

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

UNIVERSITE FERHAT ABBAS-SETIF

MEMOIRE

Présenté à la Faculté des Sciences
Département d'informatique
Pour L'Obtention du Diplôme de

MAGISTER

Option : Ingénierie des Systèmes Informatiques

Par

Mr . GHILAS Hamza

THEME

Classification neuro-floue des données multimodales

Soutenu le :

Devant le jury

Président :	Dr. Touahria mohamed	M.C	Université Ferhat Abbas Sétif
Rapporteur :	Dr. MOUSSAOUI Abdelouhab	M.C	Université Ferhat Abbas Sétif
Examineur :	Dr. KEZAR OKBA	M.C	Université de Biskra

Résumé

Dans ce travail nous sommes intéressés par la classification des données multimodales qui sont de natures différentes et peuvent avoir des valeurs incomparables voir impossible de les mettre à l'échelle. La classification de ces données est une tâche délicate, pour faire, nous avons choisi un réseau de neurones de type perceptron multicouches flou. Le classifieur n'opère directement sur les valeurs réelles des attributs mais sur des valeurs floues qui prennent valeurs dans l'intervalle $[0,1]$ pour remédier à la variabilité des données multimodales.

Le cancer du sein est une maladie dangereuse et sa détection précoce reste le seul moyen de sa lute. Cependant plusieurs pays ont lancé des campagnes de dépistage gratuites, de ce fait, les radiologues sont censés d'analyser des dizaines de clichés de mammographie par jour, alors le développement des méthodes d'aide au diagnostic et de classification automatique de ces radiographies est motivant. La présence des microcalcifications dans la mammographie est un signe du cancer. La b nignit  ou la malignit  de ces derni res d pend de leurs intensit s, leurs tailles et de leurs morphologies. Les caract ristiques de ces microcalcifications est un bon exemple des donn es multimodales.

Dans ce travail nous avons contribu  par:

- L'utilisation d'un classifieur neuro-flou pour la classification des microcalcification.
- La proposition d'une liste d'attributs extraite   partir des indices que le radiologue utilise pour le diagnostic des microcalcifications.
- La proposition et l'impl mentation d'un algorithme de segmentation des images mammographiques pour l'extraction des microcalcification.

Notre travail a fait l'objet d'une communication orale   R²I'2011(rencontre sur la recherche en informatique,12-14 juin, Universit  de Tizi-Ouzou, Alg rie) et d'une d monstration de logiciel dans la 1^{ re} journ e doctorale du laboratoire d'informatique d'Oran (JDLIO'2011, 31 mai et 1 juin, Universit  d'Oran, Alg rie).

La suite de ce m moire est organis e comme suit: il est commenc  par une **introduction** o  nous donnons une br ve description du domaine de classification de donn es et nous d finissons notre probl matique.

Dans le **premier chapitre** nous d finissons la classification de donn es. Nous pr sentons la notion de neurone formel et de neurone biologique, le r seau de neurones de type perceptron multicouches et sa m thode d'apprentissage(r tropropagation de l'erreur). Nous expliquons le probl me des donn es multimodales et l'utilisation des ensembles flous pour leur abstraction.

Un état de l'art sur les méthodes d'aide au diagnostic des microcalcifications est présenté dans le **chapitre 2** où l'architecture des systèmes d'aide au diagnostic et le principe de la mammographie numérique sont traités. Nous présentons aussi les méthodes appliquées dans chaque phase dans le processus d'aide au diagnostic à savoir le prétraitement, le rehaussement, la segmentation et la détection des lésions et en fin la classification des régions suspectes(cancer/normale).

Dans le **chapitre 3** les microcalcifications sont décrites d'un point de vue médicale. Nous présentons l'anatomie du sein et le principe de la mammographie. Nous focalisons notre étude dans ce chapitre sur le diagnostic des microcalcifications afin d'extraire leurs caractéristiques.

Notre contribution pour la classification des microcalcifications est présentée dans le **quatrième chapitre**. Nous présentons l'architecture de notre classifieur, la liste d'attributs proposés, l'algorithme d'extraction des microcalcifications et nous terminons le chapitre par la discussion des résultats obtenus.

Ce mémoire est clôturé par une **conclusion** générale et des perspectives, où nous situons notre contribution dans l'aide au diagnostic du cancer du sein. Il contient aussi le code source de quelques classes essentielles de notre application dans l'annexe A et les annexes B et C contiennent les articles que nous avons présenté dans les doctorales JDLIO'2011 et R²I'2011.

Table des matières

Table des Matières	i
Table des Figures	vi
Liste des Tableaux	vii
Introduction	1
1 Classification de données et les systèmes neuro-flous	3
1.1 La classification	3
1.2 Réseaux de neurones	4
1.2.1 Neurone biologique	4
1.2.2 Neurone formel	5
1.2.3 Topologies	6
1.2.4 Apprentissage	7
1.2.5 Perceptron multicouches	8
1.2.5.1 Architecture d'un perceptron multicouches	9
1.2.5.2 Frontières de décision d'un perceptron multicouches	10
1.2.5.3 Apprentissage d'un perceptron multicouches	11
1.3 Les données multimodales	13
1.4 Les ensembles flous	14
1.4.1 Définition	15
1.4.2 Caractéristiques d'un ensemble flou	15

1.4.3	Défuzzification	16
1.4.3.1	La méthode du maximum	16
1.4.3.2	La méthode du centre de gravité	17
1.5	Conclusion	17
2	Aide au diagnostic du cancer du sein: état de l'art	18
2.1	La mammographie numérique et les CADs	18
2.2	Architecture des CADs	19
2.3	Acquisition et prétraitement	20
2.4	Rehaussement	20
2.4.1	Les méthodes classiques	20
2.4.2	Approche région	20
2.4.3	Approche statistique	21
2.5	Segmentation	21
2.5.1	Méthodes statistiques	21
2.5.2	Approche région	22
2.5.3	Détection de contours	22
2.5.4	Autres méthodes	22
2.6	Détection	22
2.6.1	Template matching (appariement de modèles)	23
2.6.2	Méthodes basées sur les attributs propres aux microcalcifications . .	23
2.6.3	Méthodes basées sur l'analyse de texture	23
2.6.4	Les attributs de texture multi résolutions	24
2.6.5	Méthodes basées sur les dimensions fractales	24
2.7	Classification	24
2.7.1	Evaluation des performances d'un classifieur	25
2.7.2	Les réseaux de neurones	26
2.7.3	Vecteur support machine	28
2.7.4	K plus proches voisins (Knn)	29
2.8	Bilan	29

3	Cancer du sein	32
3.1	Anatomie du sein	32
3.2	La mammographie	34
3.2.1	Rayon X	35
3.2.2	Radiographie du sein	35
3.3	Cancer du sein	36
3.4	Diagnostic des microcalcifications	38
3.4.1	Microcalcifications groupées en foyer sans tumeur visible ou palpable	40
3.4.1.1	Calcifications typiquement bénignes	40
3.4.1.2	Microcalcifications de diagnostic difficile	42
3.4.1.3	Microcalcifications plus typiquement malignes	43
3.4.2	Microcalcifications éparses sans tumeur visible ou palpable	45
3.4.2.1	Calcifications bilatérales(dans les deux seins)	45
3.4.2.2	Calcifications unilatérales	46
3.4.3	Calcifications disposées de façon ordonnée, linéaire	46
3.4.3.1	Calcifications canalaies	47
3.5	Conclusion	49
4	Contribution et validation	50
4.1	Extraction des attributs	50
4.1.1	Attributs propres à une microcalcification	51
4.1.2	Attributs d'un foyer de microcalcifications	54
4.2	Classifieur neuro-flou	55
4.3	Implémentation	56
4.3.1	Environnement de développement	56
4.3.2	Weka	56
4.3.3	Base d'images	57
4.3.4	Organisation de l'application	59
4.3.5	Prétraitement	61
4.3.6	Segmentation	61

4.3.7	Calcul des attributs	64
4.4	Résultats et discussions	64
4.5	Conclusion	69
	Conclusion	70
	Bibliographie	72
	Annexes	79
A	Code source	79
B	Communication R2I'2011	97
C	Communication JDLIO'2011	104

Table des figures

1.1	Le neurone biologique.	5
1.2	Le neurone formel.	5
1.3	Fonctions d'activation d'un neurone formel.	6
1.4	Principaux types d'architectures et de structures d'interconnexions des réseaux de neurones.	7
1.5	Modélisation du "OU" logique avec un perceptron.	8
1.6	Problème du OU exclusif.	9
1.7	Architecture d'un PMC.	10
1.8	Influence de l'architecture d'un réseau de neurones sur les frontières de décision [5].	11
1.9	Relations entre le neurone de sortie k et les M neurones cachés.	12
1.10	Données multimodales.	13
1.11	Exemple de fonction d'appartenance.	15
1.12	Caractéristiques d'un sous-ensemble flou.	16
1.13	Défuzzification par la méthode des maxima.	16
1.14	Inconvénient de la méthode des maxima.	17
1.15	Défuzzification par la méthode du centre de gravité.	17
2.1	Architecture des CADs.	20
2.2	Courbe ROC.	26
3.1	Anatomie du sein.	33
3.2	Galactophore	34

3.3	Coupe d'un canal galactophorique	34
3.4	Incidences fondamentales en mammographie	36
3.5	Statistiques des cancers les plus fréquents chez les femmes en Algérie. . . .	37
3.6	Diagnostic des calcifications	39
3.7	Calcifications microkystiques	40
3.8	Calcifications péricanalaires de la mastite plasmocytaire	41
3.9	Liponécrose microkystique calcifiée	41
3.10	Calcifications semi-lunaires "en tasse à thé"	41
3.11	Calcifications en forme de petites perles calcaires.	42
3.12	Calcifications triangulaire à sommet orienté vers le mamelon.	43
3.13	Calcifications irrégulières à contours aigus.	44
3.14	Calcifications irrégulières à contours aigus, triangulaire à sommet orienté vers le mamelon.	44
3.15	Microcalcifications vermiculaires	45
3.16	Calcification bilatérales de dysplasie fibrokystique.	46
3.17	Calcifications pariétales, péricanalaires.	47
3.18	Calcifications de l'hyperplasie canalaire.	48
3.19	Calcifications du comédocarcinome	48
4.1	Rayon maximal et minimal d'une calcification.	51
4.2	Rectangularité.	52
4.3	Circularité.	52
4.4	Angularité.	54
4.5	Classifieur neuro-flou.	56
4.6	Fichier ics.	58
4.7	Fichier OVERLAY.	59
4.8	Diagramme de classes	60
4.9	Prétraitement.	61
4.10	Algorithme de segmentation.	62
4.11	Algorithme de segmentation(exécution).	63
4.12	Résultat de la segmentation.	64

Liste des tableaux

2.1	Résultats d'un classifieur	25
2.2	Méthodes de classification des microcalcifications	31
4.1	Les <i>cases</i> utilisé de la base DDSM.	65
4.2	Résultat de classification(a).	66
4.3	Résultat de classification(b).	67
4.4	Résultat de classification(c).	68

Introduction

La classification est l'attribution d'une classe spécifique et connue auparavant à un objet donné. Cette attribution a besoin d'une certaine abstraction pour pouvoir extraire des généralités à partir des exemples dont on dispose. Pour une machine, la classification appelée aussi généralisation, discrimination et etc, consiste à attribuer une classe à un individu à partir d'un ensemble de caractéristiques (attributs) par rapport à un modèle établi après apprentissage. L'apprentissage automatique peut être défini comme l'extraction automatique d'une connaissance ou d'un modèle à partir d'une collection d'éléments de données.

Plusieurs techniques sont utilisées pour l'apprentissage et la classifications automatique. L'imitation de l'intelligence naturelle a donné naissance à de nouvelles méthodes et algorithmes de classification. En particulier les réseaux de neurones qui sont issues du fonctionnement du cerveau humain. Il existe plusieurs variantes des réseaux de neurones, dans ce mémoire nous nous'intéresserons aux réseaux multicouches qui sont un outil puissant et sont plus adaptés à la classification supervisée.

L'apprentissage d'un réseaux de neurones est une tâche d'extraction de connaissance à partir des données. L'efficacité du réseau dépend fortement de la nature des données. Dans le cas où les données sont de natures différentes, multivaluées, et multimodales, leur abstraction induit à l'imprécision et l'ambiguïté. Par exemple lors de la description de la taille d'un objet, la formalisation des notions: grand, petit, moyen, géant et etc, n'est pas prise en compte par les méthodes de classification classiques. A cet effet l'introduction des ensembles flous pour la représentation des attributs de classification peut avoir beaucoup d'intérêts.

L'objectif de ce projet est d'appliquer une nouvelle stratégie de classification basée sur les réseaux de neurones et les ensembles flous pour la classification d'un corpus de données multimodales. Nous avons choisi d'appliquer notre approche sur le diagnostic automatique du cancer du sein. L'objet de notre contribution est la classification des *microcalcifications*

(signe radiologique du cancer du sein). Les microcalcifications ont des propriétés très différentes qui déterminent la nature de la pathologie(bénigne/maligne), à savoir(la taille, la forme, l'orientation,...). L'extraction de ces caractéristiques est un élément déterminant pour la classification, à cet effet nous impliquons un radiologue expert pour la détermination des attributs de classification. Ces attributs peuvent être complémentaires et parfois même contradictoires. Cette richesse d'information n'est pas traitée par les algorithmes classiques c'est pourquoi nous proposons d'appliquer une approche neuro-floue.

Classification de données et les systèmes neuro-flous

L'extraction des connaissances à partir des données est un domaine récent qui a prouvé son efficacité face à l'intelligence artificielle classique. La classification de données est l'une des tâches les plus importantes dans cette discipline, elle s'applique à la prévision, l'estimation, l'aide à la décision et etc. Dans ce chapitre nous définissons la classification de données, les réseaux de neurones qui sont un outil puissant pour l'apprentissage automatique. Nous présentons les problèmes de variabilité et d'hétérogénéité des données ainsi que l'utilisation des ensembles flous pour l'abstraction de ces données.

1.1 La classification

La classification naturelle consiste à ranger des objets dans des catégories bien définies. Pour le *data mining*, la classification est une tâche d'extraction de connaissances à partir des données, qui consiste de même à déterminer la classe d'un objet à partir de ses caractéristiques, elle est appelée aussi apprentissage supervisé. Un algorithme qui effectue la tâche de classification est appelé classifieur.

Un classifieur est conçu dans deux phases: une étape de construction d'un modèle représentant les classes à prédire (Apprentissage) et une phase d'utilisation où le classifieur reçoit des éléments inconnus pour deviner leurs classes (test). La qualité d'un classifieur est mesurée par le taux d'erreur, qui est le taux des exemples mal classés parmi les éléments de test. Il existe plusieurs variantes de classifieur dont les plus populaires sont: la méthode des K plus proches voisins (KNN), les arbres de décision, les classifieurs bayésiens, les machines à vecteur support et les réseaux de neurones.

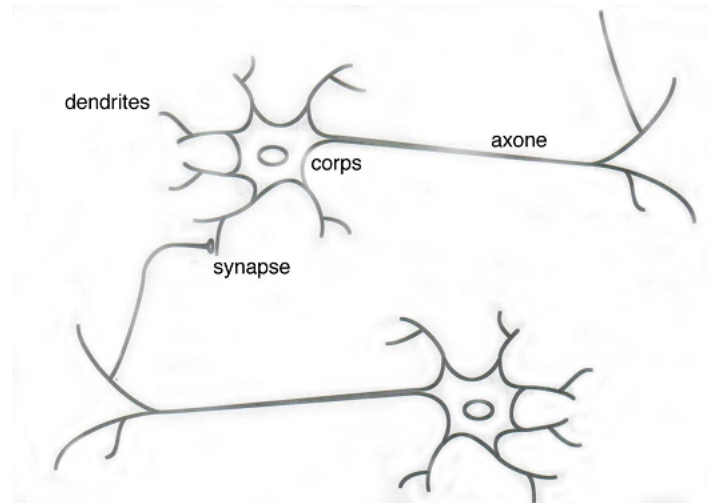
1.2 Réseaux de neurones

Un réseau de neurones est un assemblage de constituants élémentaires interconnectés, appelés “neurones”, inspirés du fonctionnement du cerveau humain. Les neurones réalisent chacun un traitement simple mais dont l’ensemble en interaction fait émerger des propriétés complexes. Les réseaux de neurones ont prouvé leur efficacité dans plusieurs domaines: l’informatique, la robotique, l’industrie, l’économie et etc.

Il existe plusieurs modèles de réseaux de neurones, chaque type de réseaux est destiné à une certaine classe de problèmes à traiter, dans cette section nous présentons les éléments de base des réseaux de neurones à savoir le neurone biologique et formel. Nous nous intéresserons au réseaux multicouches qui sont adaptés à la classification de données.

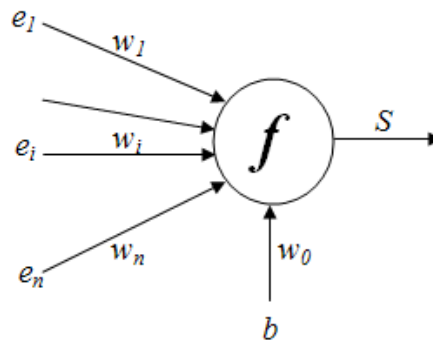
1.2.1 Neurone biologique

Le cerveau humain se compose d’environ 10^{12} neurones (mille milliards), avec 1000 à 10000 synapses (connexions) par neurone. Ce dernier possède trois principales composantes: les dendrites, le corps cellulaire et l’axone (figure 1.1). 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 un espèce d’intégrateur en accumulant des charges électriques. Lorsque le neurone devient suffisamment excité (lorsque la charge accumulée dépasse un certain seuil), par un processus électro-chimique 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 le dendrite d’un autre neurone s’appelle la synapse. L’arrangement spatial des neurones et de leur axone, ainsi que la qualité des connexions synaptiques individuelles qui déterminent la fonction précise d’un réseau de neurones biologiques. C’est en se basant sur ces connaissances que le modèle mathématique a été défini [1].

FIG. 1.1 – *Le neurone biologique.*

1.2.2 Neurone formel

La figure 1.2 montre la structure d'un neurone artificiel. Un neurone formel est un processeur élémentaire (corps du neurone). Il reçoit un nombre variable d'entrées en provenance d'autres neurones (dendrites). À chacune de ces entrées est associée un poids w (weight) représentatif de la force de la connexion (synapses). Chaque processeur élémentaire est doté d'une sortie unique, qui peut se ramifier ensuite pour alimenter un nombre variable de neurones (axones).

FIG. 1.2 – *Le neurone formel.*

Formellement la sortie S d'un neurone est vue comme une fonction numérique de la somme d'un certain nombre (n) d'entrées e_i pondérées par des poids (weight) w_i . Cette fonction (f de l'équation 1.1) est appelé fonction d'activation du neurone.

Cette modélisation est introduite par J. Mc Culloch et W. Pitts en 1943 [2], les premiers qui ont montré que des réseaux de neurones formels simples peuvent réaliser des

fonctions logiques et arithmétiques complexes. Les réseaux de neurones sont différenciés d'une typologie à une autre, par la façon dont leurs neurones sont interconnectés et le mode d'apprentissage.

$$S = f \left(\sum_{i=0}^{i=n} w_i e_i + w_0 b \right) \quad (1.1)$$

Fonction d'activation: le fonctionnement du neurone est centré autour de sa fonction d'activation appelée aussi fonction de transfert ou fonction seuil. Il existe plusieurs types de cette fonction dont les plus utilisées sont illustrées dans la figure 1.3

Nom de la fonction	Relation d'entrée/sortie	Graphique
seuil	$a = 0 \quad \text{si } n < 0$ $a = 1 \quad \text{si } n \geq 0$	
seuil symétrique	$a = -1 \quad \text{si } n < 0$ $a = 1 \quad \text{si } n \geq 0$	
linéaire	$a = n$	
linéaire saturée	$a = 0 \quad \text{si } n < 0$ $a = n \quad \text{si } 0 \leq n \leq 1$ $a = 1 \quad \text{si } n > 1$	
linéaire saturée symétrique	$a = -1 \quad \text{si } n < -1$ $a = n \quad \text{si } -1 \leq n \leq 1$ $a = 1 \quad \text{si } n > 1$	
linéaire positive	$a = 0 \quad \text{si } n < 0$ $a = n \quad \text{si } n \geq 0$	
sigmoïde	$a = \frac{1}{1+\exp^{-n}}$	
tangente hyperbolique	$a = \frac{e^n - e^{-n}}{e^n + e^{-n}}$	

FIG. 1.3 – Fonctions d'activation d'un neurone formel.

1.2.3 Topologies

La topologie décrit la structure d'interconnexion entre les neurones qui composent le réseau, on parle aussi d'architecture. Contrairement à l'architecture du cerveau, les neurones peuvent être arrangés par couches. On distingue alors une couche de neurones d'entrée, des couches cachées et la couche de neurones de sortie. Dans les **réseaux non bouclés** (figure 1.4 a-b-c), les entrées des neurones d'une couche sont les sorties des neurones de la couche précédente et les entrées des neurones de la première sont reliées au monde extérieur. Dans les réseaux à **connexions locales** (figure 1.4 c-d-e), les neurones d'une couche ne sont pas tous forcément reliés à la couche suivante.

Les réseaux non bouclés propagent donc l'information de l'entrée vers la sortie (*feed-forward*). Dans les **réseaux bouclés** appelés aussi réseaux récurrents (figure 1.4 d-e), on retrouve le même type d'interconnexion entre neurones, mais avec des possibilités de retours en arrière (*feedback*). Ces réseaux sont ainsi utilisés pour la modélisation et la commande dynamique de processus, les rétroactions permettent à ce type de réseaux de présenter un comportement temporel.

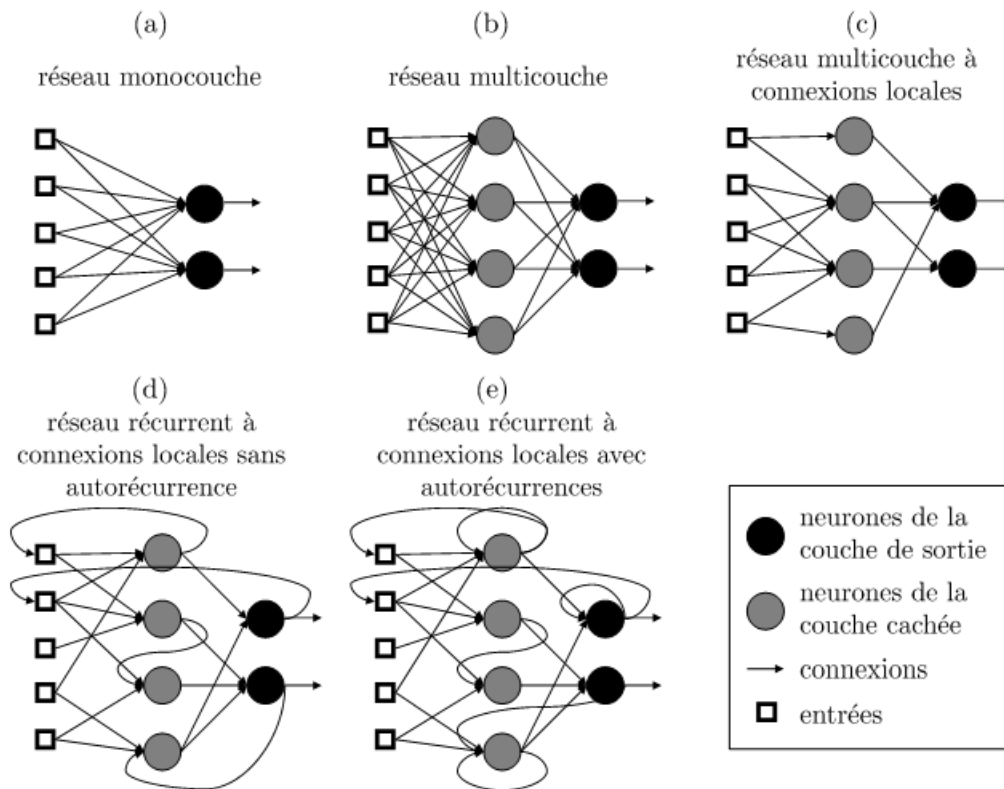


FIG. 1.4 – Principaux types d'architectures et de structures d'interconnexions des réseaux de neurones.

Les réseaux liés à notre problématique (classification) sont les réseaux non bouclés, c'est pourquoi nous concentrerons notre attention sur ce type de réseaux et plus particulièrement sur les perceptrons multicouches.

1.2.4 Apprentissage

L'apprentissage d'un réseau de neurones consiste à déterminer les poids des liaisons synaptiques. Il s'effectue en modifiant l'intensité (poids) des liaisons entre neurones en fonction d'un certain nombre d'exemple appelés "exemples d'apprentissage". Il existe deux types d'apprentissage: supervisé et non supervisé.

Dans l'apprentissage non supervisé, on ne connaît pas quelle doit être la sortie du réseau. Le réseau cherche à détecter des points communs entre les observations présentées et adapte les poids afin de donner une sortie équivalente pour des observations d'entrées proches. Cette approche est utilisée notamment dans des applications de *clustering*, où l'on cherche à regrouper les observations les plus similaires.

L'apprentissage supervisé utilise un professeur pour guider le réseau vers la solution recherchée. Pour une entrée propagée dans le réseau, le professeur compare la réponse désirée à celle obtenue par le réseau pour calculer un signal d'erreur et le réseau essaye de minimiser cette erreur itérativement.

1.2.5 Perceptron multicouches

Le premier réseau de type perceptron est inventé par Rosenblatt(1958) [3], qui a commencé par l'utilisation des réseaux à une seule couche et non récurrent pour l'estimation de fonctions. Le neurone de la figure 1.5 réalise la fonction "OU" logique. A partir d'observations étiquetées (contexte d'apprentissage supervisé), un perceptron est capable d'ajuster les poids d'un neurone, afin de converger vers une configuration apte à réaliser des opérations de classification ou de généralisation.

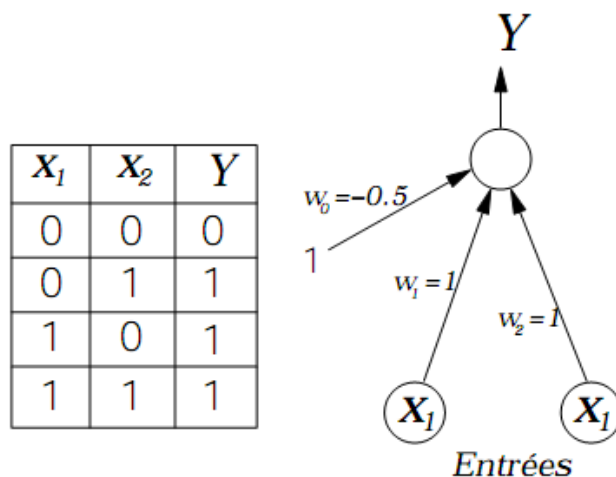


FIG. 1.5 – Modélisation du "OU" logique avec un perceptron.

Le perceptron est un modèle simple à mettre en oeuvre, mais il est incapable d'apprendre les classes d'un ensemble de données qui n'est pas linéairement séparable, la figure 1.6 est une illustration du problème du "OU exclusif". Cette limite a fortement ralenti les recherches dans ce domaine, en désintéressant la communauté scientifique. La solution de ce problème est l'utilisation de plusieurs couches successives (une couche d'entrée, une ou plusieurs couches cachées et une couche de sortie). Rosenblatt avait déjà pris conscience

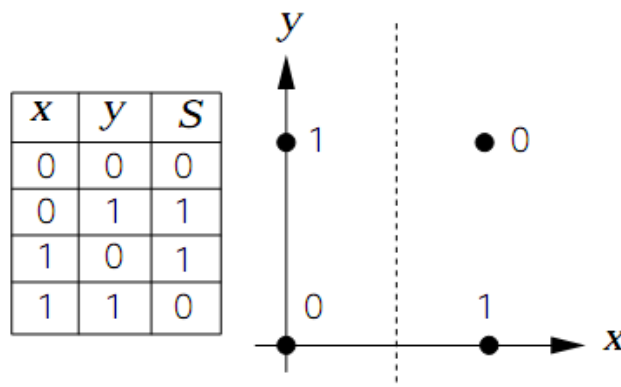


FIG. 1.6 – Problème du OU exclusif.

de la nécessité de couches cachées dans un réseau, afin d'élargir les capacités du perceptron, mais l'algorithme d'apprentissage de ces topologies n'existait pas. Ainsi, Rumelhart et al (1986) [6] ont publié un nouvel algorithme d'apprentissage, appelé l'algorithme de rétropropagation de l'erreur (*backpropagation*), qui permet l'apprentissage et donc l'optimisation des paramètres des réseaux de neurones à plusieurs couches.

1.2.5.1 Architecture d'un perceptron multicouches

L'architecture des perceptrons multicouches (PMC et en anglais MLP pour *multi-layer perceptron*) est fondée sur le perceptron, où des couches de neurones, dites couches cachées, sont ajoutées entre les entrées et la couche de sortie (figure 1.7). L'intérêt de cette évolution est de surpasser les limitations du perceptron, afin d'optimiser la capacité d'apprendre les frontières des classes qui ne sont pas linéairement séparables.

La sortie d'un neurone caché (a_i) est obtenue par une combinaison linéaire des p variables d'entrées et d'un biais, toujours pondérées par des poids, telle que :

$$a_i = \sum_{j=1}^p w_{ij}^{(1)} x_j + w_{i0}^{(1)} \quad (1.2)$$

$$= \sum_{j=0}^p w_{ij}^{(1)} x_j \quad (1.3)$$

où $w_{ij}^{(1)}$ représente un poids de la première couche (la couche cachée) associant l'entrée x_j au neurone i . La sortie de chaque unité de la couche cachée z_i est obtenue par une fonction d'activation f_1 tel que $z_i = f_1(a_i)$. Les sorties du réseau ne considèrent plus directement les entrées mais les sorties des M neurones cachés z_i , alors on retrouve la formulation suivante:

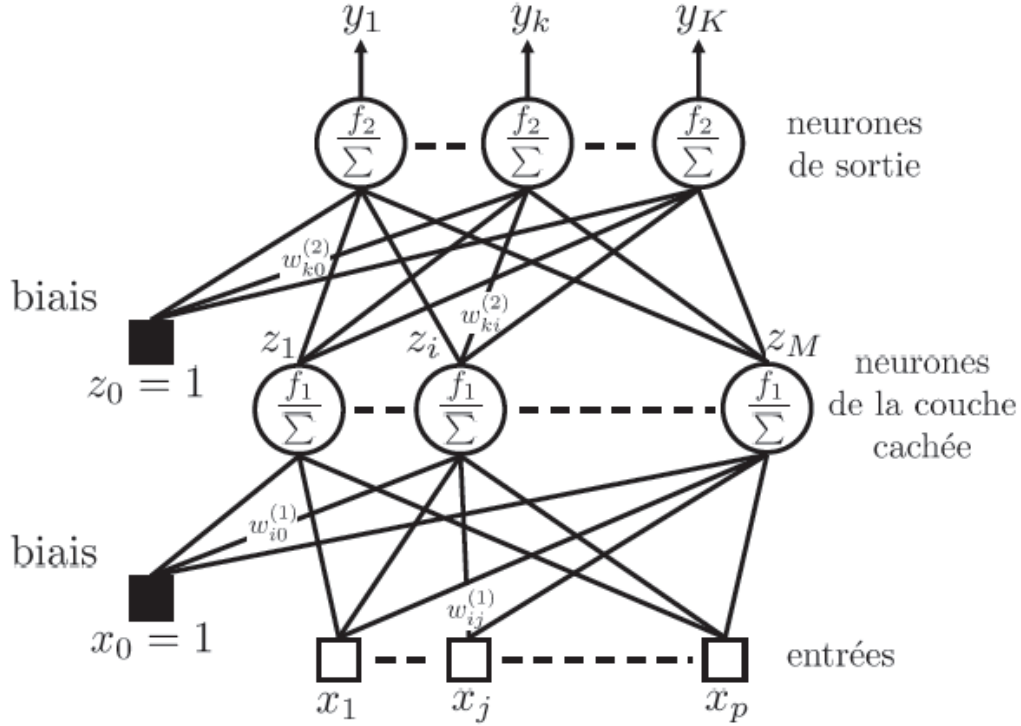


FIG. 1.7 – Architecture d'un PMC.

$$a_k = \sum_{i=1}^M w_{ki}^{(2)} z_i + w_{k0}^{(2)} \quad (1.4)$$

$$= \sum_{i=0}^M w_{ki}^{(2)} z_i \quad (1.5)$$

De même les sorties y_k sont obtenues par une fonction d'activation f_2 , l'expression de la sortie est $y_k = f_2(a_k)$. Ainsi, en combinant les relations (1.3) et (1.5), nous obtenons directement les sorties du réseau de la figure 1.7 en fonction des entrées :

$$y_k = f_2 \left(\sum_{i=1}^M w_{ki}^{(2)} f_1 \left(\sum_{j=0}^p w_{ij}^{(1)} x_j \right) \right) \quad (1.6)$$

Dans un perceptron multicouche, nous remarquons que l'information circule des entrées vers les sorties, d'où l'appellation *feedforward*.

1.2.5.2 Frontières de décision d'un perceptron multicouches

Les frontières de décision séparant les classes dans un réseau simple couche sont linéaires. L'augmentation de couches, par l'insertion de couches cachées implique une augmentation de la complexité des régions identifiant les différentes classes, comme montré à

la figure 1.8. Ainsi trois couches de neurones (figure 1.8c) sont nécessaires pour représenter des régions disjointes.

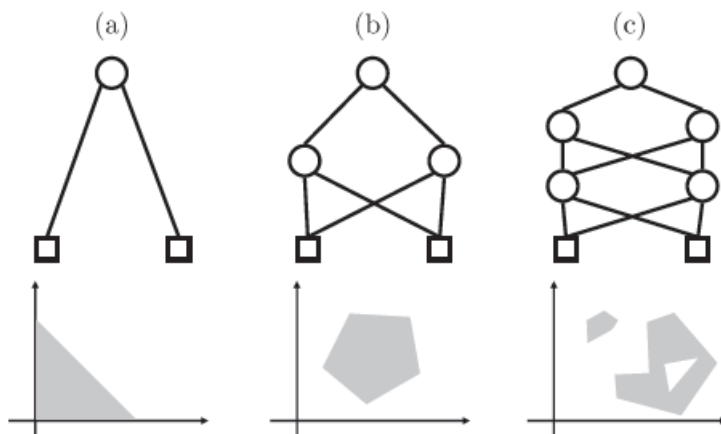


FIG. 1.8 – Influence de l'architecture d'un réseau de neurones sur les frontières de décision [5].

1.2.5.3 Apprentissage d'un perceptron multicouches

L'apprentissage des perceptrons multicouches fut leur limitation jusqu'à l'apparition de l'algorithme rétropropagation de l'erreur publié par Rumelhart et al en 1986 [6]. Cet algorithme est basé sur le principe de l'apprentissage supervisé, alors il utilise un ensemble d'exemples dont leurs classes sont connues (base d'apprentissage). L'algorithme consiste à propager une observation de l'entrée du réseau à travers les couches de neurones afin d'obtenir les valeurs de sorties. Celles-ci comparées aux sorties désirées. Elles fournissent alors les erreurs permettant d'adapter les poids des neurones. Sans couche cachée, la connaissance de l'erreur des neurones de sortie permet un calcul direct de la correction et rend l'adaptation des poids de ces uniques neurones aisée. Cependant, dans un réseau à couches cachées, les sorties désirées des neurones cachés sont inconnues, il demeure alors impossible de connaître les erreurs de ces neurones.

L'intuition qui permet de résoudre cette difficulté et qui donna naissance à la rétropropagation, fut la suivante : l'activité d'un neurone est liée aux neurones de la couche précédente. Ainsi, l'erreur d'un neurone de sortie est dû aux neurones cachés de la couche précédente proportionnellement à leur influence, donc en fonction de leur activation et des poids qui relient les neurones cachés au neurone de sortie. D'après la figure 1.9, on cherche à obtenir les contributions des M neurones cachés qui ont donné l'erreur du neurone de sortie k .

La procédure de la rétropropagation consiste alors à propager dans le réseau l'erreur (obtenue lors de la propagation d'une entrée) de la couche de sortie vers les neurones

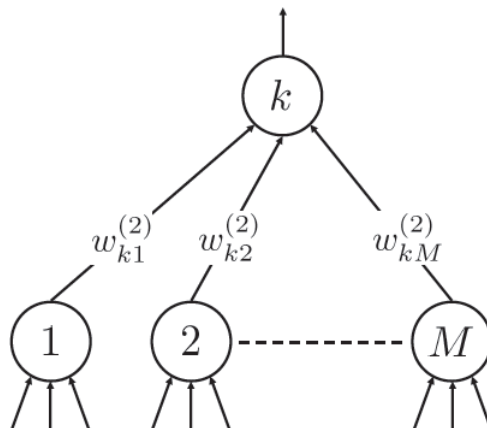


FIG. 1.9 – Relations entre le neurone de sortie k et les M neurones cachés.

cachés puis à ceux de la couche d'entrée. Les étapes d'apprentissage d'un perceptron à une seule couche cachée sont présentées ci-dessous [1]:

1. Appliquer un vecteur $x = (x_1, x_2, \dots, x_p)$ à la couche d'entrée (choisi de la base d'apprentissage).
2. Calculer les valeurs des entrées des neurones de la couche cachée:

$$a_i = \sum_{j=1}^p w_{ij}^{(1)} x_j + w_{i0}^{(1)}$$

3. Calculer les sorties des neurones de la couche cachée:

$$z_i = f_1(a_i)$$

4. Calculer les valeurs d'entrées des neurones de la couche de sortie:

$$a_k = \sum_{i=1}^M w_{ki}^{(2)} z_i + w_{k0}^{(2)}$$

5. Calculer les sorties du réseau:

$$y_k = f_2(a_k)$$

6. Calculer le terme d'erreur des neurones de sortie:

$$\delta_k^{(2)} = (d_k - y_k) f^{(2)'}(a_k)$$

Où d_k et y_k sont respectivement la sortie désirée et la sortie réelle du neurone k correspondante à l'entrée x .

7. Mettre à jour les poids des neurones de la couche de sortie:

$$w_{ki}^{(2)}(t+1) = w_{ki}^{(2)}(t) + \alpha \delta_k^{(2)} z_i$$

α est le pas d'apprentissage.

8. Calculer le terme d'erreur des neurones cachés:

$$\delta_i^{(1)} = f_i^{(1)}(a_i) \sum_{k=1}^K \delta_k^{(2)} w_{ki}$$

9. Mettre à jour les poids des neurones de la couche cachée:

$$w_{ij}^{(1)}(t+1) = w_{ij}^{(1)}(t) + \alpha \delta_i^{(1)} x_j$$

Ces étapes sont mises en oeuvre par un algorithme d'apprentissage qui essaye de minimiser l'erreur globale du réseau (somme des erreurs de chacun des neurone de sortie). Un tel algorithme commence par l'initialisation des poids des neurones par des valeurs aléatoires, puis les étapes citées ci-dessus sont réitérées jusqu'à satisfaction d'un certain critère d'arrêt (nombre d'itérations ou un seuil de l'erreur).

1.3 Les données multimodales

Avec le développement des technologies de stockage et de numérisation de l'information, les algorithmes de classification sont de plus en plus susceptibles d'opérer sur des ensembles de données de taille énorme en terme d'espace mémoire d'un côté et de la différence dans la nature des données d'un autre côté. Une donnée peut être le résultat d'un fait, une mesure physique, une déclaration d'un expert humain, une description d'un phénomène, et etc. Ces données peuvent avoir des valeurs différentes, contradictoires, ambiguës, hétérogènes et parfois complémentaires, pour cela ces informations sont qualifiées de multimodales. Pour mieux situer le problème des données multimodales nous présentons un exemple illustratif dans la figure 1.10.

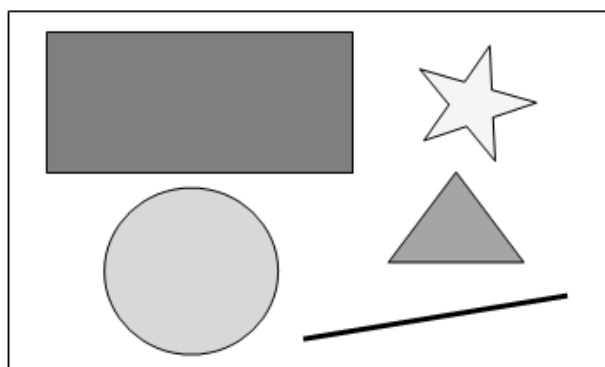


FIG. 1.10 – *Données multimodales.*

Si, nous demandons à une personne de décrire les objets de la figure 1.10, il produira des déclarations dépendantes de ses compétences. Par exemple s'il décrit la taille

des objets ses déclarations seront de type *petit*, pour le triangle, *moyen* pour le cercle, *grand* pour le rectangle et etc. Concernant l'intensité des objets, les descriptions prennent les formes suivantes: *claire*(cercle), *plus claire*(étoile), *sombre*(triangle), *noire*(rectangle). Cette description n'est pas unique et peut être différente d'une personne à une autre.

Le traitement de ces données par un algorithme de classification nécessite leurs projection dans un espace numérique. Les tailles de ces objets seront exprimées avec une quantité d'une unité de mesure(cm, mm, micron,...). L'intensité est exprimée en générale par un nombre entier prenant valeur dans l'intervalle $[0,n]$. Cette abstraction engendre des confusions, dans les valeurs d'une même propriété et entre les valeurs des différentes caractéristiques. Un objet ayant une taille de 5cm est il petit, grand ou moyen? Un objet ayant une intensité de 125 est t'il sombre ou claire? et encore qu'elle est la signification de 125 par rapport à 5 cm?

D'autre caractéristiques peuvent n'exister que chez certains éléments d'un corpus de données. Si nous introduisons la propriété d'aire(surface) pour les objets de la figure 1.10, cette propriété n'aura pas de sens pour le segment de droite.

L'extraction des caractéristiques pour l'aide au diagnostic médicale fait un exemple qui génère des données multimodales tel que l'analyse des clichés de mammographies, des images IRM et etc. Dans ce mémoire nous nous intéressons au diagnostic automatique des microcalcifications (mammographie) ayant des caractéristiques complexes (leurs morphologies, leurs intensités et etc).

Une bonne prise en compte de ces données induit vers une richesse d'informations que d'ambiguïtés. L'utilisation des ensembles flous pour l'abstraction et la modélisation de ce type d'informations peut prendre en charge cette richesse d'information. Les ensembles flous ont prouvé leurs efficacité dans le traitement des informations incertaines et incomplètes. La notion de fonction d'appartenance est très importante du point de vue de sa sémantique, elle est bien adaptée pour la représentation des expressions humaines. Pour ce fait l'usage des ensembles flous pour la normalisation des valeurs des attributs de classification que nous cherchons à appliquer au microcalcifications est motivant.

1.4 Les ensembles flous

La théorie des ensembles flous(*fuzzy sets* en anglais) fut développée à partir de 1965 par Lotfi A.Zadeh, professeur à l'université de Californie à Berkeley, généralisant la théorie des ensembles classiques. Dans la nouvelle théorie de Zadeh, un élément peut plus ou moins appartenir à un certain ensemble. Les imprécisions et les incertitudes peuvent ainsi être modélisées. Cette théorie est née suite au besoin de modéliser des situations intermédiaires.

1.4.1 Définition

Soit X un ensemble, un sous-ensemble flou A de X est défini par une fonction f_A sur X à valeurs dans l'intervalle $[0,1]$, appelée fonction d'appartenance [4].

Exemple: Une fonction caractéristique possible pour définir le sous-ensemble flou "Avoir une vingtaine d'années" figure 1.11

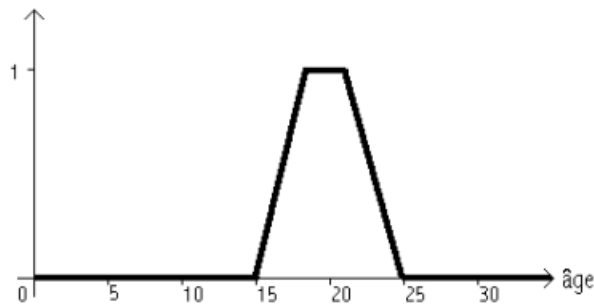


FIG. 1.11 – *Exemple de fonction d'appartenance.*

Un problème des applications pratiques est de définir ces fonctions, on fait généralement appel à des méthodes statistiques ou à l'avis d'un expert. La définition des fonctions d'appartenance pour un certain ensemble de données est appelée dans la littérature la "fuzzification".

1.4.2 Caractéristiques d'un ensemble flou

Soit A un sous-ensemble flou de X , les caractéristiques de A sont:

– *support* de A : $\text{supp}A = \{x \in X, f_A(x) \neq 0\}$.

– *hauteur* de A : $h(A) = \sup_{x \in X} f_A(x)$,

A est dit normalisé si $h(A) = 1$, les ensembles flous sont généralement utilisés dans leurs formes normalisés.

– *noyau* de A : $\text{noy}A = \{x \in X, f_A(x) = 1\}$

– *cardinalité* de A : $|A| = \sum_{x \in X} f_A(x)$.

Exemple : Caractéristique de l'exemple de la figure 1.12.

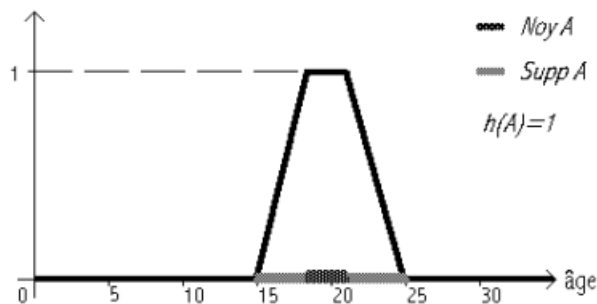


FIG. 1.12 – Caractéristiques d'un sous-ensemble flou.

1.4.3 Défuzzification

Un algorithme de classification flou fournit un résultat flou, ayant une valeur appartenant à l'intervalle $[0,1]$. La restitution de la forme naturelle de l'information est assurée par une phase dite de "défuzzification". Il existe plusieurs méthodes de défuzzification dont nous présentons quelques unes.

1.4.3.1 La méthode du maximum

Elle consiste à choisir comme solution défuzzifiée l'abscisse du maximum de la fonction d'appartenance des solutions.

Si plusieurs conviennent on peut par exemple utiliser une variante, la méthode de la *moyenne des maximas*, qui consiste à prendre comme solution la moyenne des abscisses des maximas (figure 1.13).

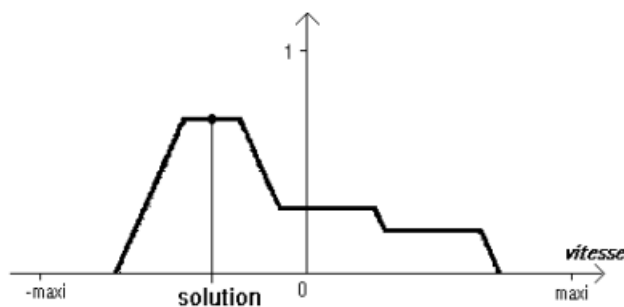


FIG. 1.13 – Défuzzification par la méthode des maximas.

L'avantage de cette méthode est qu'elle ne nécessite pas une grande puissance de calcul. Cependant, elle présente un grand inconvénient: de faibles variations du sous-ensemble flou peuvent entraîner des sauts importants du signal de sortie comme l'illustre la figure 1.14.

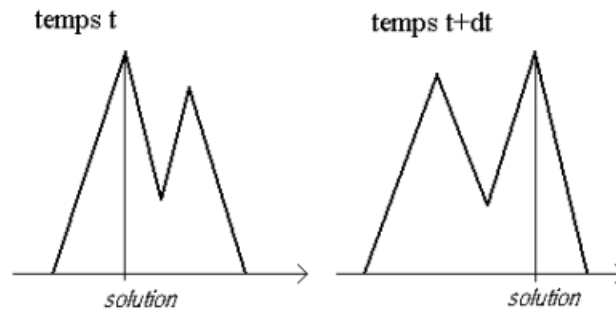


FIG. 1.14 – Inconvénient de la méthode des maxima.

1.4.3.2 La méthode du centre de gravité

Cette méthode est la plus utilisée et donne généralement de bons résultats. Elle prend comme solution l'abscisse du centre de gravité des solutions (figure 1.15).

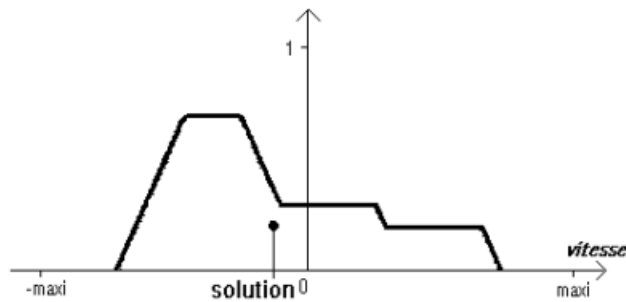


FIG. 1.15 – Défuzzification par la méthode du centre de gravité.

Par comparaison avec la méthode du maximum, les résultats fournis sont bien plus stables. Mais en contrepartie, elle exige une plus grande puissance de calcul.

1.5 Conclusion

Dans ce chapitre nous avons abordé la notion de données multimodales et les difficultés liées à leurs modélisation et abstraction. La classification par réseaux de neurones et une technique très puissante et l'utilisation des ensembles flous à côté de ce type de classifieur est une hybridation prometteuse pour la classification des données multimodales.

Aide au diagnostic du cancer du sein: état de l'art

L'aide à la décision est un domaine qui vise à concevoir des outils logiciels pour aider un décideur à analyser un problème ou une situation et lui fournir des solutions appropriées. La conception de tels outils pour l'aide au diagnostic du cancer du sein connus sous le nom de CAD (Computer Aided Diagnostic) a fait l'objet de plusieurs travaux de recherche depuis une vingtaine d'années [7, 8], malgré ces efforts, ces systèmes d'aide au diagnostic ne connaissent pas encore leur succès dans le marché. Dans ce chapitre nous présentons l'architecture des CADs et le principe de la mammographie numérique, ainsi que les méthodes appliquées dans chaque phase dans le processus d'aide au diagnostic à savoir le prétraitement, le rehaussement du contraste, la segmentation et la détection des lésions et en fin la classification des régions suspectes(cancer/normale), nous finissons ce chapitre par un bilan où nous présentons un tableau récapitulatif.

2.1 La mammographie numérique et les CADs

Les avancées technologiques dans le domaine des couches minces ont permis de mettre en œuvre des capteurs d'images sensibles aux rayons X. Ces capteurs ont rendu possible la réalisation des systèmes de radiologie digitale qui produisent directement une image plutôt qu'un film . L'image radiologique est disponible après quelques secondes, sans attendre le développement qui est nécessaire pour les clichés. Cette technologie présente comme avantages : la rapidité, la facilité d'archivage, de duplication et de transmission par réseau (télémédecine¹).

1. Prestation de soins à distance par le biais des télécommunications.

Dans le contexte de la mammographie numérique, les méthodes d'aide à la détection et à la classification de zones suspectes deviennent de plus en plus importantes. Une fois la chaîne d'acquisition est entièrement numérique, l'introduction d'un système d'aide au diagnostic ne nécessite aucun acte supplémentaire pour le médecin (tel l'introduction d'un film dans un scanner) ce qui permet un vrai gain de temps. Des systèmes CAD existent déjà sur le marché, comme [9]:

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

Le principe de ces trois systèmes est le même. Ils sont conçus pour identifier et marquer des régions d'intérêts, ils utilisent les techniques du traitement d'image couplés avec les techniques de fouille de données et de l'intelligence artificielle.

Concevoir un système de détection et de classification automatique des lésions en mammographie ne doit pas être compris comme une tentative de remplacer le médecin. La question n'est pas de se substituer en médecin, mais de lui proposer des outils performants qui l'aident dans son travail d'analyse des clichés de mammographie.

2.2 Architecture des CADs

La multiplicité des tentatives pour la conception des systèmes d'aide au diagnostic par les chercheurs fait que leurs architectures ne soient pas uniques, et dépendent de la nature de l'aide proposé par chaque chercheur. Des travaux qui visent le rehaussement des zones suspectes, d'autre pour la segmentation des images de mammographie, ainsi des travaux pour la détection des régions suspectes (présence des microcalcifications, opacités) puis la classification et la prise de décision.

H.D chang [7] présente dans son état de l'art sur les CADs l'organigramme de la figure 2.1, qui est issu de la démarche suivie par les radiologues lors de l'analyse des clichés des mammographies, acquisition d'image, détection des régions suspectes, et en fin la décision (malignité/bénignité). La détection du cancer du sein consiste à trouver des signes de malignité qui sont les opacités et les microcalcifications, dans la suite de ce chapitre nous nous intéressons aux microcalcifications.

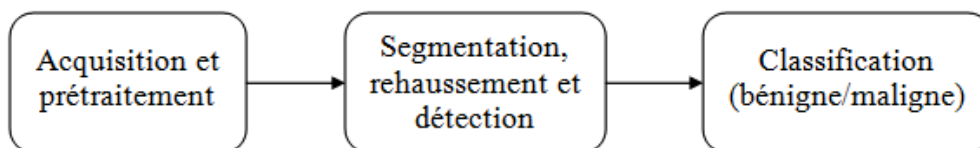


FIG. 2.1 – Architecture des CADs.

2.3 Acquisition et prétraitement

Cette phase est nécessaire pour tout système qui inclut le traitement d'image. Dans le cas des CADs, l'acquisition des images de mammographie est faite soit directement avec une radiographie digitale qui produit des images numériques, ou par numérisation des clichés classiques avec des scanners. Le prétraitement consiste à préparer les images pour le traitement, il comprend la suppression du bruit.

2.4 Rehaussement

Les microcalcifications sont plus brillantes que leurs voisinages mais sont difficilement distinguables par l'œil humaine et surtout dans les tissus denses. Le rehaussement du contraste augmente la lisibilité des images, il consiste à mettre en évidence les régions d'intérêt (microcalcifications) par manipulation du contraste et de l'intensité. On distingue dans la littérature plusieurs approches pour le rehaussement des images de la mammographie: les méthodes classiques et les approches région et statistique [7].

2.4.1 Les méthodes classiques

Les méthodes classiques de rehaussement d'images sont appliquées pour les microcalcifications, elles modifient la totalité de l'image. Nous citons ici quelques exemples tel que, l'augmentation du contraste, l'égalisation des histogrammes [10], rehaussement par les masques de convolution [11,12]. Ces méthodes rehaussent la totalité de l'image y compris le bruit et l'arrière plan et non seulement les microcalcifications, ceci est leur inconvénient majeur, alors ces méthodes sont inadaptées pour le cas des microcalcifications.

2.4.2 Approche région

Cette approche est appliquée pour la segmentation et le rehaussement d'images en générale. C'est une méthode qui peut rehausser une image sans l'introduction d'artefacts. Dans [13] l'auteur montre que la méthode peut identifier les microcalcifications dans les

images des seins denses où le contraste est peu différent entre les microcalcifications et les tissus.

2.4.3 Approche statistique

Il existe deux types de méthodes statistiques, selon leurs modifications exercées sur l'image, des méthodes qui augmentent le contraste des zones suspectes [14, 15], et d'autre suppriment l'arrière plan et le bruit [16–20]. Ces méthodes utilisent des informations statistiques des distributions (histogrammes) des niveaux de gris pour détecter des microcalcifications à rehausser. Certains chercheurs ont utilisé les ensembles flous pour la représentation des valeurs des attributs [21–23]. Dans [16, 17], les auteurs utilisent la soustraction d'une image obtenue par un filtre passe-bas de l'image originale pour supprimer l'arrière plan.

2.5 Segmentation

La segmentation en générale, consiste à diviser une image I en un ensemble de régions S_i indépendantes tel que :

$$\cup S_i = I \text{ et } S_i \cap S_j \neq \emptyset, i \neq j \quad (2.1)$$

Dans le cas des microcalcifications, la segmentation consiste à diviser l'image en des régions normales et d'autres pouvant contenir des microcalcifications. La segmentation a deux objectifs, d'une part, elle permet d'aider le radiologue à localiser les zones contenant des microcalcifications, et d'autre part, c'est une tâche nécessaire pour appliquer une méthode de classification qui décide de la bénignité ou la malignité des segments. Toute les méthodes de segmentation d'images sont appliquées sur les microcalcifications, nous citons quelques unes dans ce qui suit.

2.5.1 Méthodes statistiques

Ces méthodes utilisent des informations statistiques locales et globales, tirées de l'histogramme de l'image tel que la moyenne, l'écart type, etc. Dans [24, 25] le seuillage des histogrammes est utilisé, l'élément déterminant pour ces méthodes est le calcul du seuil, qui est souvent difficile à choisir. Une fois le seuil est calculé ces méthodes deviennent très simples, et ne nécessitent pas trop de calculs, mais ne donnent pas de bons résultats

quand l'histogramme de l'image ne contient pas de crêtes, et donnent des segments non contigües.

2.5.2 Approche région

L'approche région regroupe les pixels de propriétés similaires dans des régions homogènes. Cette approche part par un ensemble de régions initiales appelés "germes", qui grandissent par incorporation des pixels les plus similaires suivant un critère de similarité [26, 27]. Le fonctionnement des méthodes de cette approche est meilleur lorsque le critère de similarité est facile à définir, elles dépendent aussi de l'ensemble de départ et du critère d'arrêt. Ces méthodes sont coûteuses en temps de calcul et d'espace mémoire.

2.5.3 Détection de contours

La détection de contours est largement appliquée pour la segmentation des mammographies, plusieurs opérateurs ont été appliqués tel que le gradient, le laplacien morphologique, la dilatation, l'érosion, et des filtre plus compliqués avec plusieurs éléments structurants [18, 28, 29]. Les résultats de ces méthodes dépendent de la résolution de l'image et du choix des éléments structurants.

2.5.4 Autres méthodes

D'autres méthodes ont trouvé aussi leurs places dans la segmentation des mammographies tel que l'analyse multi résolution (la transformé en ondelette) [30, 31], le pouvoir discriminant de cette méthode permet de séparer les objets fins(microcalcifications) des objets larges(arrière plan). La nature floue des données calculées sur les images mammographiques et la forme irrégulière des microcalcifications induisent l'utilisation de la logique floue pour exprimer la variabilité et l'imprécision [22].

2.6 Détection

La détection des microcalcifications est un problème de reconnaissance de forme et peut se réduire en un problème de classification binaire(zone normale/microcalcification). Les méthodes de détection utilisent un ensemble d'attributs et un algorithme de classification, elles diffèrent les unes des autres selon le type des attributs et du classifieur utilisé. Les méthodes de détection sont répertoriées dans [7].

2.6.1 Template matching (appariement de modèles)

C'est la méthode classique de la reconnaissance de forme, elle est très simple à mettre en œuvre. Elle consiste à comparer un objet avec des patterns stockés dans une base. Cette méthode est appliquée dans [32] pour détecter les microcalcifications de grandes tailles. Son succès dépend de la nature du problème traité, elle est inadaptée pour les microcalcifications à cause de la grande variabilité dans la taille et la formes de ces dernières, ceci augmente la taille de la base des modèles, ce qui implique une consommation de la puissance de calcul.

2.6.2 Méthodes basées sur les attributs propres aux microcalcifications

Beaucoup de chercheurs ont utilisé les caractéristiques individuelles des microcalcifications tel que l'aire, la compacité, la circularité, l'épaisseur, l'orientation, le contraste, la luminosité, etc. Ces attributs sont calculés directement sur les images réelles et ne nécessitent aucun traitement supplémentaire, ceci présente un avantage pour ces méthodes. Dans [33], 7 attributs parmi 29 sont choisis pour évaluer différents classifieurs, le classifieur linéaire donne de meilleurs résultats. L.Zhang et al [34] utilisent deux types d'attributs, spatiales et morphologiques pour la réduction des faux positifs.

2.6.3 Méthodes basées sur l'analyse de texture

Quatre types d'attributs de texture sont utilisés:

- *Spatial gray level dependence (SGLD) features*, calculés à partir de la matrice de cooccurrence $P(i, j, d, \theta)$ qui représente la probabilité d'occurrence d'une paire de niveaux de gris (i, j) séparés par un vecteur de déplacement d et un angle de rotation θ , les attributs les plus utilisés sont l'homogénéité, le contraste, l'entropie, l'énergie, et la régularité [35–37].
- *Surround region dependence (SRD) features*, calculés à partir de la matrice de dépendance des régions voisines [15, 38], des sommes directionnelles pondérées sont calculées : horizontale, verticale, diagonale et etc.
- *Gray level run length (GLRL) features*, extraits à partir de la matrice $R(\theta) = g(i, j | \theta)$ qui indique la fréquence d'une suite de longueur j , de pixels alignés dans la direction θ ayant le niveau de gris i [15].
- *Gray level difference (GLD) features*, basés sur la probabilité d'occurrence de deux pixels ayant une différence de niveau de gris donnée et séparés par un déplacement

$(\Delta x, \Delta y)$, des attributs comme: le contraste, le second moment angulaire, l'entropie, peuvent être calculés de la matrice GLDM [15, 39].

2.6.4 Les attributs de texture multi résolutions

La transformé en ondelette est un outil très puissant pour l'analyse multi résolutions et largement appliquée à l'analyse de texture des mammographies [40–42]. Des attributs sont calculés à différent niveaux de la transformé, les plus utilisés sont l'énergie, l'entropie et la norme des coefficients d'ondelette. Les filtres de Gabor peuvent générer plusieurs images à partir de l'image originale en agissant sur les paramètres du filtre, pour chaque image filtrée des attributs de texture sont calculés [43, 44].

2.6.5 Méthodes basées sur les dimensions fractales

Des solutions utilisant les dimensions fractales² des segments d'une image mammographique sont appliquées pour la détection des microcalcifications. La dimension fractale est une valeur numérique qui caractérise la rugosité³ d'une image, les régions lisses ont la dimension fractale inférieur à celle des zones rugueuses. La dimension fractale est calculée de plusieurs façons, par exemple dans [45], l'image est représentée dans un espace 3D, deux dimensions sont les coordonnées du pixel dans l'image et la troisième dimension est son intensité. Le problème revient donc à calculer la surface des "montagnes" que contient le graphe de l'image.

2.7 Classification

Cette phase est l'étape la plus sensible dans le processus d'aide au diagnostic, elle consiste à exploiter un ensemble d'indices pour produire une décision. La malignité ou la bénignité des microcalcifications est très difficile à distinguer, surtout pour les radiologues novices. Pour concevoir des algorithmes de classification fiables, les chercheurs ont pris le chemins des techniques d'apprentissage supervisé. Cette problématique a fait l'objet de plusieurs travaux de recherche que nous allons présenter dans cette section. Nous présenterons les outils utilisés pour évaluer un classifieur, les algorithmes de classification ainsi les attributs utilisés par chaque algorithme et les résultats obtenus.

2. Un fractale est une forme géométrique récursive.

3. Métrique qui reflète l'état de surface d'un objet.

2.7.1 Evaluation des performances d'un classifieur

Une région d'une image peut être cancéreuse(positive) ou normale(négative), et le résultat d'un algorithme de classification est soit correct(vrai) ou incorrect(faux). Ceci dit que 4 résultats sont possibles comme résultats d'un classifieur qui sont présentés dans la table 2.1.

		La véritable classe	
		malade	normale
La classe prédite	malade	VP (vrai positif)	FP(faux positif)
	normale	FN (faux négatif)	VN(vrai négatif)

TAB. 2.1 – Résultats d'un classifieur

Dans la littérature, plusieurs métriques sont utilisés pour évaluer un classifieur. Le taux de reconnaissance ou bien le taux des exemples bien classés, est calculé par l'équation 2.2

$$\text{Taux de reconnaissance} = (VP + VN) / \text{le nombre totale d'exemples de teste.} \quad (2.2)$$

Cette métrique n'explique pas les taux relatifs des faux positifs et faux négatifs, c'est-à-dire deux algorithmes ayant le même taux de reconnaissance peuvent avoir les taux de vrais positifs et négatifs différents. Pour palier à ce problème deux indices sont introduits, la "sensitivité" et la "spécificité" qui représentent respectivement le taux des vrais positifs(VP/positifs) et le taux des vrai négatifs(VN/négatifs). Dans le cas d'un classifieur à sortie continue la sensibilité et la spécificité dépendent directement du seuil de discrimination, en variant ce dernier on peut obtenir différentes valeurs pour ces indices.

Pour contourner cet inconvénient la courbe ROC(Receiver Operating Characteristics) [46] est utilisée, c'est un graphe à deux dimensions, tel qu'en abscisses les taux de faux positifs(1-spécificité) et en ordonnées les taux de vrai positifs(sensitivité), calculés pour différentes valeurs du seuil de discrimination. L'aire de la surface au dessous de la courbe ROC est utilisé comme métrique d'évaluation. L'algorithme ayant un aire plus élevé est le meilleur. La figure 2.2 est un exemple de cette courbe, elle montre que M1 est meilleur que M2.

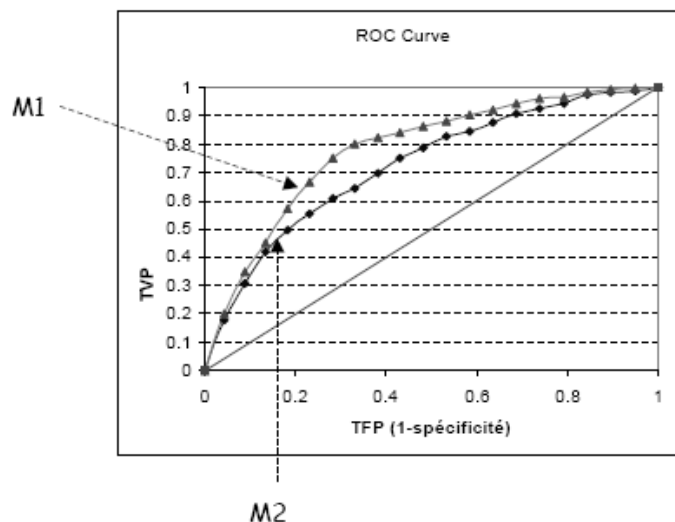


FIG. 2.2 – Courbe ROC.

2.7.2 Les réseaux de neurones

Dans [47], Verma propose une solution très simple, il utilise des images marquées par un radiologue expert, qui détecte et marque des zones suspectes dans l'image. L'expert entoure les régions d'intérêt par un cercle et mentionne son centre, puis un algorithme est utilisé pour l'extraction des zones encerclées (contenant des microcalcifications). Les valeurs des niveaux de gris des pixels sont utilisées comme entrées pour un réseau de neurone (feedforward neural network) à une seule couche cachée. Verma utilise deux algorithmes d'apprentissage différents, *error backpropagation (EBP)* par correction d'erreur qui est un algorithme itératif et *direct solution method (DSM)* une méthode non itérative. Il varie le nombre de neurones dans la couche cachée pour atteindre un taux de reconnaissance de 87%. Le système peut être utilisé pour donner une deuxième opinion pour le radiologue en cas de doute.

Verma et Zakos [48] utilisent un algorithme flou pour détecter des régions suspectes et un réseau de neurones à rétropropagation d'erreur avec trois couches, une couche d'entrée, une couche cachée et une couche de sortie. Le nombre de neurones dans les couches est ajusté selon le vecteur d'entrée. 14 attributs propres aux régions d'intérêt (niveau de gris, énergie, entropie, ...) sont testés pour définir le vecteur d'entrée pour le RdN. Ce vecteur est choisi d'une manière itérative, à chaque fois, un attribut est rajouté au vecteur, si le taux de reconnaissance augmente, l'attribut est retenu si non il est rejeté. 8192 combinaisons ont été testées pour enfin retenir la combinaison qui a donné le meilleur résultat de 89.9% de cas correctement classés.

J.C. Fu et al [49] proposent un système à deux niveaux, détection et classification.

Lors de la première phase, des opérateurs morphologiques sont appliqués pour rehausser l'image afin de détecter des régions suspectes (avec présence de faux positifs), la deuxième phase consiste à classer les régions d'intérêt en deux classes, bénigne ou maligne. L'algorithme SFS (*sequential forward search*) est utilisé pour sélectionner la combinaison optimale parmi 61 attributs. Cet algorithme commence par un ensemble vide d'attributs, à chaque itération l'attribut qui minimise l'erreur quadratique moyenne est choisi. La classification est réalisée avec deux classifieurs différents, un réseau de neurones de type *General Regression Neural Network* (GRNN) et un SVM (*support vecteur machine*). L'algorithme SFS réduit la taille du vecteur d'entrée et augmente l'aire au dessous de la courbe ROC de 97.01% à 98.00% pour SVM, et de 96.00% à 97.80% pour GRNN.

Yateen Chitre et al [50] utilisent la méthode de seuillage sur le contraste des pixels pour la segmentation et la détection des microcalcifications. Deux types de caractéristiques sont utilisées pour la classification, des attributs de texture calculés à partir de la matrice de cooccurrence (le second moment angulaire, corrélation, moyenne, l'écart type, etc.) et des attributs calculés à partir des segments (le nombre de calcifications, l'écart type de niveau de gris, la distance moyenne inter-microcalcifications, etc). Ces attributs sont injectés dans 6 différentes architectures de réseaux de neurones à rétropropagation d'erreurs. Ils obtiennent une moyenne de 0.6 pour l'aire au dessous des courbes ROC. Et dans [51] ils utilisent les attributs de texture, extraits à partir de la matrice de cooccurrence $H(y_q, y_r, d)$, ils introduisent la décomposition en ondelettes et calculent des attributs à chaque niveau (énergie et entropie à différents niveaux de la transformé en ondelettes). Ces derniers représentent le vecteur d'entrée pour un réseau de neurones à fonction de base radiale (RBF), ses performance sont comparées avec celles d'un perceptron multicouches (BNN). La courbe ROC montre que le RBF est plus performant avec un aire au dessous de la courbe de 0.74.

Yulei Jiang et al [52] ont conçu un classeur neurone dans le but d'assister le radiologue pour réduire le nombre de faux positifs (effectivement réduire le nombre d'exams par biopsie), ils utilisent 8 attributs (aire, circularité du segment, le nombre de microcalcifications dans un cluster, etc.) avec un réseau de neurones à rétropropagation d'erreurs. Pour évaluer les performances de leur système, ils le comparent avec un groupe de radiologues. Ils utilisent deux bases d'images, une contenant des images *difficult-to-detect* et obtiennent les aires ROC de 0.92 pour leur système et 0.89 pour les radiologues. Pour l'autre base contenant des images *extremely difficult-to-detect* les aires de 0.80 pour leur modèle et 0.62 pour les radiologues sont obtenus.

Dani Kramer [53] utilise trois types d'attributs, des attributs calculés à partir de la matrice de cooccurrence, des attributs calculés avec les coefficients de la transformé en

ondelette, et il propose une combinaison des deux. A chaque niveau de la transformé en ondelette les matrices de cooccurrence sont calculées pour les différents coefficients de la transformé ainsi les attributs de Haralick [54] pour chaque matrice. Un totale de 187 attributs est obtenu. Dani Kramer utilise l'algorithme SFS (*sequential forward selection*) pour la sélection des attributs qui discriminent au mieux l'espace des données. La classification est faite avec deux classifieurs différents, avec un réseau de neurone (*feed-forward neural network*) le taux de reconnaissance est de 90%, et l'algorithme des K plus proches voisins (Knn) tel qu'un taux de reconnaissance variant entre 70.08 % et 83.12% est obtenu.

Dans [55] Verma opte pour un algorithme génétique pour la sélection des attributs les plus significatifs. Il utilise la base d'images DDSM [81] (*Digital Database for Screening Mammography*), les images sont marquées par des radiologues experts, alors Verma n'a pas implémenté un algorithme de segmentation, il a utilisé les informations du fichier "OVERLAY"⁴ pour récupérer les pixels des zones suspectes. 14 attributs (nombre de pixels, niveau de gris moyens, l'écart type, l'énergie, l'entropie, etc.) sont calculés pour chacune des régions d'intérêt. L'algorithme génétique est utilisé avec un codage binaire (0 l'attribut n'est pas pris en compte, 1 l'attribut participe à la classification). A chaque itération un réseau de neurone est entraîné et testé avec les attributs sélectionnés et le taux de reconnaissance est utilisé comme fonction de fitness par l'algorithme génétique. Plusieurs combinaisons d'attributs sont sélectionnées avec un taux de reconnaissance qui dépasse 85% pour les réseaux de neurones. Verma a utilisé aussi d'autres classifieurs comme l'analyse discriminante qui donne un taux de reconnaissance de 70% et la régression logistique avec un taux de 60%. L'entraînement des réseaux de neurones à chaque itération de l'algorithme génétique pour la sélection des attribut nécessite trop de calculs, Verma introduit dans [56] un algorithme génétique pour l'apprentissage des RdN.

2.7.3 Vecteur support machine

Liyang Wei et al proposent dans [57] un SVM pour la classification des microcalcifications où la fonction de séparation est fixée à l'avance et calculée dans la phase apprentissage. Les exemples de teste sont classés avec la même fonction de séparation. Dans [58] Liyang Wei et al proposent un SVM adaptatif, ils s'inspirent du principe des Knn (K plus proches voisins), pour chaque exemple à classer la fonction de séparation est adaptée avec les propriétés de ses voisins. 12 attributs sont choisis parmi 18 avec un algorithme de sélection simple appelé *sequential backward selection (SBS)*, (la densité, la distance moyenne entre microcalcifications, l'écart type, la taille, le nombre de microcalcifications, etc.). Ces attributs constituent le vecteur d'entrée pour le classifieur. L'algorithme est testé sur

4. Fichier accompagnant les images de la base DDSM, qui contient les coordonnées des lésions.

une base d'images collectées par le département de radiologie de l'université de Chicago. Liyang Wei et al ont obtenu un aire au dessous de la courbe ROC de 0.77 pour un SVM classique et de 0.82 pour leur SVM adaptatif.

2.7.4 K plus proches voisins (Knn)

Les classeurs à base d'exemples représentatifs (K plus proches voisins) ont trouvé leurs place dans la classification des microcalcifications. Dani kramer [59] analyse la texture des régions d'intérêt extraites par des radiologues à partir des images de la base Nijmegen⁵. Il applique la transformé en ondelette (Daubechies 6, Daubechies 20 et Biorthogonal 2.8) à 4 niveaux et à chaque niveau, les attributs (énergie, entropie, et la racine carré de la norme) sont calculés pour les différents coefficients d'ondelette, en plus de ces attributs il calcule les matrices de cooccurrence de chaque région d'intérêt à chaque niveau de la transformé en ondelette pour les quatre angles (0° , 45° , 90° , 135°), puis 8 attributs parmi les 14 attributs de Haralick [54] (homogénéité, contraste, etc.) sont calculés pour chaque matrice. D.Kramer obtient 187 attributs au totale, il utilise un algorithme pour la sélection des attributs pertinents pour la classification. L'algorithme des K plus proches voisins est appliqué avec la méthode d'apprentissage *leaveout-n*, parmi les N exemples disponibles, $N-n$ sont utilisés pour l'apprentissage et les n restant pour le test, les performance de l'algorithme sont calculées par la moyenne des taux d'erreurs obtenu on variant le nombre d'exemple de teste n . Le classifieur est testé pour les deux types d'attributs séparément et pour leur combinaison. Cette dernière donne les meilleurs résultats, un taux de reconnaissance qui dépasse 90% pour les trois transformées en ondelettes.

2.8 Bilan

Dans ce chapitre nous avons présenté un état de l'art sur les systèmes d'aide au diagnostic du cancer du sein (CADs). Nous avons présenté l'architecture de ces systèmes, les travaux munis dans chaque phase et nous avons focalisé notre recherche bibliographique sur les techniques de classification des microcalcifications qui fera l'objet de notre étude. Le tableau 2.2 est un récapitulatif de la section précédente.

5. Base d'images mammographiques. Department of Radiology, University Hospital Nijmegen, The Netherlands.

Type de classeur	Type d'apprentissage	Attributs	Sélection des attributs	Résultats	Référence
Réseau de neurones	Supervisé à rétropropagation d'erreurs	Valeur des niveaux de gris des pixels		Taux de reconnaissance 87%	[47]
		Calculés de l'image originale (statistique des niveaux de gris)	Sélection itérative par teste	Taux de reconnaissance 89%	[48]
		Calculés de l'image originale (statistique des niveaux de gris)	Par un Algorithme génétique	Taux de reconnaissance 85%	[55]
		Statistique des niveaux de gris et analyse de texture		Aire au dessous de ROC 0.6	[50]
		Calculés de l'image originale (statistique des niveaux de gris)		Aire au dessous de ROC 0.92	[52]
	Réseaux à fonction de base radiale RBF	Analyse de texture et Analyse multi résolution (transformé en ondelette)		Aire au dessous de ROC 0.74	[51]

Type de classifieur	type d'apprentissage	Attributs	Sélection des attributs	Résultats	Référence
	feed-forward neural network	Analyse de texture, Analyse multi-résolution et une combinaison	SFS(sequential forward selection)	Taux de reconnaissance 90%	[53]
	General Regression Neural Network (GRNN)	Analyse de texture et des domaines spectrale et spatiale	SFS(sequential forward selection)	0.978 Aire au dessous de ROC	[49]
Vecteur support machine	Fonction de séparation adaptative	Statistiques des niveaux de gris	sequential backward selection (SBS)	0.82 Aire au dessous de ROC	[58]
K plus proche voisins	leaveout-n	Attribut de Haralick à chaque niveau de la transformé en ondelette.		Taux de reconnaissance 90%	[59]

TAB. 2.2 – Méthodes de classification des microcalcifications

On remarque que les classifieurs les plus utilisés sont les réseaux de neurones, ceci est dû à leur pouvoir d'apprentissage (mémorisation) et de discrimination. On remarque aussi que les attributs utilisés par tous les auteurs ne dépendent pas de la nature du problème à savoir les microcalcifications, se sont des caractéristiques connues dans tous les domaines du traitement d'image, tel que l'analyse de texture(homogénéité, contraste, ...), l'analyse multi-résolution (transformation en ondelette). Ceci induit un inconvénient commun pour toutes les méthodes que nous avons présenté. L'expérience du radiologue n'est prise en compte que dans la phase apprentissage des classifieurs, chose qui est insuffisante. Le choix des attributs est un élément déterminant pour la classification alors l'implication du radiologue dans la détermination et le choix des attributs peut avoir beaucoup d'intérêts, cette idée fera l'objet de notre étude.

Cancer du sein

Dans ce chapitre nous présentons l'aspect médicale du sein pour situer l'application que nous cherchons à développer dans ce contexte. Nous allons présenter l'anatomie du sein et la technique de son exploration(la mammographie). Nous allons focaliser notre étude sur le diagnostic des calcifications pour comprendre les critères et la démarche qu'utilise un radiologue pour leur diagnostic.

3.1 Anatomie du sein

L'anatomie du sein est largement traitée dans la littérature [60,61], le sein est un organe pair et globuleux situé en avant et en haut du thorax. Il contient la glande mammaire (qui se développe au moment de la puberté). 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. La figure 3.1 est une représentation simple de l'anatomie du sein.

Le sein contient de l'extérieur vers l'intérieur:

- La peau qui englobe le sein
- Le ligament de Cooper avec les crêtes de Duret
- La matrice conjonctive qui contient :
 - Les galactophores
 - Les lobules
 - Les artères
 - Les veines

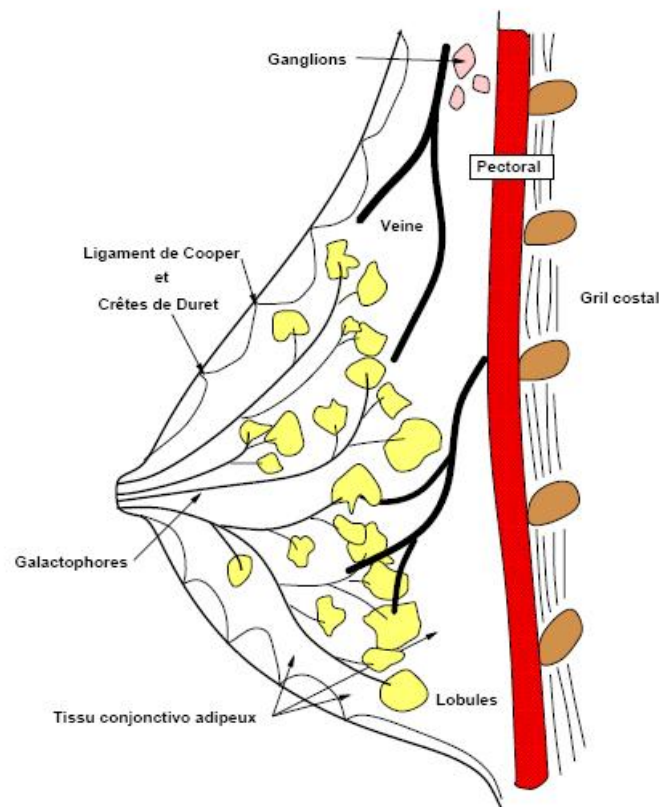
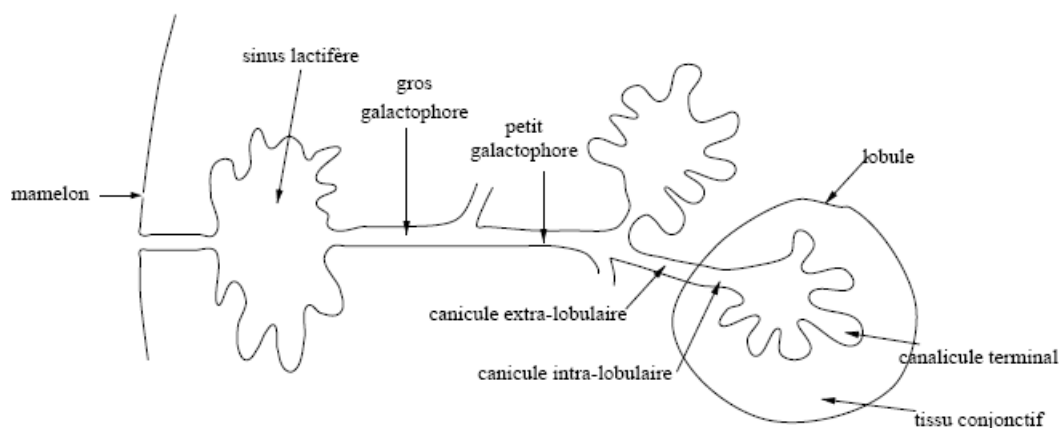


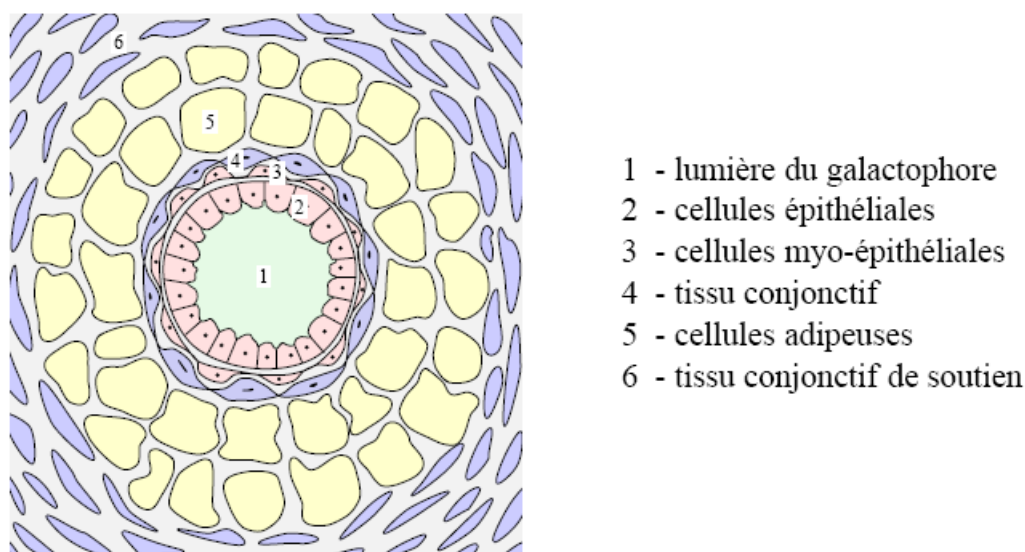
FIG. 3.1 – Anatomie du sein.

- Les nerfs
- Les ganglions
- Le pectoral et les côtes.

Pour expliquer le fonctionnement du sein, il faut regarder l'arbre galactophorique qui partent du mamelon. La structure simplifiée d'un canal galactophorique est montrée sur la figure 3.2. En partant d'un orifice galactophorique au niveau du mamelon, le canal galactophorique principal se dilate d'abord pour former une cavité appelée sinus lactifère. Ensuite, il rétrécit et se sépare en une arborescence de moyens et de petits canaux galactophoriques qui se terminent dans des lobules.

FIG. 3.2 – *Galactophore*

Un agrandissement d'une coupe au travers d'un canal galactophorique est représenté schématiquement dans la figure 3.3. L'intérieur du canal galactophorique est tapissé par les cellules épithéliales qui sont responsables de la fonction sécrétoire (production du lait). À l'extérieur de ce cylindre, il y a une membrane de cellules myo-épithéliales et du tissu conjonctif. Cet ensemble est lui-même entouré d'une couche constituée des cellules adipeuses. Ce canal galactophorique est situé dans le tissu conjonctif de soutien.

FIG. 3.3 – *Coupe d'un canal galactophorique*

3.2 La mammographie

La mammographie est une radiologie par les rayons X (section 3.2.1), elle permet une projection du volume du sein sur un plan (cliché) grâce aux différentes atténuations de

ses composants par les rayons X. C'est la technique la plus performante et la plus reproductible pour la radiographie du sein, elle permet de mettre en évidence des anomalies (opacités, calcifications). Mais souffre d'un inconvénient majeur due à cette projection, la superposition des différents tissus traversés par le même rayon et projetés sur un seul endroit du détecteur. Cette superposition est une source d'incertitude, puisqu'il n'est plus possible de distinguer sur l'image, des sur-densités réelles qui correspondent à une région radio-opaque dans le tissu et des superpositions de plusieurs tissus avec des densités relativement faibles. Des travaux de recherche se sont intéressés par la reconstruction 3D des images mammographiques [62].

3.2.1 Rayon X

Les rayons X sont comme la lumière, une forme de rayonnement électromagnétique à haute fréquence dont la longueur d'onde est comprise approximativement entre 5 picomètres et 10 nanomètres. Ils se déplacent dans le vide à la vitesse de la lumière. L'énergie de ces photons va de quelques eV (électron-volt), à plusieurs dizaines de MeV (Mega eV). C'est un rayonnement ionisant utilisé dans de nombreuses applications dont l'imagerie médicale et la cristallographie.

Les rayons X ont été découverts en 1895 par le physicien allemand Wilhelm Röntgen, qui a reçu pour cela le premier prix Nobel de physique, il les nomma ainsi car ils étaient d'une nature inconnue (la lettre X désigne l'inconnue en mathématiques).

Les rayons X sont facilement absorbés par la matière dure (matière solide dense constituée d'éléments lourds) par rapport aux matières moins denses, ceci permet d'explorer la structure anatomique interne des corps, sur un film produit par les rayons X, la luminance augmente selon la dureté de la matière bombardée, les régions brillantes représentent la matière dure (tel que les os).

3.2.2 Radiographie du sein

La mammographie est appliquée dans plusieurs cas, dépistage (prédiction), diagnostic et dans les intervention chirurgicale. Le dépistage est un moyen efficace pour lutter contre le cancer du sein. Plusieurs pays ont lancé des campagnes de dépistage, en Algérie les femme âgées de plus de 40 ans bénéficient d'un examen gratuit tout les deux ans. Deux images sont prises par sein et analysées par deux radiologues.

En cas de suspect (soit détecté par un dépistage ou par la patiente elle même) la patiente passe à une radiographie de diagnostic, dans ces conditions le but de la mammographie est d'analyser plus précisément une région du sein, et l'objectif devient alors

la classification d'une lésion plutôt que sa détection, une alternative à la mammographie, l'échographie est souvent appliquée en diagnostic, parcequ'elle permet de distinguer certains détails qui ne sont pas visibles sur un cliché de mammographie.

La mammographie est appliquée aussi lors d'une intervention chirurgicale pour localiser et calculer la position en trois dimensions d'une lésion à base de simples calculs de la géométrie.

En raison de la convexité du sein, il est impossible d'explorer sa totalité avec une seule incidence, pour ceci différentes incidences sont nécessaires lors d'une mammographie (qu'elle soit de dépistage ou de diagnostic), les incidences fondamentales sont l'incidence oblique externe, de face et profil interne (figure 3.4), d'autres incidences peuvent être prises dans des cas particuliers tel que les clichés adaptés, les clichés localisés les clichés agrandis ou localisés agrandis.

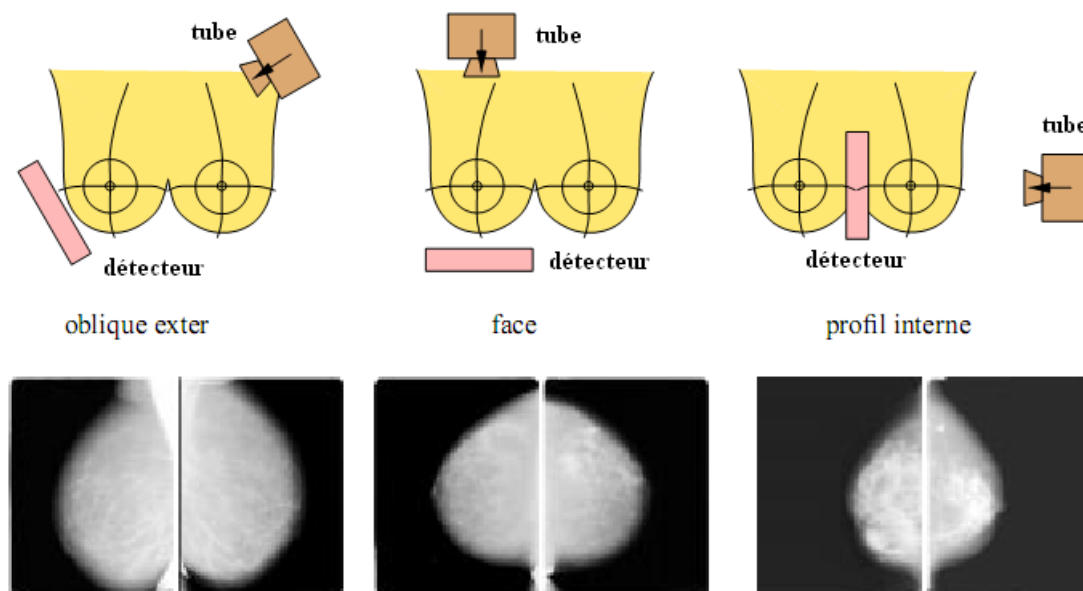


FIG. 3.4 – Incidences fondamentales en mammographie

3.3 Cancer du sein

Le mot cancer est un terme général appliqué à un grand groupe de maladies qui peuvent toucher n'importe quelle partie du corps. Cette maladie se développe quand une cellule, dont l'information génétique a été modifiée par des irradiations ou par des substances cancérogènes sans que les mécanismes de défense de l'organisme n'arrivent à la mettre sous contrôle. Dans la plupart des pays, un grand effort est fait aujourd'hui pour obtenir des statistiques fiables, afin d'en savoir plus sur les causes et les facteurs de

risques du cancer. De telles statistiques sont présentées dans la figure 3.5. Elle montrent le nombre de nouveaux cas et la mortalité pour les différents types de cancer chez les femmes en Algérie selon IARC (The International Agency for Research on Cancer) [79]. Les statistiques montrent que le cancer du sein est le cancer le plus courant chez la femme 28.3% et qu'il est aussi la première cause de mortalité par cancer 21.8%.

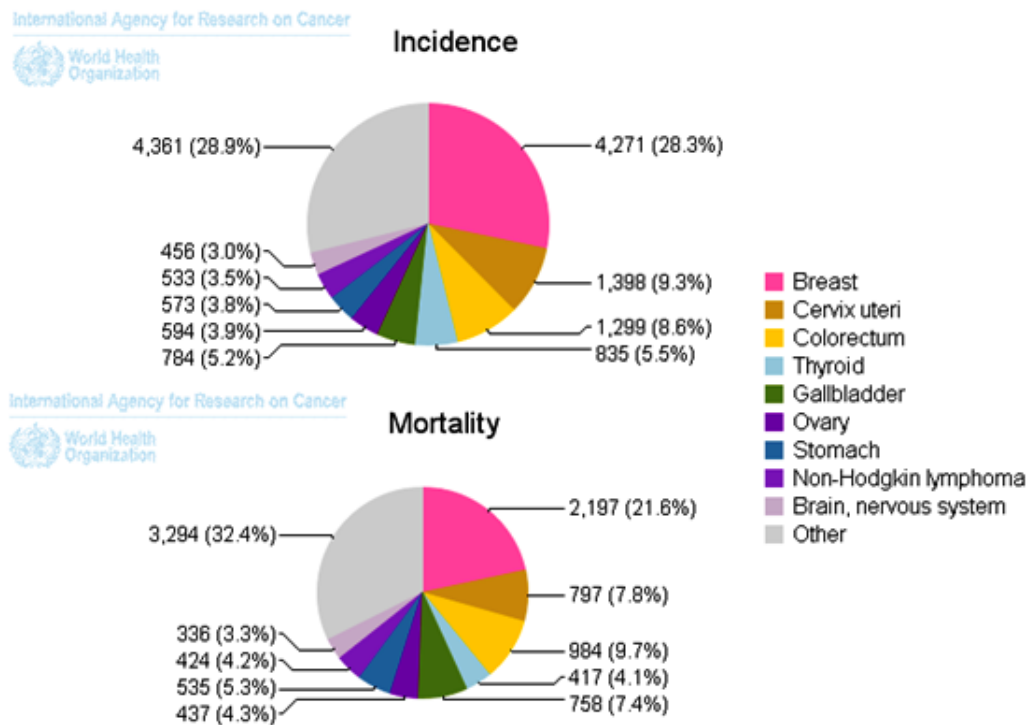


FIG. 3.5 – Statistiques des cancers les plus fréquents chez les femmes en Algérie.

La chance de survie après dix ans, tous cancers du sein confondus, est de l'ordre de 50%. Par contre, si le cancer est détecté très tôt dans son développement, c'est-à-dire quand il n'est pas encore envahissant, le taux de guérison après traitement devient proche de 100%. Cela explique l'importance de sa détection précoce. Les méthodes de détection sont l'autoexamen par la femme elle-même et des examens réguliers par le médecin.

Les signes radiologiques du cancer du sein sont : les microcalcifications et les opacités. Les microcalcifications jouent un rôle très important dans le diagnostic, elles sont formées par des dépôts calciques fortement radio-opaques et apparaissent comme des petites taches brillantes dans les mammographies, parfois elles sont visibles à la loupe. En effet la distinction des microcalcifications bénignes et malignes est très difficile, dans tel cas, la vérification par biopsie est inévitable. La classification automatique des microcalcifications peut réduire le nombre d'interventions chirurgicales qui sont dans la majorité des cas inutiles. Dans la suite de ce chapitre nous présentons le diagnostic des différents types

des microcalcifications afin d'élaborer une liste d'attributs pertinents pour la classification automatique de ces dernières.

3.4 Diagnostic des microcalcifications

Le diagnostic a pour but d'isoler les critères distinctifs qui permettent de reconnaître les lésions néoplasiques¹ à un stade précoce, infraclinique, tout en réduisant au minimum le nombre d'explorations chirurgicales. Cet objectif constituait la préoccupation de Leborgne dès 1949, qui opposait les foyers de calcifications malignes "innombrables, punctiformes, ressemblant à de fins grains de sel", aux calcifications bénignes du fibro-adménome, plus grosses, mieux visibles, parfois périphériques, capsulaires.

Assez rapidement il est apparu que cette classification était un peu trop ambiguë et souffrait de nombreuses exceptions, c'est ainsi que GeloiCohen qui en 1953 [63] souscrivait entièrement aux critères de Leborgne, fut beaucoup moins décisif quelques années plus tard [64].

Michel Benmussa et al [61] considèrent que malgré l'ambiguïté de certains cas limites, il est possible d'arriver très fréquemment au diagnostic exact en confrontant les données de l'examen clinique à l'étude de la répartition, de la disposition, du siège, du nombre, des dimensions et de l'aspect des microcalcifications ainsi que des anomalies radiologiques éventuellement associées (voir l'arbre de décision de la figure 3.6). A quelques exceptions près, ces critères sont communs à la grande majorité des auteurs [65–71].

1. Se dit du processus de formation d'un nouveau tissu.

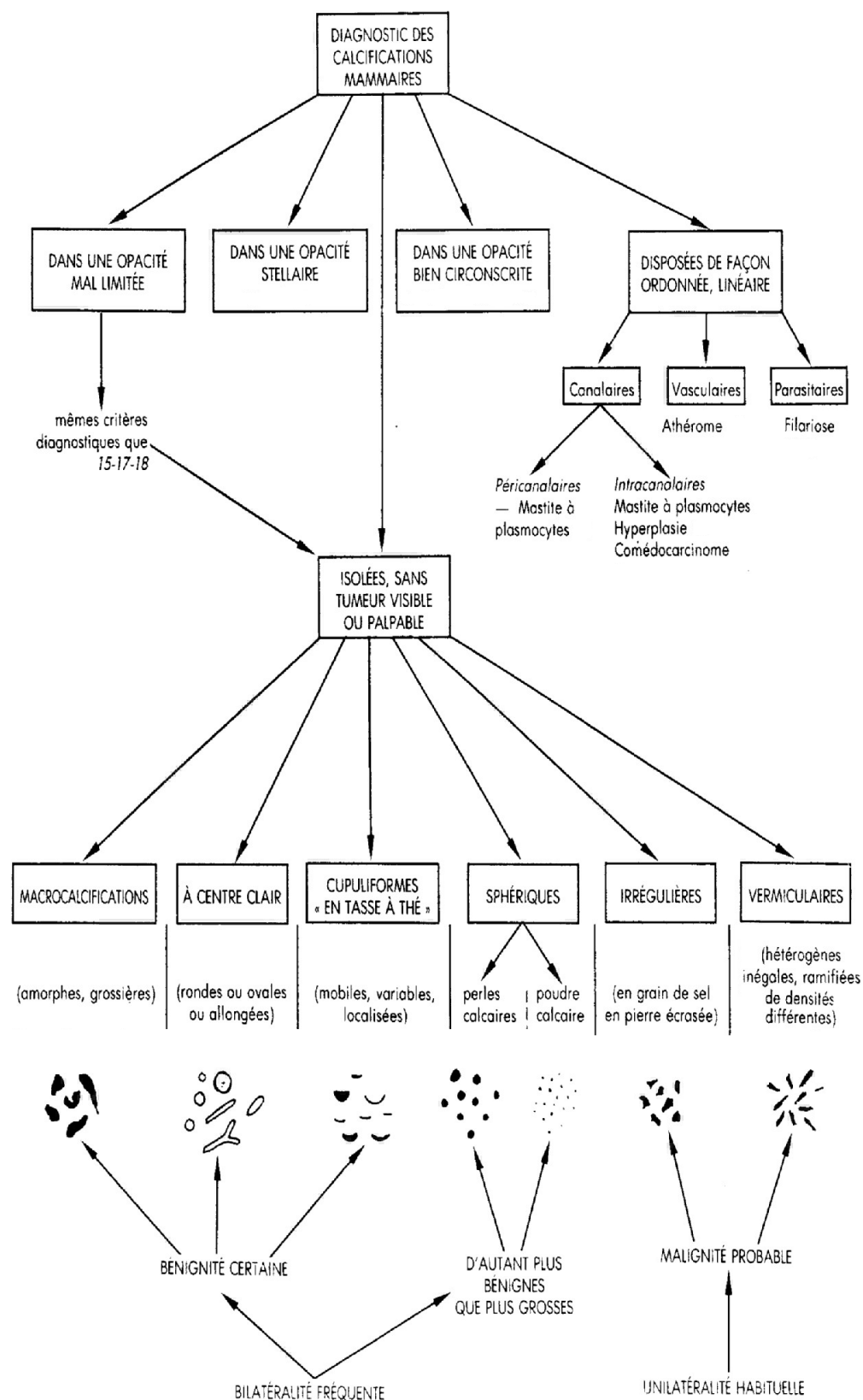


FIG. 3.6 – Diagnostic des calcifications

3.4.1 Microcalcifications groupées en foyer sans tumeur visible ou palpable

Elles sont très polymorphes. Certaines sont typiquement bénignes et ne nécessitent aucune exploration complémentaire. D'autres sont ambiguës ou franchement suspectes ce qui impose une biopsie et un examen anatomo-pathologique.

3.4.1.1 Calcifications typiquement bénignes

- **Macrocalcifications:** par convention on appelle macrocalcifications les calcifications dont les dimensions sont supérieures à 1 mm [61]. Elles ne posent aucun problème de diagnostic et sont toujours bénignes.
- **Calcifications à centre clair:** qu'elles soient rondes, ovales ou allongées et quelles que soient leurs dimensions, ces calcifications sont toujours bénignes. Les calcifications rondes à centre clair, d'origine lobulaire, se forment à la périphérie des microkystes ou des microgalactocèles (figure 3.7), leur pathogénie est la même que celle des calcifications péricanalaires ovales ou allongées de la mastite plasmocytaire, qui leur sont d'ailleurs très souvent associées (figure 3.8). Ces calcifications peuvent figurer dans une liponécrose microkystique calcifiée qui est caractérisée par l'existence de gouttelettes de graisse nécrosée, radiotransparentes, cerclées par un fin liséré calcaire (figure 3.9).

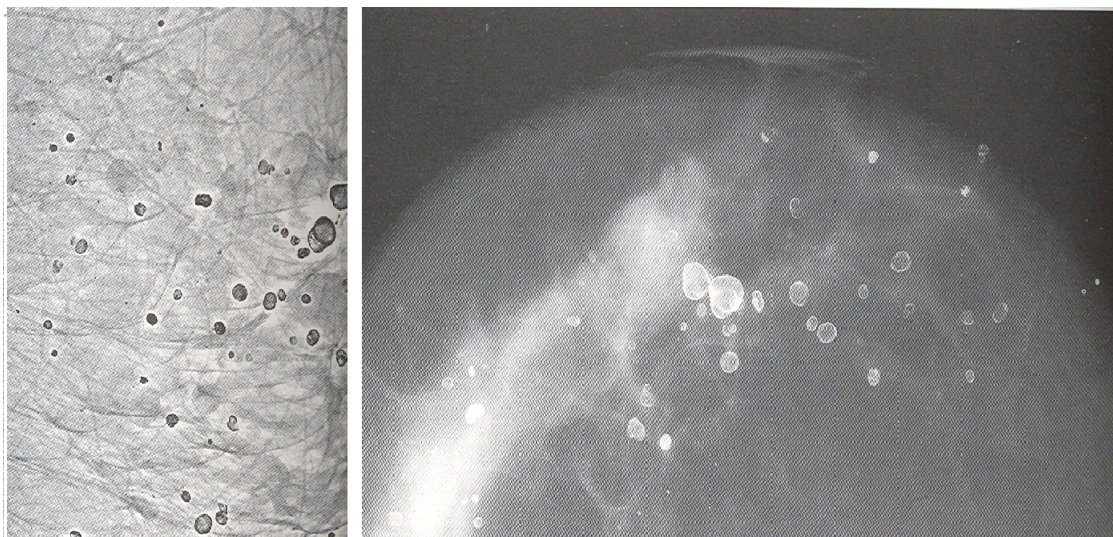


FIG. 3.7 – *Calcifications microkystiques*

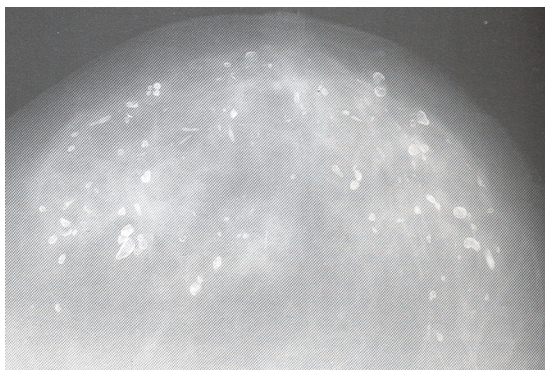


FIG. 3.8 – *Calcifications péricanalaires de la mastite plasmocytaire*

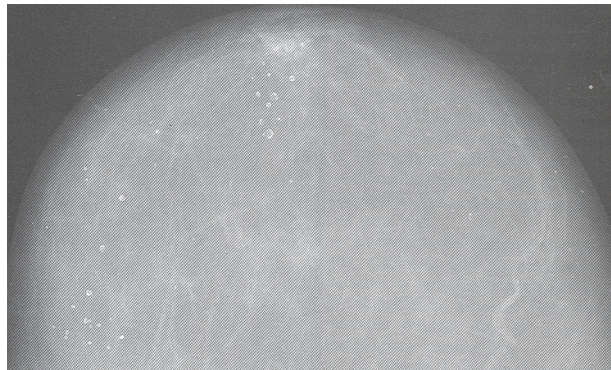


FIG. 3.9 – *Liponécrose microkystique calcifiée*

- **Calcifications semi-lunaires "en tasse à thé"**: ces images très particulières, liées à la sédimentation de calcaire en suspension au fond de microkystes, ont été décrites par Lanyi [72,73]. Elles existeraient chez 4% des femmes radiographiées [74]. On ne les repère que sur l'incidence de profil, à condition d'enregistrer celle-ci la patiente étant en situation debout ou en position assise, le faisceau de rayons X étant horizontal. Sur ce film les particules calcaires déposées au pôle inférieur de la cavité microkystique moulent son contour et forment de petites opacités linéaires ou curvilignes, à concavité supérieure (figure 3.10) ou semilunaires, "en tasse à thé" à limite supérieure horizontale. Sur l'incidence face, faisceau de rayons X vertical, les sédiments calcaires intramicrokystiques sont plus difficiles à repérer. Leur projection sur le film est arrondie, de faible tonalité. Leur contour est estompé car leur épaisseur diminue du centre vers leur périphérie.

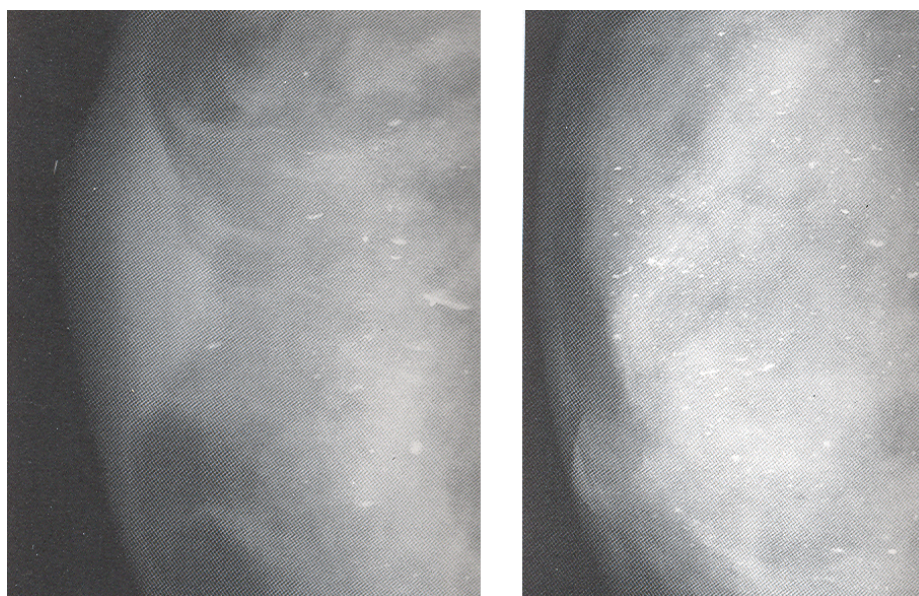


FIG. 3.10 – *Calcifications semi-lunaires "en tasse à thé"*

3.4.1.2 Microcalcifications de diagnostic difficile

Calcifications rondes, régulières. Leurs dimensions et leur morphologie sont déterminées par celles des cavités où elles se forment. La taille des microcalcifications qui présentent un cancer varie de quelques dixièmes à un millimètre environ. En présence des plus volumineuses, en forme de petites perles calcaires de densité homogène, à contours réguliers, le diagnostic de lésion bénigne est quasi certain (figure 3.11). Le diagnostic étiologique n'est ambigu que dans la mesure où les calcifications sont si petites qu'il n'est plus possible d'apprécier leur forme on ne peut alors préjuger de leur siège lobulaire ou intracanalalaire. Afin de limiter le nombre d'explorations chirurgicales négatives, il faut rechercher d'autres indices orientant vers l'une ou l'autre hypothèse.

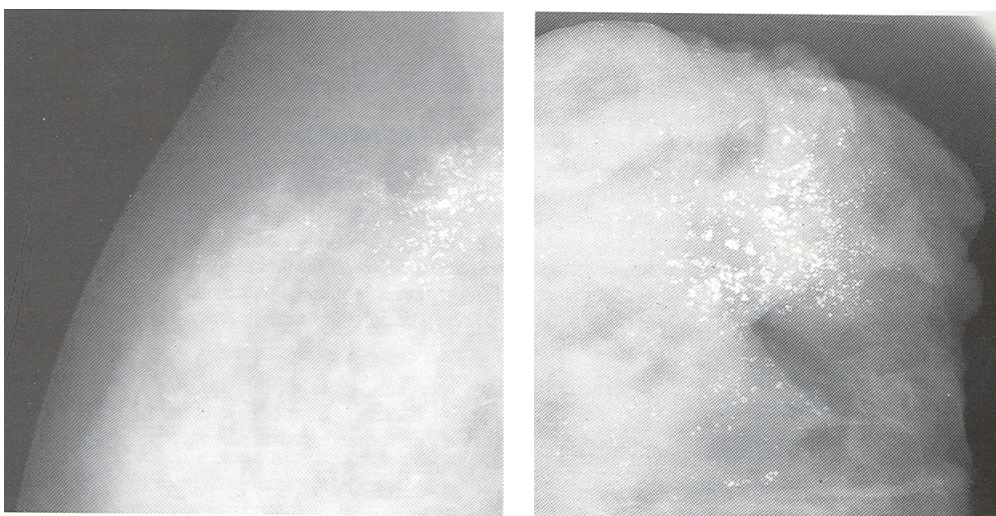


FIG. 3.11 – *Calcifications en forme de petites perles calcaires.*

- Quand les microcalcifications occupent une surface étendue, à contours irréguliers ou surtout triangulaire à *sommet orienté vers le mamelon*, on peut penser qu'elles sont intracanalaires et peuvent indiquer la présence d'un cancer (figure 3.12).



FIG. 3.12 – *Calcifications triangulaire à sommet orienté vers le mamelon.*

- Certains auteurs considèrent que les calcifications *très nombreuses, étroitement groupées* (plus de 10 dans une aire de 5 mm de diamètre) sont en majorité malignes [65, 69, 75]. Mais ces calcifications dépendent de l'aire où elles s'inscrivent et lorsqu'elles sont régulières, rondes ou ovalaires, elles sont probablement lobulaires (bénignes) [76].
- L'existence de plusieurs foyers de microcalcifications dans un même sein n'est pas un signe rassurant, 15% des cancers non palpables sont plurifocaux, 20% mesurent plus de 20 mm [77].
- L'extrême variabilité des dimensions et de la densité des calcifications est un signe en faveur du cancer [65, 66].
- La stabilité de l'image mammographique ne permet pas d'affirmer que les calcifications sont bénignes, même si la surveillance porte sur une longue période. Gershon-Cohen rapporte l'observation d'un cancer surveillé pendant plus de 5 ans, sans changement radiologique notable [78].

3.4.1.3 Microcalcifications plus typiquement malignes

Ce sont les calcifications *irrégulières*, granuleuses, et les calcifications *vermiculaires*. Elles sont en très grande majorité d'origine maligne. Elles s'observent, en règle générale, dans des carcinomes canauxaux bien différenciés. Même quand elles s'étendent à un territoire assez vaste, représentant plusieurs centimètres sur les mammographies, le cancer

reste fréquemment "in situ", strictement intragalactophorique, et donc de pronostic favorable.

- **Microcalcifications irrégulières, granuleuses:** ce sont des calcifications anguleuses, "en grain de sel", à contours aigus, de densités différentes, inégales entre elles, comparées par Hoeffken [66] aux éclats d'une pierre écrasée au marteau (figure 3.13,3.14).

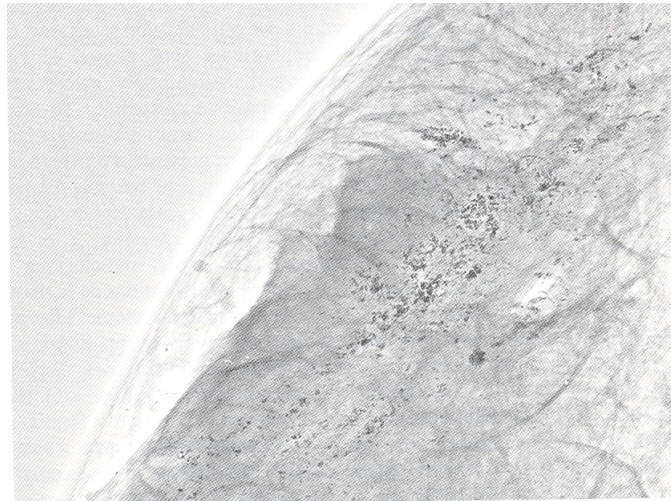


FIG. 3.13 – *Calcifications irrégulières à contours aigus.*

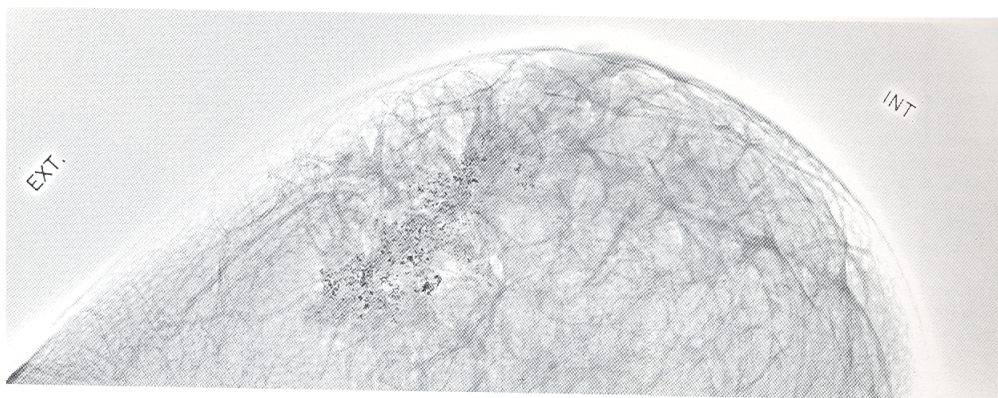


FIG. 3.14 – *Calcifications irrégulières à contours aigus, triangulaire à sommet orienté vers le mamelon.*

- **Microcalcifications vermiculaires:** (figure 3.15) ce sont des variantes des précédentes. Elles leur sont très communément associées. En forme de courts bâtonnets, curvilignes, parfois ramifiées, moulées sur un embranchement canalaire, elles ont un aspect irrégulier. De même que les calcifications "granuleuses", elles présentent des différences de densité les unes par rapport aux autres et à l'intérieur même de chaque calcification [71].

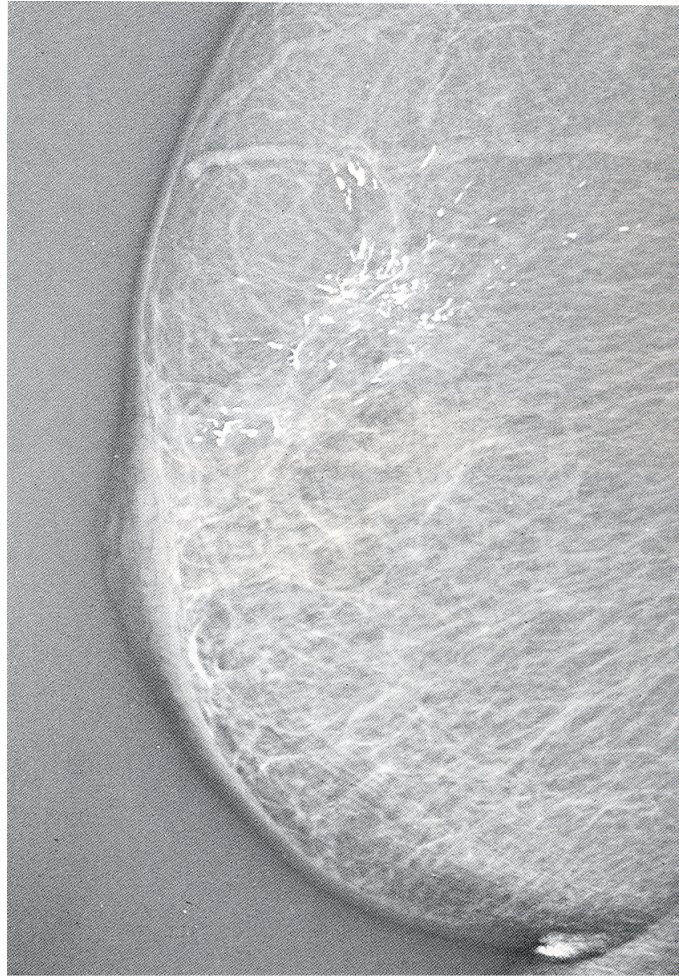


FIG. 3.15 – *Microcalcifications vermiculaires*

3.4.2 Microcalcifications éparses sans tumeur visible ou palpable

Quand les calcifications, innombrables, couvrent la totalité de l'aire mammaire ou un secteur très étendu, les critères morphologiques décrits si-dessus gardent évidemment leur valeur diagnostique, plusieurs types de lésions pouvant d'ailleurs coexister.

3.4.2.1 Calcifications bilatérales(dans les deux seins)

La bilatéralité et la diffusion des microcalcifications sont très en faveur d'une affection mammaire bénigne. Il faut savoir cependant que le comédo-carcinome peut être bilatéral, les calcifications mammaires sont alors de type malin.

- **Dysplasie fibrokystique:** c'est le processus pathologique le plus fréquent. Cet état se caractérise par de multiples modifications de l'architecture mammaire, cette anomalie est connue cliniquement sous le terme de mastose. La variante histologique

où les microcalcifications sont les plus communément observées est *l'adénose sclérosante*. Les calcifications sont arrondies, régulières, lobulaires, dispersées dans un parenchyme dense, fibroscléreux (figure 3.16). L'examen clinique de ces seins "mastotiques" est généralement malaisé.

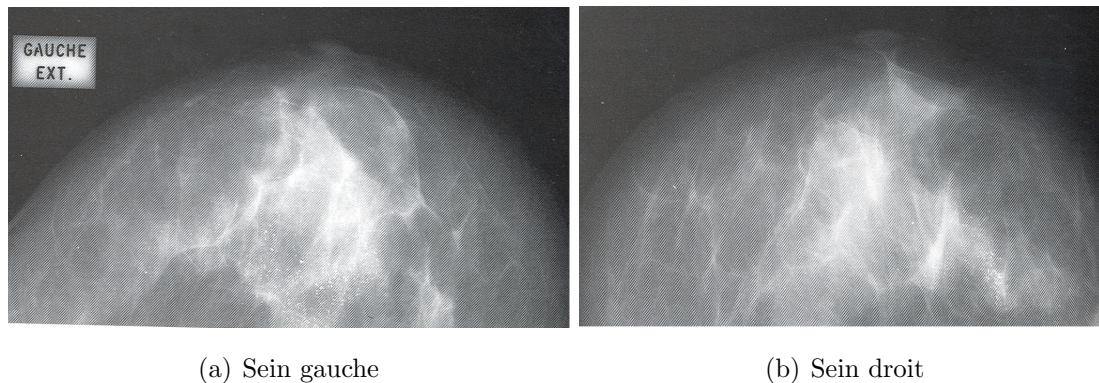


FIG. 3.16 – *Calcification bilatérales de dysplasie fibrokystique.*

- **Calcifications dermiques:** (proches de la peau) leur diagnostic est généralement évident en raison de leur siège superficiel et du contexte clinique. Quand une patiente, atteinte d'une dermite mammaire est traitée par des applications de pommade opaque aux rayons X, celle-ci produit des artefacts ressemblant à des microcalcifications diffuses, de faible tonalité. Certains tatouages peuvent également simuler des calcifications mammaires.

3.4.2.2 Calcifications unilatérales

Les calcifications éparses, unilatérales, peuvent être liées à une adénose sclérosante mais la première hypothèse à envisager en leur présence est un carcinome canalaire, même si leur morphologie est à priori rassurante. L'existence d'une galactorragie est un argument clinique majeur en faveur d'une étiologie maligne. Dans la plupart des cas les calcifications sont manifestement néoplasiques, irrégulières ou vermiculaires. L'examen anatomopathologique montre en général un comédocarcinome.

3.4.3 Calcifications disposées de façon ordonnée, linéaire

Les calcifications ordonnées linéairement sont souvent des calcifications canalaire

3.4.3.1 Calcifications canalaire

Elles se forment dans la paroi ou dans la lumière des galactophores.

- **Calcifications pariétales, péricanalaires** (figure 3.17) ces calcifications allongées, épaisses, à centre clair, orientées vers le mamelon et alignées sur le trajet d'un réseau excréteur peuvent mesurer plusieurs centimètres. Elles sont toujours bénignes.

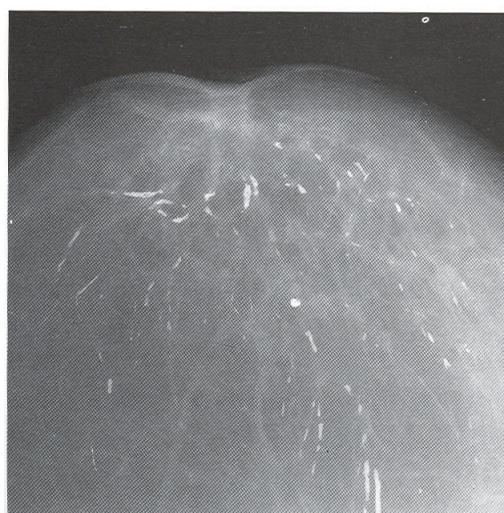


FIG. 3.17 – *Calcifications pariétales, péricanalaires.*

- **Calcifications intracanalaires**, elles moulent la lumière des galactophores, plus elles sont grosses et moins elles sont suspectes.
 - *Calcifications de la mastite plasmocytaire* (figure 3.17), plus fines et plus régulières que les calcifications péricanalaires, elles leur sont pratiquement toujours associées et se forment dans les dépôts intragalactophoriques épaissis produits par cette maladie sécrétante. Ces lithiases intragalactophoriques sont denses, longues, régulières, en segments d'aiguille brisée, parfois ramifiées, et adoptent la disposition anatomique du réseau excréteur. L'affection, pratiquement toujours bilatérale, est suspectée cliniquement en présence d'une ombilication mamelonnaire ancienne ou d'antécédents d'écoulements persistants.
 - *Calcifications de l'hyperplasie canalaire*. Elles sont assez polymorphes, associant des éléments linéaires en bâtonnets et des calcifications punctiformes, disposées en chapelet le long des galactophores (figure 3.18).

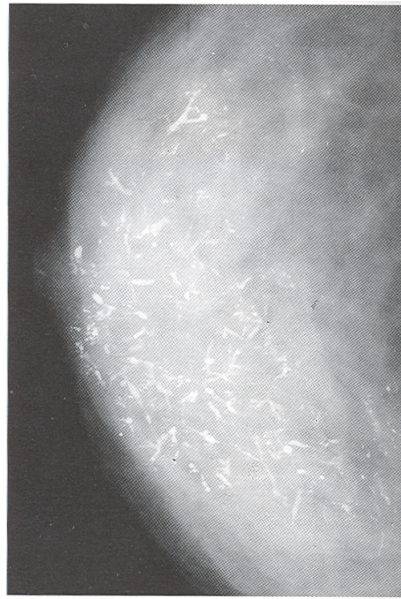


FIG. 3.18 – *Calcifications de l'hyperplasie canalaire.*

- *Calcifications du comédocarcinome.* Ces calcifications sont très nombreuses et adoptent pour la plupart une morphologie typiquement maligne, granuleuse ou vermiculaire, mais leur particularité essentielle est de dessiner les arborisations du réseau excréteur où chemine le processus tumoral. Certaines calcifications s'alignent en file indienne, grand axe orienté vers le mamelon, d'autres soulignent en pointillé le trajet du canal habité (figure 3.19 a). Les plus typiques, très fines, sablonneuses, s'agrègent et réalisent un véritable moule des segments galactophoriques dilatés (figure 3.19 b).

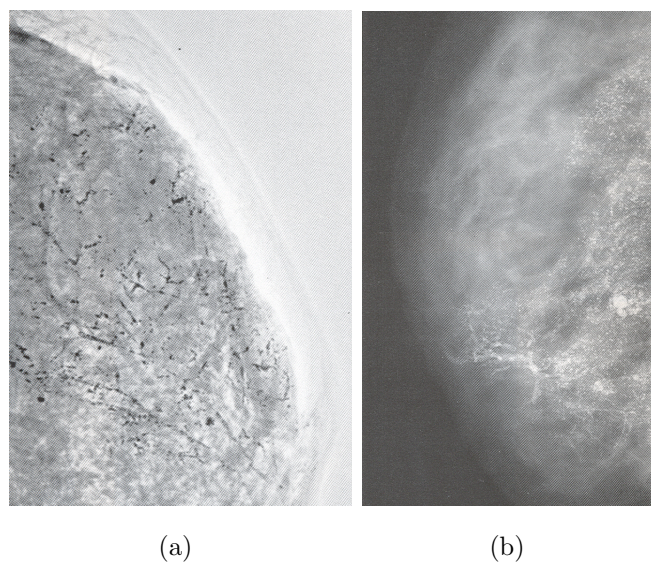


FIG. 3.19 – *Calcifications du comédocarcinome*

3.5 Conclusion

Dans ce chapitre nous avons présenté les différents types de microcalcifications existants et leurs diagnostics. Il est remarquable que la discrimination des calcifications bénignes et malignes est une tâche complexe et précise. L'extraction des attributs pour la classification automatique de ces dernières doit se baser sur les concepts médicales et les caractéristique réelles des calcifications.

Contribution et validation

Le diagnostic des microcalcifications est une tâche très complexe, ce qui complique davantage son automatisisation. Cette problématique est largement traitée dans la littérature (chapitre 3). Dans ce chapitre nous présentons notre contribution qui se porte sur la phase classification du processus d'aide au diagnostic automatique (CAD). Nous avons proposé une architecture neuro-floue pour la classification des microcalcifications en utilisant une liste d'attributs extraite par implication des spécialistes du diagnostic et d'analyse des clichés de mammographies. Dans ce chapitre nous présentons la liste des attributs que nous avons proposé, l'architecture du classifieur neuronal et la représentation de nos attributs par les ensembles flous.

4.1 Extraction des attributs

L'extraction des attributs est un élément déterminant pour la classification de données en générale et des microcalcifications en particulier. La recherche de cet ensemble est une tâche très complexe et surtout lorsqu'il s'agit des données non maîtrisées par les concepteurs des classifieurs, tel que le diagnostic des microcalcifications. Les auteurs utilisent généralement un large ensemble d'attributs issus des techniques du traitement d'image classique, puis ils appliquent un algorithme d'optimisation pour sélectionner une combinaison d'attributs qui donne le meilleur résultat. Dans ce cas l'expérience de l'expert humain n'est prise en compte que dans la phase apprentissage du classifieur.

Notre contribution est basée sur l'implication des spécialistes du diagnostic et d'analyse des clichés de mammographies afin de proposer une liste d'attributs pertinents dans le but d'améliorer la classification des microcalcifications. Pour atteindre notre objectif (liste d'attributs), nous nous sommes basés sur le diagnostic des radiologues expérimentés, qui se basent généralement sur le diagnostic de Michel Benmussa que nous avons présenté dans

le chapitre 3. Nous avons proposé une liste d'attributs qui reflètent les indices utilisés par ces radiologues pour le diagnostic des microcalcifications. Cette liste contient deux sortes d'attributs, des attributs propres à une microcalcification et des attributs concernant un foyer regroupant plusieurs microcalcifications, que nous allons détailler dans la suite de cette section.

4.1.1 Attributs propres à une microcalcification

C'est un ensemble de caractéristiques concernant une microcalcification isolée. Ces attributs concernent la taille, l'intensité, la morphologie de la microcalcification et etc.

1. L'aire

La taille d'une microcalcification est très importante pour le diagnostic, les microcalcifications ayant une taille supérieure à un millimètre sont de bénignité certaine [23]. L'aire d'une microcalcification est calculée par le nombre de pixels et la résolution de l'image.

2. Rayon maximal et rayon minimal

Ils sont respectivement la longueur du plus long et du plus court segment de droite délimités par le contour de la microcalcification et passant par son centre, figure 4.1. Les attributs R_{min} et R_{max} permettent de distinguer les calcifications canalaire des calcifications lobulaires.



FIG. 4.1 – Rayon maximal et minimal d'une calcification.

3. Rectangularité

La rectangularité est définie dans la littérature par le rapport entre l'aire de la calcification et l'aire du rectangle minimal qui l'enveloppe, dans ce cas la rectangularité ne prend pas en compte la différence entre la longueur et la largeur d'une calcification. La figure 4.2 montre que les deux calcifications (a) et (b) ont la même rectangularité.

$$rectangularite(a) = \frac{\pi R_{min} R_{max} / 4}{R_{min} R_{max}} = \pi / 4, \quad rectangularite(b) = \frac{\pi (R/2)^2}{R^2} = \pi / 4$$

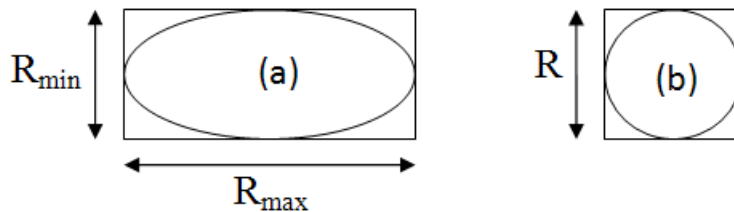


FIG. 4.2 – Rectangularité.

Pour pallier à ce problème nous calculons la rectangularité par le rapport entre le rayon maximal et le rayon minimal, la calcification est autant rectangulaire que ce rapport est proche de zéro.

$$Rectangularite = R_{min}/R_{max}. \quad (4.1)$$

4. Circularité interne et circularité externe

La circularité a un poids important dans la description morphologique des microcalcifications, elle est calculée par le rapport entre le cercle minimal qui enferme la microcalcification (C_{ext} dans la figure 4.3) et son aire, ce rapport souffre d'un inconvénient tel illustré dans la figure 4.3.

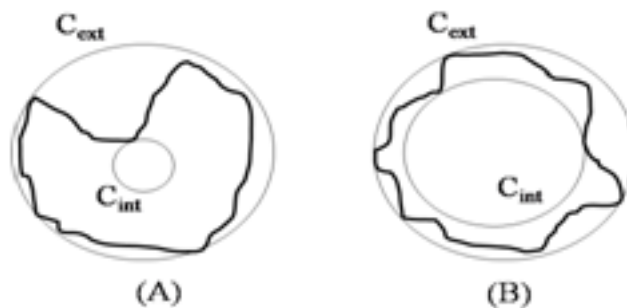


FIG. 4.3 – Circularité.

Circularité(A) = circularité(B), alors que la microcalcification(B) est plus circulaire que (A). Pour remédier à cette lacune nous avons introduit la circularité interne et la circularité externe.

– Circularité interne

La circularité interne est calculée par le rapport entre l'aire du cercle maximal qu'enferme la microcalcification et son aire.

$$Circulariteinterne = aire(Mc)/aire(Cint) \quad (4.2)$$

– **Circularité externe**

La circularité externe est calculée par le rapport entre l'aire du cercle minimal qui enveloppe la microcalcification et l'aire de celle-ci.

$$Circularitexterne = aire(Cext)/aire(Mc) \quad (4.3)$$

5. Contraste de la microcalcification

Par définition, le contraste d'une image désigne et quantifie la différence entre les parties claires et foncées d'une image. Dans notre cas les images sont en niveau de gris et le contraste va exprimer la brillance de la microcalcification par rapport à son voisinage, il est calculé par le rapport entre l'intensité moyenne(Im) de la microcalcification, et l'intensité moyenne des pixels voisins.

$$Contraste(Mc) = Im(Mc)/Im(voisinage) \quad (4.4)$$

6. Contraste du contour

Le diagnostic des calcifications à centre claire est de bénignité certaine [23]. Cette caractéristique est prise en compte par l'attribut contraste du contour qui est calculé comme suit :

$$Contraste(contour) = Im(contour)/Im(Mc) \quad (4.5)$$

7. L'écart type des niveaux de gris

L'écart type est calculé par la formule qui suit, avec N est le nombre de pixels de la calcification, I_i est l'intensité du pixel i et Im est l'intensité moyenne. Cette attribut exprime la variabilité des niveaux de gris des pixels d'une calcification.

$$E = \frac{1}{N} \sum_{i=0}^{i=N} (I_i - Im)^2 \quad (4.6)$$

8. Angularité

Les calcifications anguleuses, à contour aigus sont en majorité de malignité certaine. Cette caractéristique est prise en compte par l'attribut angularité qui est le nombre d'angles que contient le contour de la calcification. Un algorithme de détection des points forts est applicable pour le calcul de l'angularité (ex : Angularité(A)=15).

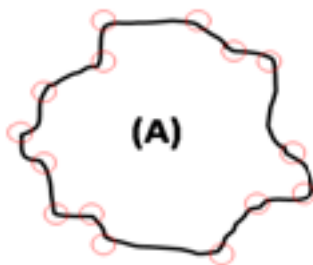


FIG. 4.4 – Angularité.

4.1.2 Attributs d'un foyer de microcalcifications

Les informations concernant un groupe de calcifications offrent des caractéristiques importantes et surtout pour les microcalcifications de diagnostic difficile.

1. Nombre de calcifications

Certains auteurs considèrent que les calcifications très nombreuses, étroitement groupées (plus de 10 dans une aire de 5 mm de diamètre) sont en majorité malignes [24-26].

2. Densité des calcifications dans un foyer

C'est le nombre de calcifications dans un foyer par rapport à la surface qu'elles occupent.

$$Densite = \frac{\text{nombre de calcifications}}{\text{aire du foyer}} \quad (4.7)$$

3. Nombre de foyers dans le sein

L'existence de plusieurs foyers de microcalcifications dans un même sein n'est pas un signe rassurant, 15% des cancers non palpables sont pluri-focaux [27].

4. L'écart type des tailles des microcalcifications

L'extrême variabilité des dimensions et de la densité des calcifications est un signe en faveur du cancer [24].

$$E_t = \frac{1}{N} \sum_{i=0}^{i=N} (Ti - Tm)^2 \quad (4.8)$$

Avec T_i est la taille de la microcalcification i , T_m est la taille moyenne des microcalcifications et N est le nombre de microcalcifications dans le foyer.

5. L'écart type des intensités des microcalcifications

$$E_{int} = \frac{1}{N} \sum_{i=0}^{i=N} (I_i - I_m)^2 \quad (4.9)$$

Avec I_i est l'intensité moyenne de la microcalcification i , I_m est la moyenne des I_i et N est le nombre des microcalcifications dans le foyer.

6. L'écart type des distances inter-microcalcifications

$$Ed = \frac{1}{N} \sum_{i=0}^{i=N} (D_i - D_m)^2 \quad (4.10)$$

Avec D_i est la distance entre la calcification i et sa plus proche calcification voisine, D_m est la moyenne des D_i et N est le nombre de calcifications dans le foyer.

7. La bilatéralité

Cet attribut est binaire. Il prend la valeur 1, si des calcifications existent dans les deux sein, sinon il prend la valeur 0. La bilatéralité et la diffusion des microcalcifications sont très en faveur d'une affection mammaire bénigne [23].

8. Distance du foyer de la peau

Le diagnostic des calcifications proches de la peau est généralement évident en raison de leur siège superficiel. Quand une patiente atteinte d'une dermite mammaire est traitée par des applications de pommade opaque aux rayons X, celle-ci produit des artefacts ressemblants à des microcalcifications, de faible tonalité. Certains tatouages peuvent également simuler des calcifications mammaires. En raison de la projection du sein sur un plan, une calcification proche de la peau peut apparaître au milieu du sein, pour ceci cette distance doit être calculées sur les deux incidences (profil et face) et la valeur minimum sera prise en compte.

4.2 Classifieur neuro-flou

Le choix d'un classifieur dépend directement du champ d'application, de la nature des données et de la classe à prédire. La classification des microcalcifications est un problème de discrimination binaire(maligne/bénigne), qui consiste à déterminer la classe d'un individu(calcification) en fonction d'un ensemble de caractéristiques qu'on appelle attributs. Notre choix est un réseau de neurones artificiels de type perceptrons multi-couche.

Les données sur lesquelles opère notre classifieur à savoir les caractéristiques (attributs) des microcalcifications sont multimodales, c'est-à-dire des données ayant des valeurs différentes(nombre de calcifications, écart type, circularité et etc). Ces données peuvent être complémentaire et parfois même contradictoires. C'est pourquoi nous proposons, à cet effet d'utiliser les ensembles flous pour prendre en compte cette richesse d'information d'un côté et de rapprocher les valeurs des attributs aux expression qu'utilisent les experts du domaine d'un autre côté(petit, grand, dense, épars, ...). L'idée est de normaliser les

attributs pour prendre valeurs dans l'intervalle $[0,1]$. La figure 4.5 illustre l'architecture de notre classifieur

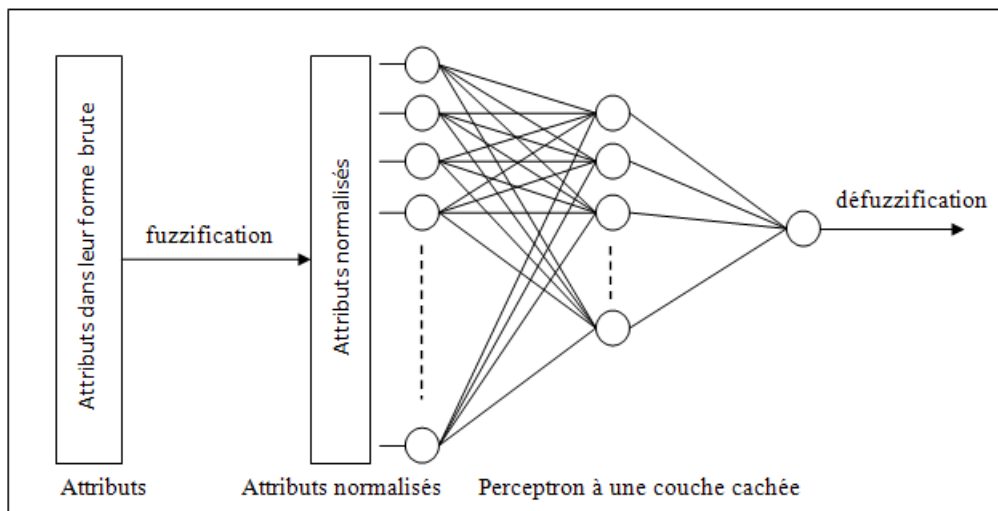


FIG. 4.5 – *Classifieur neuro-flou.*

4.3 Implémentation

Dans cette section nous présentons les outils techniques et algorithmiques utilisés pour l'implémentation et l'évaluation de notre solution à savoir la base d'images utilisée, le pré-traitement effectué sur ces images pour l'extraction des régions d'intérêt, l'algorithme de segmentation et détection des microcalcifications et nous présentons les résultats obtenus.

4.3.1 Environnement de développement

Pour le test de notre approche, nous avons utilisé l'environnement Windows, et le langage de programmation java avec l'IDE eclipse. L'implémentation du classifieur est réalisé avec l'API weka. L'ensemble des tests s'est porté sur un sous ensemble de la base DDSM(Digital Database for Screening Mammography).

4.3.2 Weka

Weka(Waikato Environment for Knowledge Analysis) [80] est une boîte à outils complète pour l'apprentissage automatique et la fouille de données développée avec java à l'université de Waikato, New Zealand. Weka est un logiciel gratuit disponible sous la licence GNU(General Public License).

Weka supporte différentes tâches d'analyse de données, prétraitement, segmentation, classification, régression et visualisation. Cette API peut être utilisée via son interface graphique ou en lignes de commandes, comme on peut faire appelle à ses différentes classes à partir de nos programme java.

Nous avons choisi d'implémenter notre algorithme avec Weka pour ses multiples avantages:

- Gratuit et disponible.
- Offre un package complet pour les réseaux de neurones.
- Utilise un format simple pour la représentation de données(fichier ARFF).
- Implémenté en java, un langage orienté objet très puissant et gratuit.

Représentation de données: Un ensemble de données(dataset) est vue par Weka comme un ensemble à deux dimensions équivalent à une table relationnelle, cet ensemble est implémenté dans Weka par la classe *weka.core.Instances*. Un dataset est une collection d'exemples, chaque exemple est une instance de la classe *weka.core.Instance*. Un objet de type *Instance* rassemble un ensemble d'attributs qui peuvent être de type *nominal*(prend ses valeurs dans un ensemble fini),*numeric*(nombre réel ou entier) ou de type *string*(chaîne de caractères). Les données d'un objet *Instances* sont exportées dans un fichier texte au format ARFF propre à weka. Dans notre cas le dataset est lu à partir d'un fichier ARFF qui est créés dans la phase calcul des attributs de notre application(voir section 4.3.7).

4.3.3 Base d'images

L'évaluation d'un algorithme de classification nécessite l'utilisation d'un référentiel appelé en informatique "*benchmark*". Un benchmark de mammographies est caractérisé par le nombre de patientes traitées, la densité optique, c-à-dire le nombre de bits par pixel(l'intervalle des niveaux de gris) et la résolution des images qui est exprimée en micron, elle désigne la taille représenté par un pixel sur le cliché réel. Dans le domaine de l'aide au diagnostic du cancer du sein il existe plusieurs bases d'images, MIAS¹(Mammographic Image Analysis Society), DDSM²(Digital Database for Screening Mammography), AMDI³(Indexed Atlas of Digital Mammograms),...

Pour le test de notre approche nous avons choisie la base d'images DDSM(Digital Database for Screening Mammography) [81], une base de mammographies digitales très

1. <http://peipa.essex.ac.uk/info/mias.html>

2. <http://marathon.csee.usf.edu/Mammography/Database.html>

3. <http://www.lcc.ufu.br/amdi/>

riche, développée en collaboration de Massachusetts General Hospital, University of South Florida, Sandia National Laboratories et Washington University School of Medicine. Cette base est destinée à la communauté scientifique pour contribuer au développement d’algorithmes d’aide à la décision et d’apprentissage automatique dans le domaine du diagnostic et détection automatiques du cancer du sein, son utilité consiste dans sa richesse d’informations, elle regroupe plus de 2500 dossiers de patientes, chaque dossier contient 4 images au format LJPEG(deux incidences par sein), un fichier .ics au format ASCII, un fichier .OVERLAY pour chaque image contenant des anomalies et un fichier image .16_PGM regroupant les 4 radios et présentant un aperçu rapide. Les dossiers de patientes(case) sont organisé dans 4 types de volumes, *Normal*, *Cancer*, *Benign* et *Benign without callback*.

– **Le fichier .ics (Exemple : A-1001-1.ics)**

```
ics_version 1.0
filename A-1001-1
DATE_OF_STUDY 29 6 1990
PATIENT_AGE 66
FILM
FILM_TYPE REGULAR
DENSITY 3
DATE_DIGITIZED 16 12 1997
DIGITIZER DBA 21
SEQUENCE
LEFT_CC LINES 3721 PIXELS_PER_LINE 1681 BITS_PER_PIXEL 16 RESOLUTION 42 OVERLAY
LEFT_MLO LINES 5311 PIXELS_PER_LINE 2116 BITS_PER_PIXEL 16 RESOLUTION 42 OVERLAY
RIGHT_CC LINES 3991 PIXELS_PER_LINE 1771 BITS_PER_PIXEL 16 RESOLUTION 42 NON_OVERLAY
RIGHT_MLO LINES 4756 PIXELS_PER_LINE 1996 BITS_PER_PIXEL 16 RESOLUTION 42 NON_OVERLAY
```

FIG. 4.6 – *Fichier ics.*

Le fichier ics contient un ensemble d’informations sur la patiente, son âge, date de l’examen, la densité des tissus, la date de numérisation des clichés, le type du scanner utilisé et les 4 dernières lignes sont des informations nécessaires pour la lecture des images LJPEG qui ne contiennent pas d’entête. Chaque ligne contient la taille de l’image(nombre de lignes et colones), le nombre de bits par pixel, la résolution de l’image en micron et la présence d’un fichier OVERLAY est indiquée. Dans le fichier présenté ci-dessus les incidences LEFT_CC et LEFT_MLO contiennent des fichiers OVERLAY.

– **Le fichier OVERLAY (Exemple : A_1001_1.LEFT_CC.OVERLAY)**


```
TOTAL_ABNORMALITIES 1
ABNORMALITY 1
LESION_TYPE CALCIFICATION TYPE PLEOMORPHIC DISTRIBUTION CLUSTERED
ASSESSMENT 4
SUBTLETY 2
PATHOLOGY MALIGNANT
TOTAL_OUTLINES 1
BOUNDARY
1192 1962 7 7 7 7 7 6 6 6 6 6 7 7 7 7 7 6 6 6 6 7 7 7 7 7 6 6 6 6
0 0 0 0 0 7 7 7 7 7 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 . . . . 7 7 7 7 7 7 7 #
```

FIG. 4.7 – *Fichier OVERLAY.*

Un fichier OVERLAY est associé à chaque image présentant des zones anormales, la première ligne désigne le nombre de zones anormales dans l'image, ensuite pour chaque anormalité des informations utiles exprimées par un radiologue expérimenté sont signalées, tel que le type de lésion, la pathologie(maligne/bénigne) et une chaîne de code après le mot clé "BOUNDARY" désignant le contour de la zone anormale dessiné par le radiologue(section 4.3.5).

- Le format LJPEG

Les images de la base DDSM sont sauvegardées dans le format Lossless JPEG [?], un format compressé sans perte de données, le fichier originale est restitué après décompression. Une fois l'image est décompressée, un fichier brut est obtenu, il contient les valeurs des pixels dans des mots de 16 bits, il n'inclut pas d'informations d'entête, à cet effet le fichier ics est nécessaire pour la connaissance de la taille de l'image et le nombre de bits significatifs de chaque pixel.

4.3.4 Organisation de l'application

L'ensemble des classes de notre application sont montrées dans le diagramme de classe simplifié de la figure 4.8. Notre application est centrée sur la classe *case.java* qui correspond à un dossier de patiente représenté dans la base DDSM par un dossier case (ex: case1001). Cette classe est instanciée pour chaque cas de la base dans le but de calculer les attributs pour la classification. Une méthode nommée *getNomAttribut()* retourne la valeur de chacun des attributs. La classe *Incidence.java* regroupe les informations concernant une image, la largeur (*width*), la hauteur (*height*), le nombre de bits significatifs et un tableau de type *Cluster*. La classe *Cluster.java* correspond à une région de l'image contenant des calcifications qui est extraite à l'aide du fichier OVERLAY. La classe *Calcification.java* implémente les méthodes qui calculent les attributs propres à une calcification. *Segmenter.java* est la classe qui implémente l'algorithme de segmentation, elle utilise la classe *Vertex.java* qui représente un sommet. La classe *Calculator.java* implémente des

méthodes de calcul tel que $\min()$, $\max()$. La classe *LJPEGReader.java* prend en charge la lecture des images JPEG, elle permet la lecture de toute ou une région de l'image avec la méthode *readRectangle(int x0, int y0, int rectHeight, int rectWidth)*. La classe *ArffManager.java* prend en charge la gestions des fichiers ARFF.

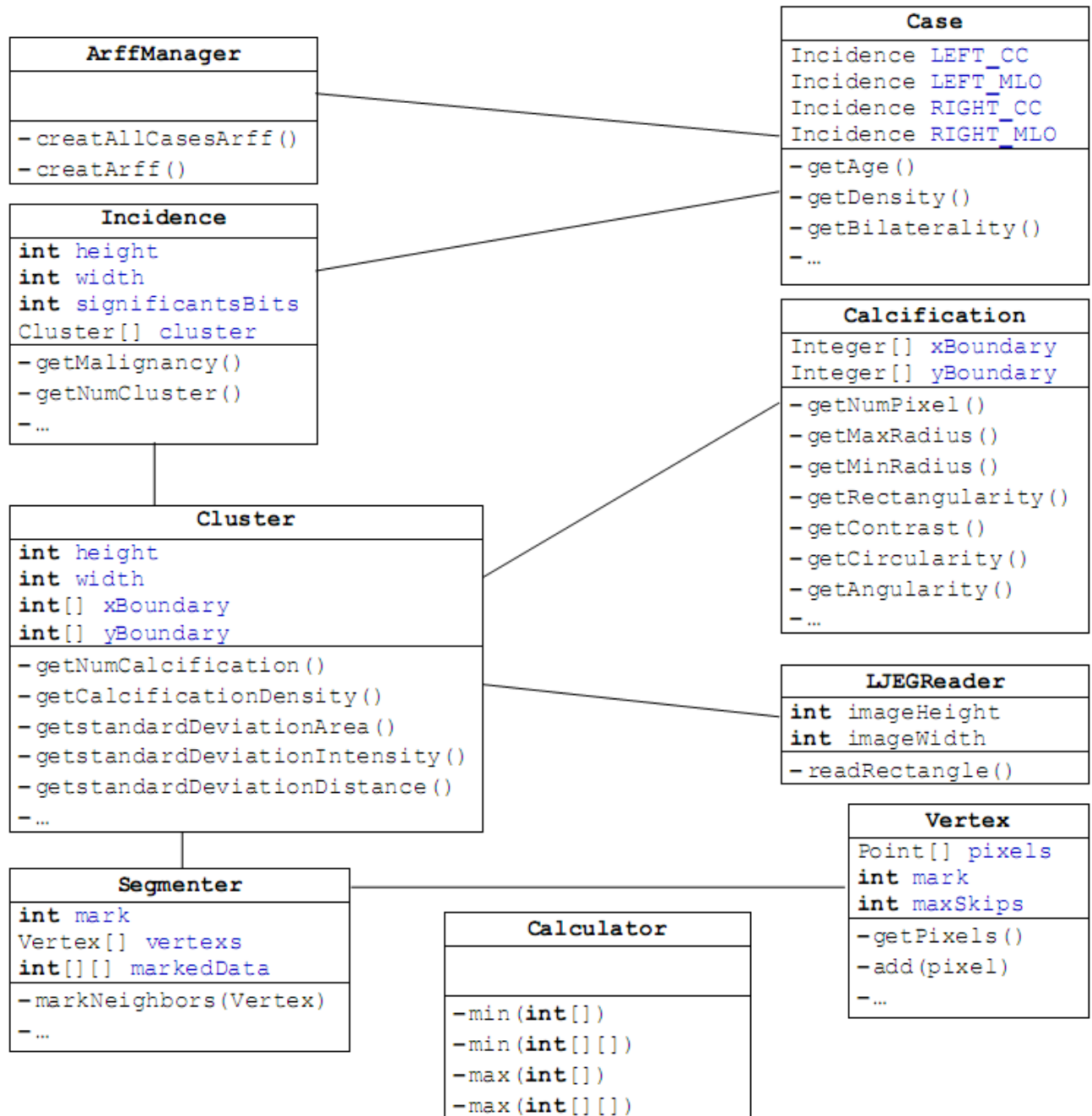


FIG. 4.8 – Diagramme de classes

4.3.5 Prétraitement

Les images de la base DDSM sont numérisées avec des résolutions entre 42 et 50 micron, la numérisation d'un cliché de 22cm sur 8cm produit une image de 5311 par 2116 pixel. Les pixels des images JPEG sont codés sur 16 bits ce qui donne une image d'environ 22 Mo. Le traitement d'un nombre important de ce type d'images est très coûteux en terme d'espace mémoire, à cet effet nous avons appliqué notre algorithme de segmentation uniquement sur une portion de l'image qui contient des calcifications(cluster), les coordonnées du contour d'une telle région sont codés dans le fichier OVERLAY avec une chaîne d'entiers après le mot "BOUNDARY". Les deux premiers nombre de la chaîne correspondent aux coordonnées du premier pixel du contour, x et y respectivement. Le reste de la chaîne(entier entre 0 et 7) montre le déplacement dans l'image pour dessiner le contour tel illustré par la figure4.9.

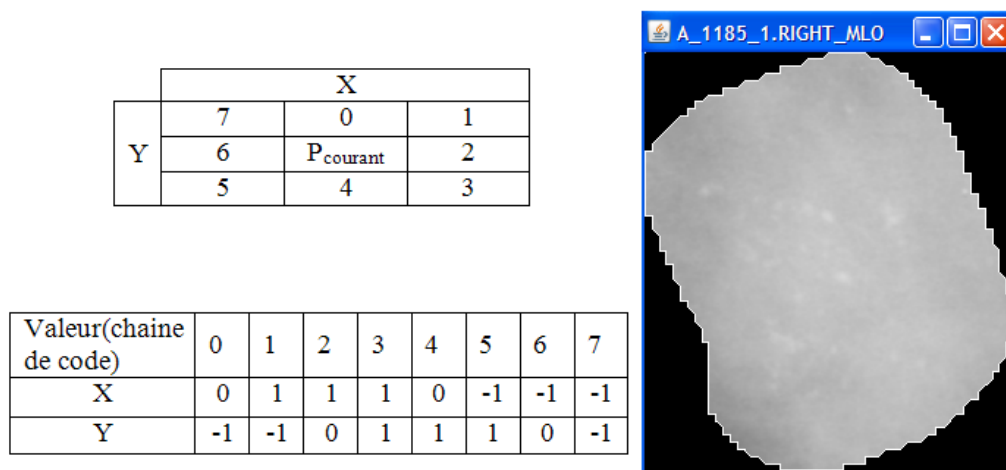


FIG. 4.9 – Prétraitement.

Les coordonnées du contour sont calculés dans le constructeur de la classe *Cluster.java* et sauvés dans les tableaux `xBoundary[]` et `yBoundary[]` (voir Annexe A).

4.3.6 Segmentation

Les calcifications sont caractérisées par la différence d'intensité par rapport à l'arrière plan, elle apparaissent dans les mammographies comme des petits points brillants. Pour extraire ces calcifications de l'image, nous proposons un algorithme hybride, l'idée est de combiner les méthodes de segmentation région(*region growing*) avec le seuillage des niveaux de gris. L'algorithme détecte les calcifications par incorporation itérative, un seuil des niveaux de gris est décrémenté à chaque itération et les pixels voisins d'une

calcification ayant leurs intensités supérieures à seuil lui sont incorporés, si le seuil est décrémenté sans incorporer de pixels, alors la calcification a atteint sa frontière.

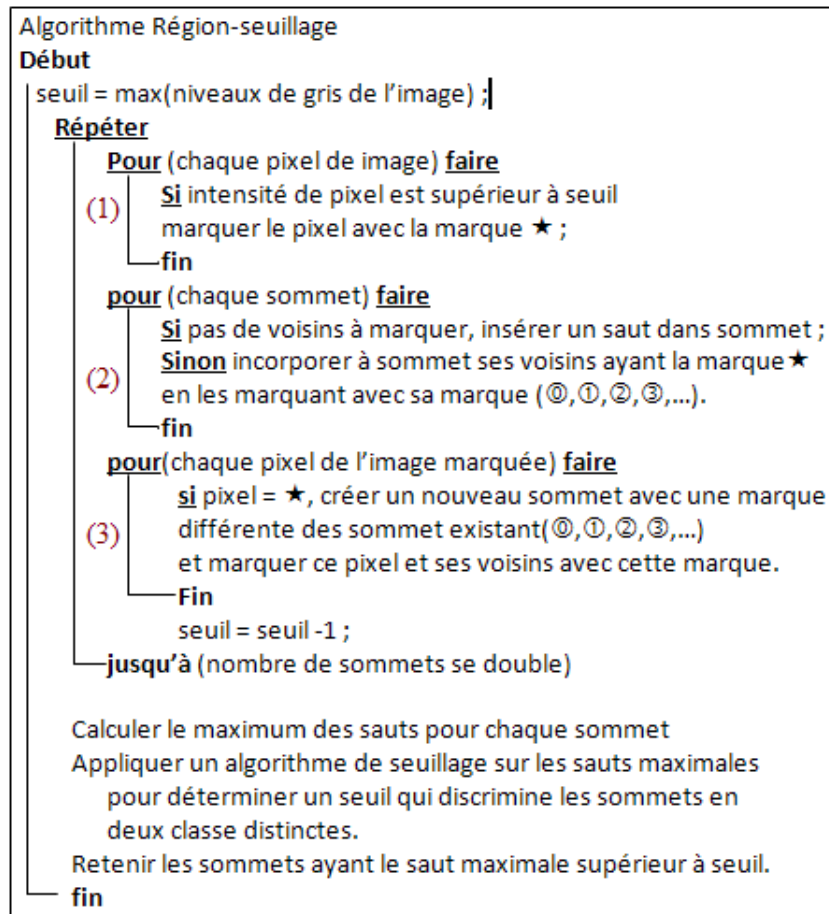


FIG. 4.10 – *Algorithme de segmentation.*

Fonctionnement: L'algorithme commence par l'initialisation du seuil au maximum des niveaux de gris de l'image, puis ce seuil est décrémenté itérativement jusqu'à ce que le nombre de sommets créés augmente brusquement, c.à.d l'algorithme crée des sommets qui appartiennent à l'arrière plan de l'image.

A chaque itération l'algorithme exécute les trois boucles (1), (2) et (3). Dans la première boucle, l'algorithme recherche les pixels ayant leurs intensités supérieures à seuil, ces pixels seront marqués avec une étoile (image(a), figure4.11). Lors de la première itération, l'algorithme exécute directement la boucle(3) sans passer par(2), car initialement aucun sommet n'est créé, pour chaque ensemble convexe de pixels marqués avec une étoile, un sommet est créé et ses pixels seront marqués avec une marque dans l'ensemble $\{1, 2, 3, \dots, n\}$ (image(b), figure4.11).

Dans les itérations qui suivent, l'algorithme procède de la même façon, la boucle(1) marque les pixels ayant leurs intensités supérieures à seuil. La boucle(2) effectue l'incorpora-

tion des pixels aux sommets, pour chaque sommet, les pixels voisins ayant la marque étoile sont incorporés à ce sommet en les marquant avec sa marque (images(c,d), figure4.11), si aucun pixel n'est incorporé au sommet alors un saut est inséré dans le sommet. La boucle(3) crée les nouveaux sommets, après l'incorporation des pixels marqués (boucle(2)), s'il existe des pixels marqués avec l'étoile, l'algorithme crée un nouveau sommet avec chaque ensemble convexe de ces pixels (images(e,f), figure4.11).

Les longueurs des sauts dans les sommets est l'élément clé de notre algorithme. Si un sommet contient un saut de longueur n cela veut dire que le seuil des niveaux de gris était décrétement de n , sans incorporer aucun pixel au sommet. Les sommets ayant des longueurs de saut importantes sont considérés comme des calcifications qui seront représentés par les pixels incorporés avant l'apparition du saut maximale. Cet algorithme est implémenté dans la classe *Segmenter.java* (Annexe A).

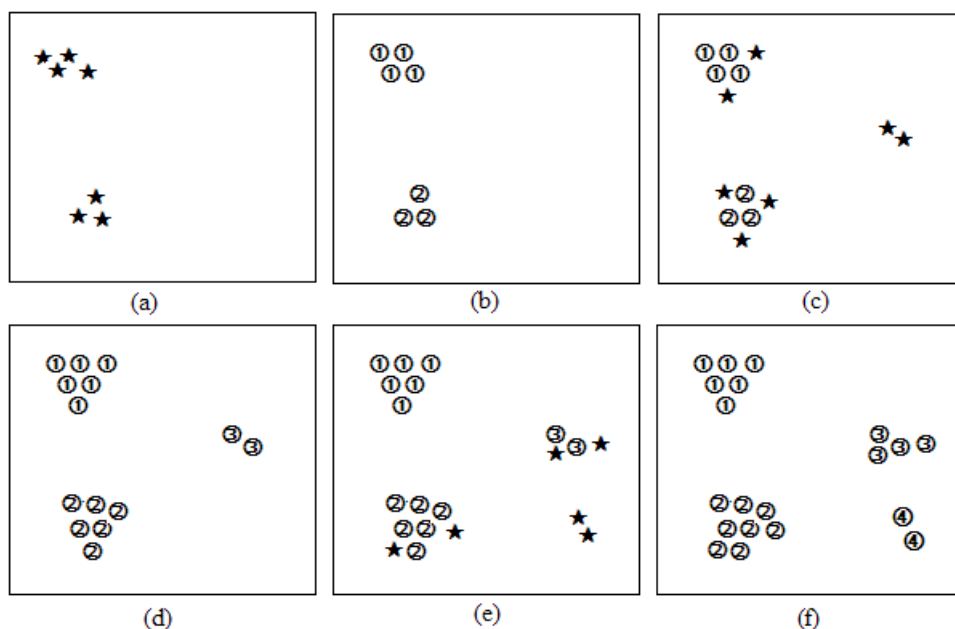
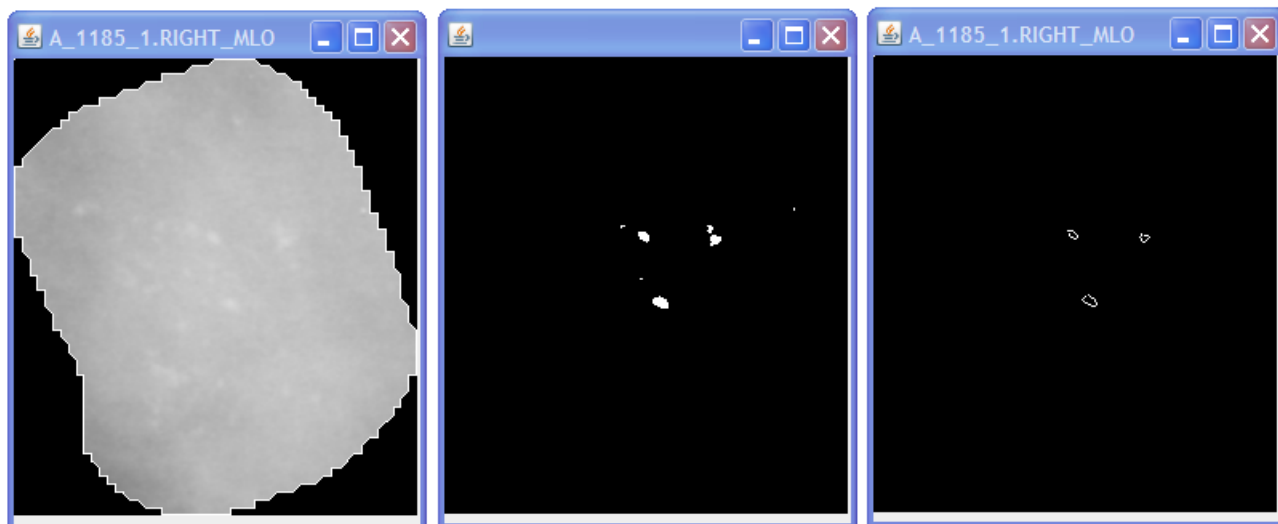


FIG. 4.11 – *Algorithme de segmentation(exécution)*.

Une fois la segmentation est terminée, la méthode *getCalcifications()* de la classe *Cluster.java*, effectue l'extraction des calcification de l'image segmentée en calculant leurs contours. La figure4.12 montre le résultat de la segmentation appliquée à "case1185" du volume "cancer_06".

FIG. 4.12 – *Résultat de la segmentation.*

4.3.7 Calcul des attributs

Les attributs que nous avons proposés ci-dessus sont calculés pour chaque calcification et chaque foyer de calcification détectés dans les images, mais une seule valeur par attribut est utilisable pour un cas de la base DDSM. Par exemple, l'attribut *aire* doit avoir une seule valeur pour refléter la taille de toutes les calcifications de l'image. Pour ceci nous avons procédé comme le radiologue. Le médecin radiologue fait un jugement sur la taille des calcifications de toute l'image (petites, grandes, fines,...). Dans notre cas nous calculons les attributs pour chaque calcification et foyer de calcification, puis la moyenne des valeurs de l'attributs est retenue.

Les attributs propres à une calcification sont calculés dans la classe *Calcification.java*. La classe *Cluster.java* effectue le calcul des attributs d'un foyer de calcification, et calcule aussi les moyennes des attributs des calcification. Les valeurs finales des attributs sont calculées dans la classe *Incidence.java*, qui calcule à son tours les moyennes des valeurs des attributs calculés pour chaque cluster. Ces valeurs sont sauvegardées dans un fichier ARFF par la méthode *creatAllCasesArff()* de la Classe *ArffManager.java* (voir Annexe A).

4.4 Résultats et discussions

Nous avons utilisé un sous-ensemble de la base DDSM sélectionné manuellement. L'ensemble d'apprentissage et des tests doit contenir chacun des types de lésion (Bénigne,

Cancer, Benign without callback). Les cas utilisés pour l'apprentissage et le test de notre approche sont répertoriés dans le tableau 4.1.

Volumes	Cases
cancer_04	case1001, case1010, case1045
cancer_06	case1108, case1113, case1131, case1133, case1141, case1148, case1152, case1153, case1167, case1185, case1188, case1201, case1212
cancer_07	case1214, case1215, case1219, case1220, case1223, case1238, case1245, case1248, case1257
cancer_08	case1282, case1415, case1470, case1500, case1508, case1535
benigne_01	case0239, case0224, case0246, case3091, case3094, case3095, case3104, case3111, case3117, case3119, case3124, case3125, case3127, case3130
benigne_08	case1565, case1705, case1706, case1707, case1713, case1734, case1735
bwc_01	case3159, case3161, case3162, case3163, case3164

TAB. 4.1 – Les cases utilisé de la base DDSM.

Les attributs de classification sont calculés à partir de l'image résultat de l'algorithme de segmentation, par conséquent, les résultats de classification dépendent fortement de l'efficacité de la segmentation. Alors nous avons effectué l'apprentissage et les tests du classifieur avec différentes configurations de l'algorithme de segmentation, en variant le critère d'arrêt pour le seuil des niveaux de gris et le taux des calcifications prise en compte après segmentation. Les résultats obtenus pour les différentes configurations sont montrés dans les tables 4.2 à 4.4.

Paramètres de segmentation		Paramètres et résultats de classification			
Seuil des niveaux de gris	Calcification prise en compte	Type d'attributs	Test	Apprentissage	Taux de classification
96%	100%	Calcification	30%	70%	47,36%
			70%	30%	35,08%
		Foyer	30%	70%	56,14%
			70%	30%	43,85%
		Tous	30%	70%	61,04%
			70%	30%	59,64%
96%	75%	Calcification	30%	70%	40,35%
			70%	30%	28,07%
		Foyer	30%	70%	38,59%
			70%	30%	24,56%
		Tous	30%	70%	57,89%
			70%	30%	45,61%
96%	50%	Calcification	30%	70%	22,80%
			70%	30%	36,84%
		Foyer	30%	70%	12,28%
			70%	30%	14,03%
		Tous	30%	70%	17,54%
			70%	30%	31,57%

TAB. 4.2 – Résultat de classification(a).

Paramètres de segmentation		Paramètres et résultats de classification			
Seuil des niveaux de gris	Calcification prise en compte	Type d'attributs	Test	Apprentissage	Taux de classification
94%	100%	Calcification	30%	70%	42,10%
			70%	30%	38,59%
		Foyer	30%	70%	26,31%
			70%	30%	24,56%
		Tous	30%	70%	33,33%
			70%	30%	42,10%
94%	60%	Calcification	30%	70%	57,89%
			70%	30%	50,87%
		Foyer	30%	70%	54,38%
			70%	30%	52,63%
		Tous	30%	70%	64,91%
			70%	30%	50,87%
94%	30%	Calcification	30%	70%	40,35%
			70%	30%	36,84%
		Foyer	30%	70%	35,08%
			70%	30%	31,57%
		Tous	30%	70%	59,64%
			70%	30%	49,12%

TAB. 4.3 – Résultat de classification(b).

Paramètres de segmentation		Paramètres et résultats de classification			
Seuil des niveaux de gris	Calcification prise en compte	Type d'attributs	Test	Apprentissage	Taux de classification
92%	100%	Calcification	30%	70%	26,31%
			70%	30%	29,82%
		Foyer	30%	70%	10,52%
			70%	30%	14,03%
		Tous	30%	70%	22,80%
			70%	30%	29,82%
92%	50%	Calcification	30%	70%	26,31%
			70%	30%	12,28%
		Foyer	30%	70%	28,07%
			70%	30%	21,05%
		Tous	30%	70%	38,59%
			70%	30%	29,82%
92%	20%	Calcification	30%	70%	57,89%
			70%	30%	52,63%
		Foyer	30%	70%	45,61%
			70%	30%	42,10%
		Tous	30%	70%	54,38%
			70%	30%	38,59%

TAB. 4.4 – Résultat de classification(c).

Les résultats obtenus montrent que le taux de classification est relatif aux paramètres de l'algorithme de segmentation. Nous avons construit le fichier ARFF qui contient les valeurs des attributs pour plusieurs combinaisons différentes des paramètres de l'algorithme de segmentation. Nous présentons ici les résultats obtenus pour les configurations suivantes:

- Décrémentement du seuil des niveaux de gris jusqu'à ce qu'il atteigne 96% du maximum des niveaux de gris, et prise en compte des taux de 100%, 75% et 50% des classifications détectées après segmentation. Cette configuration a donné un meilleur résultat de 59,65% d'images bien classées (tableau 4.2).
- Seuil des niveaux de gris décrémenté jusqu'à 94% et les pourcentages des calcifications prises en compte de 100%, 60% et 30%. Cette configuration a donné le meilleur résultat par rapport à toutes les configurations que nous avons testé, 64,91%

(tableau 4.3).

- Le tableau 4.2, présente les résultats de classification, pour un minimum du seuil des niveaux de gris de 92% et la selection des taux de 100%, 50% et 20%, de clacification considérées pour la classificaton. Le meilleur résultat de classification est un taux de 57,89%.

Notre système de classification a données des résultats plus stables avec un seuil des niveaux de gris atteignant 94% du maximu des niveaux de gris de l'image. Le meilleur résultats est donné par l'utilisation des deux types d'attributs(attributs propre à une calcification isolée et les attributs d'un foyer de calcification), avec prise en compte de 60% des calcifications détectées par l'algorithme de segmentation.

4.5 Conclusion

Dans ce chapitre nous avons exposé notre contribution qui inclut la proposition d'une liste d'attributs issue de l'expérience des experts du domaine. Cette liste d'attribut est injecté dans un classifieur neuro-flou. Les résultats présenté dans ce chapitre sont satisfaisant par rapport à la complexité du problème à savoir le diagnostic des microcalcifications.

Conclusion et perspectives

Dans ce mémoire nous avons utilisé un réseau de neurones de type perceptron multicouches pour la classification d'un corpus de données multimodales. Nous avons choisi d'opérer sur les images mammographiques. Notre objectif est de classer les microcalcifications (bénignes/malignes), signe radiologique du cancer du sein. Nous avons utilisé la notion de fonction d'appartenance des ensembles flous pour la représentation des valeurs des attributs.

Le diagnostic automatique des microcalcifications est une tâche très complexe. Elle nécessite de réunir des compétences de plusieurs disciplines. Comme nous avons montré dans l'architecture d'un système d'aide au diagnostic au chapitre 2, un système CAD est un ensemble de trois étapes indissociables, "prétraitement", "segmentation" et "classification". La réalisation d'un tel système implique la maîtrise des méthodes de segmentation d'images d'un côté et de la classification de données d'un autre côté.

Dans le but d'améliorer le taux de classification des microcalcifications nous avons proposé une liste d'attributs issus d'une étude de recueil d'informations de point de vue médicale. Le calcul de ces attributs a nécessité l'extraction des microcalcifications de l'arrière plan de l'image mammographique, pour cela nous avons proposé un algorithme de segmentation simple basé sur le seuillage des niveaux de gris.

Nos contributions présentées dans ce mémoire se résument dans les points suivants:

- Nous avons utilisé un classifieur neuro-flou pour la classification d'un corpus de données multimodales.
- Nous avons proposé et implémenté un algorithme de segmentation des images mammographiques pour l'extraction des microcalcifications.
- Nous avons proposé une liste d'attributs issus de l'expérience des radiologues experts.

Perspectives:

- Valider les attributs que nous avons proposé en les utilisant avec d'autre attributs sur un même ensemble d'images et encore sur des base d'images différentes.
- Introduire un optimiseur génétique pour élire parmi nos attributs ceux ayant le meilleure pouvoir de discrimination.
- Généraliser l'algorithme de segmentation pour opérer sur la totalité de l'image sans utilisation du fichier OVERLAY.

Bibliographie

- [1] James A.Freeman, David M.Skapura, Neural Networks Algorithms: Applications and Programming Techniques, Addison-Wesley Publishing Company, ISBN 0-201-51376-5, 1991.
- [2] McCulloch WS, Pitts W. A logical calculus of the ideas imminent in nervous activity. Bull Math Biophys 1943;5:115-33.
- [3] Rosenblatt F, The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review 65,1958,386-408.
- [4] Asli Celikyilmaz, I. Burhan Türksen, Modeling Uncertainty with Fuzzy Logic With Recent Theory and Applications, Springer, ISBN 978-3-540-89923-5.
- [5] C. M. Bishop. Neural networks for pattern recognition. Oxford University Press, 1995.
- [6] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation, Parallel Distributed Processing, volume 1, pp 318-362. Cambridge, MA : MIT Press, 1986.
- [7] H.D. Cheng, Xiaopeng Cai, Xiaowei Chen, Liming Hu, Xueling Lou, Computer-aided detection and classification of microcalcifications in mammograms: a survey, Pattern Recognition 36 (2003) 2967 – 2991.
- [8] H.D. Cheng, Juan Shan, Wen Ju, Yanhui Guo,Ling Zhang, Automated breast cancer detection and classification using ultrasound images:Asurvey, Pattern Recognition 43 (2010) 299 - 317.
- [9] J.Stines, A Noël, L.Lévy, B.Séradour, P.Heid, La mammographie numérique : technique, applications et apport de l'aide informatisé au diagnostic, Journal de Radiologie Vol 83, avril 2002,pp 581-590.
- [10] N. Karssemeijer, Adaptive noise equalization and image analysis in mammography, Information Processing in Medical Imaging: 13th International Conference, IPMI 93, AZ, USA, 1993, pp. 472–486.

- [11] M.B. McSweeney, P. Sprawls, R.L. Egan, Enhanced image mammography, *AJR* 140 (1983) 9–14.
- [12] H.P. Chan, C.J. Vyborny, H. McaMahon, C.E. Metz, K. Doi, E.A. Sickles, Digital mammography: ROC studies of the effects of pixel size and unsharp-mask filtering on the detection of subtle microcalcifications, *Invest. Radiol.* 22 (7) (1987) 581–589.
- [13] W.M. Morrow, R.B. Paranjape, R.M. Rangayyan, J.E.L. Desautels, Region-based contrast enhancement of mammograms, *IEEE Trans. Med. Imag.* 11 (3) (1992) 392–406.
- [14] M.N. Gurcan, Y. Yardimci, A.E. Cetin, R. Ansari, Detection of microcalcifications in mammograms using higher order statistics, *Signal Process. Lett.* 4 (8) (1997) 213–216.
- [15] J.K. Kim, H.W. Park, Statistical textural features for detection of microcalcifications in digitized mammograms, *IEEE Trans. Med. Imag.* 18 (3) (1999) 231–238.
- [16] X. Zhou, R. Gordon, Detection of early breast cancer: an overview and future prospects, *Crit. Rev. Biomed. Eng.* 17 (1989) 203–255.
- [17] H.P. Chan, K. Doi, C.J. Vyborny, K.L. Lam, R.A. Schmidt, Computer-aided detection of microcalcifications in mammograms—methodology and preliminary clinical study, *Invest. Radiol.* 23 (9) (1988) 664–671.
- [18] J. Dengler, S. Behrens, J.F. Desaga, Segmentation of microcalcifications in mammograms, *IEEE Trans. Med. Imag.* 12 (4) (1993) 634–642.
- [19] H. Li, K.J. Ray Liu, S.C.B. Lo, Fractal modeling and segmentation for the enhancement of microcalcifications in digital mammograms, *IEEE Trans. Med. Imag.* 16 (6) (1997) 785–798.
- [20] L.N. Mascio, J.M. Hernandez, C.M. Logan, Automated analysis for microcalcifications in high resolution digital mammograms, *SPIE Image Process.* 1898 (1993) 472–479.
- [21] H.D. Cheng, H.J. Xu, Fuzzy approach to contrast enhancement, 14th International Conference on Pattern Recognition, Brisbane, Australia, August 1998, pp. 17–22.
- [22] H.D. Cheng, Y.M. Lui, R.I. Freimanis, A novel approach to microcalcification detection using fuzzy logic technique, *IEEE Trans. Med. Imag.* 17 (3) (1998) 442–450.
- [23] H. Cheng, Y. Lui, R. Freimanis, A novel approach to microcalcification detection using fuzzy logic technique, *IEEE Trans. Med. Imag.* 17 (3) (1998).
- [24] Y. Wu, K. Doi, M.L. Giger, R.M. Nishikawa, Computer-aided detection of microcalcifications in digital mammograms, *Invest. Radiol.* 9 (1988) 664–674.
- [25] D.H. Davies, D.R. Dance, The automatic computer detection of subtle calcifications in radiographically dense breasts, *Phys. Med. Biol.* 37 (1992) 1385–1390.

- [26] K.S. Woods, C.C. Doss, K.W. Bowyer, J.L. Solka, C.E. Priebe, W.P. Kegelmeyer, Comparative evaluation of pattern recognition techniques for detection of microcalcifications in mammography, *Int. J. Pattern Recognition Artif. Intell.* 7 (1993) 1417–1436.
- [27] I.N. Bankman, T. Nizialek, I. Simon, O.B. Gatewood, I.N. Weinberg, W.R. Brody, Segmentation algorithms for detecting microcalcifications in mammograms, *IEEE Trans. Med. Imag.* 1 (2) (1997) 141–149.
- [28] D. Zhao, M. Shridhar, D.G. Daul, Morphology on detection of calcifications in mammograms, 1992 IEEE International Conference on Acoustics, Speech and Signal Processing, III, March 23–26, 1992, pp. 129–132.
- [29] H.R. Jin, Extraction of microcalcifications from mammograms using morphological filter with multiple structuring elements, *System Comput. Jpn.* 24 (11) (1993) 66–74.
- [30] L.P. Clarke, M. Kallergi, W. Qian, H.D. Li, R.A. Clark, M.L. Silbiger, Tree-structured non-linear filter and wavelet transform for microcalcification segmentation in digital mammography, *Cancer Lett.* 77 (1994) 173–181.
- [31] W. Qian, M. Kallergi, L.P. Clarke, H.D. Li, P. Venugopal, Tree structured wavelet transform segmentation of microcalcifications in digital mammography, *Med. Phys.* 22 (8) (1995) 1247–1253.
- [32] S. Morrison, L.M. Linnett, A model based approach to object detection in digital mammography, *IEEE Int. Conf. Image Process.* 2 (1999) 182–186.
- [33] S.H. Nam, J.Y. Choi, A methods of image enhancement and fractal dimension for detection of microcalcifications in mammogram, *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. 20 (2), 1998, pp. 1009–1012.
- [34] L. Zhang, W. Qian, R. Sankar, D. Song, R. Clark, A new false positive reduction method for MCCs detection in digital mammography, *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 7–11, 2001, pp. 1033–1036.
- [35] C.Y. Enderwich, E.M. Tzanakou, Classification of mammographic tissue using shape and texture features, *Proceedings of the 19th International Conference-IEEE/EMBS*, 1997, pp.810–813.
- [36] R.J. Ferrari, A.C.P.L.F. de Carvalho, P.M.A. Marques, A.F. Frere, Computerized classification of breast lesions: shape and texture analysis using an artificial neural network, *Image Process. Appl.* (1999) 517–521.
- [37] A.P. Dhawan, Y. Chitre, C. Kaiser-Bonasso, M. Moskowita, Analysis of mammographic microcalcifications using grey-level image structure features, *IEEE Trans. Med. Imag.* 15 (3) (1996) 246–259.

- [38] C.S. Lee, J.K. Kim, H.W. Park, Computer-aided diagnostic system for breast cancer by detecting microcalcification, SPIE 3335 (1998) 615–626.
- [39] J.S. Geronimo, D.P. Hardin, P.R. Massopust, Fractal functions and wavelet expansions based on several scaling functions, J. Approx. Theory 78 (1994) 373–401.
- [40] S. Yu, L. Guan, A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films, IEEE Trans. Med. Imag. 19 (2) (2000) 115–126.
- [41] A.P. Dhawan, Y. Chitre, C. Bonasso, K. Wheeler, Radial-basis-function based classification of mammographic microcalcifications using texture features, IEEE Proceedings of the 1995 IEEE Engineering in Medicine and Biology 17th Annual Conference and 21st Canadian Medical and Biological Engineering Conference, September 20–23, 1995, pp. 535–536.
- [42] H.S. Zadeh, S.P. Nezhad, F.R. Rad, Texture feature extraction methods for microcalcification classification in mammograms, Proc. SPIE 3979 (2000) 982–989.
- [43] T. Bhangale, U.B. Desai, U. Sharma, An unsupervised scheme for detection of microcalcifications on mammograms, IEEE International Conference on Image Processing, September 10–13, 2000, pp. 184–187.
- [44] G.L. Rogova, P.C. Stomper, C. Ke, Microcalcification texture analysis in a hybrid system for computer aided mammography, SPIE 3661 (1999) 1426–1433.
- [45] C.B. Caldwell, S.J. Stapleton, D.W. Holsworth, R.A. Jong, Characterization of mammographic parenchymal pattern by fractal dimension, Phys. Med. Biol. 35 (2) (1990) 235–247.
- [46] Tom Fawcett, ROC Graphs: Notes and Practical Considerations for Researchers, HP Laboratories, 2004.
- [47] B. Verma, A neural network based technique to locate and classify microcalcifications in digital mammograms, IEEE International Conference on Neural Networks—Conference Proceedings, Vol. 3, May 4–9, 1998, pp. 1790–1793.
- [48] B. Verma, J. Zakos, A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques, IEEE Trans. Inform. Technol. Biomed. 5 (1) (2001) 46–54.
- [49] J.C. Fu, S.K. Lee, S.T.C. Wong, J.Y. Yeh, A.H. Wang, H.K. Wu, Image segmentation feature selection and pattern classification for mammographic microcalcifications, Computerized Medical Imaging and Graphics 29 (2005) 419–429.
- [50] Y. Chitre, A.P. Dhawan, M. Moskowitz, Artificial neural network based classification of mammographic microcalcifications using image structure and cluster features, Proceedings of the Annual Conference on Engineering in Medicine and Biology, Vol. 15, 1993, pp. 298–309.

- [51] A.P. Dhawan, Y. Chitre, C. Bonasso, K. Wheeler, Radial-basis-function based classification of mammographic microcalcifications using texture features, IEEE Proceedings of the 1995 IEEE Engineering in Medicine and Biology 17th Annual Conference and 21st Canadian Medical and Biological Engineering Conference, September 20–23, 1995, pp.535–536.
- [52] Y. Jiang, R.M. Nishikawa, D.E. Wolverton, C.E. Metz, R.A. Schmidt, K. Doi, Computerized classification of malignant and benign clustered microcalcifications in mammograms, Proceedings of 19th International Conference-IEEE/EMBS, October 30–November 2, 1997, pp. 521–523.
- [53] Dani Kramer, Farzin Aghdasi, Texture analysis techniques for the classification of microcalcifications in digitized mammograms, Proceedings of the 1999 Fifth IEEE AFRICON Conference Electrotechnical Service for Africa, September 28 – October 1, 1999, pp. 395–400.
- [54] Robert M. Haralick, K. Shanmugam, Its' Hak Dinstein, textural features for image classification, IEEE transactions on systems, Vol SMC-3, No 6 Novembre 1973, pp.610–621.
- [55] Ping Zhang , Brijesh Verma , Kuldeep Kumar, Neural vs. statistical classifier in conjunction with genetic algorithm based feature selection, Pattern Recognition Letters 26 (2005) 909–919.
- [56] Brijesh Verma, Ping Zhang, A novel neural-genetic algorithm to find the most significant combination of features in digital mammograms, Applied Soft Computing 7 (2007) 612–625.
- [57] L. Wei, Y. Yang, R.M. Nishikawa, A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications, IEEE Trans. Med. Imaging 24(3)(2005)371–380.
- [58] Liyang Wei, Yongyi Yang, Robert M. Nishikawa, Microcalcification classification assisted by content-based image retrieval for breast cancer diagnosis, Pattern Recognition 42(2009)1126–1132.
- [59] D. Kramer, F. Aghdasi, Classification of microcalcifications in digitised mammograms using multiscale statistical texture analysis, Proceedings of the South African Symposium on Communications and Signal Processing, September 7–8, 1998, pp. 121–126.
- [60] Jean-louis lamarque, Le sein radiodiagnostic clinique, médecine et sciences internationales, ISBN 2-86439-031-0, 1981.
- [61] Henri Tristan, Michel Benmussa, Jacques Bokobsa, Mammographie et échographie mammaire de l'image au diagnostic, Flammarion Medecine-Sciences, 1989.

- [62] B. Martins, A. Smith, Z. Jing, A. Lazare and J.M. Artonne, Mammographie 3D : reduction de la dose et visualisation des structures a faible contraste, *Journal de Radiologie* Volume 87, Issue 10, October 2006, Page 1271.
- [63] GERSHON-COHEN J, INGLEBİ H, HERMEL MB. Roentgenographic diagnosis of calcification in carcinoma of the breast. *JAAIA*, 1953, 152: 676-677.
- [64] GERSHON-COHEN J, BERGER SM, CURCIO BM. Breast cancer with microcalcifications : diagnostic difficulties. *Radiology*, 1966, 87: 613-622.
- [65] EGAN RL, Mc SWEENEY MB, SEWELL CW. Intramammary calcifications without an associated mass in benign and malignant diseases. *Radiology*, 1980, 137: 1-7.
- [66] HOEFFKEN W, LANYI M. Röntgenuntersuchung der Brust. Stuttgart, Georg Thieme Verlag, 1973.
- [67] LANYI M. Microcalcifications in the breast : a blessing or a curse? A critical review. *Diagn Imag Clin Mcd*, 1985, 54: 126- 145.
- [68] LE GAL M, DURAND JC, LAURENT M, PELLIER D. Conduite à tenir devant une mammographie révélatrice de microcalcifications groupées, sans tumeur palpable. *Nouv. Presse Mcd.*, 1976, 5:1623-1627.
- [69] LE GAL M, CHAVANNE G, PELLIER D. Valeur diagnostique des microcalcifications groupées découvertes par mammographie (A propos de 227 cas avec vérification histologique et sans tumeur palpable), *Bull Cancer (Paris)*, 1984, 71: 57-64.
- [70] SICKLES EA. Mammographic detectability of breast microcalcifications. *AJR*, 1982, 139: 913-918.
- [71] TABAR L, DEAN PB. Teaching atlas of mammography. New York, Thieme-Stratton, 1985.
- [72] Lyt M. Differentialdiagnose der Mikroverkalkungen. Die verkalkte mastopathische Microcyste. *Radiologe*, 1977, 17: 217-218.
- [73] LANYI M, CITOLER P. Differentialdiagnostik der Mikroverkalkungen: die kleinzystische (blunt duct) Adenose. *Fortschr. Röntgenst*, 1981, 134: 225-231.
- [74] SICKLES EA, ABELE JS. Milk of calcium within tiny benign breast cysts. *Radiology*, 1981, 141: 655-658.
- [75] MUIR BB, LAMB J, ANDERSON TJ, KIRKPATRICK A. Microcalcification and its relationship to cancer of the breast : experience in a screening clinic. *Clin Radiol*, 1983, 34: 193-200.
- [76] LANYI M. Differentialdiagnose der Mikroverkalkungen Röntgenbild-analyse von 60 intraductalen Karzinomen, das "Dreieckprincip". *Radiologe*, 1977, 17: 213-216.

- [77] SICKLES EA. Mammographic features of 300 consecutive non palpable breast cancers. *AJR*, 1986, 146: 66 1-663.
- [78] GERSHON-COHEN J, LORNA S Yiu, BERGER SM. The diagnostic importance of calcareous patterns in roentgenography of breast cancer. *AJR*, 1962, 88: 1117-1125.
- [79] The International Agency for Research on Cancer, the World Health Organization. URL : <http://www.iarc.fr>
- [80] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; *SIGKDD Explorations*, Volume 11, Issue 1. <http://www.cs.waikato.ac.nz/ml/weka/>
- [81] Current status of the Digital Database for Screening Mammography, Michael Heath, Kevin Bowyer, Daniel Kopans, W. Philip Kegelmeyer, Richard Moore, Kyong Chang, and S. MunishKumaran, in *Digital Mammography*, 457-460, Kluwer Academic Publishers, 1998, *Proceedings of the Fourth International Workshop on Digital Mammography*.

Code source

Cette annexe contient le code java de quelques classes de notre application à savoir la classe *Cluster.java* qui permet d'extraire les régions d'intérêt de l'image, la classe *Segmenter.java*, qui implémente l'algorithme de segmentation et la classe *ArffManager.java* dont la construction du fichier ARFF est implémentée.

Communication R2I'2011

Dans cette partie nous présentons un article qui a fait l'objet d'une communication orale dans les rencontres sur la recherche en informatique, R²I'2011, 12-14 juin 2011, université de Tizi-Ouzou, Algérie. Dans cet article nous avons résumé l'essentiel de notre contribution.

Article:

Hamza GHILAS, Meriem GAGAOUA, Abdelouahab MOUSSAOUI, "*Classification des Microcalcifications: Nouvelle Approche pour l'Extraction des Attributs*", "R²I'2011", 12-14 juin 2011, Tizi-Ouzou, Algérie, 2011.

Classification des Microcalcifications: Nouvelle Approche pour l'Extraction des Attributs

Hamza GHILAS ⁽¹⁾, Meriem GAGAOUA ⁽²⁾, Abdelouahab MOUSSAOUI ⁽³⁾

Département informatique

Université Ferhat Abas, Sétif

Algérie

hamzaghilas@yahoo.fr ⁽¹⁾ gagauameriem@yahoo.fr ⁽²⁾

mousaoui.abdel@gmail.com ⁽³⁾

Résumé. L'apparition des systèmes de mammographie digitale ont rendu l'utilisation des systèmes d'aide au diagnostic plus importante. La classification des microcalcifications détectées dans les mammographies joue un rôle très important dans ces systèmes. Dans cet article, nous proposons une liste d'attributs issus de l'expérience des radiologues experts. Ces attributs sont une projection des informations qu'utilise le radiologue pour l'analyse d'une image mammographique. L'ensemble des tests s'est effectué sur la base d'images DDSM avec un classifieur neuro-flou.

Mots clés: Data mining, cancer du sein, extraction des attributs, classification des microcalcifications, aide au diagnostic.

Abstract. With the appearance of the digital systems of mammography, the computer aided diagnosis systems became more important. The microcalcifications classification plays a very important part in these systems. In this article, we propose a list of attributes resulting from the experiment of the expert radiologists. These attributes are a projection of information which the radiologist uses for the analysis of a mammographic films. The whole of tests was carried out on the DDSM database with a fuzzy-neural classifier

Key Words: Data mining, breast cancer, feature extraction, calcifications classification, computer aided diagnosis.

1 Introduction

Le cancer du sein est une maladie dangereuse, sa détection précoce est la meilleure solution de lutte, en organisant des compagnies de dépistage. Cependant la mammographie numérique et les systèmes d'aide au diagnostic (CAD, Computer Aided Diagnostic) présentent beaucoup d'intérêts. Les signes radiologiques du cancer du sein sont : les microcalcifications (Mcs) et les opacités. Les Mcs sont formées par des dépôts calciques fortement radio-opaques et la distinction entre les Mcs bénignes et les Mcs malignes est très difficile, dans tel cas, la vérification par biopsie est inévitable. La classification automatique des Mcs peut réduire le nombre d'interventions chirurgicales inutiles. La classification des Mcs est l'étape la plus sensible dans le processus d'aide au diagnostic. Elle consiste à exploiter un ensemble d'indices pour produire une décision. Les solutions proposées dans la littérature varient les unes des autres par le type du classifieur utilisé d'une part et le type d'attributs d'autre part.

Dans la suite de cet article, nous nous intéressons à la classification des Mcs. Nous présentons un état de l'art sur les attributs utilisés pour la classification des Mcs, puis nous présentons la liste d'attributs que nous avons proposé et nous finissons cet article par l'évaluation de notre approche.

2 Etat de l'art

Les attributs utilisés dans la littérature sont généralement choisis aléatoirement parmi les attributs de traitement d'images classique, puis une méthode d'optimisation est appliquée pour déterminer la combinaison d'attributs qui donne le meilleur résultat. Selon l'état de l'art de H.D.CHENG [1], ces attributs sont répartis dans trois catégories, analyse de texture, analyse multi-résolution et statistiques des niveaux de gris.

Tableau 1. Attributs des microcalcifications.

Attributs	Détails	Références
Attributs de texture multi-résolution	Attributs calculés des matrices de cooccurrences à chaque niveau de la transformé en ondelette.	[2],[3]
Attributs de la transformée en ondelettes	Energie, entropie, et norme, extrait pour chaque sous-image.	[4],[5]
Indices de Haralick	Le second moment angulaire, l'entropie, etc.	[6-10],[13]
Attributs de microcalcification isolé	L'intensité, la moyenne, la variance, l'écart type, etc.	[14],[13],[9-12]
Attributs d'un foyer de microcalcifications	Nombre de calcifications, distance inter-calcification, etc.	[13],[10-12]

3 Extraction des attributs

L'ensemble des attributs est un élément déterminant pour la classification de données en général et des Mcs en particulier. La recherche de cet ensemble est une tâche très complexe et surtout lorsqu'il s'agit des données non maîtrisées par les concepteurs des classifieurs, tel que le diagnostic des Mcs. Notre contribution est basée sur l'implication des spécialistes du diagnostic et d'analyse des clichés mammographiques afin de proposer une liste d'attributs pertinents dans le but d'améliorer la classification des Mcs. Pour atteindre notre objectif (liste d'attributs), nous avons organisé des séances de travail avec des radiologues expérimentés et nous avons proposé une liste d'attributs issus des indices utilisés par ces radiologues pour le diagnostic des Mcs. Cette liste contient deux sortes d'attributs, des attributs propres à une microcalcification (Mc) et des attributs concernant un foyer regroupant plusieurs Mcs.

3.1 Attributs propres à une microcalcification

L'aire. La taille des Mcs est très importante pour le diagnostic, les Mcs ayant une taille supérieure à un millimètre sont de bénignité certaine [15].

Rayon maximal et rayon minimal. Il sont respectivement la longueur du plus long et du plus court segment de droite délimité par le contour de la Mc et passant par son centre. Les attributs Rmin et Rmax permettent de distinguer les calcifications canalaies des calcifications lobulaires.



Fig.1 Rayon maximal et minimal d'une calcification.

Rectangularité. La rectangularité est calculé par le rapport suivant :

$$\text{Rectangularité} = R_{min} / R_{max} \quad (1)$$

Circularité interne et circularité externe. La circularité a un poids important dans la description morphologique des Mcs, elle est calculée par le rapport entre le périmètre de la Mc et son aire, ce rapport souffre d'un inconvénient tel illustré dans la figure 2.

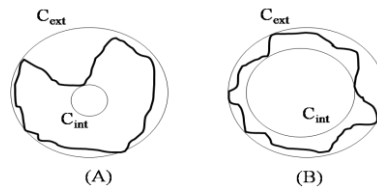


Fig.2 Circularité interne et circularité externe.

Circularité(A) = circularité(B) , alors que la Mc (B) est plus circulaire que la Mc (A). Pour remédier à cette lacune nous avons introduit la circularité interne et la circularité externe.

Circularité interne. La circularité interne est calculée par le rapport entre l'aire du cercle maximal que la Mc enferme et l'aire de la Mc.

$$\text{Circularité interne} = \text{aire}(Mc) / \text{aire}(C_{int}) \quad (2)$$

Circularité externe. La circularité externe est calculée par le rapport entre l'aire du cercle minimal qui enveloppe la Mc et l'aire de la Mc.

$$\text{Circularité externe} = \text{aire}(\text{Cext}) / \text{aire}(\text{Mc}) \quad (3)$$

Contraste de la microcalcification. Il exprime la brillance de la Mc par rapport à son voisinage, il est calculé par le rapport suivant :

$$\text{Contraste (Mc)} = \text{Im}(\text{Mc}) / \text{Im}(\text{voisinage}) \quad (4)$$

Avec Im est l'intensité moyenne.

Contraste du contour. Le diagnostic des calcifications à centre clair est de bénignité certaine [15]. Cette caractéristique est prise en compte par l'attribut contraste du contour qui est calculé comme suit :

$$\text{Contraste(contour)} = \text{Im}(\text{contour}) / \text{Im}(\text{Mc}) \quad (5)$$

L'écart type des niveaux de gris. Cet attribut exprime la variabilité des niveaux de gris des pixels de la Mc.

$$E = \frac{1}{N} \sum_{i=1}^{i=N} (I_i - I_m)^2 \quad (6)$$

Avec N est le nombre de pixels de la Mc, I_i est l'intensité du pixel i et I_m est l'intensité moyenne.

Angularité. Les calcifications anguleuses, à contour aigus sont en majorité de malignité certaine. Cette caractéristique est prise en compte par l'attribut angularité qui est le nombre d'angle que contient le contour de la Mc.

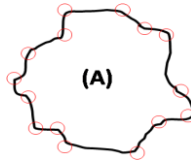


Fig.3 Angularité.

3.2 Attributs d'un foyer de microcalcifications

Nombre de calcifications. Les calcifications très nombreuses, étroitement groupées (plus de 10 dans une aire de 5 mm de diamètre) sont en majorité malignes [16-18].

Densité des calcifications dans un foyer. C'est le nombre de calcifications dans un foyer par rapport à la surface qu'elles occupent.

$$D(\text{Mc}) = \frac{\text{nombre de calcifications}}{\text{nombre de pixels du foyer}} \quad (7)$$

Nombre de foyers dans le sein. L'existence de plusieurs foyers de microcalcifications dans un même sein n'est pas un signe rassurant, 15% des cancers non palpables sont pluri-focaux [19].

L'écart type des tailles des microcalcifications. L'extrême variabilité des dimensions et de la densité des calcifications est un signe en faveur du cancer [24].

$$E = \frac{1}{N} \sum_{i=1}^{i=N} (T_i - T_m)^2 \quad (8)$$

Avec T_i est la taille de la Mc i , T_m est la taille moyenne des Mcs et N est le nombre de Mcs dans le foyer.

L'écart type des intensités des microcalcifications

$$E = \frac{1}{N} \sum_{i=1}^{i=N} (I_i - I_m)^2 \quad (9)$$

Avec I_i est l'intensité moyenne de la Mc i , I_m est la moyenne des I_i et N est le nombre des Mcs dans le foyer.

L'écart type des distances inter-microcalcifications

$$E = \frac{1}{N} \sum_{i=1}^{i=N} (D_i - D_m)^2 \quad (10)$$

Avec D_i est la distance entre la Mc i et sa plus proche Mc voisine, D_m est la moyenne des D_i et N est le nombre de Mcs dans le foyer.

La bilatéralité. Cet attribut est binaire. Il prend la valeur 1, si des calcifications existent dans les deux seins, sinon il prend la valeur 0. La bilatéralité et la diffusion des microcalcifications sont très en faveur d'une affection mammaire bénigne [15].

4 Validation

Pour le test de la pertinence des attributs que nous avons proposé, nous avons utilisé un classifieur neuro-flous. Un perceptron multicouches avec des entrées prenant des valeurs dans l'intervalle $[0,1]$, que nous avons implémenté avec Weka, une API java pour le fouille de données. L'ensemble des tests est effectué sur la base d'images DDSM [20] (Digital Database for Screening Mammography). L'apprentissage du classifieur sur chacun des sous ensembles d'attributs a donné les résultats présentés dans le tableau suivant :

Tableau 2. Résultat de classification.

Type d'attributs	Apprentissage	Test	Taux de classification
Attributs propre à une Mc.	1/3	2/3	42%
	2/3	1/3	60,20%
Attributs d'un foyer de Mcs.	1/3	2/3	35,12%
	2/3	1/3	49,56%
Tout les attributs	1/3	2/3	65,03%
	2/3	1/3	82,21%

Les résultats de classification montrent que les attributs concernant les calcifications isolées ont un pouvoir de discrimination plus important que les attributs d'un foyer de Mcs, et ceci quelque soit la répartition de la base d'images pour l'apprentissage et les tests. La réunion des deux types d'attributs a donné des résultats plus satisfaisant.

5 Conclusion

La liste d'attributs proposée dans cet article est issue d'une étude de recueil d'informations du point de vue médical. Ces attributs peuvent être utilisés comme vecteur d'entrée pour n'importe quel type de classifieur (Support Vecteur Machine, K plus proche voisins, Réseau de neurones, Arbres de décision, Système immunitaire artificiels et etc).

Nous envisageons d'utiliser nos attributs à la fois avec les autres types d'attributs (texture, ondelettes, ...) et d'appliquer un optimiseur génétique pour élire les attributs qui donnent de bons résultats. Ceci nous permettra d'évaluer le pouvoir de discriminations de nos attributs par rapport à ceux connues déjà dans la littérature.

Références

1. H.D. Cheng, Xiaopeng Cai, Xiaowei Chen, Liming Hu, Xueling Lou, Computer-aided detection and classification of microcalcifications in mammograms: a survey, *Pattern Recognition* 36, 2003, 2967 – 2991.
2. D. Kramer, F. Aghdasi, Classification of microcalcifications in digitised mammograms using multiscale statistical texture analysis, *Proceedings of the South African Symposium on Communications and Signal Processing*, 1998, pp. 121–126.
3. Dani Kramer, Farzin Aghdasi, Texture analysis techniques for the classification of microcalcifications in digitized mammograms, *Proceedings of the 1999 Fifth IEEE AFRICON Conference Electrotechnical Service for Africa*, 1999, pp. 395–400.
4. A.P. Dhawan, Y. Chitre, C. Bonasso, K. Wheeler, Radial-basis-function based classification of mammographic microcalcifications using texture features, *IEEE Proceedings of the 1995 IEEE Engineering in Medicine and Biology 17th Annual Conference and 21st Canadian Medical and Biological Engineering Conference*, 1995, pp.535–536.
5. C.M. Kocur, S.K. Gogers, L.R. Myers, T. Burns, M. Kabrisky, J.W. Holmeister, K.W. Bayer, J.M. Steppe, Using neural networks to select wavelet features for breast cancer diagnosis, *IEEE Eng. Med. Biol.* 15 (3), 1996, 95–102.
6. A.P. Dhawan, Y. Chitre, M. Moskowitz, Artificial neural network based classification of mammographic microcalcifications using image structure features, *Proc. SPIE* 1905, 1993, 820–831.
7. A.P. Dhawan, Y. Chitre, C. Bonasso, K. Wheeler, Radial-basis-function based classification of mammographic microcalcifications using texture features, *IEEE Proceedings of the 1995 IEEE Engineering in Medicine and Biology 17th Annual Conference and 21st Canadian Medical and Biological Engineering Conference*, 1995, 535–536.
8. C.M. Kocur, S.K. Gogers, L.R. Myers, T. Burns, M. Kabrisky, J.W. Holmeister, K.W. Bayer, J.M. Steppe, Using neural networks to select wavelet features for breast cancer diagnosis, *IEEE Eng. Med. Biol.* 15 (3), 1996, 95–102.
9. A.P. Dhawan, Y. Chitre, M. Moskowitz, E. Gruenstein, Classification of mammographic microcalcification and structural features using an artificial neural network, *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. 13(3), 1991, 1105 –1106.
10. H.S. Zadeh, S.P. Nezhad, F.R. Rad, Shape-based and texture-based feature extraction for classification of microcalcification in mammograms, *Proc. SPIE* 4322, 2001, 301–310.
11. Y. Jiang, R.M. Nishikawa, D.E. Wolverton, C.E. Metz, R.A. Schmidt, K. Doi, Computerized classification of malignant and benign clustered microcalcifications in mammograms, *Proceedings of 19th International Conference-IEEE/EMBS*, 1997, 521–523.
12. A.C. Patrocínio, H. Schiabel, R.H. Benatti, C.E. Goes, F.L.S. Nunes, Investigation of clustered microcalcification features for an automated classifier as part of a mammography CAD scheme, *Proceedings of the 22nd Annual EMBS International Conference*, 2000, 1203–1205.
13. Y. Chitre, A.P. Dhawan, M. Moskowitz, Artificial neural network based classification of mammographic microcalcifications using image structure and cluster features, *Proceedings of the Annual Conference on Engineering in Medicine and Biology*, Vol. 15, 1993, pp. 298–309.
14. B. Verma, J. Zakos, A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques, *IEEE Trans. Inform. Technol. Biomed.* 5 (1), 2001, 46–54.
15. Henri Tristan, Michel Benmussa, Jacques Bokobsa, *Mammographie et échographie mammaire de l'image au diagnostic*, Flammarion Medecine-Sciences, 1989.
16. LE GAL M, CHAVANNE G, PELLIER D. Valeur diagnostique des microcalcifications groupées découvertes par mammographie (A propos de 227 cas avec vérification histologique et sans tumeur palpable), *Bull Cancer (Paris)*, 1984, 71: 57-64.
17. MUIR BB, LAMB J, ANDERSON TJ, KIRKPATRICK A. Microcalcification and its relationship to cancer of the breast : experience in a screening clinic. *Clin Radiol*, 1983, 34: 193-200.
18. EGAN RL, Mc SWEENEY MB, SEWELL CW. Intramammary calcifications without an associated mass in benign and malignant diseases. *Radiology*, 1980, 137: 1-7.
19. SICKLES EA. Mammographic features of 300 consecutive non palpable breast cancers. *AJR*, 1986, 146: 66 1-663.
20. Michael Heath, Kevin Bowyer, Daniel Kopans, Richard Moore and W. Philip Kegelmeyer, *The Digital Database for Screening Mammography*, *Proceedings of the Fifth International Workshop on Digital Mammography*, M.J. Yaffe, ed., 212-218, Medical Physics Publishing, 2001. ISBN 1-930524-00-5.

Communication JDLIO'2011

Dans cette communication nous avons présenté une partie de notre implémentation qui prend en charge la préparation des images de la base DDSM pour l'extraction des attributs et la classification.

Article:

Hamza GHILAS, Meriem GAGAOUA, Abdelouahab MOUSSAOUI, *"Prétraitement des Images Mammographiques de la Base DDSM"*, 1^{ère} journées doctorales du laboratoire d'informatique d'Oran, "JDLIO'2011", les 31 Mai et 01 Juin, Oran-Algérie, 2011.

Prétraitement des Images Mammographiques de la Base DDSM

Hamza GHILAS ⁽¹⁾, Meriem GAGAOUA ⁽²⁾, Abdelouahab MOUSSAOUI ⁽³⁾

Département d'informatique

Université Ferhat Abas, Sétif

Algérie

hamzaghilas@yahoo.fr ⁽¹⁾ gagaouameriem@yahoo.fr ⁽²⁾

moussaoui.abdel@gmail.com⁽³⁾

1 Introduction

Dans cet article nous présentons un logiciel pour le prétraitement des images de la base d'images mammographiques DDSM. L'objectif de ce travail est de préparer les données de la base DDSM pour être utilisées par un algorithme. Notre logiciel permet de visualiser les images de cette base qui ont des résolutions entre 42 et 50 microns et l'extraction des régions d'intérêts. Ce logiciel est destiné à être utilisé par les chercheurs utilisant la base DDSM pour tester leurs algorithmes. Nous avons développé ce logiciel avec le langage objet java, ce qui nous offre la portabilité et la gratuité.

1.1 Domaine

Le cancer du sein est une maladie très dangereuse du fait que ses causes sont toujours inconnues. La détection précoce est le meilleur moyen pour lutter contre cette maladie, alors certains pays ont lancé des campagnes de dépistage et le besoin d'automatisation de cette tâche a émergé. La recherche dans l'aide au diagnostic et le diagnostic automatique de ce cancer est un axe de recherche très actif. L'évaluation

des résultats d'un système d'aide au diagnostic est relatif à la base d'image utilisé appelée *benchmark*. Dans la littérature plusieurs base d'images mammographiques sont développées : MIAS(Mammographic Image Analysis Society), DDSM(Digital Database for Screening Mammography), AMDI (Indexed Atlas of Digital Mammograms). Le benchmark le plus utilisé dans la littérature est la base DDSM, elle est caractérisé par sa richesse d'informations.

DDSM(Digital Database for Screening Mammography) [1] est une base de mammographies digitales très riche, développée en collaboration de Massachusetts General Hospital, University of South Florida, Sandia National Laboratories et Washington University School of Medicine.. L'avantage de la base DDSM consiste dans sa richesse d'informations, elle regroupe plus de 2500 dossiers de patientes, chaque dossier contient 4 images au format LJPEG (deux incidences par sein), un fichier .ics au format ASCII, un fichier .OVERLAY pour chaque image contenant des anomalies et un fichier image .16 PGM regroupant les 4 radios et présentant un aperçu rapide. Les dossiers de patientes(case) sont organisé dans 4 types de volumes, *Normal, Cancer, Benign et Benign without callback*. Les images de la base DDSM sont prises à partir de cas réels et vérifiés par biopsie. Les images sont analysées par des radiologues experts, si une image comporte des anomalies, elle est accompagnée d'un fichier texte avec l'extension OVERLAY. Ce fichier contient les détails de l'anomalie, la nature de la pathologie(maligne/bénigne) et les coordonnées du contour de la région suspecte tracé par un radiologue.

1.2 Objectifs

Le software doit permettre d'afficher les image de la base DDSM qui ne sont pas lisibles par les outils de lecture des formats usuels(jpg, bmp, png,...). L'utilisateur doit avoir la possibilité d'effectuer une recherche dans la base. Le logiciel doit aussi extraire les régions d'intérêts dont leurs contours sont mentionnés dans les fichiers OVERLAY.

1.3 Outils de développement

Le développement de notre s'est déroulé sous windows XP. Nous avons utilisé le langage objet gratuit java avec l'IDE *eclipse* pour l'implémentation des différentes classes.

2 Architecture de l'application

Notre logiciel dispose d'une interface graphique tel illustré par la figure.1. L'interface dispose de trois zones, la zone gauche est un arbre qui permet d'explorer et de naviguer dans le contenus de la base, la zone médiane permet d'afficher les différent fichiers de la base et une zone de recherche qui permet d'effectu  une recherche dans la base avec un ensemble de mots cl s.

2.1 Etapes de d roulement de l'application

Au lancement de l'application le chemin d'acc s aux diff rents fichiers de la base DDSM doit  tre indiqu  pour  tre sauvegard  dans un fichier de configuration. Un clic double sur un fichier de l'arbre permet de l'afficher. L'extraction d'une r gions d'int r t se fait par appelle de la m thode *Incidence.getCluster()*. L'affichage des r gions d'int r ts d'une image se fait par le menu contextuel.

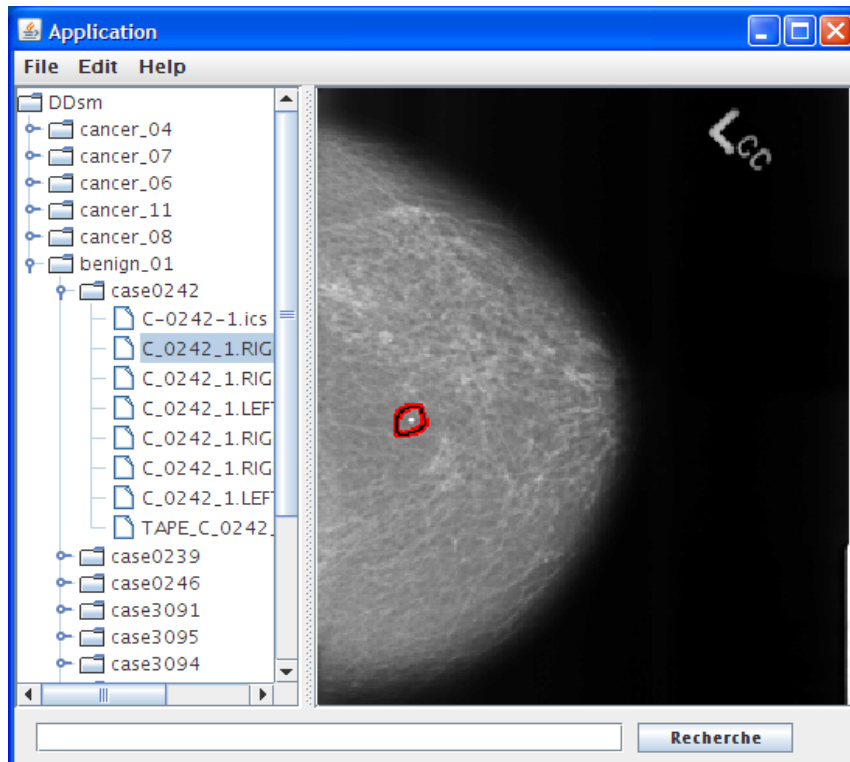


Fig.1 Interface Graphique du logiciel.

3 Conclusion et perspectives

Notre logiciel peut être utilisé par les chercheurs qui travaillent sur les méthodes d'aide au diagnostic du cancer du sein. l'implémentation du logiciel avec le langage objet java lui offre la facilité d'extension par les chercheurs pour le développement de nouveaux algorithmes pour la segmentation, la reconnaissance et la classification.

4 Références bibliographiques

- [1] Current status of the Digital Database for Screening Mammography, Michael Heath, Kevin Bowyer, Daniel Kopans, W. Philip Kegelmeyer, Richard Moore, Kyong Chang, and S. MunishKumaran, in Digital Mammography, 457-460, Kluwer Academic Publishers, 1998, Proceedings of the Fourth International Workshop on Digital Mammography.