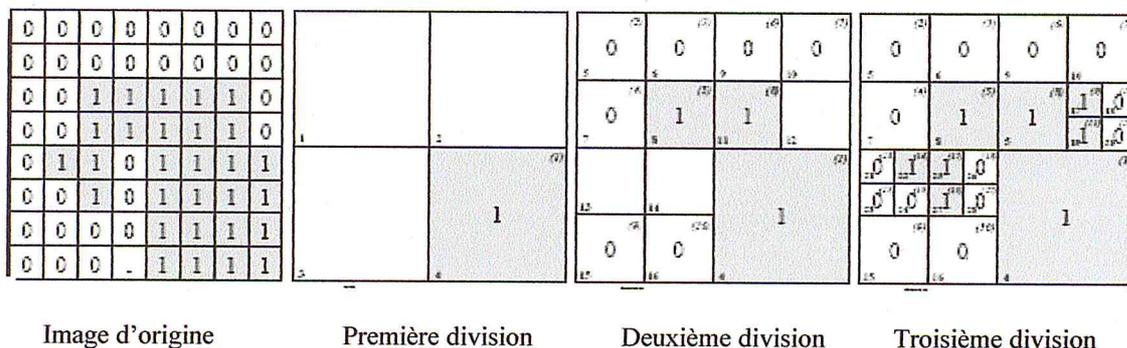


Figure II.2 : Segmentation par division de régions

II.2.2 Segmentation par croissance de régions

La croissance de régions [3], est une méthode particulière de la segmentation. Elle vise à grouper les points adjacents dont les attributs varient d'une façon négligeable. D'un point de vue topologique, la segmentation par croissance de régions correspond à une partition de l'image I en L régions connexes R_i , $i=1, \dots, L$ telle que l'union des L supports de ces régions redonne le support de l'image originale. Le problème fondamental lié à la croissance de régions, est la définition de l'attribut A et du prédicat d'uniformité $P(A, R_i)$ qui est vrai dans chaque région $i=1, \dots, L$ et qui est faux pour l'ensemble de deux régions adjacentes R_i et R_j .

$$P[A, R_i] = \text{VRAI}, i=1, \dots, L.$$

$$P[A, R_i \cup R_j] = \text{FAUX}, i \neq j \text{ et } R_i \text{ adjacent à } R_j.$$

Un des points attractifs de la segmentation par croissance de régions, est lié au concept même de la région. En effet, plus une région est grande, plus la croissance statistique des propriétés de ses points est précise.

II.2.3 Segmentation par fusion de régions

Une région est un ensemble de points n -connexes de même niveau de gris, qui seront 4-connexes dans l'exemple suivant (cf. Figure II.3). La seule fonction requise, est de pouvoir énumérer les pixels qui composent une région. Il est utile pourtant de mémoriser, pour chaque région, le nombre de pixels, la somme de leurs valeurs (pour calculer la moyenne) et d'autres informations qui serviront à déterminer rapidement des critères de fusion. Lors de la fusion de deux régions, l'une est choisie arbitrairement pour "absorber" l'autre. Les données de la région absorbante sont alors "augmentées" de celles de la région absorbée.

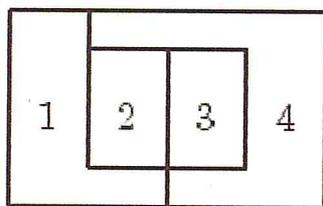


Figure II.3 : esquisse de la segmentation en 4 régions

Deux régions sont adjacentes, si au moins un pixel de l'une est voisin direct (pour la "4-connexité") d'un pixel de l'autre. Cela définit très naturellement une structure de graphe non orienté. La figure 8 donne un exemple, les adjacences sont alors $(1 \leftrightarrow 2)$, $(1 \leftrightarrow 4)$, $(2 \leftrightarrow 3)$,

$(2 \leftrightarrow 4)$ et $(3 \leftrightarrow 4)$.

La mise à jour des adjacences, lors d'une fusion constitue l'une des difficultés principales de ce sujet. On étudiera sur la figure 8, les conséquences d'une absorption de la région 4 par la région 1. Les adjacences restantes sont alors $(1 \leftrightarrow 2)$, $(1 \leftrightarrow 3)$ et $(2 \leftrightarrow 3)$. Le coût de mise à jour doit être linéaire dans le nombre d'adjacences concernées, c'est-à-dire, celles des deux régions à fusionner ainsi que celles de toutes leurs voisines directes [8].

II.2. 4. Segmentation par division fusion

La segmentation par division fusion, regroupe les deux types d'approches précédentes. Tout d'abord, l'image est divisée en régions homogènes qui respectent des critères globaux, puis les régions adjacentes qui répondent à des critères locaux, sont fusionnées [7].

II.2. 5. Discussion

Les méthodes de segmentation en régions d'image présentées ci-dessus, sont adaptées aux images en niveaux de gris et aux images couleurs. Dans le cas de l'image couleur, elles considèrent une région comme un ensemble de pixels connexes ayant une couleur proche d'une valeur moyenne. Ces méthodes analysent directement l'image matricielle et font intervenir l'information couleur multidimensionnelle dans la phase de décision.

II. 3 Morphologie mathématique [9][10]

La morphologie mathématique est une science de la forme et de la structure. Le principe de base de la morphologie mathématique, est de comparer les objets que l'on veut analyser à un autre objet de référence, de taille et de forme connues, appelé **élément structurant**. En quelque sorte, chaque élément structurant fait apparaître l'objet sous une nouvelle forme. Fondamentalement, la morphologie mathématique a un caractère ensembliste et ceci correspond à la démarche dont nous disposons pour analyser quantitativement une image.

La structure d'une stratégie de segmentation par morphologie mathématique, se décrit par trois étapes séquentielles : la simplification, l'extraction de marqueurs, la croissance [7].

- ✓ **La simplification** : La simplification de l'image, consiste en une phase de prétraitements dont le but est de simplifier les images. La simplification peut consister en un lissage de l'image, grâce à une régularisation réalisée par un filtre. Des opérateurs connexes basés, sur la reconstruction morphologique permettent d'éliminer certains détails ou objets selon des critères de taille, de forme, ou de couleur.
- ✓ **L'extraction** : L'extraction de marqueurs est l'étape d'initialisation de la croissance : c'est une étape orientée vers une extraction grossière des objets. Une connaissance a priori sur l'information des images à traiter et sur les caractéristiques des objets dans l'image, permet de choisir des opérateurs de traitement d'images. Ces opérateurs permettent, en exploitant les caractéristiques des objets, d'extraire partiellement ou totalement les objets sans toutefois être précis au niveau de la localisation. Pour les images couleurs de microscopie, la principale information a priori se situe au niveau de la couleur des objets. Pour cette raison, l'extraction des marqueurs se doit donc d'utiliser la couleur ou une composante couleur.
- ✓ **La croissance** : A partir des marqueurs obtenus à l'étape précédente, on utilise l'une des principales transformations de morphologie mathématique : la ligne de partage des eaux. Cette dernière réalise une croissance et permet d'obtenir les régions finales correspondant aux objets segmentés. Cette étape doit

également faire appel aux informations a priori disponibles sur les objets et en particulier leur couleur. L'utilisation de la couleur dans le calcul de la ligne de partage des eaux, permettra d'obtenir des résultats qui correspondront mieux aux changements significatifs présents dans les images.

II.3.1 Concepts de base de la morphologie mathématique

Cette théorie est fondée sur les opérations d'érosion et de dilatation. Ces opérations répétées, permettent l'ouverture et la fermeture.

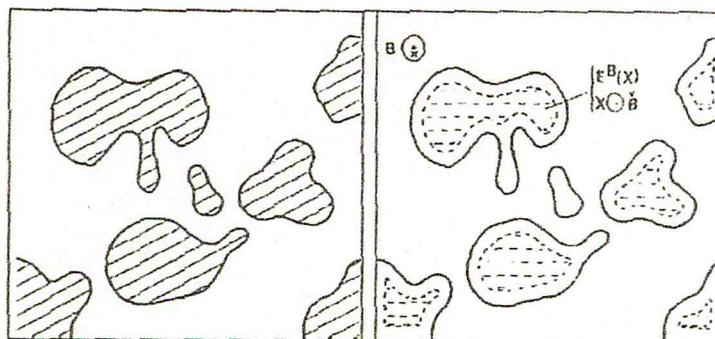
1. L'érosion

Soit X l'ensemble à éroder, et B un élément structurant de géométrie simple, par exemple un cercle. L'érosion consiste à faire déplacer B de telle sorte que son centre occupe toutes les positions x de l'espace. Pour chaque position, on pose la question : est-ce que B est complètement inclus dans X ?

L'ensemble des positions x correspondant à une réponse positive forme le nouvel ensemble Y, appelé érodé de X par B, noté $X \ominus B$ (cf. Figure II.4).

Cet ensemble satisfait l'équation

$$X \ominus B = \{ x \text{ dans } X : B \subset X \} \quad \text{II.1}$$



a : Ensemble d'objets X b : Erodé de X

Figure II.4 Erosion de (a) par un élément structurant circulaire

2. La Dilatation

L'opération de dilatation se définit de manière analogue à l'érosion. En prenant le même élément structurant B, on pose pour chaque point x la question "Bx touche-t-il l'ensemble X ?", c'est à dire, y a-t-il une intersection non vide entre Bx et X ?

L'ensemble des points de l'image (cf. Figure II.5) correspondant aux réponses positives forme le nouvel ensemble Y des dilatés de X, noté comme suit :

$$y = x \oplus B = \{ x \text{ dans } I : B \cap X \neq \emptyset \} \quad \text{II.2}$$

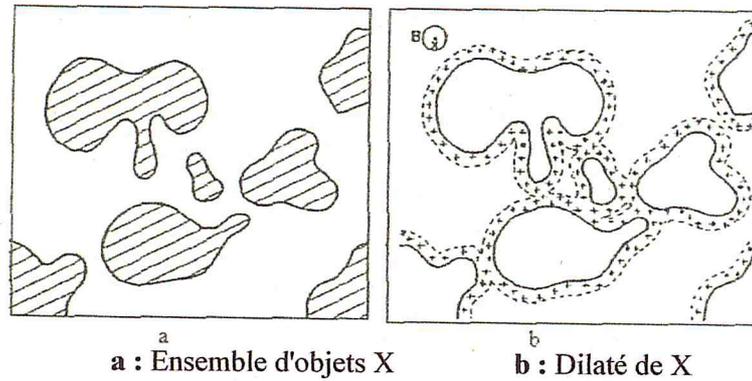


Figure II.5. Dilatation de (a) par un élément structurant circulaire

3. Ouverture

Puisque l'érosion et la dilatation sont des transformations itératives, il est possible d'effectuer sur un ensemble X une érosion, puis de dilater l'ensemble érodé par un même élément structurant B. Le résultat est nommé l'ouvert de X par B, noté comme suit :

$$O^B(X) = (X \ominus B) \oplus B \quad \text{II.3}$$

Après une ouverture, on ne retrouve pas l'ensemble de départ; L'ensemble ouvert est plus régulier et moins riche en détails que l'ensemble X initial (cf. Figure II.6). La transformation par ouverture adoucit les contours, coupe les isthmes étroits, supprime les petites îles et les caps étroit.

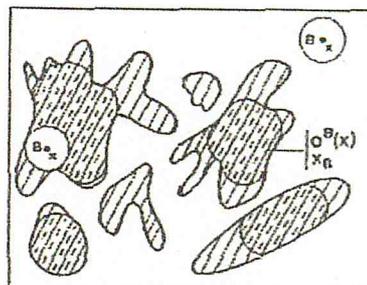


Figure II.6 : Ouverture d'un ensemble d'objets X par un élément structurant circulaire

4. Fermeture

La fermeture est l'opération "inverse" de l'ouverture, c'est à dire que l'on applique tout d'abord une dilatation puis une érosion (toujours en gardant le même élément structurant) Le résultat est nommé le fermé de X par B, noté comme suit :

$$F^B(X) = (X \oplus B) \ominus B$$

II.4

Un ensemble fermé est également moins riche en détails que l'ensemble initial (cf. Figure II.7). La transformation par fermeture bouche les canaux étroits, supprime les petits lacs et les golfes étroits.

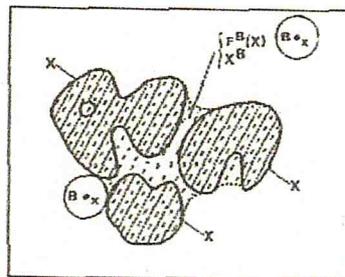


Figure II.7 : fermeture d'un ensemble d'objets X par un élément structurant

II.3.2 Ligne de partage des eaux (watershed) [7][10][11]

Cette technique consiste à faire grossir simultanément toutes les régions jusqu'à ce que l'image soit entièrement segmentée.

Cette technique tire son nom d'une analogie avec la géophysique. On peut en effet, considérer les valeurs d'intensité des pixels d'une image comme une information d'altitude. Dans ce cas on peut représenter cette image (appelée carte d'élévation), comme un terrain en trois dimensions. Le principe est alors, de remplir progressivement d'eau chaque bassin du terrain. Chaque bassin illustre une région.

Lorsque l'eau monte et que deux bassins se rejoignent, la ligne de rencontre (i.e. la ligne de partage des eaux) est marquée comme une ligne de frontière entre 2 les régions (cf. Figure II.8).

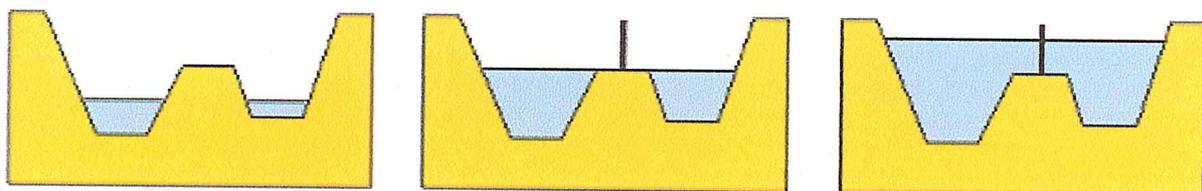


Figure II.8 : ligne de partage des eaux entre 2 bassins

L'algorithme se compose de deux étapes:

1. Générer une carte d'élévation à partir de l'image de départ;
2. Remplir progressivement les bassins.

➤ **Carte d'élévation**

Dans la carte d'élévation, on assigne l'altitude la plus élevée (HMAX) aux pixels ayant un fort gradient ainsi qu'aux bords de l'image. Cela nous donne les lignes de crête.

La valeur d'altitude des autres pixels, est calculée en fonction de l'éloignement de la ligne de crête: plus on s'éloigne de la ligne de crête, plus l'altitude doit décroître. La formule pour calculer la valeur d'altitude d'un pixel est donc :

$$\text{Altitude(Pixel)} = \text{HMAX} - \text{distance (Pixel, ligne_de_crête)} \quad \text{II.5}$$

➤ **Remplissage des bassins**

Il existe de nombreux algorithmes permettant de réaliser le remplissage des bassins. Nous détaillons ici un algorithme utilisant la technique de croissance de régions. Dans cet algorithme (cf. Figure II.9), le remplissage des bassins est un procédé itératif qui consiste à élever progressivement le niveau de l'eau de zéro à HMAX. A chaque itération, on doit remplir les bassins existants (extension des régions) et éventuellement créer les nouveaux bassins (nouvelles régions).

```

/** Action 1 : Etendre les régions existantes */

Pour chaque Région « [R] » dans la liste « [Régions] »
    Faire croître « [R] » jusqu'à l'altitude NIVEAU
Fin Pour

/** Action 2 : créer les nouvelles régions */

Pour chaque pixel « P » à l'altitude NIVEAU
    Si « P » n'est pas associé à une région Alors
        Créer une nouvelle région « [R] » dans la liste
    « [Régions] »
        Ajouter le pixel « P » dans la région « [R] »
        Faire croître « [R] » jusqu'à l'altitude NIVEAU
    Fin Si
Fin Pour
Fin Pour

```

Figure II.9 : algorithme de la ligne de partage des eaux

Discussion

Les outils de la morphologie mathématique ont été développés au départ pour le traitement d'images binaires en faisant appel à la morphologie mathématique ensembliste. Leur utilisation a été ensuite étendue aux images en niveaux de gris: on parle à ce moment, de morphologie mathématique fonctionnelle. Désormais, certains opérateurs fonctionnent sur les images couleurs comme la ligne de partage des eaux (vu précédemment).

II.4 Segmentation en contours

L'approche contour est l'approche duale de l'approche région: elle consiste à rechercher les frontières des régions caractérisées par une discontinuité visuelle. Ces méthodes sont basées sur la recherche de variations locales significatives. Largement étudiée pour les images en niveaux de gris, la recherche de contours repose principalement sur le calcul des dérivées [8].

Pour une image monochrome, le passage d'une région à une autre, se traduit par un changement rapide de niveaux de gris (un gradient élevé dans la direction normale au contour). Le gradient se caractérise par sa direction et sa norme.

Pour déterminer les positions de forts gradients, soit on calcule la dérivée première pour ne conserver que les régions de plus forte pente, soit on calcule les passages par zéro de la dérivée seconde. Les contours ainsi obtenus, forment rarement des frontières fermées, il faut par la suite appliquer des algorithmes de fermeture de contours. Pour les images couleurs, on identifie deux stratégies principales :

- les méthodes marginales qui cherchent les contours indépendamment dans chaque composante couleur avant de fusionner les contours extraits,
- les méthodes vectorielles qui reposent sur la représentation vectorielle de la couleur [8].

II.4.1 La Segmentation par contours actifs [12]

Un contour actif est un ensemble de points qu'on va tenter de déplacer pour leur faire épouser une forme. Il s'agit d'une technique d'extraction de données utilisée en traitement d'images.

La courbe représentant le contour est défini sous la forme paramétrique de l'Équation suivante :

$$C = (x(s, t), y(s, t)) \quad \text{II.6}$$

L'idée de cette méthode, est de déplacer les points pour les rapprocher des zones de fort gradient tout en conservant des caractéristiques comme la courbure du contour ou la répartition des points sur le contour ou d'autres contraintes liées à la disposition des points.

Au démarrage de l'algorithme, le contour est disposé uniformément autour de l'objet à détecter puis il va se rétracter pour en épouser au mieux ses formes. De la même manière, un contour actif peut aussi se dilater et tenter de remplir une forme, il sera alors situé à l'intérieur de celle-ci au démarrage de l'algorithme.

A chaque itération, l'algorithme va tenter de trouver un meilleur positionnement pour le contour pour minimiser ses dérives par rapport aux contraintes utilisées. L'algorithme s'arrêtera, lorsqu'il ne sera plus possible d'améliorer le positionnement ou simplement quand le nombre maximum d'itérations aura été atteint. On utilise les notions d'énergies interne et externe pour caractériser respectivement; la forme du contour et tous les éléments qui lui sont propres et, le positionnement du contour sur l'image en tenant compte des lignes de gradient.

➤ Phases de l'algorithme

L'algorithme suit les phases ci-dessous à chaque itération:

- calcul des énergies interne et externe, caractérisant le contour lui-même et son positionnement sur l'image;
- pour chaque point du contour, détermination d'une nouvelle position, sur laquelle le contour devrait mieux minimiser les écarts de contraintes;
- arrangement du contour pour qu'il respecte des contraintes d'écartement entre les points, de régularité de points...etc.

➤ L'énergie interne

L'énergie interne ne dépend pas de l'image ni de la forme à détourer, elle ne dépend que des points du contour. Elle regroupe des notions comme la courbure du contour ou la régularité d'espacement des points. En effet, le contour doit conserver une forme arrondie en minimisant les dérivées d'ordre 1, 2, ... et doit empêcher un point de se détacher trop loin du reste du contour.

Idéalement, l'énergie interne est minimale pour un cercle où tous les points sont régulièrement espacés.

➤ L'énergie externe

L'énergie externe correspond à l'impact du contour sur l'image. Pour la calculer, il faut considérer l'opposé de la valeur de son gradient (ou de toute autre représentation mettant en jeu les contours à épouser) en chaque point du contour.

Cette énergie externe doit théoriquement être minimale si le contour épouse parfaitement la forme à extraire.

On ne considère l'opposé du gradient que pour avoir une énergie externe minimale à la convergence de l'algorithme.

On pourra utiliser une représentation en gradient et en flot pour ajouter de l'information dans les zones uniformes où le gradient est nul, de manière à guider le snake vers le bord le plus proche.

➤ Utilisation des deux énergies

Chaque position du contour actif donne une énergie interne et une énergie externe dont la somme doit être minimisée et influencera les mouvements des points du contour actif. La grande difficulté de l'utilisation des snakes réside dans le choix des pondérations à donner à chaque énergie.

Traditionnellement il est d'usage d'utiliser un paramètre pour l'énergie interne et un autre pour l'énergie externe, mais un réglage plus fin peut s'opérer en ajoutant des paramètres différents pour des aspects différents de l'énergie interne ou de l'énergie externe.

On pourra ainsi avoir :

- un paramètre pour pondérer l'énergie dégagée par la mauvaise courbure du snake;
- un autre pour pondérer l'énergie dégagée par le non respect de l'écartement régulier entre ses points;
- un autre pour pondérer l'énergie dégagée par le positionnement sur les lignes de gradient de l'image.

➤ Modifications du contour

Après avoir calculé l'énergie globale dégagée par le contour et par son positionnement sur l'image, il convient de déterminer comment le faire évoluer pour minimiser cette énergie.

Pour cela, une méthode simple et intuitive est d'observer les pixels voisins immédiats de chaque point du contour pour déterminer pour chacun d'eux l'énergie globale du snake, chaque meilleur voisin devenant un point du contour.

Il est nécessaire que le contour possède toujours suffisamment de points pour être sûr de bien calculer son énergie globale, en particulier son énergie externe qui sera plus précise en tenant compte d'avantage de points. C'est la raison pour laquelle il peut s'avérer pertinent de rajouter ou de supprimer des points à chaque itération si des contraintes ne sont pas suffisamment respectées. Par exemple, on pourra rajouter un point au snake si ses voisins sont trop éloignés. A l'inverse, on pourra supprimer un point s'il est trop près de ses voisins. La figure II.10, illustre un exemple de l'application des contours actifs.

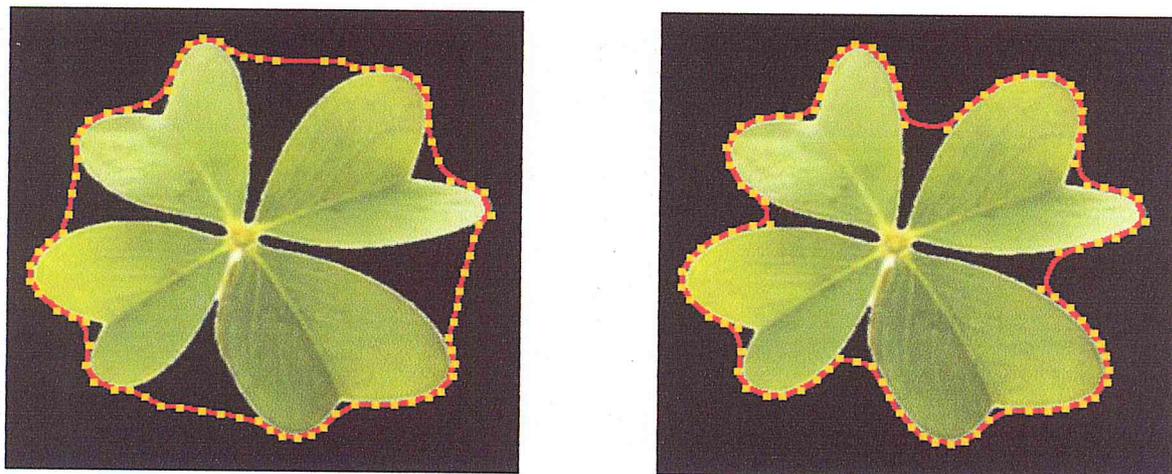


Figure II.10 : illustrations des grandes étapes du déroulement d'un contour actif

II.4.2 La segmentation par les ensembles de niveaux (levels sets) [13]

La méthode des Ensembles de Niveaux, est une méthode de simulation numérique utilisée pour l'évolution des courbes et des surfaces dans les domaines discrets. C'est une représentation variationnelle des contours, qui conduisent à des solutions qui évoluent au cours du temps dans l'image, régies par un critère global.

Cette méthode permet de faire évoluer une courbe paramétrique fermée $C(p)$ suivant une équation du type :

$$\frac{\partial C}{\partial t} = FN \quad \text{II.7}$$

Où t est le temps, F est la vitesse d'évolution et N est la normale unitaire à la courbe.

Chaque point de la courbe C évolue suivant la direction normale à la courbe avec une vitesse F .

L'avantage principal de cette méthode, est la possibilité de gérer automatiquement le changement de topologie de la courbe en évolution.

La courbe C peut être divisée en deux ou trois courbes. Inversement, plusieurs courbes peuvent fusionner et devenir une seule courbe.

La méthode des Ensembles de Niveaux, est capable de réaliser ce changement en utilisant une fonction d'ordre supérieure :

$$\phi : R^{N+1} \rightarrow R^N \quad \text{II.8}$$

C étant le niveau zéro de la courbe d'ordre supérieur. Il est représenté selon:

$$C(t) = \{s \in R^2 \mid \phi(s, t) = 0\} \quad \text{II.9}$$

Le niveau zéro (C) peut se diviser tandis que la fonction ϕ reste continue.

➤ Principe de la méthode

Le principe des contours actifs est de faire évoluer une courbe. On a vu dans la section précédente une formulation dynamique qui fait intervenir la vitesse d'évolution du contour. On va donc s'intéresser à la façon de faire évoluer la courbe et plus généralement à la notion de propagation de fronts.

➤ Avantages

- Prise en compte des changements de topologies automatiques grandeurs géométriques intrinsèques (normales entrante/sortante, courbure) faciles à calculer;
- extension à la 3D simple: il suffit d'ajouter une coordonnée à l'équation d'évolution de la fonction Φ : on a alors un volume $\Phi(x, y, z, t)$;
- Discrétisation de Φ avec une grille définit dans le domaine de l'image, i.e valeurs possibles pour (x, y) ;
- Utilisation des méthodes numériques connues pour calculer les dérivées.

➤ Inconvénients

L'implémentation implique plusieurs problèmes:

1. On doit construire une fonction initiale $\Phi(x, y, z, t=0)$ de manière à ce que son niveau zéro corresponde à la position initiale du contour;
2. L'équation d'évolution n'est dérivée qu'au *niveau zéro*, la fonction vitesse v n'est donc pas définie (en générale) pour les autres *niveaux*;
3. La déformation constante, peut causer la formation de coins saillants sur le modèle initialement lisse. Ce coin, peut perturber les déformations successives, puisque la définition de la normale devient ambiguë.
4. il est parfois nécessaire de recalculer la fonction distance par rapport au niveau zéro.

II.5 L'approche coopérative [12]

Plusieurs chercheurs s'accordent pour affirmer que la coopération entre les segmentations par régions et par contours, contribuent à une meilleure prise en compte des caractéristiques des entités de l'image et, par conséquent, à une meilleure segmentation.

En effet, les algorithmes combinant les techniques de segmentation basées sur les régions et celles basées sur les contours prennent avantage de la nature complémentaire de l'information sur la région et sur le contour. Ainsi, une segmentation par coopération régions-contours peut être exprimée, comme une entraide entre ces deux concepts afin d'améliorer le résultat final. Elle exploite les avantages de ces deux types de segmentation, pour aboutir à un résultat de segmentation plus précis et plus fidèle que celui obtenu à l'aide d'une seule technique.

La segmentation d'images par coopération régions-contours peut se faire en fonction de trois approches:

1. La coopération séquentielle

Il s'agit d'une coopération (cf. Figure II.11) dans laquelle l'une des techniques de segmentation (région ou contour) est réalisée en premier lieu; son résultat va être exploité par l'autre technique pour renforcer la définition des critères ou des paramètres de la segmentation.

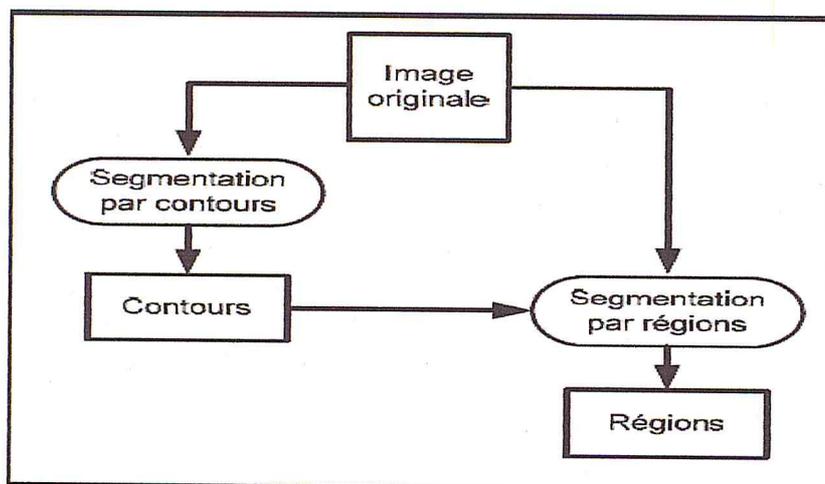


Figure II. 11: schéma de la coopération séquentielle.

2. La coopération des résultats

Les deux types de la segmentation seront réalisés indépendamment; la coopération concernera leurs résultats qui seront intégrés afin d'atteindre une meilleure segmentation;

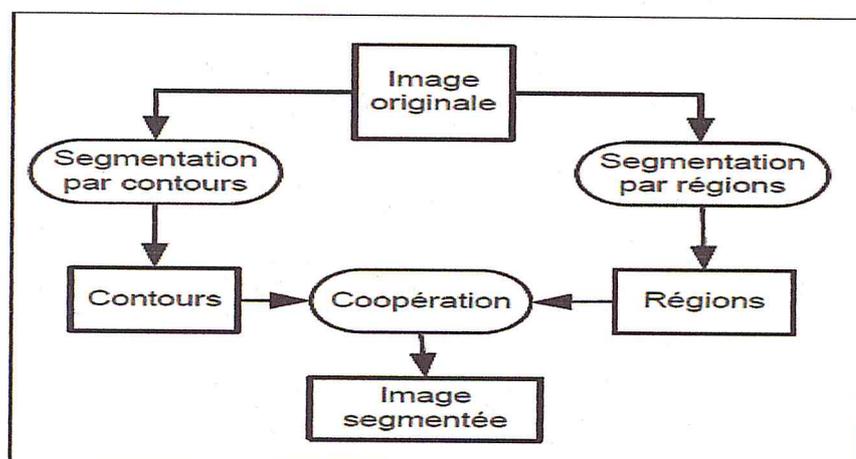


Figure II. 12 : synoptique de la coopération des résultats

3. La coopération mutuelle

la coopération mutuelle: les deux types de la segmentation, coopéreront mutuellement au cours de leur processus d'exécution

Les deux types de segmentations coopéreront mutuellement au cours de leur processus d'exécution.

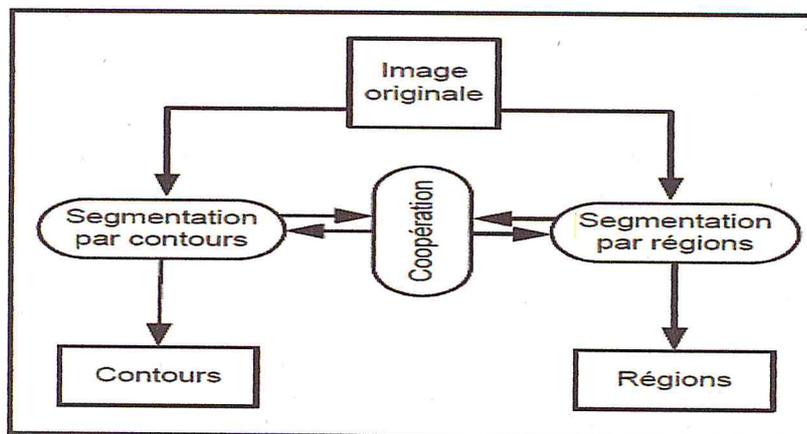


Figure II.13 : synoptique de la coopération mutuelle

II.6 Parcours de quelques travaux en imagerie cytologique

Nous présentons dans cette partie quelques travaux parcourus dans le cadre de la segmentation en imagerie cytologique, à partir desquels, nous allons extraire notre méthodologie.

I.6.1 Travail de Lezoray [6][7]

Ce travail présente une méthode morphologique de segmentation d'images cytologiques couleurs. Cette méthode est basée sur la ligne de partage des eaux utilisant une fonction de potentiel couleur combinant l'information locale et globale afin de séparer les cytoplasmes des noyaux.

L'utilisation de la couleur dans le calcul de la ligne de partage des eaux, permettra d'obtenir des résultats, qui correspondront mieux aux changements significatifs présents dans les images.

Les points clés de la segmentation par L.P.E sont:

- l'extraction de marqueurs des régions à segmenter (les régions à l'intérieur de chaque objet);
- la détermination de l'image sur laquelle on calcule la L.P.E;
- Le calcul de la L.P.E.

La ligne de partage des eaux trouve les lignes de séparation de plus grande intensité dans une image (cf. Figure II.14).

La figure II.15 illustre un exemple de l'extraction des cytoplasmes et des noyaux par la LPE selon le système semi automatique développé par Lezoray.

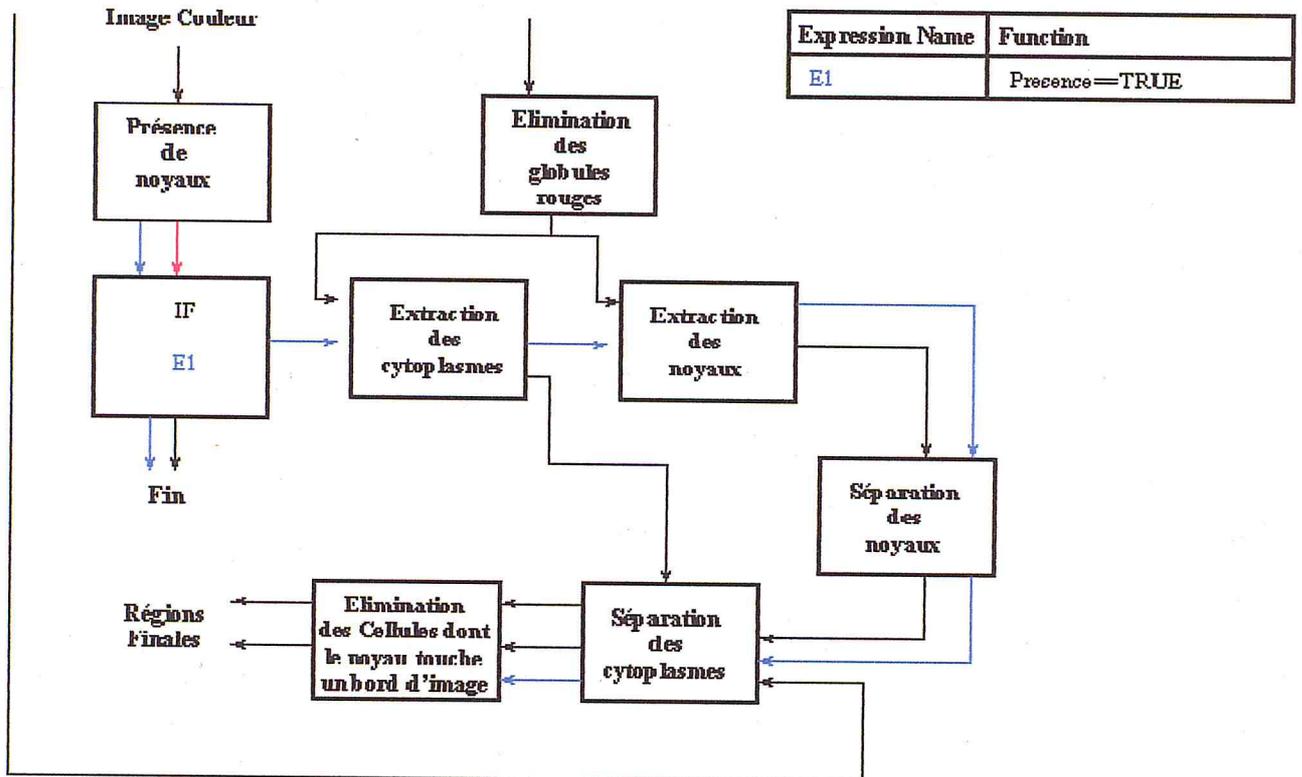
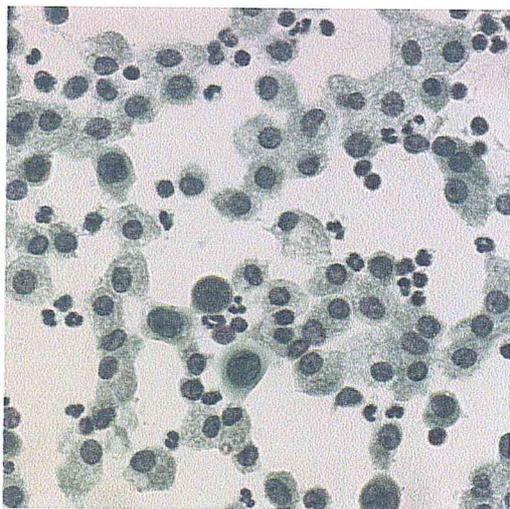
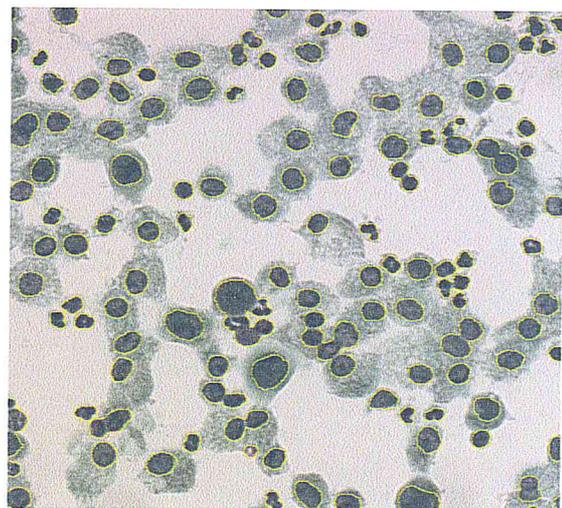


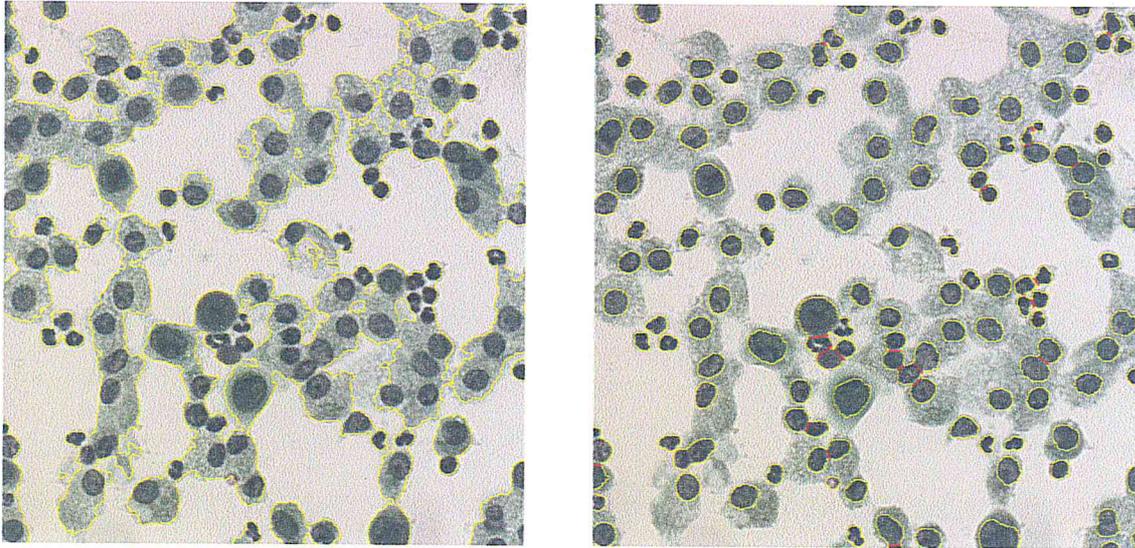
Figure II. 14: stratégie de segmentation d'images couleurs de cytologie des séreuses.



a : Image Couleur de cellules provenant de liquides d'épanchement des séreuses.



b : Image des contours nucléaires avant séparation.



c : Image des contours cytoplasmiques

d : Image des contours nucléaires après séparation.

Figure II. 15: Résultats de la segmentation d'images couleurs de cytologie des séreuses [6][7]

Les travaux de Ghandour [14] de l'université de Toulouse, se sont aussi basés sur la détection de structures pathologiques à partir d'images microscopiques, suivant des traitements de la morphologie mathématique.

2. Travaux de LRDSI [15][16][17]

Le laboratoire de recherche pour le développement des systèmes informatisés s'est aussi inspiré des travaux de l'université de Caen, pour l'analyse d'images cytologiques des séreuses et du sein.

Le premier travail [15], s'inscrit parmi les méthodes faisant appel aux techniques d'analyse de données multidimensionnelles qui considèrent que les régions sont définies comme des ensembles de pixels connexes appartenant à une même classe de pixels. La méthode adoptée ici procède en trois étapes :

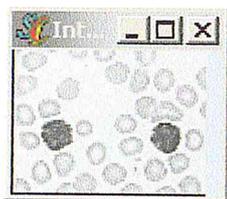
- Détermination des régions uniformes par l'identification des pics significatifs de l'histogramme d'homogénéité. Cette étape prend en compte aussi bien l'information locale que globale,
- Analyse de l'histogramme de la caractéristique Hue (teinte) de chaque région obtenue lors de la première étape,

➤ Fusion des sous régions pour éviter la sursegmentation. Le critère de fusion s'appuie sur la distance entre deux couleurs et elle est calculée dans l'espace CIE(L*a*b*).

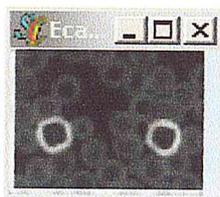
La méthode décrite ci-dessus a été tout d'abord validée sur des images de référence (Lena, maison , ...etc.); son application à des images cytologiques a engendré des résultats de bonne qualité; l'algorithme a permis de mettre en évidence après l'étape de segmentation toutes les structures (noyaux, globules rouges et cytoplasme); l'opération de fusion n'a conservé que les noyaux (cf .Figure II.16).



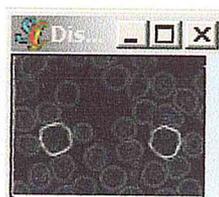
A : Image cytologique de la leucémie



B : Intensité



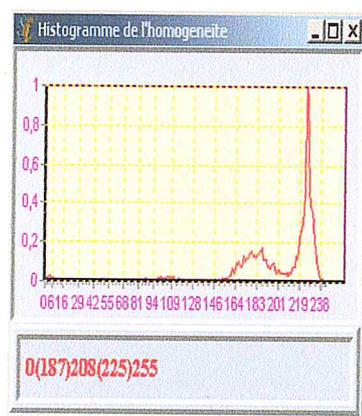
c : Ecart type



d : Discontinuité



h : Homogénéité



E : Histogramme d'homogénéité



f : régions uniformes

Figure II.16 : résultats de la segmentation par analyse d'histogramme [15]

Cette méthode de segmentation tient en compte aussi bien de l'information locale que globale ; elle bénéficie des avantages de deux méthodes : analyse d'histogramme qui considère qu'une région est constituée par un ensemble de pixels connexes qui appartiennent à une même classe , et fusion de régions ou une région est un ensemble de pixels connexes ayant une couleur proche d'une valeur moyenne.

Le **second travail** est basé sur la ligne de partage des eaux. Il permet d'extraire les cytoplasmes et les noyaux pour une analyse quantitative en vue d'une classification connexionniste non supervisée des cellules suivant le modèle de Kohonen.

L'extraction de l'information pertinente est effectuée suivant les opérations de la morphologie mathématique en se basant sur l'information couleur.

Pour déterminer les marqueurs du cytoplasme et du noyau; le gradient couleur est utilisé. Ces marqueurs, permettent d'extraire uniquement les régions cytoplasmiques et nucléaires. La figure suivante donne les différentes phases de segmentation pour les régions cytoplasmiques (cf. Figure II.17), ainsi que pour les régions nucléaires (cf. Figure II.18).



Image cellulaire

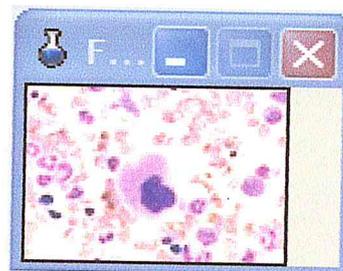


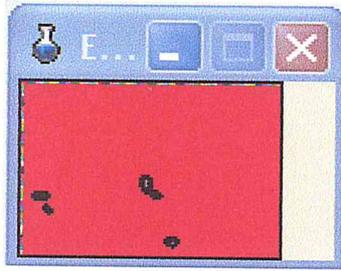
Image filtrée



Marqueurs cytoplasmiques



Gradient couleur



Marqueurs étiquetés



Régions cytoplasmiques

Figure II.17 : Segmentation des cytoplasmes



Image cellulaire

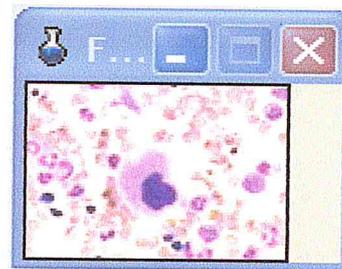


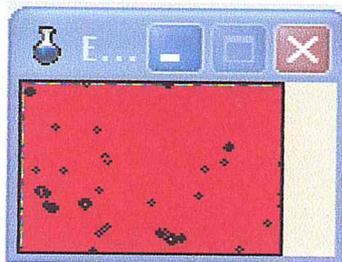
Image filtrée



Marqueurs nucléaires



Gradient couleur



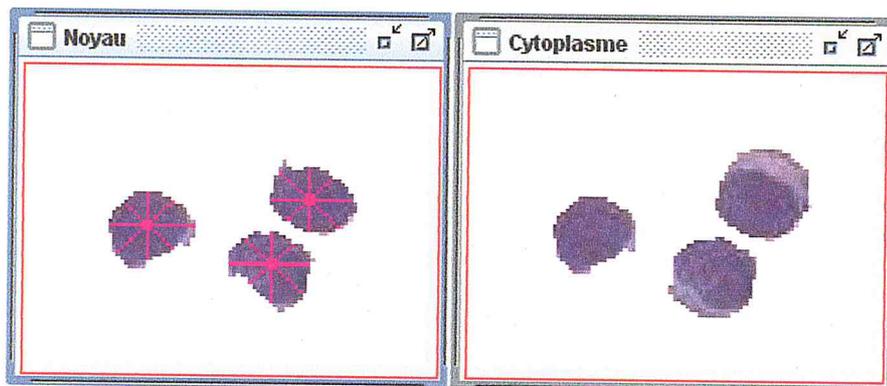
Marqueurs étiquetés



Régions nucléaires

Figure II.18 : segmentation des noyaux

Le troisième travail [17] repose sur une version améliorée de la LPE (cf. Figure II.19). Celle-ci utilise une fonction de potentiel couleur combinant l'information locale et globale.



(g) : extraction des cytoplasmes (h) : extraction des noyaux

Figure II.19: Segmentation cytoplasmique et nucléaire avec précision du centre de gravité.

La LPE améliorée, est appréhendée suivant l'analyse de l'histogramme où nous nous sommes plus particulièrement intéressés au choix de l'espace couleur et à l'obtention de marqueurs. Notre méthode de segmentation permet ainsi, l'extraction de l'image des cytoplasmes et celle des noyaux

3. Parcours de Glory [18]

Glory s'est basée sur la segmentation d'images de microscopie en couleur pour l'analyse automatique de cultures cytologiques. Ce projet doctoral traite de l'analyse d'images appliquée aux images de microscopie en couleur pour caractériser les cultures cellulaires. Les principaux axes de recherche de cette thèse sont l'évaluation de la qualité des segmentations d'images en couleur, la classification des noyaux par une méthode non-paramétrique et la séparation des agrégats nucléaires.

Le critère de qualité pour la segmentation des noyaux selon Glory suppose les conditions ci-dessous:

- les régions doivent être homogènes,
- les régions doivent être simples et sans trou à l'intérieur,
- les régions adjacentes doivent avoir des caractéristiques de region significativement différentes,
- les contours doivent être lisses et précis.

La stratégie adoptée par Groy, pour séparer les noyaux est donnée par les figures ci-dessous :

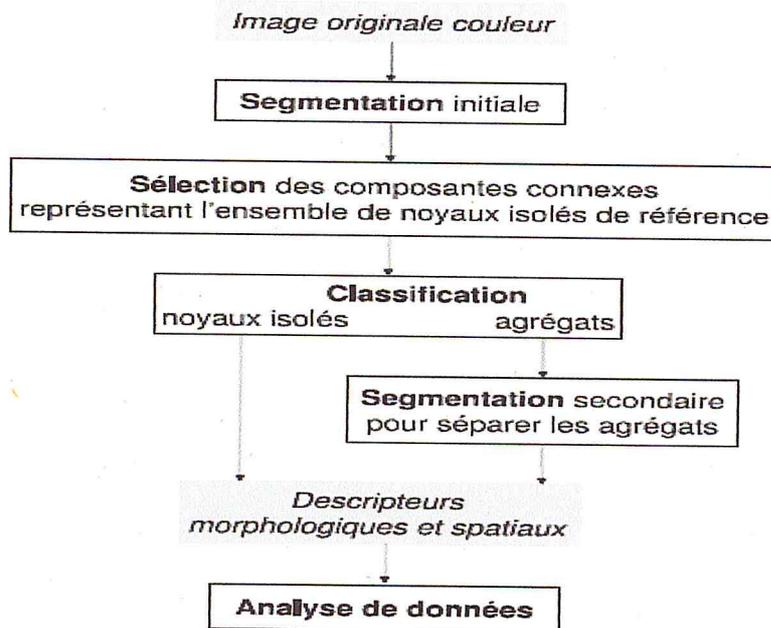


Figure II.20 : la segmentation des noyaux basée sur la couleur

Schéma de classification

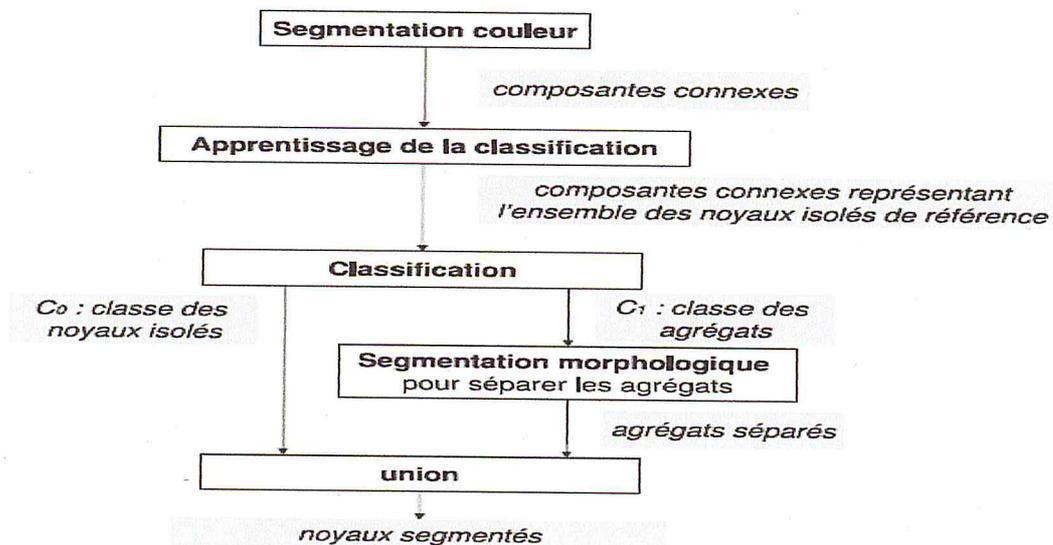


Figure II.21 : schéma de la stratégie appliquée aux noyaux issus de la segmentation couleur.

La séparation des agrégats en noyaux individuels a engendré dans les travaux de Glory, les résultats suivants:

La séparation

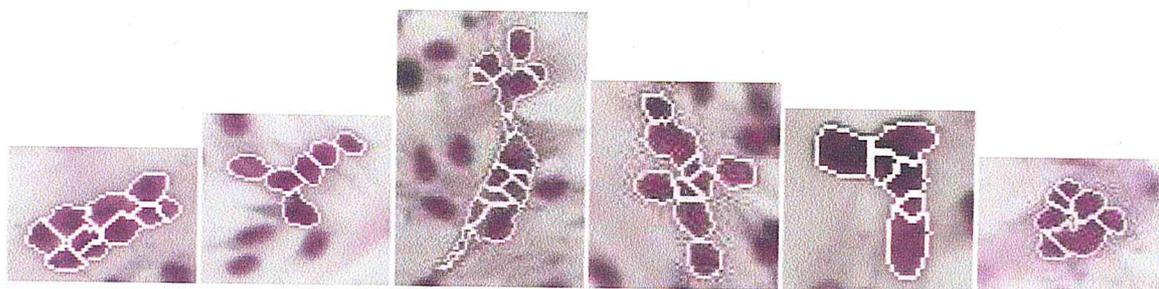
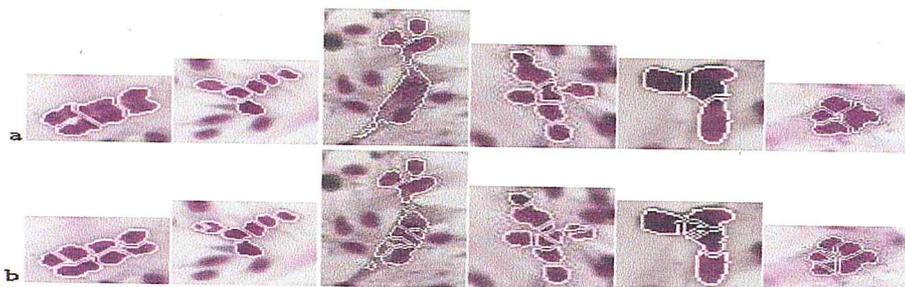


FIG. 9.5: Résultat du processus itératif de séparation

La fusion

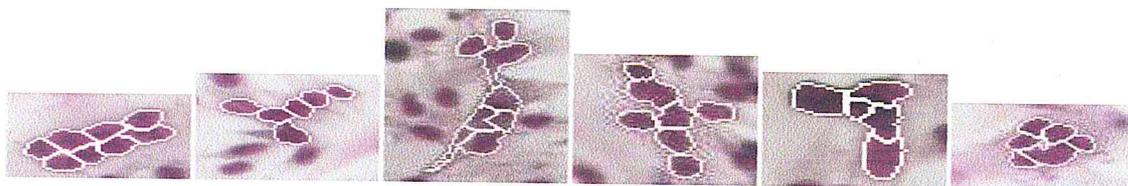


FIG. 9.7: Résultat du processus de fusion

Figure II.22 : Résultats des traitements de séparation et de fusion

Discussion

La méthode ne permet pas de segmenter les noyaux, quand la densité cellulaire est trop importante. De plus, les agrégats ne sont pas toujours correctement segmentés et le processus de division des agrégats est assez lent.

4. Résumé du travail de Boujelben [13]

Le travail de recherche appréhendé par Boujelben, se rapporte à la détection des calcifications dans une image mammographique suivant les ensembles de niveaux.

La première phase, consiste à appliquer sur l'image une transformation logarithmique, pour améliorer le contraste des régions proches du contour de la tumeur (cf. Figure II.23).

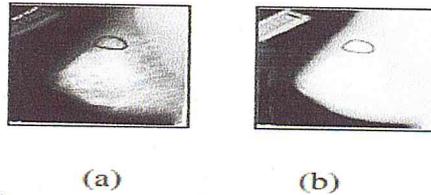
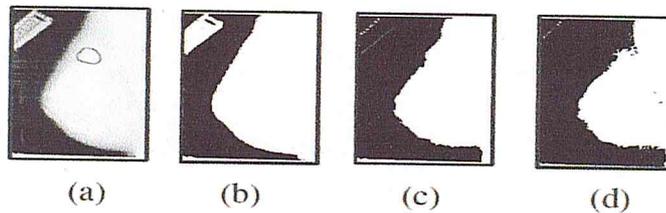


Figure 2. Logarithmic enhancement result: (a) original image; (b) enhanced image.

Figure II.23 : résultats de la transformation logarithmique

La seconde étape consiste en la binarisation de l'image, pour recueillir l'enveloppe de la glande mammaire (cf. Figure II.24).



Binarization result: (a) Enhanced image; (b) Otsu's method; (c) Maximum-entropy principle method; (d) Maximum correlation criterion method.

Figure II.24 : résultats de la binarisation

La troisième étape consiste à déterminer l'orientation de la tumeur (cf. Figure II.25).

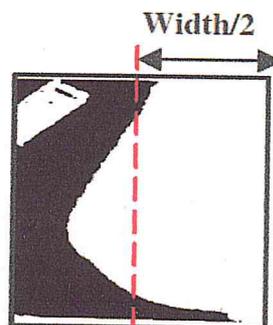


Figure II.25 : résultats de l'orientation.

La troisième étape vise la séparation du buste par rapport au fond de l'image, en utilisant l'étiquetage des composantes connexes (cf. Figure II.26). :

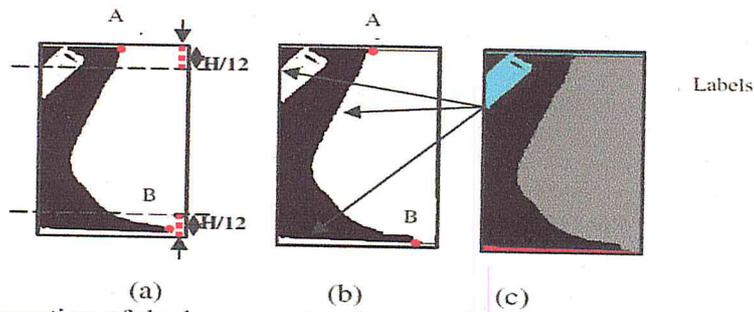


Figure 5. Separation of the breast region from the background: (a) Identification of top and bottom points; (b) Drawing of two lines for separation; (c) Using of the Connected Component Labelling algorithm.

Figure II.26 : séparation de la glande mammaire du fond

L'étape qui suit, vise à la sélection des régions (cf. Figure II.27).

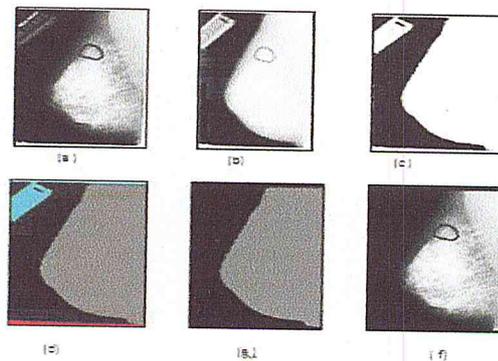


Figure 6. Breast region extraction: (a) Original image; (b) Logarithmic enhancement; (c) Binarization with Otsu's thresholding; (d) Separation of the breast region; (e) Selection of the largest label; (f) Final result of breast extraction.

Figure II.26 : résultat final du fond

La détection des régions d'intérêt, se fait dans ce cas en appliquant les ensembles de niveaux.

La figure suivante, illustre l'application de l'approche des ensembles des niveaux sur une image présentant une tumeur maligne. utilisant pour l'initialisation, le niveau de gris

maximum..

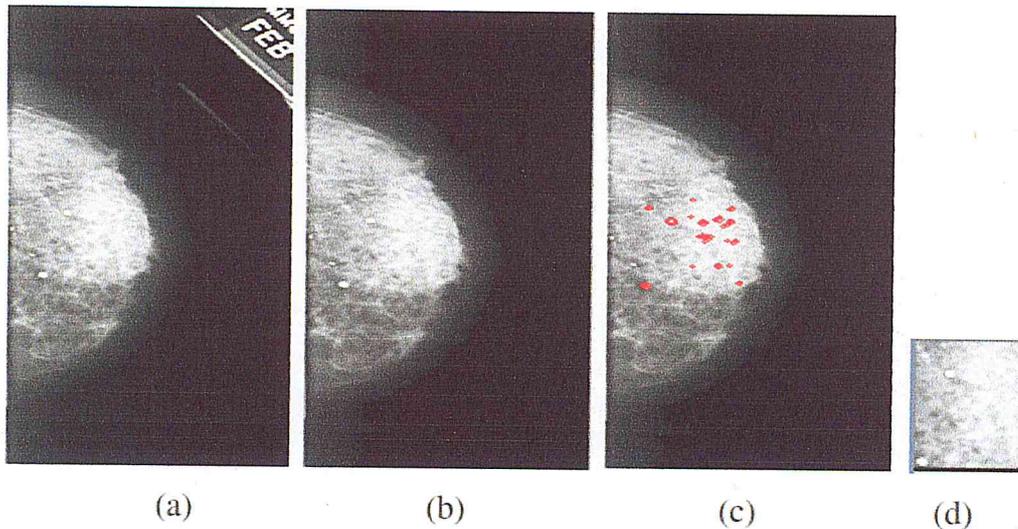


Figure II. 27: détection de la région d'intérêt: (a) image originale, (b) détection de la glande mammaire, (c) application des ensembles de niveaux, (d) Région détectée.

Cette approche, permet de guider le radiologiste dans son interprétation.

5. Travaux d'Oger : Indexation automatique d'images numériques application aux images histopathologiques du cancer du sein et hématologiques de leucémies lymphoïdes chroniques [5]

Les lésions précoces découvertes lors du dépistage sont souvent mal connues et/ou de très petite taille ce qui rend délicat le diagnostic histopathologique.

Un problème similaire est rencontré en hématologie avec la pratique de plus en plus répandue des examens sanguins systématiques et la difficulté d'identification de cellules suspectes et d'évènement rares au sein d'un frotti sanguin.

Il est par conséquent très important d'apprécier dans quelle mesure la microscopie numérique et les outils d'analyse automatique des images pourront dans l'avenir aider ces spécialistes dans l'accomplissement de leurs tâches quotidiennes.

Le présent travail de thèse, mené dans cette optique, se fonde sur l'utilisation de lames virtuelles des préparations histologiques et cytologiques, acquises à basse ou à haute résolution.

Il consiste à développer et à tester une série d'outils d'aide au diagnostic basés sur l'indexation automatique des images. Cependant, l'utilisation des lames virtuelles implique la manipulation d'une masse de données très importante, qui constitue un frein pour traiter, analyser et même visualiser les images de manière classique.

Oger a testé tout d'abord la pertinence d'une analyse globale des images, puis d'une analyse locale de celles-ci, accompagnées d'une réduction de dimension des données par diverses méthodes, dont l'analyse spectrale.

Le système est mis en œuvre à propos de deux localisations dont l'incidence constitue un problème de santé publique, les tumeurs mammaires et la leucémie lymphoïde chronique.

II. 7 Choix de l'espace et approches adoptées

En s'inspirant du travail de Glory, nous pouvons établir le synoptique de la méthodologie de la détection des structures tumorales dans les images cytologiques, en recherchant l'espace le plus approprié.

Nous présentons notre synoptique du choix de l'espace couleur suivant la figure II.28.

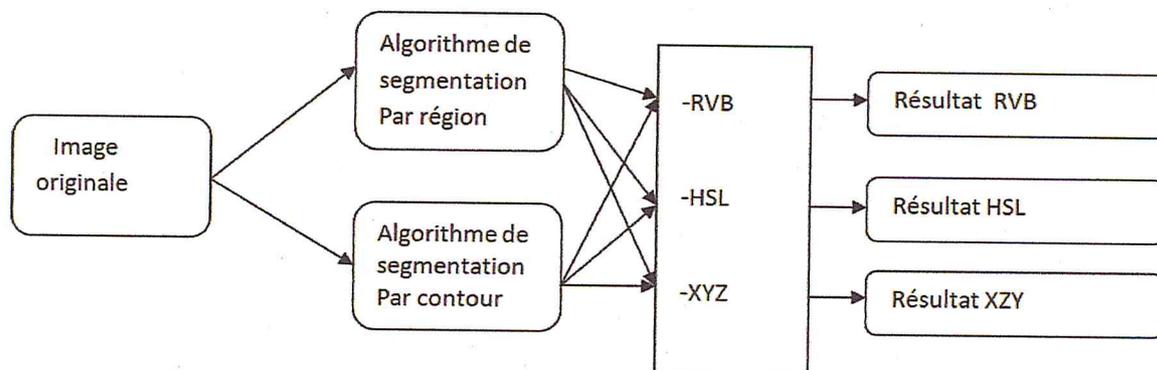


Figure II.28 : Synoptique du choix de l'espace couleur

Le choix de l'approche de la segmentation en régions, est adopté par l'utilisation de la LPE, relativement aux résultats de bonne qualité obtenus par les travaux de Lezoray et les travaux de LRDSI, puisque l'avantage principal de la morphologie mathématique est l'analyse des formes.

La segmentation en contours, est appréhendée par l'approche des ensembles de niveaux, pour l'obtention du contour de la forme. Une coopération séquentielle des résultats sera effectuée, d'une façon séquentielle, pour profiter des avantages des deux approches et extraire ainsi, les régions significatives.

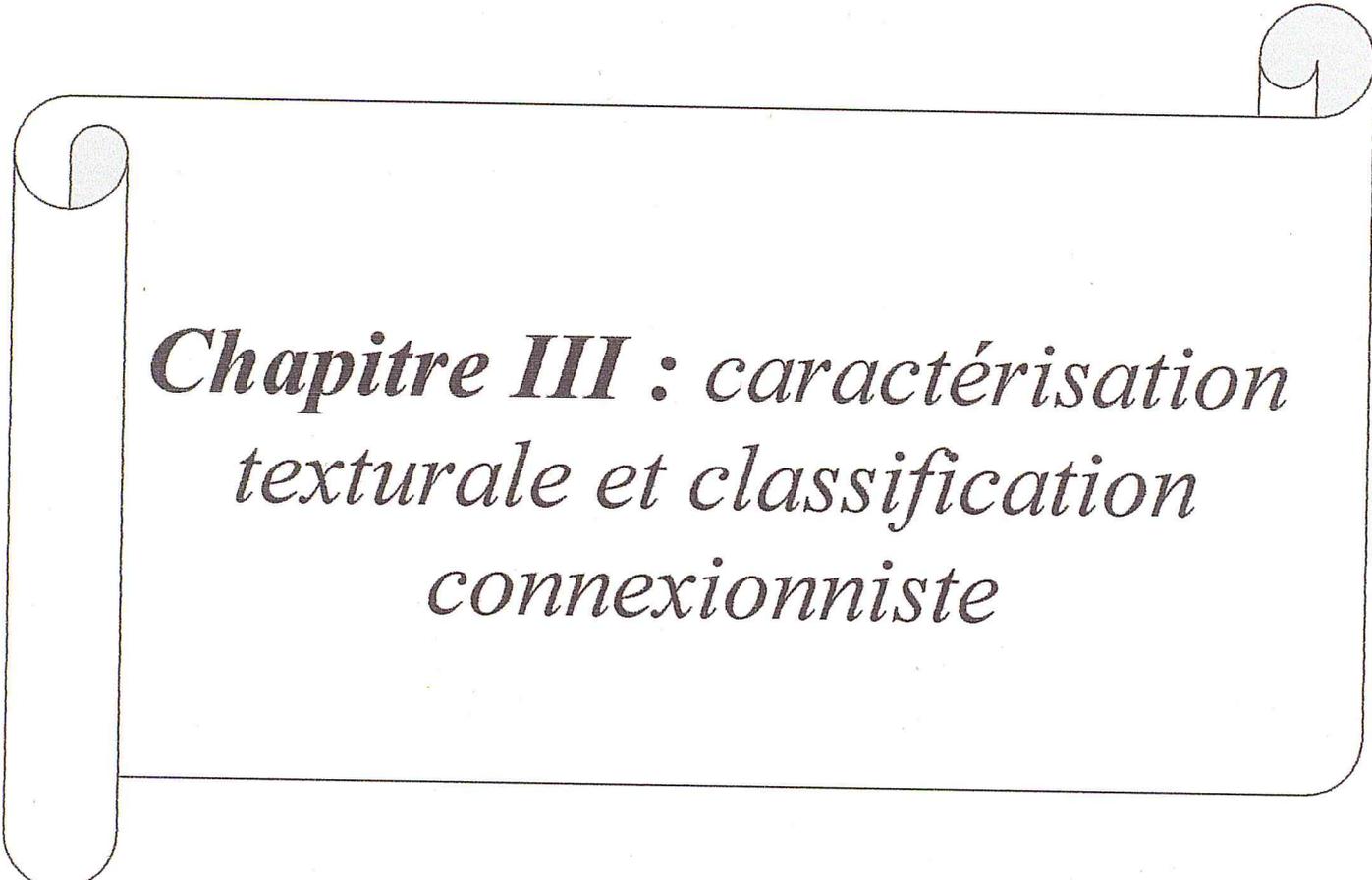
II.8 Conclusion

Dans ce chapitre, nous avons étudié les approches de base de la segmentation, nécessaires à l'analyse des techniques optimales utilisées dans le cadre de ce mémoire. Qu'elle soit par

région, contour, ou suivant les traitements de la morphologie mathématique, la segmentation permet l'extraction de l'information utile en prenant en compte les caractéristiques des régions d'intérêt.

En comparaison avec une segmentation par régions ou par contours, le résultat de la segmentation coopérative intègre les deux types d'information (régions et contours) afin de permettre une meilleure prise en considération des caractéristiques des objets de l'image. Les zones homogènes, ainsi que les transitions entre elles, sont respectées simultanément.

Dans le chapitre suivant, nous abordons la caractérisation texturale et topologique des régions ainsi que la classification connexionniste.



***Chapitre III : caractérisation
texturale et classification
connexionniste***

III.1 Introduction

Ce chapitre introduit un point très important de notre travail: le problème de l'extraction de caractéristiques. Par ce terme, nous entendons l'ensemble des méthodes permettant d'extraire des informations relatives à la texture et à la topologie, mais aussi et surtout au contenu structurel de l'image. Cette caractérisation nous dirige vers la classification connexionniste, étape finale de notre système.

L'analyse de texture regroupe un ensemble de techniques mathématiques permettant de quantifier les différents niveaux de gris présents dans une image en termes d'intensité ou de rugosité et leur distribution. Plusieurs méthodes d'analyse de texture existent et peuvent être classées en deux grandes catégories: les méthodes structurelles et les méthodes statistiques [19]. Les méthodes structurelles permettent de décrire la texture en définissant les primitives et les "règles" d'arrangement qui les relie. Les méthodes statistiques étudient quant à elles, les relations entre un pixel et son voisinage. Elles sont définies selon divers ordres: 1, 2, 3, ..., n (l'ordre d'un paramètre de texture pouvant être perçu comme étant le nombre de pixels pris en compte dans une itération lors de l'évaluation de ce paramètre).

L'analyse de la topologie, repose sur la caractérisation de la géométrie suivant une description spatiale des pixels [19].

III. 2 Qu'est ce qu'une texture ? [19]

Dans le domaine du traitement de l'image et de la vision, il n'existe pas de définition satisfaisante de la texture. Les définitions mathématiques construites à partir de propriétés statistiques sont soit trop générales et imprécises soit trop restrictives pour s'adapter à la diversité des cas rencontrés. La définition présentée n'est pas opérationnelle et prête sur ce point à critique. Elle s'appuie sur une constatation expérimentale : une texture est un champ de l'image qui apparaît comme un domaine cohérent et homogène, c'est à dire formant un tout pour un observateur. La figure (III.1) illustre la diversité de quelques textures.

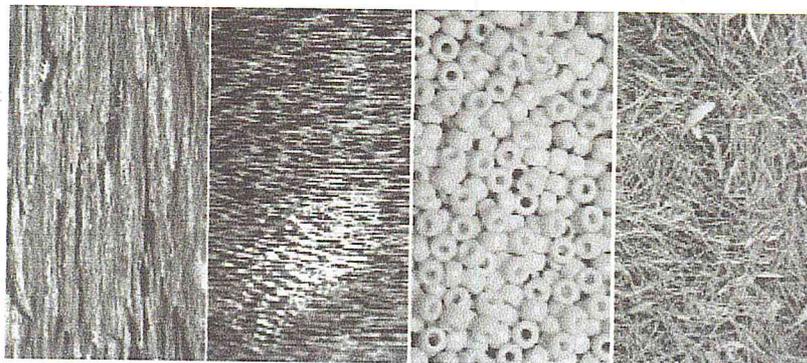


Figure III.1 : Quelques exemples de textures naturelles : écorce d'arbre, poil court, perles, tapis d'aiguilles de pin et de feuilles.

III.2.1 Distribution aléatoire ou régulière

Une première constatation que l'on fait en examinant des textures naturelles, est le rôle particulier que joue l'aléatoire dans la texture. On distingue assez naturellement deux modèles extrêmes de textures, entre lesquels se positionnent un peu toutes les textures [16][19] :

1-les textures régulières: dans lesquelles la périodicité du motif est évidente: grilles, murs, tissus, etc...

2-les textures aléatoires: pour lesquelles la distribution des intensités n'est l'objet d'aucune régularité apparente : sable, nuages, herbe, foule.

La première famille sera bien décrite par des approches fréquentielles ou des approches structurelles dans lesquelles on associera un motif et des règles de placement sur un pavage régulier.

III.2.2 Les échelles des textures

Une autre particularité très importante des textures, est qu'elles présentent généralement plusieurs niveaux d'échelle :

- ✓ A petite échelle (pour des détails fins), on observe souvent un objet élémentaire qui constitue la base. Une autre particularité très importante des textures est qu'elles présentent généralement plusieurs de types (perle, aiguille de pin, cf. Figure III.1). Cette primitive peut être très régulière géométriquement (perle), ou photométriquement (aiguille de pin). Elle peut être au contraire relativement variable (personne dans une foule, caillou sur une plage, nuage). Elle possède une taille et des propriétés statistiques caractéristiques (qui joueront sur sa fonction de corrélation, . . .). Cette composante élémentaire de la texture a tendance naturellement à disparaître lorsqu'on observe le champ global de la texture.
- ✓ A plus grande échelle (donc pour une analyse plus grossière), la texture apparaît comme une juxtaposition plus ou moins régulière des motifs ci-dessus. Cette juxtaposition possède ses propres lois d'isotropie, de périodicité, de régularité et induit donc d'autres lois statistiques (corrélation, densité de puissance) qui se combinent à celles du motif de base.

Une bonne analyse de texture donnera des informations sur ces deux composantes également.

III.2.3 Caractéristiques de la texture

Les caractéristiques de base les plus utilisées pour identifier les textures sont:

- ❖ La grossièreté: une texture grossière possède des primitives larges: il existe alors peu de variation entre l'intensité d'un pixel et celles de ses pixels voisins,
- ❖ le contraste: une texture possède un contraste élevé si les différences d'intensité entre primitives sont importantes,

- ❖ la complexité: une texture possède plusieurs types de primitives, dans ce cas le contenu d'information présent dans la texture est important,
- ❖ la force: plus la force est élevée et plus les primitives sont facilement définissables et visibles.

III.3 Paramètres texturaux [19]

L'analyse de texture regroupe un ensemble de techniques mathématiques permettant de quantifier les différents niveaux de gris présents dans une image en terme d'intensité et de distribution.

Cette méthode d'analyse a d'abord été utilisée pour la segmentation des images satellitaires; ce n'est qu'au début des années 70 que les premières applications biomédicales sont apparues: on peut citer les travaux de Sutton sur le poumon [19].

Ces techniques apparaissent plus sensibles que l'œil aux variations de texture. En effet, l'ensemble 'œil cerveau' est seulement capable d'apprécier un niveau de complexité de premier et second ordre. L'analyse de texture permet d'accroître le niveau d'informations extraites de l'image puisqu'il existe des méthodes statistiques d'ordre supérieur à deux.

L'étude s'effectue sur des images numérisées. L'analyse de texture consiste à calculer, à l'aide de la valeur des pixels (intensité) de chaque région, un certain nombre de paramètres mathématiques caractéristiques de la texture à étudier.

Dans les méthodes de premier ordre, qui correspondent à une description de l'histogramme des niveaux de gris, il n'y a pas d'informations sur la localisation du pixel. Il est donc nécessaire d'utiliser des méthodes d'ordre supérieur pour une analyse plus précise.

L'ordre des méthodes est donné par le nombre de pixels mis en jeu dans le calcul des paramètres. On retiendra, en particulier les matrices de cooccurrence.

➤ La méthode de matrice de cooccurrence

Elle permet de déterminer la fréquence d'apparition d'un "motif" formé de deux pixels séparés par une certaine distance d , dans une direction particulière θ par rapport à l'horizontale. Afin de limiter le nombre de calculs, on prend généralement comme valeurs 0° , 45° , 90° , 135° , 180° et 1 pour la valeur de d .

La taille de la matrice est $N_g * N_g$, où N_g correspond au maximum des niveaux de gris de l'image. Pour ne pas avoir une taille de matrice trop élevée, on choisira le plus souvent $N_g=8$, 16 ou 32. A chaque direction θ et pour chaque valeur de d , correspond une matrice de cooccurrence $\phi(d, \theta)$.

On définit généralement les matrices symétriques de cooccurrence. Elles sont construites à partir des relations suivantes :

$$\phi(d, 0^\circ) = \phi^t(d, 180^\circ)$$

$$\phi(d, 45^\circ) = \phi^t(d, 225^\circ)$$

$$\phi(d, 90^\circ) = \phi^t(d, 270^\circ)$$

III.1

$$\phi(d, 135^\circ) = \phi^t(d, 315^\circ)$$

Par exemple, la matrice symétrique associée à la direction 0° sera de la forme :

$$S_0(d) = \frac{1}{2} [\phi(d, 0^\circ) + \phi(d, 180^\circ)] \quad \text{III.2}$$

Une fois la matrice symétrique réalisée, il est possible d'en extraire une quinzaine de paramètres. Ils contiennent des informations sur la finesse, la directionnalité et la granularité de la texture.

Pour une texture grossière, les valeurs de la matrice sont concentrées sur la diagonale principale. Au contraire, pour une texture fine, les valeurs de la matrice seront dispersées. En effet, pour une telle texture il existe beaucoup de transitions de niveaux de gris. Nous distinguons dans ce qui suit, les paramètres texturaux :

1. La moyenne

Elle est donnée par la relation suivante;

$$\text{MOY} = \sum_i \sum_j P(i, j) \quad \text{III.3}$$

où $p(i, j)$ correspond aux éléments de la matrice de cooccurrence; c'est à dire à la probabilité de passer d'un pixel de niveau de gris i à un pixel de niveau de gris j .

2. La variance

Elle est représentée par la formule suivante;

$$\text{VAR} = \sum_i \sum_j (i - \text{Moy})^2 P(i, j) \quad \text{III.4}$$

Elle caractérise la distribution des niveaux de gris autour de la valeur moyenne déterminée précédemment.

3. L'énergie, moment angulaire d'ordre deux

Ce paramètre mesure l'énergie (cf. Equation III.5) de l'image. L'énergie a une valeur d'autant plus faible qu'il y a peu de zones homogènes: dans ce cas, il existe beaucoup de transitions de niveaux de gris.

$$E = \sum_{i,j} (P(i,j))^2 \quad \text{III.5}$$

4. Le contraste

Ce paramètre mesure les variations locales des niveaux de gris. Si elles sont importantes (c'est à dire s'il existe peu de régions homogènes), alors le contraste sera élevé. Ce paramètre permet aussi de caractériser la dispersion des valeurs de la matrice par rapport à sa diagonale principale.

$$\text{CONT} = \sum_i \sum_j (i-j)^2 P(i,j) \quad \text{III.6}$$

Une texture apparaît plus nette lorsque le contraste est plus élevé, les transitions entre niveaux de gris sont alors plus visibles.

5. Corrélation

Ce paramètre (cf. Equation III.7) a une grande valeur quand les valeurs, sont uniformément distribuées dans la matrice de cooccurrence et une faible valeur dans le cas contraire.

$$\text{COR} = \sum_i \sum_j (i - \mu_x)(j - \mu_y) P(i,j) \quad \text{III.7}$$

Où μ_x et μ_y représentent les moyennes respectivement des lignes et des colonnes de la matrice.

6. Corrélation normalisée

Celle-ci (cf. Equation III.8) permet de déterminer si certaines colonnes de la matrice sont égales. Plus les valeurs sont uniformément distribuées dans la matrice et, plus la corrélation est importante.

$$\text{CORN} = \sum_{i,j} \frac{ijP(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad \text{III.8}$$

où σ_x et σ_y représentent les écarts types respectivement des lignes et des colonnes de la matrice.

7. L'entropie

Celle-ci (cf. Equation III.9) mesure la complexité de l'image. Lorsque les valeurs de la matrice sont presque toutes égales, l'entropie est élevée.

$$\text{ENT} = - \sum_i \sum_j P(i,j) \log p(i,j) \quad \text{III.9}$$

Elle permet de caractériser le degré de granulation de l'image. Plus l'entropie est élevée et plus la granulation est grossière.

9. L'homogénéité ou moment différentiel inverse

Ce paramètre (cf. Equation III.10) a un comportement inverse du contraste. Plus la texture possède de régions homogènes et plus le paramètre est élevé. Il est basé sur le calcul de la relation suivante ;

$$\text{HOM} = \sum_{i,j} \frac{1}{1 + (i-j)^2} P(i,j) \quad \text{III.10}$$

Ce paramètre indique une mesure de l'uniformité des niveaux de gris de l'image.

10. L'écart type

Ce paramètre (cf. Equation III.11) mesure la répartition des niveaux de gris autour de la moyenne.

$$\text{ET} = \sqrt{\left(\sum_{i=0}^{I-1} \sum_{j=0}^{I-1} (P(i,j,d,\theta)(i-M)) \right)} \quad \text{III.11}$$

III.4 Paramètres topologiques [19]

1. Compacité

La compacité (cf. Equation III.11) est le rapport entre le carré du périmètre de la forme et l'aire de la région.

$$C_o = \frac{P^2}{A}, \quad \text{III.12}$$

2. Excentricité

Elle définit la distribution spatiale le long de l'axe de l'ellipse. Elle est déterminée selon le moment central (cf. Equation III.13).

$$\frac{(\mu_{02} - \mu_{20})^2 + 4\mu_{11}}{A} \quad \text{III.13}$$

Où μ_{pq} représente le moment central et A l'aire de la région

$$\mu_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (x - \bar{x})^p (y - \bar{y})^q \quad \text{III.14}$$

Il existe d'autres paramètres que nous n'utilisons pas dans le cadre de notre travail. Nous nous sommes basées, à cet effet, sur la caractérisation des pathologies cytologiques mammaires, suivant la détermination des attributs, de l'entropie, de l'écart type, de l'homogénéité, du contraste, de l'énergie et de la corrélation ainsi que de la compacité et de l'excentricité.

III. 5 Classification connexionniste [17][20][21]

De la biologie au connexionnisme, l'objectif de cette partie, est d'étudier les réseaux de neurones, pour établir notre choix concernant la classification des pathologies mammaires.

III.5.1 Fonctionnement général d'un réseau de neurones

1. Qu'est-ce qu'un réseau de neurones ?

Tout d'abord, ce que l'on désigne habituellement par réseau de neurones, est en fait un réseau de neurones artificiels basé sur un modèle simplifié de neurone. Ce modèle permet certaines fonctions du cerveau, comme la mémorisation associative, l'apprentissage par l'exemple, le travail en parallèle. Cependant, le neurone artificiel est loin de posséder toutes les capacités du neurone biologique. Les réseaux de neurones biologiques (cf. Figure III.2) sont ainsi beaucoup plus compliqués que les modèles mathématiques et informatiques. En effet, le neurone biologique possède trois principales composantes : les dendrites, le corps cellulaire et l'axone. Les dendrites forment un maillage de récepteurs nerveux qui permettent d'acheminer vers le corps du neurone des signaux électriques en provenance d'autres neurones. Celui-ci agit comme intégrateur en accumulant des charges électriques. Lorsque le neurone devient suffisamment excité (lorsque la charge accumulée dépasse un certain seuil), par un processus électrochimique, il engendre un potentiel électrique qui se propage à travers son axone pour éventuellement venir exciter d'autres neurones. Le point de contact entre l'axone d'un neurone et la dendrite d'un autre neurone s'appelle la synapse. Il semble que c'est l'arrangement spatial des neurones et de leur axone, ainsi que la qualité des connexions synaptiques individuelles qui détermine la fonction précise d'un réseau de neurones biologique. C'est en se basant sur ces connaissances que le modèle mathématique a été défini.

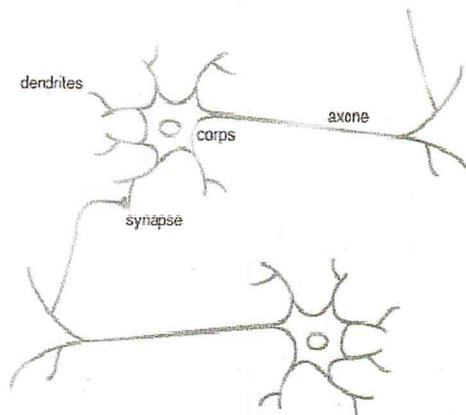


Figure III.2 : Modèle d'un neurone biologique [20]

2. Modélisation d'un neurone

Le fonctionnement d'un neurone artificiel peut être modélisé par le schéma suivant :

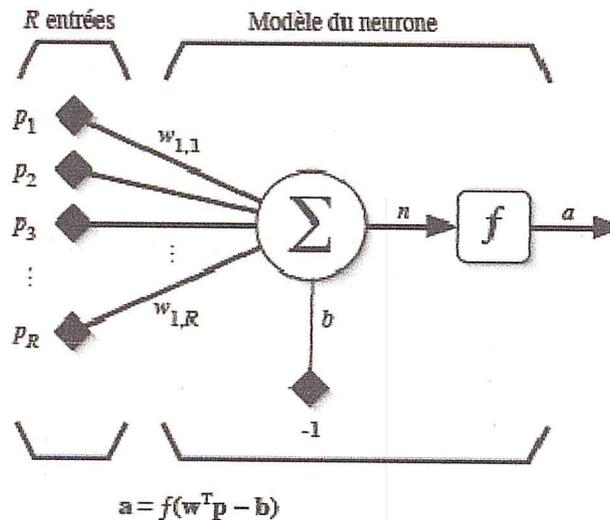


Figure III.3 : Modèle d'un neurone artificiel [21]

Le modèle mathématique d'un neurone artificiel est illustré à la figure III.4 Un neurone est essentiellement constitué d'un intégrateur qui effectue la somme pondérée de ses entrées. Le résultat n de cette somme est ensuite transformé par une fonction de transfert f qui produit la sortie a du neurone. En suivant les notations présentées à la section précédente, les R entrées du neurone correspondent au vecteur $\mathbf{P} = [p_1 p_2 \dots p_R]^T$, alors que $\mathbf{w} = [w_{1,1} w_{1,2} \dots w_{1,R}]^T$ représente le vecteur des poids du neurone. La sortie n de l'intégrateur est donnée par l'équation suivante :

$$\begin{aligned}
 n &= \sum_{j=1}^R w_{1,j} p_j - b \\
 &= w_{1,1} p_1 + w_{1,2} p_2 + \dots + w_{1,R} p_R - b,
 \end{aligned}
 \tag{III.15}$$

que l'on peut aussi écrire sous forme matricielle :

$$n = \mathbf{w}^T \mathbf{p} - b.
 \tag{III.16}$$

Cette sortie correspond à une somme pondérée des poids et des entrées moins ce qu'on nomme le biais b du neurone. Le résultat n de la somme pondérée s'appelle le niveau d'activation du neurone. Le biais b s'appelle aussi le seuil d'activation du neurone. Lorsque le niveau d'activation atteint ou dépasse le seuil b , alors l'argument de f devient positif (ou nul). Sinon, il est négatif [21].

Un poids d'un neurone artificiel représente donc l'efficacité d'une connexion synaptique. Un poids négatif vient inhiber une entrée, alors qu'un poids positif vient l'accentuer.

Lorsqu'on ajoute au modèle formulé par l'équation III.16 la fonction d'activation f , on obtient la sortie du neurone :

$$a = f(n) = f(\mathbf{w}^T \mathbf{p} - b). \quad \text{III.17}$$

En remplaçant \mathbf{w}^T par une matrice $\mathbf{W} = \mathbf{w}^T$ d'une seule ligne, on obtient la forme suivante:

$$a = f(\mathbf{W}\mathbf{p} - b). \quad \text{III.18}$$

L'équation III.18 permet d'introduire, un modèle plus compact que celui de la figure III.3. La figure III.4, illustre celui-ci. On y représente les R entrées comme un rectangle noir (le nombre d'entrées est indiqué sous le rectangle). De ce rectangle sort le vecteur \mathbf{p} dont la dimension matricielle est $R \times 1$. Ce vecteur est multiplié par une matrice \mathbf{W} qui contient les poids (synaptiques) du neurone. Dans le cas d'un neurone simple, cette matrice possède la dimension $1 \times R$. Le résultat de la multiplication correspond au niveau d'activation qui est ensuite comparé au seuil b (un scalaire) par soustraction. Finalement, la sortie du neurone est calculée par la fonction d'activation f . La sortie d'un neurone est toujours un scalaire.

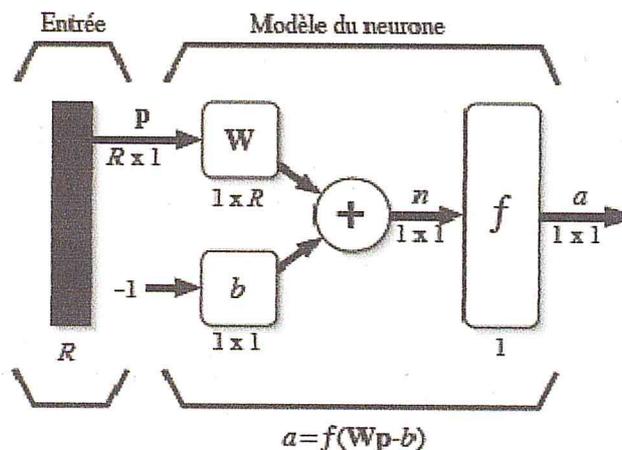


Figure III.4 : Représentation matricielle du modèle d'un neurone artificiel [21]

3. Connectivité

La connectivité des réseaux est la manière dont les neurones, sont connectés entre eux. Elle peut être totale (tous les neurones sont connectés entre eux) ou organisée par couche (les neurones d'une couche ne sont connectés qu'à la couche précédente en entrée et à la couche suivante en sortie). Il existe des réseaux monocouches ou multicouches [22].

4. Apprentissage

Une caractéristique des réseaux de neurones, est leur capacité à apprendre. Mais cette connaissance n'est pas acquise dès le départ. La plupart des réseaux de neurones apprennent par l'exemple en suivant un algorithme d'apprentissage. Il y a deux algorithmes principaux: lors d'un apprentissage supervisé, les résultats corrects (c'est-à-dire les valeurs que l'on désire que le réseau obtienne en sortie) sont fournis au réseau, si bien que celui-ci peut ajuster ses poids de connexions pour les obtenir. Après l'apprentissage, le réseau est testé en lui donnant seulement les valeurs d'entrée mais pas les sorties désirées, et en regardant si le résultat obtenu, est proche du résultat désiré. Lors d'un apprentissage non supervisé, on ne fournit pas au réseau les sorties que l'on désire obtenir. On le laisse évoluer librement jusqu'à ce qu'il se stabilise [20] [21].

5. Calcul des poids synaptiques

La rétropropagation (cf. Figure III.5), est une méthode de calcul des poids (aussi appelés poids synaptiques du nom des synapses, terme désignant la connexion biologique entre deux neurones) pour un réseau à apprentissage supervisé qui consiste à minimiser l'erreur quadratique de sortie (somme des carrés de l'erreur de chaque composante entre la sortie réelle et la sortie désirée) [21].

III.5.2 Quelques réseaux célèbres

Il existe de nombreux réseaux de neurones actuellement. On en présente dans le cas de notre étude, les classiques [17] [20][21][22].

1. Le perceptron

C'est un des premiers réseaux de neurones, conçus en 1958 par Rosenblatt [21]. Il est linéaire et monocouche. Il est inspiré du système visuel. La première couche (d'entrée), représente la rétine. Les neurones de la couche suivante (unique, d'où le qualificatif de monocouche) sont les cellules d'association, et la couche finale représente les cellules de décision. Les sorties des neurones ne peuvent prendre que deux états (-1 et 1 ou 0 et 1). Seuls les poids des liaisons entre la couche d'association et la couche finale, peuvent être modifiés. La règle de modification des poids utilisée est la règle de Widrow-Hoff [21] : si la sortie du réseau (donc celle d'une cellule de décision) est égale à la sortie désirée, le poids de la connexion entre ce neurone et le neurone d'association qui lui est connecté n'est pas modifié. Dans le cas contraire, le poids est modifié proportionnellement. À la différence entre la sortie obtenue et la sortie désirée suivant la relation ci-dessous :

$$w \leftarrow w + k (d - s) \quad \text{III.19}$$

Où s est la sortie obtenue, d la sortie désirée et k une constante positive. En 1969, Papert et Minsky [21] ont démontré les limites du perceptron classique, incapable, par exemple de simuler la fonction ou exclusif (xor).

2. Les perceptrons multicouches (PMC)

Ils sont une amélioration du perceptron comprenant une ou plusieurs couches intermédiaires dites couches cachées, dans le sens où elles n'ont qu'une utilité intrinsèque pour le réseau de neurones et pas de contact direct avec l'extérieur. Chaque neurone n'est relié qu'aux neurones des couches directement précédente et suivante, mais à tous les neurones de ces couches (cf. Figure III.5).

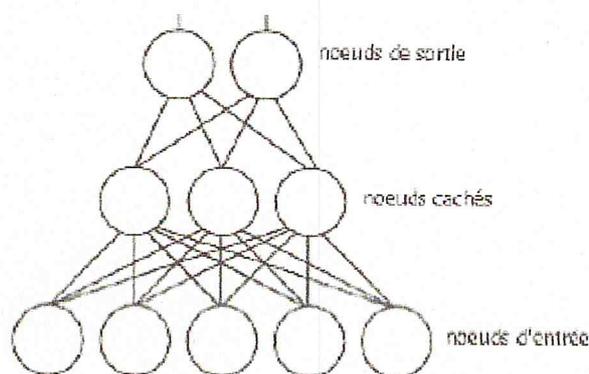


Figure III.5: Modèle des PMC

Les PMC utilisent, pour modifier leurs poids, un algorithme de rétropropagation du gradient qui est une généralisation de la règle de Widrow-Hoff [22]. Il s'agit toujours de minimiser l'erreur quadratique. On propage la modification des poids de la couche de sortie, jusqu'à la couche d'entrée. Les PMC agissent comme un séparateur non linéaire et peuvent être utilisés pour la classification.

3. Les réseaux de Hopfield

Il s'agit d'un réseau constitué de neurones à deux états (-1 et 1, ou 0 et 1), dont la loi d'apprentissage est la règle de Hebb (1949) [21], qui veut qu'une synapse améliore son activité si et seulement si l'activité de ses deux neurones est corrélée (c'est-à-dire que le poids d'une connexion entre deux neurones augmente quand les deux neurones sont activés en même temps).

4. Les réseaux de Kohonen

Contrairement aux réseaux de Hopfield [16][21] où les neurones sont modélisés de la façon la plus simple possible, on recherche ici un modèle de neurone plus proche de la réalité. Ces réseaux sont inspirés des observations biologiques du fonctionnement des systèmes nerveux de perception des mammifères.

Une loi de Hebb modifiée (tenant compte de l'oubli) est utilisée pour l'apprentissage. La connexion est renforcée dans le cas où les neurones reliés ont une activité simultanée, et diminuée dans le cas contraire (alors qu'il ne se passait précédemment rien dans ce cas). Ceci se résume par la formule suivante [14] :

$$d w / dt = k S e - B(S) w$$

III.20

5. Le modèle du perceptron

Il s'agit d'un réseau de neurones [22] inspiré du système visuel possédant deux couches de neurones: une première couche relative à la perception et une seconde liée à la prise de décision. C'est le premier système artificiel capable d'apprendre par expérience.

L'objectif initial, celui poursuivi par le cognitivisme, est l'analyse des différentes étapes du processus en jeu dans la production d'une pensée rationnelle afin de tenter d'en reproduire les aspects les plus complexes. La difficulté majeure, repose cependant sur l'impossibilité de bien discerner les différentes étapes d'un tel processus. Malgré des résultats probants, force est de reconnaître que la pensée produite reste bien pâle devant celle de l'homme, indéfectible constat de l'intelligence artificielle.

A. Notion de perceptron

➤ Le modèle

A l'image du neurone biologique (cf. figure III.2), véritable support de notre perception des sens, le perceptron se révèle être le support de la perception des informations. On peut ainsi définir un perceptron linéaire à seuil possédant un nombre quelconque d'entrées et une sortie.

Une entrée est une variable booléenne qui indique s'il y a une impulsion électrique et est associée à un poids qui représente l'amplitude de l'impulsion. La sortie est égale à 1 si la somme des poids est strictement supérieure à une valeur seuil. Les entrées représentent donc les dendrites, la sortie l'axone et le perceptron le neurone [17].

➤ L'utilisation

L'algorithme le plus simple utilisant un perceptron, est l'algorithme par correction d'erreurs. Il s'agit de présenter au perceptron différents vecteurs d'entrées associées à une variable booléenne. On appelle vecteur d'entrée, un vecteur qui initialise chaque entrée du perceptron à 0 ou 1. C'est l'ensemble des vecteurs d'entrées qu'on appelle échantillon. Le booléen représente la classe associée au vecteur. Tous les poids sont initialisés à une valeur aléatoire; cependant, au moins un poids doit avoir une valeur non nulle pour que la sortie ne soit pas toujours égale à zéro [17].

L'algorithme consiste, à présenter au perceptron un certain nombre de fois ces vecteurs d'entrée. Lorsqu'un vecteur est présenté au perceptron, les poids sont modifiés à chaque fois que la sortie est différente de la valeur de la classe. L'apprentissage est terminé une fois que tous les vecteurs ont été présentés sans modification des poids.

Il convient de remarquer qu'il s'agit effectivement d'un algorithme par correction d'erreurs car les poids ne sont modifiés que si la sortie calculée est différente de la sortie attendue. La règle de modification des poids est la suivante:

$$W_i \leftarrow W_i + (C - O) \cdot X_i \quad \text{III.21}$$

où :

W_i : représente le poids de l'entrée i

C : la classe

O : la sortie calculée

X_i : l'entrée i

➤ **Les limites de ce modèle**

Ce modèle a rapidement montré ces limites. En effet, l'algorithme exige un échantillon linéairement séparable, ce qui est rarement le cas des données que l'on veut faire apprendre artificiellement. Dans le cas où l'échantillon n'est pas linéairement séparable, l'algorithme ne se terminera pas. C'est pourquoi, les réseaux de neurones multicouches ont été mis en place.

Cette contrainte ne permettra donc pas l'apprentissage du XOR, ou exclusif, par cet algorithme basé sur la notion de perceptron. Il apparaît évidemment impossible, de séparer linéairement le XOR en 2 classes distinctes (cf. Figure III.6) [17].

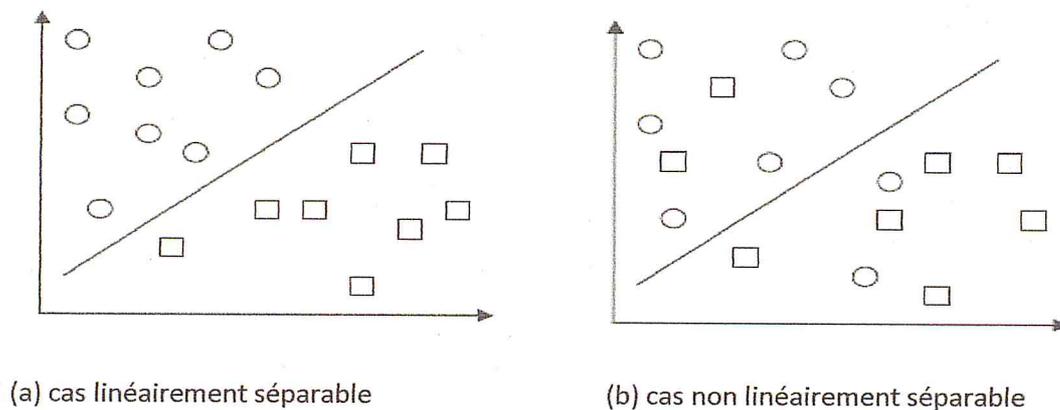


Figure III.6: Deux cas de séparation

Pour améliorer l'apprentissage du perceptron, Widrow et Hoff [21] ont proposé une nouvelle règle d'apprentissage plus proche des mécanismes d'optimisation. Il s'agit non seulement de trouver la frontière séparant les classes lorsqu'elle existe, mais aussi de trouver la meilleure solution qui minimise l'erreur lorsque cette séparation linéaire n'est pas possible.

6. Classification par le réseau PMC [17][21]

Le réseau de neurones ou perceptron multicouche PMC est un ensemble de neurones interconnectés. Cela constitue une extension du modèle de perceptron. Dans le réseau multicouche, les neurones sont disposés en couches successives. Dans notre cas, nous n'étudierons que les réseaux à 3 couches (cf. figure III.7) :

- Une couche d'entrée représentant les diverses impulsions électriques que reçoit le réseau,
- une couche cachée qui définit l'ensemble des neurones intermédiaires,
- la couche de sortie qui définit la réponse des neurones par rapport aux impulsions d'entrée.

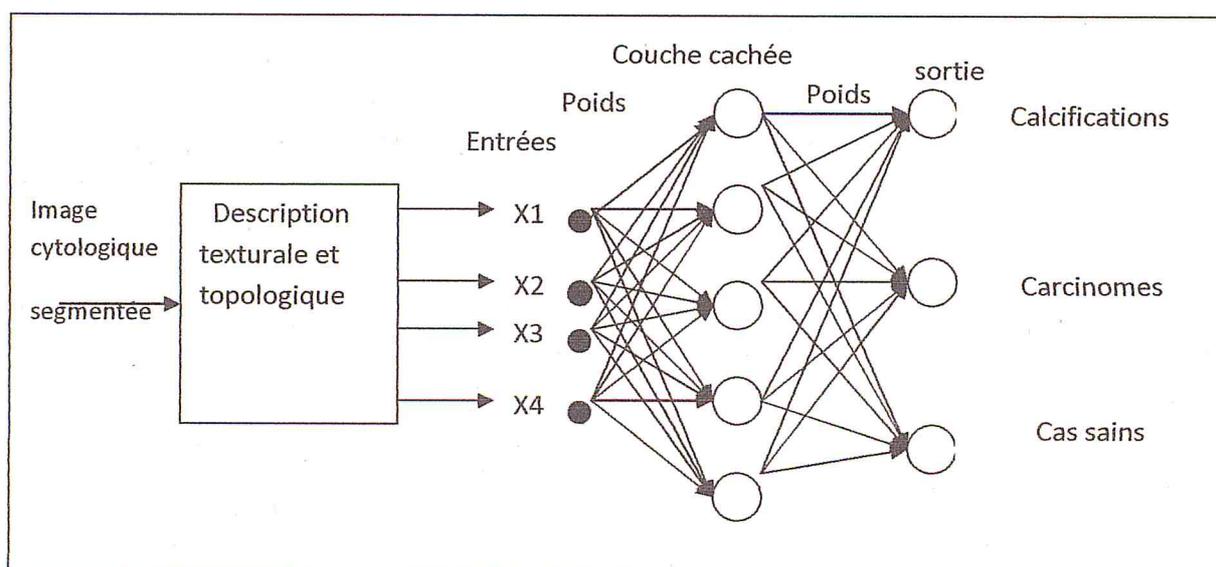


Figure III.7 : déroulement du réseau PMC

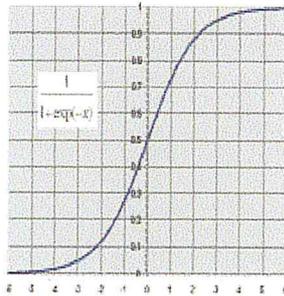
➤ Création de la base d'apprentissage

Lors de la phase d'apprentissage, (cf. Figure III.9) le neurone modifie ses poids. Nous allons maintenant expliquer comment cette modification s'effectue en amont du neurone, par la règle de rétropropagation du gradient et la règle de Hebb [21]. L'algorithme de rétropropagation du gradient, consiste à déterminer l'erreur effectuée par chaque neurone puis à modifier la valeur

des poids pour minimiser cette erreur. Il faut effectuer des rétropropagations jusqu'à ce que, l'erreur quadratique moyenne devienne inférieure à un certain seuil. L'erreur quadratique est définie comme la somme sur les sorties des carrés des différences entre l'entrée obtenue et l'entrée désirée soit :

$$E = \sum (S_i - Y_i)^2 \quad \text{III.22}$$

Le réseau est représenté par 3 couches de neurones, chaque neurone d'une couche étant relié à tous les neurones de la couche suivante par une liaison pondérée. Chaque neurone intègre les données grâce à une fonction non linéaire qui permet l'apprentissage de classes non linéairement séparables [17]. On utilise ici la fonction sigmoïde suivante :



$$s(x) = \frac{1}{1 + e^{-x}} \quad \text{III.23}$$

$$ds(x) = s(x)(1 - s(x))$$

Figure III.8 : Représentation de la fonction sigmoïde [21]

On effectue tout d'abord, une propagation de l'entrée à travers le réseau. Considérons S le vecteur de sortie obtenu après la propagation de l'entrée X, à travers le réseau et Y, le vecteur des sorties que l'on aurait voulu obtenir. Pour la couche de sortie considérons le neurone i ; on obtient donc l'erreur pour ce neurone : $S[i] - Y[i]$

On appelle p la taille de la couche cachée. Le neurone i reçoit alors les p sorties des neurones de la couche cachée modulées de leur poids respectifs. Ce sont ces poids qu'il nous faut modifier. Pour relativiser l'importance d'un poids par rapport à un autre, dans l'erreur totale il faut moduler sa modification par la sortie du neurone auquel il est rattaché [17]. Il faut maintenant utiliser le gradient de l'erreur afin de pouvoir à terme, faire converger la valeur du poids. On obtient alors la formule suivante :

La règle de modification des poids pour un exemple k :

$$W_{ij}(k) = W_{ij}(k-1) - \text{pas} \times D_i \times O_j \times ds(I_i) \quad \text{III.24}$$

$$D_i = (S_i - Y_i) \text{ pour la couche de sortie,}$$

Avec :

W_{ij} : le poids entre le neurone i d'une couche et le neurone j de la couche précédente

Pas: pas de l'apprentissage

O_j : Sortie du neurone j

I_i : entrée du neurone i : $I_i = \sum_j W_{ij} \cdot O_j$

S_i : Sortie obtenue,

Y_i : Sortie désirée,

ds : Fonction sigmoïde dérivée.

Pour la couche cachée, déterminer l'erreur commise par chaque neurone n'est pas instantanée ; puisque l'on ne dispose pas de la valeur théorique qu'il devrait y avoir. La méthode, consiste à remarquer que l'erreur effectuée par un neurone de la couche cachée est la somme des erreurs des neurones de la couche de sortie modulée par les poids respectifs des liaisons entre les deux couches. Ainsi l'erreur pour le neurone i , de la couche cachée devient [17]:

$$\sum_h W_{hi} \cdot (S[h] - Y[h]) \quad h \text{ appartenant à la couche de sortie.}$$

On obtient alors la formule de modification des poids suivante :

$$W_{ij}(k) = W_{ij}(k-1) - \text{pas} \times D_i \times O_j \times ds(I_i) \quad \text{III.25}$$

$$D_i = \sum_h D_h \cdot W_{hi} \quad \text{pour la couche cachée}$$

Après avoir présenté les étapes de l'algorithme de rétropropagation du gradient, nous résumons alors ces étapes dans la figure III. 9 :

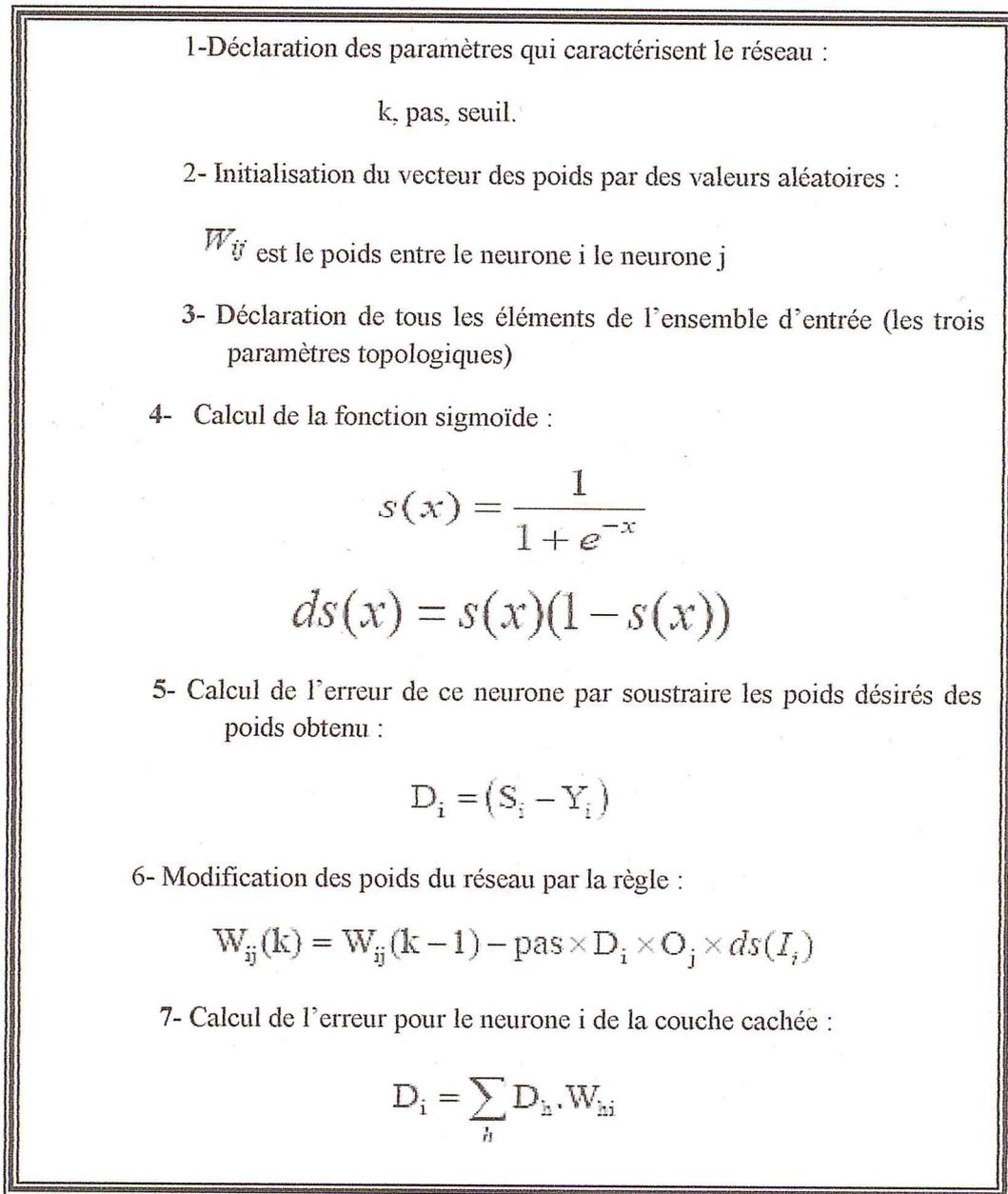


Figure III.9 : Etapes d'apprentissage du réseau PMC

Nous résumons le principe de classification par la méthode connexionniste PMC dans le schéma suivant :

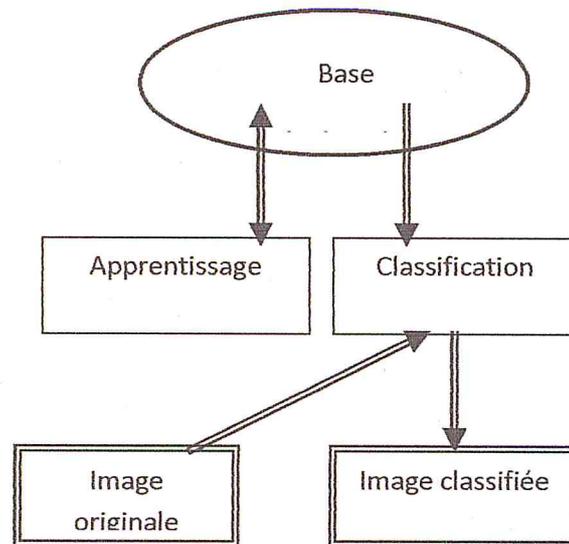


Figure III.10 : principe de classification par le PMC

Les neurones sont regroupés au sein d'une classe composée de tous les neurones d'une couche. La classe réseau, représente le réseau de neurones dans son ensemble et permet à partir d'un tableau de valeurs d'entrées, de parcourir le réseau de neurones pour déterminer la sortie associée à cette entrée [17].

Cette méthode effectue un nombre suffisant de rétropropagations, permettant à l'erreur quadratique d'être inférieure au seuil voulu [22].

III. 6 Conclusion

Dans ce chapitre, nous avons étudié une série de facteurs texturaux et topologiques, adaptés en particulier aux images médicales. La combinaison des paramètres étudiés, devrait fournir une bonne indication de la nature des formes cellulaires étudiées.

La deuxième phase que nous venons de présenter dans ce chapitre, est l'étude de quelques réseaux de neurones, en particulier le réseau du perceptron multicouche, réseau qui nous permettra de classer les pathologies mammaires rencontrées.

Le chapitre suivant, présente notre environnement de travail, les traitements réalisés sur des images cytologiques du sein.

A decorative border resembling a scroll, with rounded corners and a vertical strip on the left side. The text is centered within the scroll.

*Chapitre IV: Implémentation
et présentation de
l'application*

L'objectif de ce chapitre, est de présenter l'application développée, en vue d'une extraction et d'une classification connexionniste des structures mammaires cytologiques, pour l'aide à l'interprétation.

Les images cytologiques étudiées, présentant des cas sains, des carcinomes et des calcifications, sont expertisées par des cytopathologistes [23].

IV.1 Environnement de travail

Notre système a été développé sous environnement Windows, à l'aide de l'outil de programmation matlab de la version 7.9.

IV.1.1 Matériel utilisé

L'implantation de nos codes a été réalisée, grâce à un PC portable, ayant un processeur Intel core i3 duo de fréquence de 2.13 Ghz et d'une capacité de mémoire vive « RAM de 4 Go », d'un système d'exploitation 64 bits, d'un disque dur de 500 Go et d'une horloge de 2.27.

IV.1.2 Qu'est ce que Matlab ?

Notre système "cytocoop" a été développé sous environnement Windows, à l'aide de l'outil de programmation matlab (abréviation de " **Matrix Laboratory** ") de la version 7.9 [24], qui est un environnement informatique conçu pour le calcul matriciel. L'élément de base est une matrice dont la dimension n'a pas à être fixée. Matlab est un outil puissant qui permet la résolution de nombreux problèmes en beaucoup moins de temps. On peut se rendre compte de la puissance du logiciel en lançant la commande *demo*. Les compléments d'information peuvent être facilement obtenus, grâce à l'aide en ligne. Matlab s'élargit sur la librairie des fonctions mathématiques, l'environnement graphique, ainsi que sur une interface de développement.

L'outil de programmation matlab, présente de nombreux avantages dont;

- Programmation infiniment plus rapide pour le calcul et pour l'affichage,
- une librairie très riche,
- possibilité d'inclure un programme en C/C++,
- langage interprété: pas de compilation donc pas d'attente pour compiler,
- possibilité d'exécuter le code en dehors du programme,
- code facile à comprendre et très lisible.

IV.2 Présentation du système développé

Notre système permet une classification des principales pathologies mammaires, suivant le perceptron multicouche. Notre travail a été consacré dans une première partie, à l'extraction des régions d'intérêt des différentes pathologies étudiées, à partir de la ligne de partage des eaux s'adaptant aux images couleurs et à la segmentation par les ensembles de niveaux, pour la détection des contours des cellules. La seconde partie, s'est appuyée sur l'analyse texturale et topologique des zones d'intérêt, en vue d'une classification connexionniste pour la réalisation d'un système d'aide à la décision.

Le système général *cytocoop* adoptant notre méthodologie, est donné par l'architecture ci-dessous;

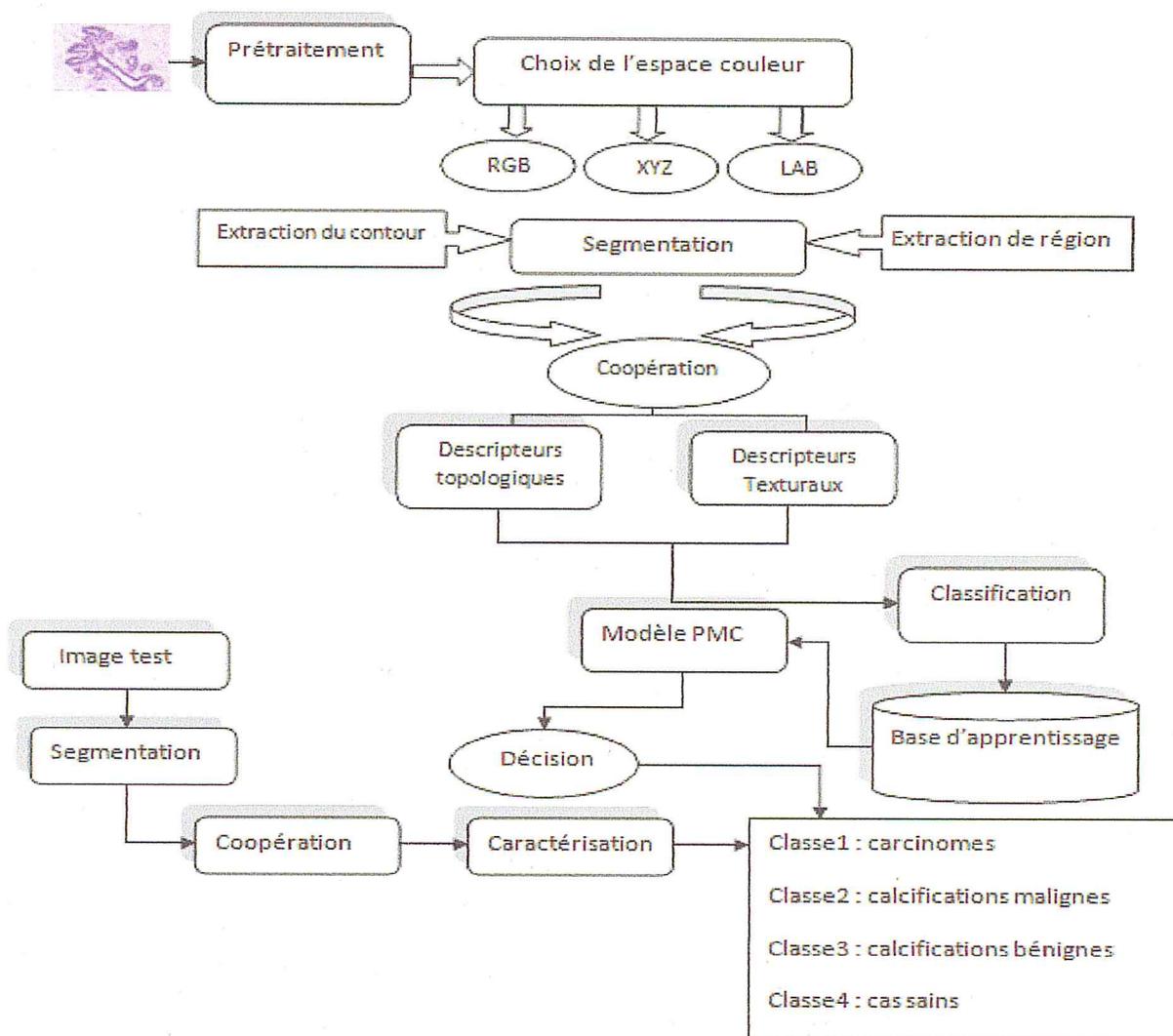
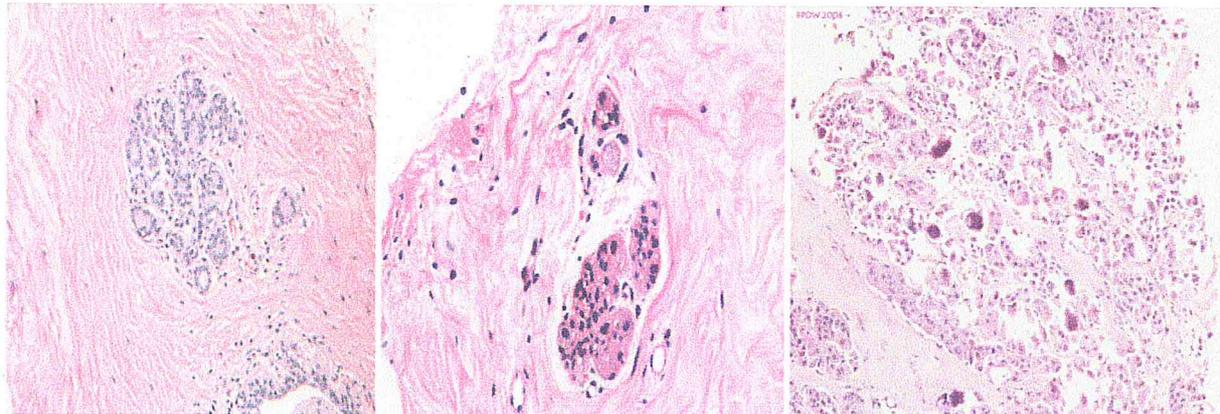


Figure IV.1 : synoptique général du système développé

IV.1.1 Images étudiées

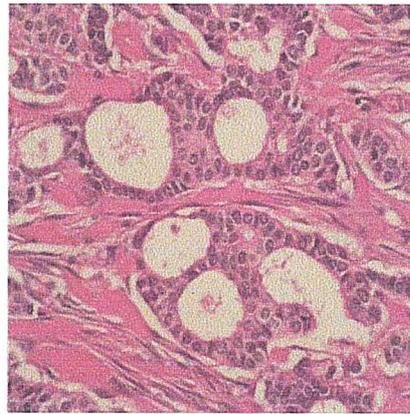
Les différentes images cytologiques mammaires issues de la base de données 'cytopathnet' [23], qui font l'objet de notre étude, illustrent principalement des calcifications malignes et bénignes, des carcinomes ainsi que des cas sains (cf. Figure IV.2). La base comprend 13 cas sains, 28 cas de carcinomes, 33 cas de calcifications bénignes et 12 cas de calcifications malignes.



Cas sain

Calcifications bénignes

Calcifications malignes



Carcinomes

Figure IV.2 : exemples d'images cytologiques mammaires

IV.2.2 Description des modules du système 'cytocoop'

➤ Prétraitement

Le terme de prétraitement est réservé aux processus, qui améliorent l'image acquise pour favoriser les analyses ultérieures suivant un filtrage passe bas. Cette étape consiste à

augmenter les informations pertinentes et à diminuer les bruits parasites. Ce module ne fait que compléter notre application, puisque les images utilisées n'ont subi aucun prétraitement.

➤ **Choix de l'espace**

Si un espace de couleurs fournit de meilleurs résultats que les autres, dans le cadre d'une étude donnée, cela ne signifie pas pour autant que cet espace de couleurs fournira systématiquement de meilleurs résultats: cela dépend du jeu d'images en couleur considéré. D'une étude à l'autre, les critères retenus ne sont pas les mêmes. Tantôt, ils font appel à des critères numériques, tantôt à des critères visuels, ce qui renvoie au problème de l'évaluation de la qualité d'un espace de couleurs relativement à l'évaluation de la qualité d'une segmentation. C'est à cet effet, que Glory [18] a effectué une évaluation de la qualité de la segmentation, suivant plusieurs facteurs, pour opter finalement pour l'espace RVB. Nous avons alors suivi son choix (cf. Figure II.28).

➤ **Segmentation**

Nous allons coopérer la segmentation par région et par contour: pour l'approche région nous avons opté pour la ligne de partage des eaux (cf. Figure II.9) et pour, l'approche contour nous avons choisi les ensembles de niveaux (cf. §II.3.2), qui prennent en compte les contours actifs actionnés par l'énergie de l'image locale appropriée.

➤ **Caractérisation**

Cette étape, permet la caractérisation texturale et topologique des différentes structures mammaires, en vue de les classifier. Les paramètres que nous avons choisis, permettent de bien discriminer les classes étudiées.

Ils sont au nombre de six descripteurs (L'entropie, l'écart type, le contraste, l'énergie, corrélation et l'homogénéité) pour la texture et de deux facteurs pour la topologie (la compacité et l'excentricité).

➤ **Classification**

La classification des structures analysées, est faite avec le perceptron multicouche, pour l'aide à la décision.

IV.3 Présentation de l'application cytocoop

Cytocoop est un logiciel de traitement d'image qui a pour but de classifier les images cytologiques couleurs en se basant sur la coopération séquentielle de la ligne de partage des eaux et des ensembles des niveaux, pour profiter des informations des formes et des régions, en vue d'une classification connexionniste des données caractérisées.

➤ **Interface du logiciel**

La fenêtre principale est présentée par la figure suivante:

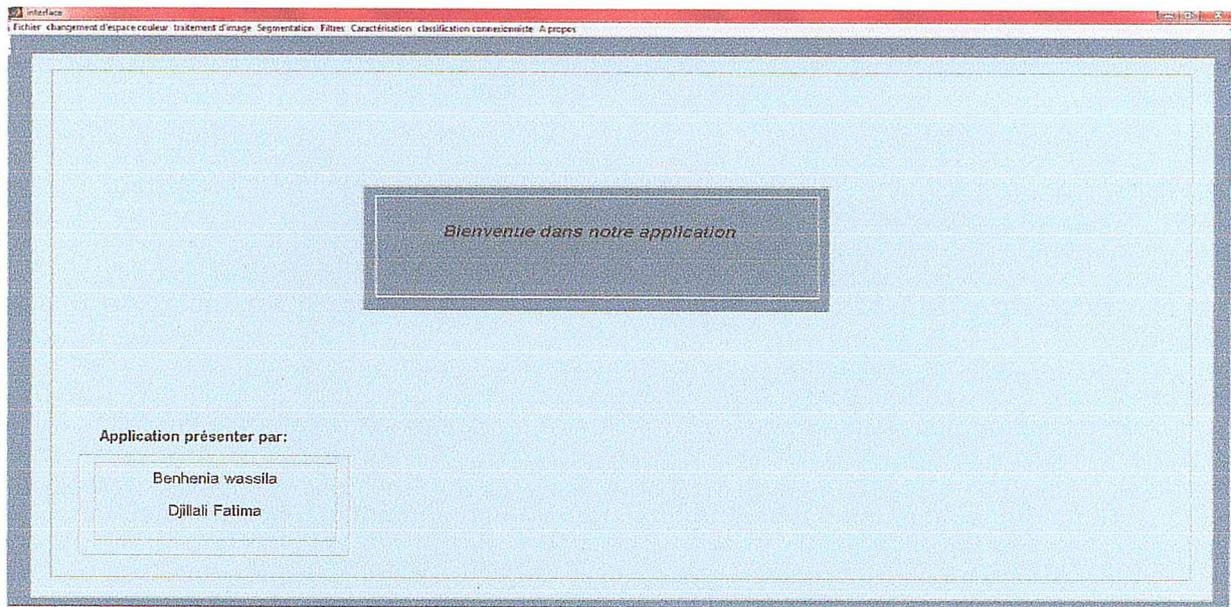


Figure IV.3: fenêtre principale

Dans la fenêtre principale, se trouve la barre du menu nécessaire à notre application. Nous décrivons dans ce qui suit les différents menus:

1. le menu « fichier »

Il contient quatre commandes (cf. Figure IV.4):

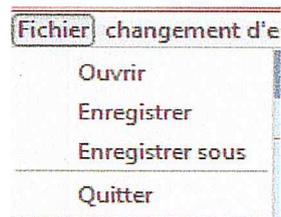


Figure IV.4 : menu « fichier »

- ✓ **Ouvrir** : l'exécution de cette commande, permet d'afficher une image de tout types de format(bmp, jpg, png... etc),
- ✓ **Enregistrer** : sauvegarder une image,
- ✓ **Enregistrer sous:** permet de sauvegarder une image sous un nom et dans un répertoire choisi par l'utilisateur,
- ✓ **Quitter** : cette commande permet de fermer l'application.

2. le menu « changement d'espace couleur »

Il contient quatre espace couleurs : HSV, YCbCr, XYZ, l*a*b*, suivant la figure ci-dessous:

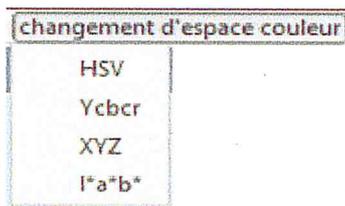


Figure IV.5 : menu « changement d'espace couleur »

3. le menu « prétraitement »

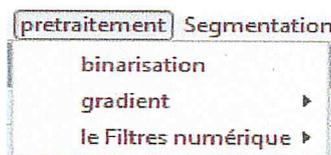


Figure IV.6 : menu « prétraitement »

Il contient trois sous menus : binarisation, le gradient et les filtres numériques suivant les figures IV.7 et IV.8.

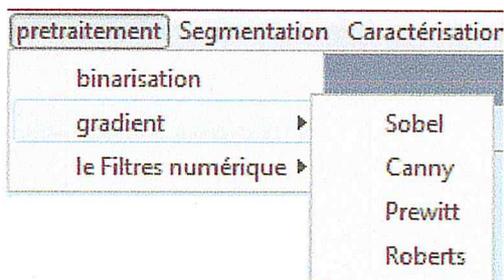


Figure IV.7 : sous menu « gradient »

Le sous menu 'gradient', présente les opérateurs dérivatifs de : Sobel, Canny, Prewitt, et Roberts.

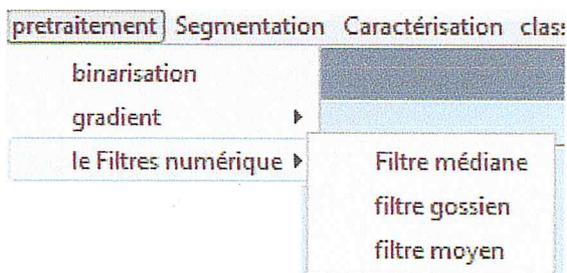


Figure IV.8 : sous menu « les filtres numériques »

Le sous menu 'les filtres numériques' contient trois opérateurs, le filtre médian, le gaussien et le moyen.

4. le menu « segmentation »

Il est composé de la segmentation par la ligne de partage des eaux, de la segmentation par les ensembles de niveaux et de la segmentation coopérative (cf. Figure IV.9).

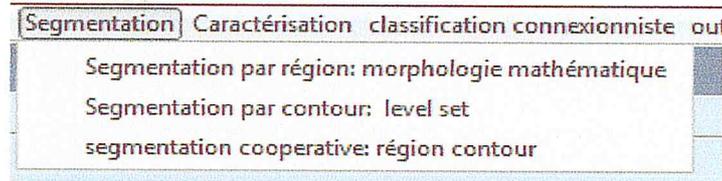


Figure IV.9 : menu « segmentation »

5. le menu « caractérisation »

Ce menu (cf. Figure IV.10) présente les différents descripteurs extraits dans l'application.

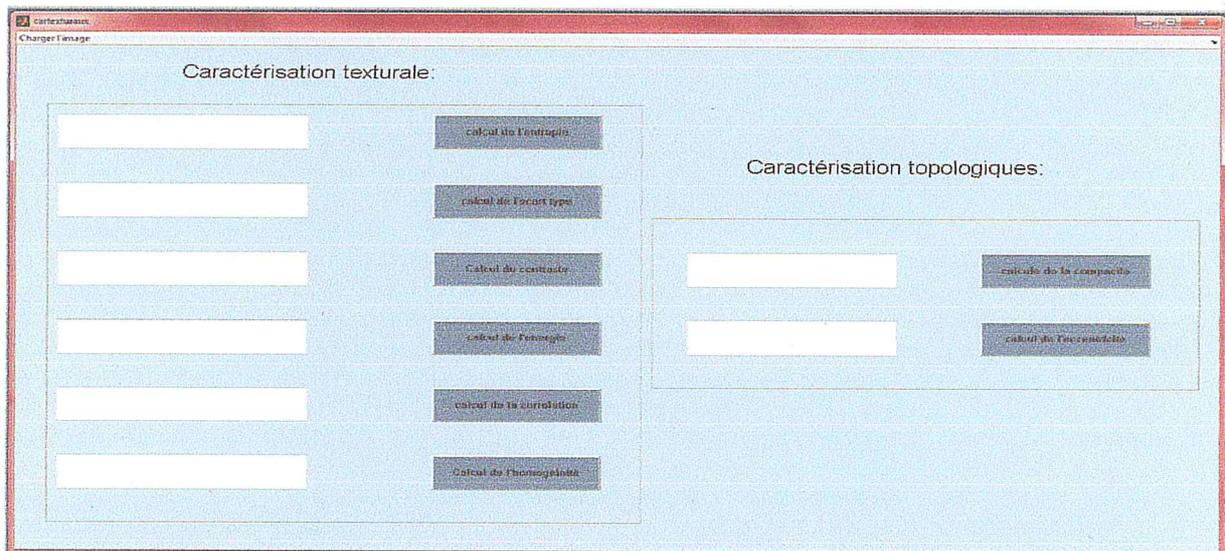
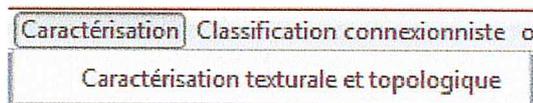


Figure IV.10 : menu « caractérisation texturale et topologique »

6. le menu « classification connexionniste »

Ce menu (cf. Figure IV.11) contient deux sous menus: la phase d'apprentissage supervisé suivant le perceptron multicouches et la classification.



Figure IV.11 : menu « l'apprentissage »

L'apprentissage comprend à son tour, trois sous menus. Nous avons effectué nos tests sur des images coopératives, segmentées avec la morphologie mathématique et avec les ensembles de niveaux. Dès qu'on clique sur l'un des sous menus, une petite interface apparait et nous montre la structure de notre réseau, le nombre d'itérations, ses performances,...etc. voici un exemple :

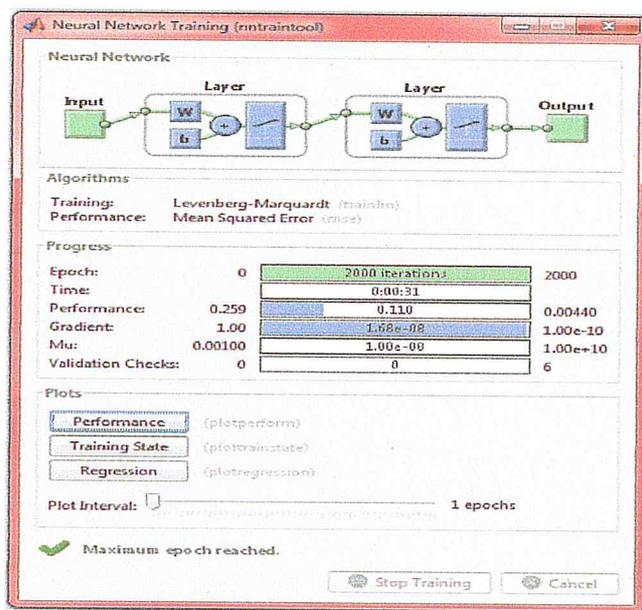


Figure IV.12 : structure du réseau

Après l'apprentissage viens l'étape de la classification connexionniste (cf. Figures IV.12 et IV.13). C'est à ce niveau, que nous effectuons les différents tests sur des images issues de la segmentation coopérative, des images segmentées avec la ligne de partage des eaux et enfin avec celles produites par les ensembles de niveaux.

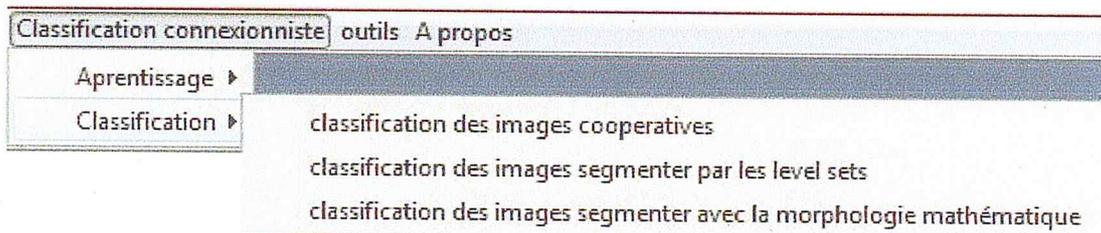


Figure IV.13 : menu « classification connexionniste »

Quand l'un des menus est exécuté, une fenêtre apparait comme la figure ci-dessous :

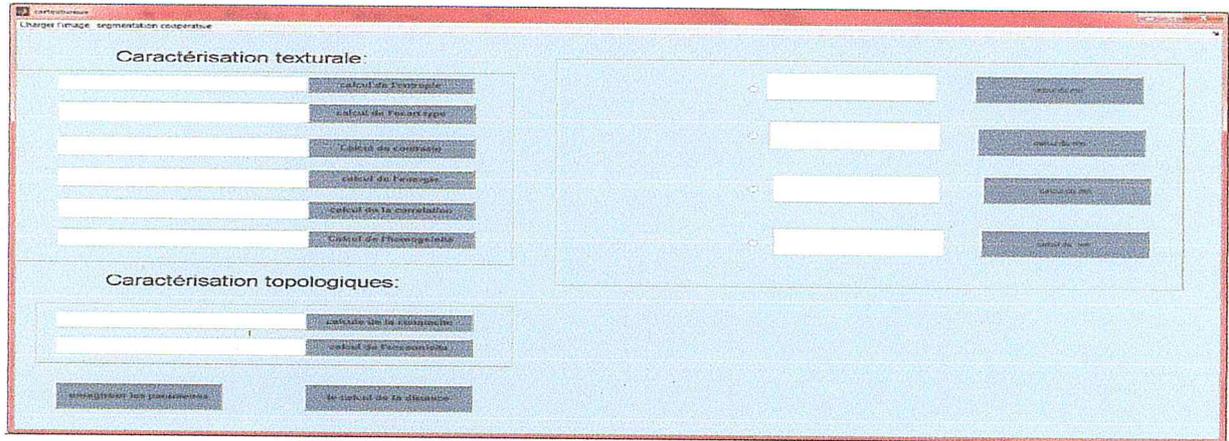


Figure IV.14 : sous menu « classification »

7. Le menu « outils »

Le menu (cf. Figure IV.14) outil six sous menus: l'image en niveau de gris, son histogramme, l'image avec chaque composante de l'espace RGB, l'historgramme des trois bandes de l'espace RGB, l'égalisation des histogrammes des composants de l'espace RVB, ainsi que les opérateurs classiques de détection de contours.

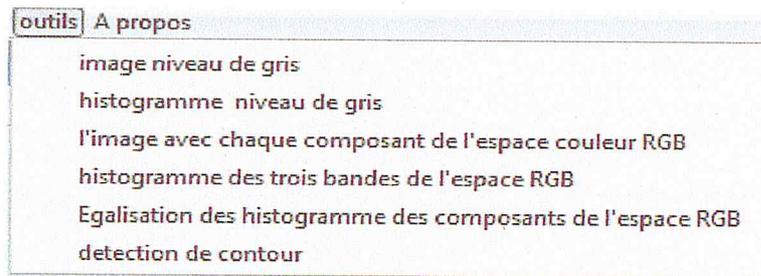


Figure IV.15 : menu « outils »

8. le menu « A propos »

Il affiche une fenêtre (cf. Figure IV.15) contenant les informations de l'application cytocoop.



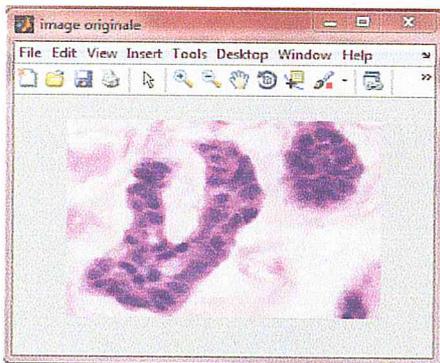
Figure IV.16 : menu « A propos »

IV.4 Présentation des résultats et discussion

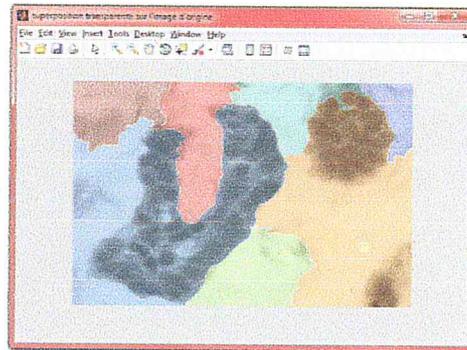
L'application 'cytocoop', nous permet de produire la carte des régions d'intérêt, par la ligne de partage des eaux, la carte des contours suivant une méthode évolutive, qui permet de déterminer le contour exact des différentes cellules pathologiques étudiées. Le système engendre aussi la coopération des deux approches, pour profiter des résultats de chaque approche. La discrimination est faite suivant les images issues de la coopération et les images obtenues par chaque approche prise séparément.

IV.4.1 Extraction des régions d'intérêt

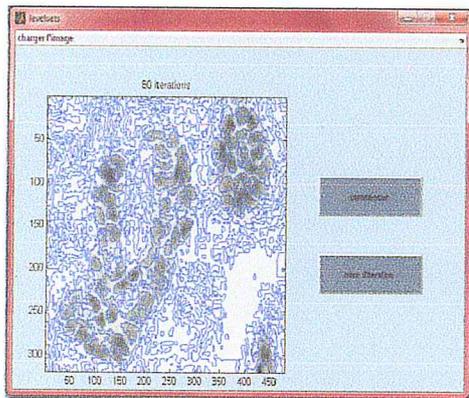
Les figures suivantes, illustrent les résultats des approches étudiées de la segmentation sur des carcinomes, des calcifications bénignes et des calcifications malignes:



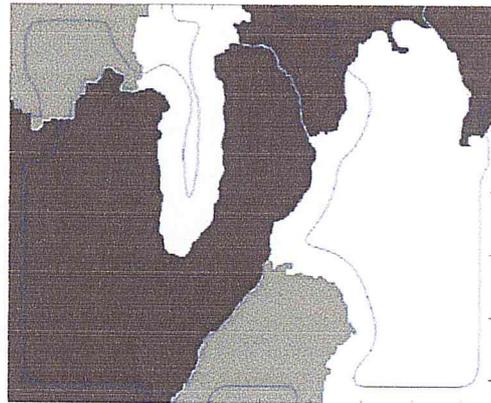
a : carcinomes



b : résultat de la LPE

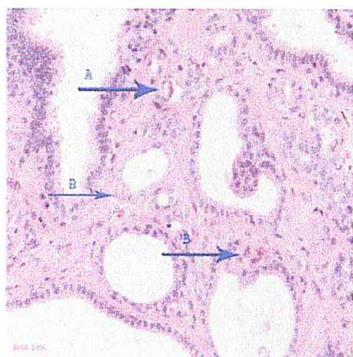


c : résultat des ensembles de niveaux

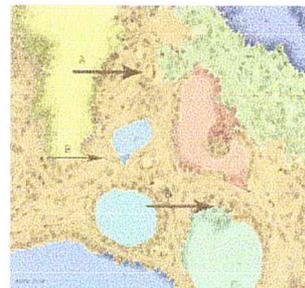


d : résultat de la segmentation coopérative

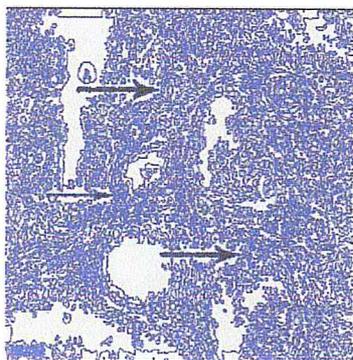
Figure IV.17 : segmentation d'une image présentant des carcinomes



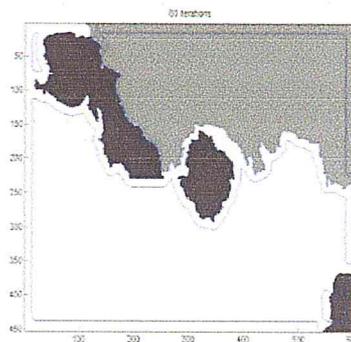
a : calcifications bénignes



b : résultat de la LPE (calcifications bénignes)

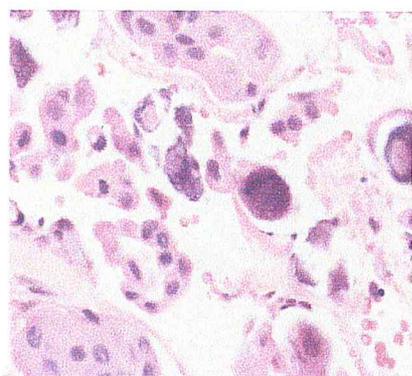


c : résultat de l'ensemble des niveaux

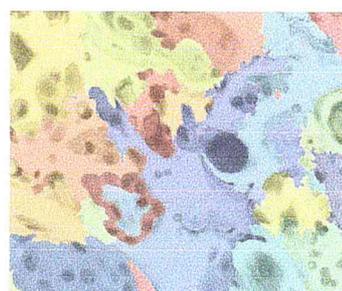


d : résultat de la segmentation coopérative

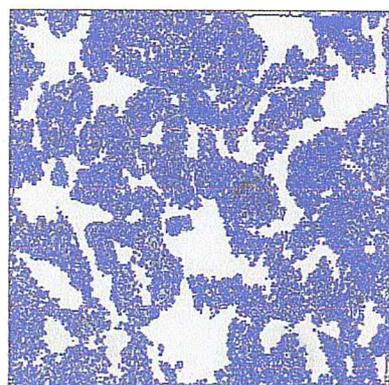
Figure IV.18 : segmentation d'une image présentant des calcifications bénignes



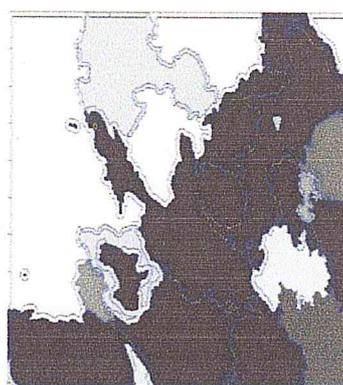
a : calcifications malignes



b : résultat de la LPE



c : résultat de l'ensemble des niveaux



d : résultat de la segmentation coopérative

Figure IV.19 : segmentation d'une image présentant des calcifications malignes

Nous obtenons enfin par la ligne de partage des eaux, l'image correspondant à l'application qui nous définit les différentes régions de l'image comme si il s'agissait d'une carte topographique. Nous remarquons d'ailleurs, sur les trois types d'images rencontrées, la délimitation des régions d'intérêt.

Selon, la méthode des ensembles des niveaux, l'évolution de la courbe suivant les énergies, a bien parcouru les différentes formes autour de 80 itérations.

IV.4.2 Caractérisation texturale et topologique

Le tableau suivant illustre les descripteurs texturaux et topologiques de huit images de la base de données cytologiques, où la classe 1 représente des carcinomes, la classe 2, illustre des cas sains, la classe 3, parcourt des calcifications bénignes et la classe 4, comprend des calcifications malignes.

Image N°	Entropie	Écart type	Homogénéité	contraste	Énergie	corrélation	
Classe1	1	4.48132	0.280234	0.801498	0.299742	0.89887	0.93533
	2	3.49928	0.385433	0.805302	0.386768	0.946094	0.969733
Classe2	3	1.93255	0.338216	0.318932	0.678	0.971533	0.982804
	4	3.21269	0.456031	0.899005	0.418361	0.957056	0.964821
Classe3	5	2.52821	0.346246	0.470062	0.41639	0.958236	0.978581
	6	2.48218	0.483098	0.419681	0.439225	0.981937	0.982114
Classe4	7	1.23421	0.258892	0.284739	0.956677	0.956677	0.98799
	8	1.86474	0.43739	0.375204	0.58104	0.980189	0.985838

Tableau IV.1 : descripteurs texturaux

Image N°	Compacité	Excentricité	
Classe1	1	0.996435	0.678876
	2	0.991384	0.939515
Classe2	3	0.99088	0.869202
	4	0.993752	0.638893
Classe3	5	0.974182	0.999998
	6	0.979177	0.999999
Classe4	7	0.986373	0.87956
	8	0.978771	0.913788

Tableau IV.2 : descripteurs topologiques

Nous remarquons dans le cas de ces descripteurs, que le désordre est plus important pour les carcinomes, qui présentent des cellules tumorales à un stade élevé relativement aux calcifications qui le sont moins.

L'extraction de paramètres texturaux et topologiques, nous permet de constituer des vecteurs caractéristiques de chacune des formes étudiées, en vue d'une classification, pour une prise de décision.

IV.4.3 La classification connexionniste

Dans notre application, nous mettons en œuvre un réseau de neurones multicouches avec une seule couche cachée (qui est un classificateur universel) dont la structure et le fonctionnement est décrit dans ce qui suit. Le classificateur est accompli par la rétropropagation du gradient de l'erreur. Le paramètre de sortie Y_k est calculé comme :

$$Y_k = f\left(\sum_{j=1}^J W'_{jk} X'_j - \theta'_k\right), \quad X'_j = f\left(\sum_{i=1}^I W_{ij} X_i - \theta_j\right),$$

Où $k = 1, 2, \dots, K$ $j = 1, 2, \dots, J$

Avec $f(.)$; fonction sigmoïde Dans les équations ci-dessus, θ_j et θ'_k sont les noeuds de compensation; W'_{jk} et W_{ij} sont les noeuds des poids, $(X, i = 1, 2, \dots, I)$ sont les paramètres d'entrée des pathologies mammaires, et I, J et K sont les nombres des noeuds à la couche d'entrée, à la

couche cachée, et aux couches de sortie, respectivement. Les poids et les compensations sont mis à jour par :

$$W'_{jk}(t+1) = W'_{jk}(t) + \eta[Y_k(1 - Y_k)(O_k - Y_k)]X'_j + \alpha[W'_{jk}(t) - W'_{jk}(t-1)],$$

$$\theta'_k(t+1) = \theta'_k(t) + \eta[Y_k(1 - Y_k)(O_k - Y_k)](-1) + \alpha[\theta'_k(t) - \theta'_k(t-1)],$$

$$W_{ij}(t+1) = W_{ij}(t) + \eta[X'_j(1 - X'_j) \sum_{k=1}^K \{Y_k(1 - Y_k)(O_k - Y_k)W'_{jk}\}]X_i + \alpha[W_{ij}(t) - W_{ij}(t-1)],$$

$$\theta_j(t+1) = \theta_j + \eta[X'_j(1 - X'_j) \sum_{k=1}^K \{Y_k(1 - Y_k)(O_k - Y_k)W'_{jk}\}](-1) + \alpha[\theta_j(t) - \theta_j(t-1)],$$

Où O_k sont les sorties désirées, α est un terme de dynamique (momentum), et η est le taux d'apprentissage. L'algorithme est répété jusqu'à ce que l'erreur absolue entre les résultats escomptés et les résultats réels de la formation des données, soit inférieure à un seuil prédéterminé (0,1 dans cette étude). L'algorithme utilisé pour l'apprentissage du réseau de neurones dans le cas de notre étude, est l'algorithme de rétropropagation du gradient d'erreur avec un terme supplémentaire appelé "momentum"

Le principe fondamental du réseau de neurones choisi, est le suivant : le réseau est "entraîné" pour associer un modèle d'entrées "vecteur d'entrées" à un modèle de sortie "vecteur de sortie". Pendant la phase d'apprentissage, les deux modèles d'entrée et de sortie sont présentés au réseau de neurones afin de lui permettre d'apprendre et de mémoriser l'allure de ces deux vecteurs.

Les principales étapes de l'apprentissage du réseau de neurones étudié, pour la conception de notre système de classification sont données par le synoptique suivant ;

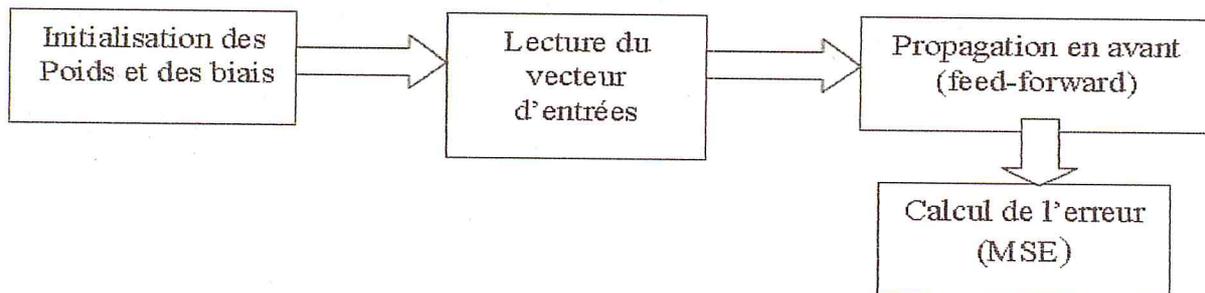


Figure IV.20 : principales étapes de l'apprentissage

Les principales étapes de l'apprentissage sont :

- Etape 1 : Initialisation des poids ;
- Etape 2 : Lecture du vecteur d'entrées et de la sortie désirée ;
- Etape 3 : Propagation en avant ou " feed-forward "
- Etape 4 : Calcul d'erreurs (MSE) : Erreur quadratique moyenne.

Discussion

Après un ensemble de tests effectués, en utilisant le Matlab 7.9, nous avons opté pour les deux paramètres (η et α) adoptant respectivement les valeurs $\eta = 0.2$ et $\alpha = 0.1 \cdot 10^{-3}$. En effet, on a constaté un meilleur taux de reconnaissance par rapport au choix d'autres valeurs des

paramètres. Le réseau de neurones adopté, est le réseau de neurones avec huit neurones dans la couche cachée.

➤ Résultats obtenus durant la phase d'entraînement

L'apprentissage est l'étape essentielle dans les systèmes de classification. Les résultats corrects (c'est-à-dire les valeurs que l'on désire que le réseau obtienne en sortie) sont fournis au réseau, si bien que celui-ci peut ajuster ses poids de connexions pour les obtenir. Après l'apprentissage, le réseau est testé en lui donnant seulement les valeurs d'entrée mais pas les sorties désirées, et en regardant si le résultat obtenu est proche du résultat désiré.

L'entraînement se fait avec une seule couche cachée, de 8 neurones et le nombre d'itérations est de 2000 avec un pas de 8.

➤ Lecture du vecteur d'entrées et de la sortie désirée

Le vecteur d'entrée se composant par les caractéristiques de forme et de texture discutées précédemment, sont employés dans l'entrée ($X_i, i = 1, 2, 3, 4$), et la sortie du réseau représente l'une des classes étudiées (calcifications bénignes, calcifications malignes, carcinomes et cas sains).

On donne : [1000]= classe1 (classe des carcinomes), [0100]= classe2 (classe des cas sains), [0010]=classe3 (classe des calcifications bénignes), [0001]=classe4 (classe des calcifications malignes).

On classe les images cytologiques en quatre classes.

Nous avons utilisé les deux tiers des vecteurs comme entrée du réseau, pour faire l'entraînement.

La figure IV.20 représente la convergence du réseau lors de l'étape d'apprentissage (entraînement en fonction du paramètre de performance du réseau (momentum)).

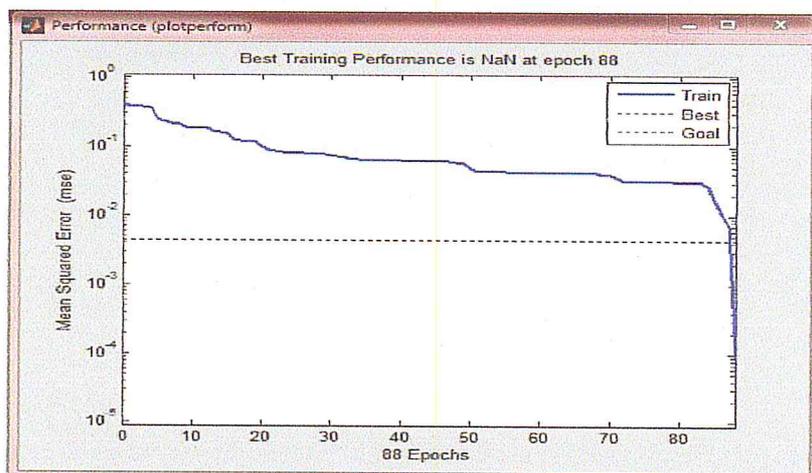


Figure IV.21 : Convergence du réseau de neurones lors de l'étape d'apprentissage

Dans ce test notre réseau a atteint une erreur quadratique moyenne de $8,64 \cdot 10^{-5}$, ce qui indique une erreur acceptable pour 88 itérations.

➤ **Résultats obtenus durant la phase de test avec les images de la segmentation coopérative**

Nous avons choisi les six premiers vecteurs de chaque classe, pour faire le test. Les résultats sont obtenus dans les tableaux suivants:

✓ *Test des vecteurs de la classe1 (classe des carcinomes)*

	Vecteur1	Vecteur2	Vecteur3	Vecteur4	Vecteur5	Vecteur6
Entropie	4.0100	3.6995	3.6195	5.1646	3.8705	3.4563
Écart type	0.3806	0.4487	0.4188	0.3699	0.2771	0.4740
contraste	0.6264	0.4718	0.5898	0.5411	0.4082	0.4111
énergie	0.9571	0.9768	0.9672	0.9620	0.9495	0.9818
corrélation	0.3614	0.3475	0.4049	0.2198	0.3622	0.3917
homogénéité	0.9559	0.9620	0.9637	0.9410	0.9556	0.9695
compacité	0.9897	0.9893	0.9863	0.9936	0.9875	0.9845
excentricité	1.0000	0.7885	1.0000	1.0000	1.0000	0.7812

Tableau IV.3 : vecteurs de la classe 1

Le réseau fait la simulation et classe les vecteurs de test de la classe1 : 'classe des carcinomes' selon le tableau IV.4 :

	Vecteur1	Vecteur2	Vecteur3	Vecteur4	Vecteur5	Vecteur6
Classe1	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998
Classe2	0.0113	0.0000	0.0110	0.0001	0.0001	0.0000
Classe3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0006
Classe4	0.0000	0.0048	0.0282	0.0000	0.0000	0.0000

Tableau IV.4 : sortie de test de la classe 1

Nous remarquons que tous les vecteurs sont classés dans la classe1, donc :

- Taux d'erreur=0%
- Taux de bonne classification = 100%

Taux d'erreur=nombre de cas mal classés / nombre total des cas

✓ *Test des vecteurs de la classe 2, (cas sain)*

	Vecteur1	Vecteur2	Vecteur3	Vecteur4	Vecteur5	Vecteur6
Entropie	3.8882	2.0236	3.3886	3.1537	2.2262	5.2820
Écart type	0.4494	0.1355	0.4438	0.2254	0.4667	0.3944
contraste	0.6881	0.1622	0.3391	0.3350	0.1700	0.5696

énergie	0.9662	0.9213	0.9824	0.9318	0.9921	0.9621
corrélation	0.3846	0.7560	0.3150	0.5097	0.5224	0.1840
homogénéité	0.9477	0.9740	0.9721	0.9546	0.9858	0.9341
compacité	0.9934	0.9529	0.9655	0.9888	0.9334	0.9949
excentricité	1.0000	0.5904	1.0000	0.8966	1.0000	1.0000

Tableau IV.5 : vecteurs de la classe 2

Le réseau fait la simulation et classe les vecteurs de test de la classe 2 : 'classe des cas sains' selon le tableau IV.6 :

	Vecteur1	Vecteur2	Vecteur3	Vecteur4	Vecteur5	Vecteur6
Classe1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
Classe2	1.0000	0.9998	0.9766	0.9793	0.9301	0.9886
Classe3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Classe4	0.0001	0.0000	0.0000	0.0000	0.0000	0.0105

Tableau IV. 6 : sortie de test de la classe 2

Nous remarquons que les 6 vecteurs sont classés dans la classe 2, ce qui indique alors :

- Taux d'erreur=0%
- Taux de bonne classification = 100%

✓ Test des vecteurs de la classe3, (calcification bénignes)

	Vecteur1	Vecteur2	Vecteur3	Vecteur4	Vecteur5	Vecteur6
Entropie	3.0226	3.7371	3.0597	2.8444	3.0036	3.4326
Écart type	0.4792	0.4163	0.3910	0.4663	0.4690	0.3591
contraste	0.3773	0.2788	0.2727	0.3671	0.3449	0.3727
énergie	0.9835	0.9841	0.9826	0.9831	0.9842	0.9707
corrélation	0.4209	0.2948	0.4240	0.4919	0.4033	0.4670
homogénéité	0.9726	0.9680	0.9734	0.9722	0.9733	0.9620
compacité	0.9811	0.9784	0.9738	0.9800	0.9872	0.9849
excentricité	0.8839	0.8393	0.7593	1.0000	0.6201	1.0000

Tableau IV. 7 : vecteurs de la classe 3

Le réseau fait la simulation et classe les vecteurs de test de la classe3 : 'classe des calcifications bénignes' selon le tableau IV.8 :

	Vecteur1	Vecteur2	Vecteur3	Vecteur4	Vecteur5	Vecteur6
Classe1	0.0000	0.0030	0.0000	0.0000	0.0000	0.0000
Classe2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Classe3	0.9801	0.9992	0.9974	0.9918	0.9998	0.9922
Classe4	0.0000	0.0000	0.0017	0.0000	0.0000	0.0000

Tableau IV. 8 : sortie de test de la classe 3

Nous remarquons que les 6 vecteurs sont classés dans la classe 3, ce qui indique alors :

- Taux d'erreur=0%
- Taux de bonne classification = 100%

✓ *Test des vecteurs de la classe4, (calcification malignes)*

	Vecteur1	Vecteur2	Vecteur3	Vecteur4	Vecteur5	Vecteur6
Entropie	3.4178	4.5073	3.6121	3.5273	3.3254	2.5364
Écart type	0.4105	0.4442	0.4804	0.3524	0.4324	0.4511
contraste	0.2834	0.5656	0.5949	0.3379	0.3777	0.4555
énergie	0.9825	0.9716	0.9743	0.9730	0.9797	0.9775
corrélation	0.3919	0.2833	0.4077	0.3044	0.3238	0.5331
homogénéité	0.9743	0.9443	0.9570	0.9602	0.9687	0.9755
compacité	0.9771	0.9925	0.9898	0.9787	0.9765	0.9846
excentricité	0.6478	1.0000	1.0000	0.5484	0.7934	0.4596

Tableau IV.9 : vecteurs de la classe 4

Le réseau fait la simulation et classe les vecteurs de test de la classe3 : 'classe des calcifications malignes' selon le tableau IV.10 :

	Vecteur1	Vecteur2	Vecteur3	Vecteur4	Vecteur5	Vecteur6
Classe1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Classe2	0.0000	0.0036	0.0003	0.0161	0.0000	0.0082
Classe3	0.0151	0.0000	0.0000	0.0000	0.0000	0.0000
Classe4	1.0000	1.0000	0.9999	1.0000	0.9968	0.9974

Tableau IV.10 : sortie de test de la classe 4

Nous remarquons que les 6 vecteurs sont classés dans la classe 4, ce qui indique alors :

- Taux d'erreur=0%
- Taux de bonne classification = 100%
- Temps d'exécution du réseau: 0:00:02

✓ **Résultat des tests avec les images segmentées par les ensembles de niveaux**

Nous avons utilisé les deux tiers des vecteurs comme entrée du réseau, pour la phase de l'entraînement.

La figure IV.21 représente la convergence du réseau lors de l'étape d'apprentissage (entraînement en fonction du paramètre de performance du réseau (momentum)).

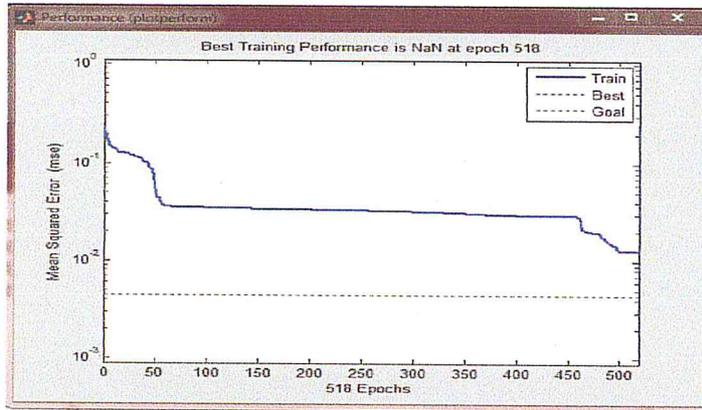


Figure IV.22 : convergence du réseau de neurones lors de l'étape d'apprentissage

Dans ce test, notre réseau a atteint une erreur quadratique moyenne de 0,0127, ce qui indique une erreur acceptable pour 518 itérations.

➤ **Résultats obtenus durant la phase de ce test**

Nous avons choisi les quatre premiers vecteurs pour faire le test, les résultats sont obtenus dans les tableaux suivants:

✓ **Test des vecteurs de la classe1 (classe des carcinomes)**

	Vecteur1	Vecteur2	Vecteur3	Vecteur4
Entropie	6.6061	6.7812	6.6675	6.9685
Ecart type	0.3744	0.3632	0.2689	0.2998
contraste	10.7403	10.2754	4.4299	5.1296
Energie	0.2337	0.1250	0.4055	0.4201
Corrélation	0.1026	0.1247	0.1947	0.1971
Homogénéité	0.5939	0.5852	0.7178	0.7272
Compacité	0.9997	0.9997	0.9991	0.9994
excentricité	1.0000	1.0000	1.0000	0.6161

Tableau IV. 11 : vecteurs de la classe 1

Le réseau fait la simulation et classe les vecteurs de test de la classe 1 : 'classe des carcinomes' selon le tableau IV.12 :

	Vecteur1	Vecteur2	Vecteur3	Vecteur4
Classe1	1.0000	1.0000	1.0000	1.0000
Classe2	0.0000	0.0000	0.0000	0.0000
Classe3	0.0000	0.0000	0.0000	0.0000
Classe4	0.0000	0.0000	0.0000	0.0000

Tableau IV.12 : test des vecteurs de la classe 1

Nous remarquons que tous les vecteurs sont classés dans la classe1, donc :

- Taux d'erreur=0%
- Taux de bonne classification = 100%

Taux d'erreur=nombre de cas mal classés / nombre total des cas

✓ *Test des vecteurs de la classe2, (cas sain)*

	Vecteur1	Vecteur2	Vecteur3	Vecteur4
Entropie	5.9508	6.5166	6.7806	5.7340
Ecart type	0.4107	0.4054	0.3638	0.3323
contraste	7.7172	6.4852	8.1883	4.4388
Energie	0.5555	0.3395	0.3580	0.6235
Corrélation	0.2509	0.2782	0.1109	0.2245
Homogénéité	0.7653	0.7134	0.6388	0.7919
Compacité	0.9993	0.9996	0.9997	0.9987
excentricité	1.0000	1.0000	0.8609	1.0000

Tableau IV.13: vecteurs de la classe 2

Le réseau fait la simulation et classe les vecteurs de test de la classe 2: 'classe des cas sains' selon le tableau IV.14 :

	Vecteur1	Vecteur2	Vecteur3	Vecteur4
Classel	0.0000	0.0000	1.0000	0.0000
Classe2	1.0000	1.0000	1.0000	1.0000
Classe3	0.0000	0.0000	0.0000	0.0000
Classe4	0.0000	0.0000	0.0000	0.0000

Tableau IV.14 : Test des vecteurs de la classe 2

Nous remarquons que tous les vecteurs sont classés dans la classel, donc :

- Taux d'erreur=0%
- Taux de bonne classification = 100%

✓ *Test des vecteurs de la classe3, (calcification bénignes)*

	Vecteur1	Vecteur2	Vecteur3	Vecteur4
Entropie	6.7720	6.6889	6.3283	6.6817
Ecart type	0.3740	0.3421	0.2718	0.3503
contraste	7.3070	5.2657	3.9113	5.4995
Energie	0.3886	0.5089	0.4619	0.5680
Corrélation	0.1699	0.2020	0.4132	0.1212
Homogénéité	0.6893	0.7489	0.8185	0.7372
Compacité	0.9996	0.9995	0.9987	0.9994
excentricité	1.0000	0.8611	0.8660	1.0000

Tableau IV.15 : vecteurs de la classe 3

Le réseau fait la simulation et classe les vecteurs de test de la classe 3: 'classe des calcifications bénignes' selon le tableau IV.16 :

	Vecteur1	Vecteur2	Vecteur3	Vecteur4
Classe1	0.0000	0.0000	0.0000	0.0000
Classe2	0.0000	0.0000	0.0000	0.0000
Classe3	1.0000	1.0000	1.0000	1.0000
Classe4	0.0000	1.0000	0.0000	0.0000

Tableau IV.16 : Test des vecteurs de la classe 3

Nous remarquons que tous les vecteurs sont classés dans la classe1, donc :

- Taux d'erreur=0%
- Taux de bonne classification = 100%

✓ *Test des vecteurs de la classe 4 (calcification malignes)*

	Vecteur1	Vecteur2	Vecteur3	Vecteur4
Entropie	6.7250	6.5876	6.6786	6.9336
Ecart type	0.2948	0.3957	0.3731	0.2678
contraste	5.1543	5.2156	7.3060	3.9238
Energie	0.4022	0.5704	0.4102	0.4593
Corrélation	0.2912	0.2315	0.1595	0.2037
Homogénéité	0.7608	0.7682	0.6863	0.7513
Compacité	0.9993	0.9991	0.9995	0.9991
excentricité	0.6152	0.8660	0.6580	0.8660

Tableau IV. 17 : vecteurs de la classe 4

Le réseau fait la simulation et classe les vecteurs de test de la classe 4: 'classe des calcifications malignes' selon le tableau IV.18 :

	Vecteur1	Vecteur2	Vecteur3	Vecteur4
Classe1	0.0000	0.0000	0.0000	0.0000
Classe2	0.0000	0.0000	0.0000	0.0000
Classe3	0.0000	0.0000	0.0000	0.0000
Classe4	1.0000	1.0000	1.0000	1.0000

Tableau IV. 18: test des vecteurs de la classe 4

Nous remarquons que tous les vecteurs sont classés dans la classe1, donc :

- Taux d'erreur=0%
 - Taux de bonne classification = 100%
 - Temps d'exécution: 0:00:08
- **Résultat des tests avec les images segmenté par la ligne de partage des eaux**
Nous avons utilise les deux tiers de nos vecteurs comme entrée du réseau, pour l'apprentissage.

La figure IV.22 représente la convergence du réseau lors de l'étape d'apprentissage (entraînement en fonction du paramètre de performance du réseau (momentum)).

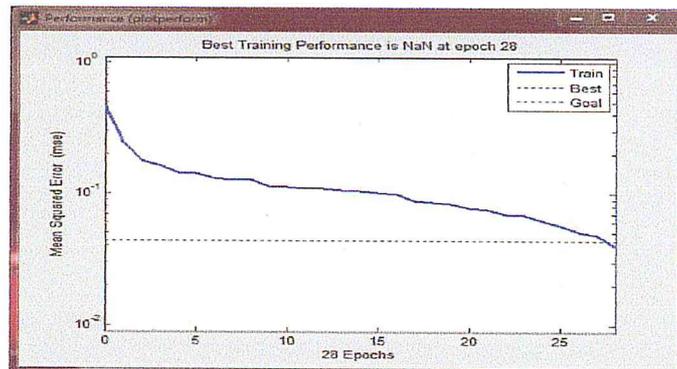


Figure IV.23 : convergence du réseau de neurones lors de l'étape d'apprentissage

Dans ce test, notre réseau a atteint une erreur quadratique moyenne de 0,0397, ce qui indique une erreur acceptable pour 28 itérations.

✓ *Test des vecteurs de la classe1 (classe des carcinomes)*

	Vecteur1	Vecteur2	Vecteur3	Vecteur4	Vecteur5	Vecteur6
Entropie	7.0542	7.3017	7.1056	6.9143	6.7643	6.8034
Ecart type	0.1363	0.1994	0.1855	0.1529	0.1507	0.1696
contraste	0.0983	0.0483	0.1669	0.1349	0.0847	0.1436
Energie	0.3915	0.2162	0.3453	0.2899	0.5978	0.3235
Corrélation	0.8843	0.9853	0.9153	0.8767	0.7860	0.8846
Homogénéité	0.9532	0.9776	0.9356	0.9352	0.9578	0.9361
Compacité	0.9989	0.9630	0.9962	0.9978	0.9981	0.9887
excentricité	0.7018	0.7767	0.6254	0.7906	0.8157	0.7507

Tableau IV.19 : vecteurs de la classe 1

Le réseau fait la simulation et classe les vecteurs de test de la classe 1: 'classe des carcinomes selon le tableau IV.20 :

	Vecteur1	Vecteur2	Vecteur3	Vecteur4	Vecteur5	Vecteur6
Classe1	0.9780	0.8482	0.8462	0.8696	0.9943	0.6458
Classe2	0.0000	0.0620	0.2072	0.0001	0.0000	0.0028
Classe3	0.0422	0.0726	0.0022	0.0000	0.0348	0.0000
Classe4	0.0075	0.0097	0.0402	0.3995	0.0055	0.2640

Tableau IV.20 : test des vecteurs de la classe 1

Nous remarquons que tous les vecteurs sont classés dans la classe1, donc :

- Taux d'erreur=0%
- Taux de bonne classification = 100%

✓ *Test des vecteurs de la classe2, (cas sain)*

	Vecteur1	Vecteur2	Vecteur3	Vecteur4	Vecteur5	Vecteur6
Entropie	7.0655	7.2057	7.3368	7.4501	7.2224	7.5296
Ecart type	0.1734	0.1783	0.1806	0.1872	0.1756	0.1989
contraste	0.1376	0.1371	0.1181	0.0754	0.0949	0.1273
Energie	0.3047	0.3017	0.3049	0.2178	0.2902	0.1900
Corrélation	0.9081	0.8818	0.9097	0.9707	0.9355	0.9529
Homogénéité	0.9449	0.9341	0.9415	0.9660	0.9533	0.9427
Compacité	0.9937	0.9993	0.9988	0.9891	0.9942	0.9968
excentricité	0.7376	0.7676	0.6306	0.3149	0.6859	0.8398

Tableau IV.21 : vecteurs de la classe2

Le réseau fait la simulation et classe les vecteurs de test de la classe 2: 'classe des cas sains' selon le tableau IV.22 :

	Vecteur1	Vecteur2	Vecteur3	Vecteur4	Vecteur5	Vecteur6
Classe1	0.0182	0.0037	0.0686	0.0142	0.0016	0.0117
Classe2	0.7944	0.9905	0.9502	0.9447	0.9323	0.9955
Classe3	0.0004	0.0010	0.0027	0.0064	0.0032	0.0027
Classe4	0.1639	0.1194	0.0526	0.0764	0.1325	0.0582

Tableau IV.22 : Test des vecteurs de la classe2

Nous remarquons que tous les vecteurs sont classés dans la classe2, donc :

- Taux d'erreur=0%
- Taux de bonne classification = 100%

✓ *Test des vecteurs de la classe3, (calcification bénignes)*

	Vecteur1	Vecteur2	Vecteur3	Vecteur4	Vecteur5	Vecteur6
Entropie	7.0615	7.2264	6.9472	6.8425	5.9761	6.5786
Ecart type	0.1656	0.1561	0.1468	0.1499	0.1395	0.1747
contraste	0.2040	0.1108	0.0773	0.0653	0.0487	0.0509
Energie	0.2344	0.2882	0.4324	0.3478	0.4077	0.3125
Corrélation	0.8734	0.9169	0.9133	0.9368	0.9433	0.9629
Homogénéité	0.9077	0.9478	0.9662	0.9683	0.9765	0.9750
Compacité	0.9991	0.9991	0.9974	0.9963	0.9921	0.9919
excentricité	0.8871	0.6457	0.8888	0.8464	0.4318	0.8775

Tableau IV.23 : vecteurs de la classe3

Le réseau fait la simulation et classe les vecteurs de test de la classe 3: 'classe des calcifications bénignes' selon le tableau IV.24 :

	Vecteur1	Vecteur2	Vecteur3	Vecteur4	Vecteur5	Vecteur6
Classe1	0.2037	0.0715	0.0169	0.0000	0.0001	0.0000
Classe2	0.2552	0.0259	0.0000	0.0000	0.0000	0.0001
Classe3	0.0000	0.0002	0.9747	0.9913	0.9908	0.9997

Classe4	0.2785	0.2748	0.0060	0.0816	0.1077	0.2057
---------	--------	--------	--------	--------	--------	--------

Tableau IV.24 : test des vecteurs de la classe3

✓ Test des vecteurs de la classe4, (calcification malignes)

	Vecteur1	Vecteur2	Vecteur3	Vecteur4	Vecteur5	Vecteur6
Entropie	7.1657	5.7561	7.0732	7.0709	7.1059	6.9333
Ecart type	0.1630	0.1497	0.1683	0.1542	0.1638	0.1669
contraste	0.2108	0.0742	0.2507	0.1376	0.0892	0.1344
Energie	0.3272	0.4060	0.2601	0.2376	0.2828	0.4186
Corrélation	0.7977	0.9231	0.7918	0.9208	0.9485	0.8515
Homogénéité	0.9004	0.9680	0.8883	0.9366	0.9646	0.9377
Compacité	0.9998	0.9530	0.9996	0.9978	0.9938	0.9953
excentricité	0.7305	0.6615	0.5170	0.8491	0.7228	0.4854

Tableau IV.25 : vecteurs de la classe 4

Le réseau fait la simulation et classe les vecteurs de test de la classe 4: 'classe des calcifications malignes selon le tableau IV.26 :

	Vecteur1	Vecteur2	Vecteur3	Vecteur4	Vecteur5	Vecteur6
Classe1	0.0470	0.0768	0.0000	0.0705	0.0000	0.0188
Classe2	0.0000	0.0000	0.0076	0.0005	0.0155	0.0000
Classe3	0.0000	0.0329	0.0000	0.0000	0.0130	0.0001
Classe4	0.9956	0.8536	1.0000	0.6016	0.5461	0.9946

Tableau IV.26 : test des vecteurs de la classe 4

Nous remarquons que tous les vecteurs sont classés dans la classe 4, donc :

- Taux d'erreur=0%
- Taux de bonne classification = 100%
- Temps d'exécution: 0:00:01

La répartition des classes de ces données est équidistribuée. Finalement nous pouvons dire que les résultats obtenus, sont très intéressants. En effet, notre système a atteint un taux de reconnaissance de: 100%.

Le système *cytocoop* analyse les différentes images étudiées, par une caractérisation topologique et texturale pour une classification connexionniste selon le modèle du perceptron multicouches. Nous présentons trois exemples des résultats obtenus, dans ce qui suit :

1/ cas de carcinomes

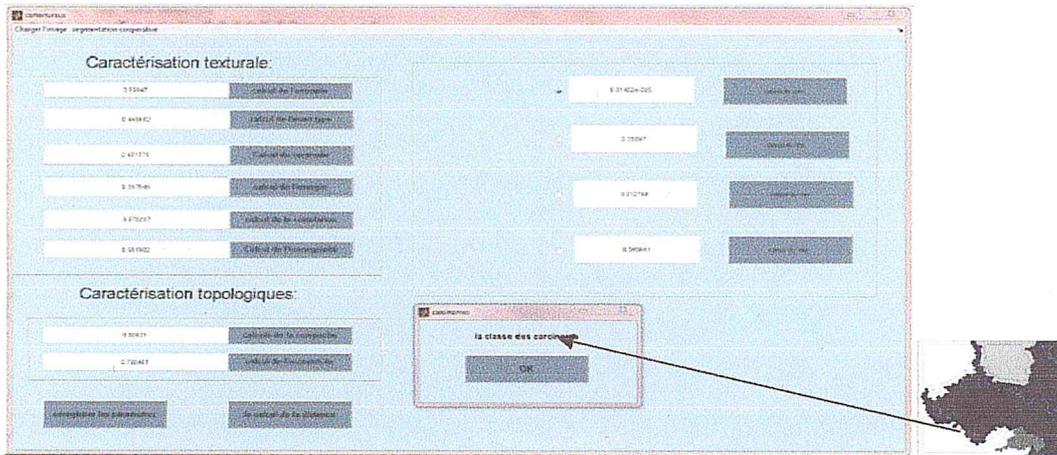


Figure IV.24 : test de la classe des carcinomes

2/ cas de cas sains

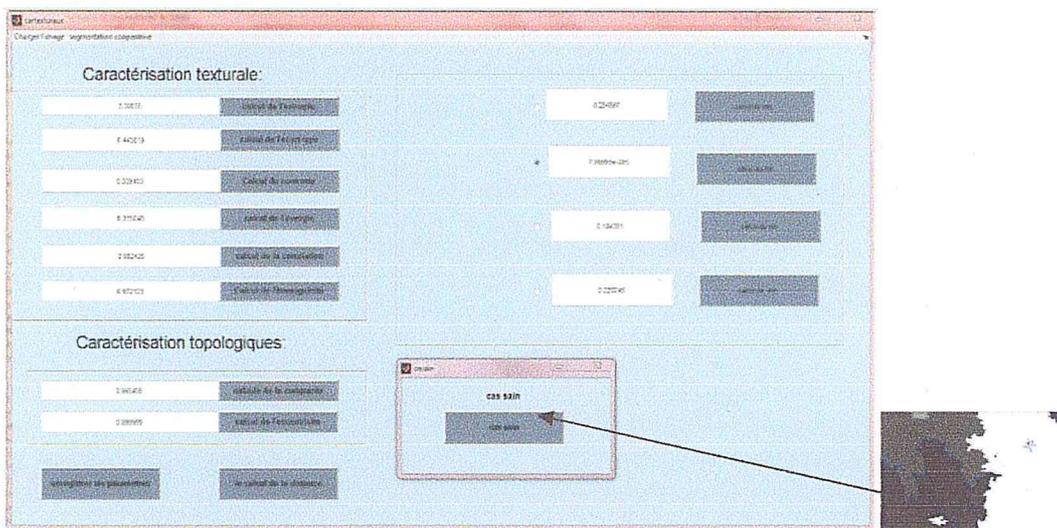


Figure IV.25 : test de la classe des cas sains

3/ cas de calcifications bénignes

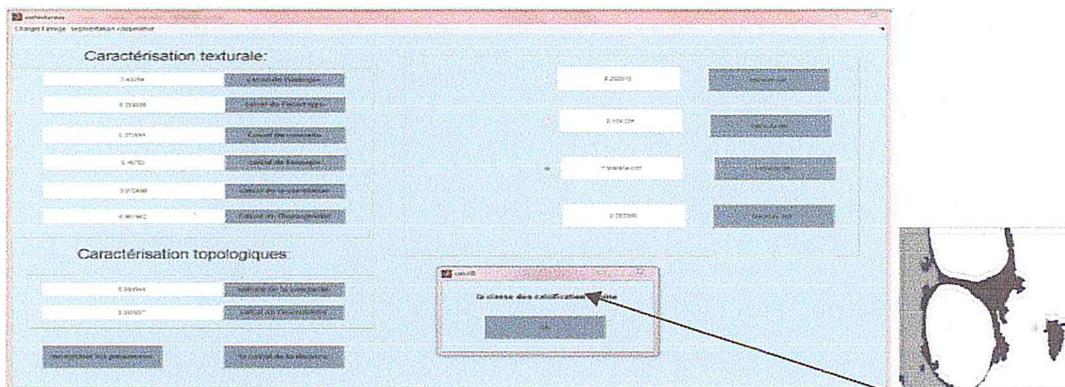


Figure IV.26 : test de la classe des calcifications bénignes

4/cas de calcifications malignes

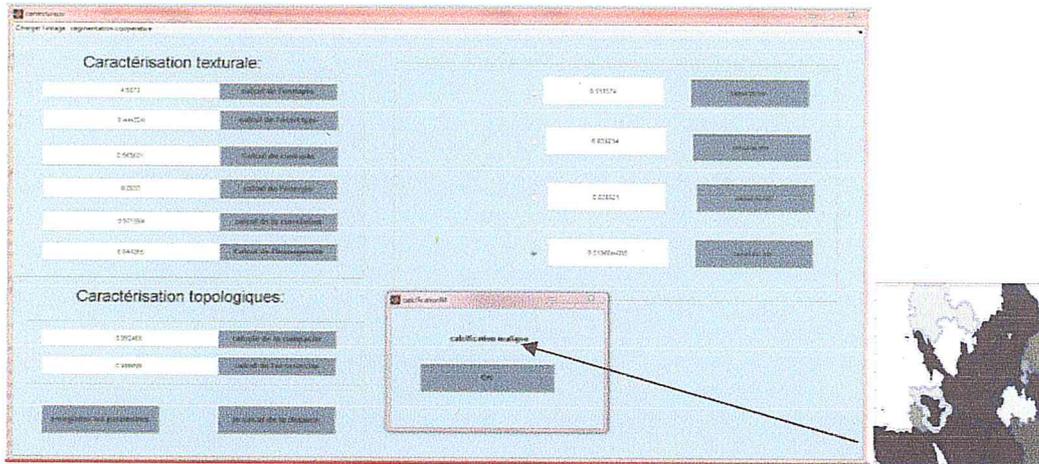


Figure IV.27 : test de la classe des calcifications malignes

IV.5 Conclusion

L'application de *Cytocoop* a engendré des résultats satisfaisants, permettant une bonne discrimination des différentes structures cytologiques mammaires étudiées. En effet, le choix des paramètres texturaux et topologiques, nous a permis d'obtenir des performances intéressantes pour classifier nos données.

La coopération de l'approche région et l'approche contour, nous a permis d'exploiter toutes les formes des zones d'intérêt, ce qui a contribué à l'exploitation de toutes caractéristiques des données.

Cette étude a confirmé la performance du perceptron multicouches, dans le domaine de la reconnaissance des pathologies mammaires, suivant un choix judicieux des attributs de caractérisation. Nous pouvons conclure que les réseaux de neurones, constituent un outil puissant permettant de résoudre le problème de classification des formes.

Conclusion générale

Notre projet a été réalisé dans le cadre d'un travail de recherche appréhendé aux laboratoires « LRDSI » (laboratoire de recherche pour le développement de systèmes informatiques) et « LATSI » (laboratoire de traitement de signal et de l'image). Son objectif principal est la détection des cellules tumorales des images cytologiques mammaires pour l'aide au diagnostic.

Une des applications principales de la cytologie au sein d'un laboratoire d'anatomie pathologique, est d'aider le cytopathologiste à caractériser au mieux les cellules qui se multiplient de manière anarchique dans les cancers. Le projet pour lequel, les chercheurs du LRDSI et de LATSI ont eu un intérêt, et a engendré ce mémoire, est la détection de cellules cancéreuses à partir d'images cytologiques couleurs mammaires pour l'aide au diagnostic. Nous avons alors, pu mettre en œuvre une méthode d'analyse d'images, basée sur une segmentation coopérative suivant la ligne de partage des eaux, pour l'extraction des régions et, les ensembles de niveaux, pour l'extraction des contours. La coopération permet ainsi, de prendre en compte les caractéristiques de toutes les régions d'intérêt.

L'étape suivante appréhendée, est l'extraction des caractéristiques qui décrivent les régions de l'image en vue d'une classification pour une prise de décision.

Il a été montré que l'analyse de forme est un moyen pratique, de caractériser les différentes anomalies mammaires rencontrées. Cette caractérisation, est faite autour de paramètres topologiques et texturaux prenant en compte la nature elle-même de la cellule.

La classification des cellules caractérisées, est faite suivant le modèle du perceptron. Nous montrons en particulier que l'algorithme du perceptron multicouche, s'avère un classifieur bien adapté pour les images couleurs. La méthode proposée, fournit de bons résultats, puisque le taux de reconnaissance est autour de 100%..

Notre classification permet de répartir nos images cytologiques, selon qu'elles sont composées de cellules saines, de carcinomes, de calcifications bénignes ou malignes.

Le processus développé peut être appliqué à divers cas de cellules cancéreuses. Le traitement d'images différentes, ne nécessitera qu'un simple ajustage des paramètres du système (taille et forme de l'élément structurant, nombre de neurones de la couche cachée dans le réseau connexionniste). L'extraction de plus de paramètres morphologiques des images cellulaires, participe à une caractérisation encore plus approfondie.

L'entreprise d'une étude de classification clinique par un autre type de réseaux de neurones d'une part et par, les modèles Bayésiens d'autre part, serait envisageable pour comparer les deux approches.

La construction du classificateur, peut être totalement adaptée à l'utilisation que l'on veut en faire. Par exemple, un médecin intéressé uniquement par des coupes provenant d'un organe donné, utilisera des images d'apprentissage caractéristiques des cellules qu'il désire traiter. Le classificateur obtenu pourra donc être très pointu et spécialisé : il suffira de lui présenter des images spécifiques en grand nombre lors de la phase d'apprentissage. Si, de plus, les images

sont très homogènes (conditions d'acquisition identiques, ...), les performances du classificateur seront encore accrues. On pourrait aussi, imaginer un praticien possédant plusieurs classificateurs, chacun étant adapté à un organe ou à un tissu particulier. Par contre, si l'on se choisit des images d'apprentissage très générales, provenant d'un ensemble d'organes différents, mais répondant bien sûr aux mêmes critères de malignité, on a alors la possibilité de construire un classificateur de type plus "généraliste".

Références bibliographiques

- [1] A.Bendib , N . Benzidane H.Guendouz , A.Abdelouahab, Dépistage du cancer du sein, Journée scientifique sur le cancer du sein, association d'aide aux malades atteints de cancer, El-Badr, CHU, F.Fanon, Blida, Février 2010
- [2] R. Ghys, Les maladies du sein, Editions Québec/Amérique, 1989
- [3] J.J Toumazet, Traitement de l'image par l'exemple, Sybex, 1990
- [4] D. Ortais, Nouvelles connaissances de l'anatomie du sein et de la physiologie et leurs conséquences dans la pratique, CHU de Nantes, 2007
- [5] M. Oger, indexation automatique d'images numériques : application aux images histopathologiques du cancer du sein et hématologiques de leucémies lymphoïdes chroniques. Université de Caen, 2008
- [6] O. Lezoray, " La réalisation d'une étude sur le tri informatique en cytologie", rapport de stage, DEA , Université de CAEN.1995
- [7] O. Lezoray, Segmentation d'images couleur par morphologie mathématique et classification de données par réseaux de neurones: application à la classification de cellules en cytologie des séreuses, thèse de doctorat, université de Caen, 2000
- [8] J.P Cocquerez et S. Philipp, "Analyse d'images : filtrage et segmentation". Edition Masson, 1998.
- [9] S. Beucher, Segmentation d'images et morphologie mathématique, thèse de doctorat, Ecole Nat. Sup. des Mines de Paris, 1990.
- [10] M.Coster et J.L.Cherman, « Traitement d'image par morphologie mathématique application au domaine médical et industriel », Thèse de doctorat, université Blaise Pascal, 1987.
- [11] M. Oger, P. Belhomme, J. Klossa, J. J Michels, A.Elmoataz, Automated region of interest retrieval and classification using spectral analysis, 9th European Congress on Telepathology and 3rd International Congress on Virtual Microscopy Toledo, Spain. May 2008
- [12] I.Sebari et D.Chen. HE, les approches de segmentation d'images par coopération régions-contours, Revue Télédétection, vol. 7, n° 1-2-3-4, p. 499-506, 2007
- [13] A. Boujelben, H. Tmar, J.Mnif, M. Abid, Automatic application level set approach in detection calcifications in mammographic image, International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 4, pp 1-14, 2011
- [14] S. Ghandour. Segmentation d'images couleurs par morphologie mathématique: application aux images microscopiques, Université de Toulouse, 2010.

- [15] A. Ramoul, Segmentation d'Images couleurs application aux images cellulaires, mémoire d'ingénieur, Université Saad Dahleb de Blida, 2007
- [16] O .Djeddi. Classification connexionniste d'images couleurs: application aux images cytologiques, mémoire d'ingénieur, Université Saad Dahleb de Blida, 2005
- [17] S. Mennad, N.Amia, Détection de pathologies dans les images cytologiques par morphologie mathématique et classification connexionniste, mémoire d'ingénieur, Université Saad Dahleb de Blida, 2007
- [18] E. Glory .Segmentation d'Images de Microscopie en Couleur pour l'Analyse Automatique de Cultures Cytologiques, thèse de doctorat, Université Descartes, Paris 5, 2005
- [19] S. Herlidou, Caractérisation tissulaire en IRM par l'analyse de texture : étude du tissu musculaire et de tumeurs intracrâniennes, Thèse de doctorat, Université de Rennes, 1999
- [20] G. Dreyfus, M. Samuelides, J.-M. Martinez, M. B. Gordon, F. Badran, S. Thiria, L. Hérault, Réseaux de neurones (méthodologies et applications), Eyrolles, 2004
- [21] S. Nissen, Création d'un réseau de neurones, Intelligence artificielle www.software20.org, 2005
- [22] Fouad Badran, Perceptrons multicouches : fonctions d'erreur, maximum de vraisemblance, modèle multi-expert, support de cours, 'http://cedric.cnam.fr/~crucianm/ml.html'
- [23] Base d'images cytologiques du sein, http://cytopathnet.org/tiki-browse_gallery.php
- [24] R. C. Gonzalez, Digital Image processing using Matlab, Pearson publication, 2005