



وزارة التعليم العالي والبحث العلمي
MINISTRE DE L'ENSEIGNEMENT
SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE
جامعة عبد الحميد بن باديس مستغانم
Université Abdelhamid Ibn Badis Mostaganem
كلية العلوم والتكنولوجيا
Faculté des Sciences et de la Technologie
DEPARTEMENT DE GENIE ELECTRIQUE



N° d'ordre: M/GE/2019

MEMOIRE

Présenté pour obtenir le diplôme de

MASTER EN ELECTRONIQUE

Option : Électronique des systèmes embarqués

Par

- HAMADOUCHE Meroua.
- NEDJAR Djamel.

Détection et classification des tumeurs dans les tissus mammaires par la machine à vecteur de support(SVM)

Soutenu le 11/07/2019 devant le jury composé de:

Président :	Mr. CHAAB.O	MCA	Université de Mostaganem
Examineur :	Mr. YAGOUBI.B	Prof	Université de Mostaganem
Examineur :	Mr. OULD MAMMAR.M	MCA	Université de Mostaganem
Encadreur :	Mme. BERRADJA.K	MCA	Université de Mostaganem
Co-encadreur :	Mr.SOLTANE.B	MCB	Université de Mostaganem

Table des matières

Introduction générale	5
1 Contexte Médicale	8
1.1 Introduction	8
1.2 L'anatomie du sien chez la femme	8
1.3 Le cancer du sein	9
1.4 Dépistage du cancer du sein	10
1.4.1 L'échographie	10
1.4.2 L'imagerie par résonance magnétique (IRM)	10
1.4.3 La mammographie	11
1.5 Diagnostic du cancer du sein assisté par ordinateur(DAOx)	13
1.5.1 L'importance des systèmes de diagnostic par ordinateur	14
1.6 Description des masses	14
1.6.1 La forme	14
1.6.2 Le contour	15
1.6.3 La densité	16
1.7 Les bases de mammographie	17
1.7.1 Mammographie Image Analysis Society Digital Mammogram Data- base(MIAS)	18
1.7.2 Digital Database For Screening Mammography(DDSM)	18
1.7.3 Banco Web LAPIMO	19
1.8 Conclusion	19
2 Traitement des mammographies	21
2.1 Introduction	21
2.2 Prétraitement des mammographies	22
2.2.1 Suppression du bruit	22
2.2.2 Amélioration de contraste	23
2.3 Segmentation des images mammographie	24
2.3.1 2.3.1 Segmentation par la ligne de partage des eaux(LPE)	24
2.4 Conclusion	27
3 Classification par SVM	29
3.1 Introduction	29
3.2 Support vector machine(SVM)	29
3.2.1 Principe de fonctionnement	30
3.3 Formalisme des SVM	31
3.4 Entraînement des SVM	32
3.5 SVM non linéaire et la fonction noyau	34

3.6	Conclusion	35
4	<i>Conception du système d'aide au diagnostic assisté par ordinateur</i>	37
4.1	Introduction	37
4.2	Problématique	37
4.3	Base de données	38
4.4	Système d'aide au diagnostic assisté par ordinateur proposé	38
4.4.1	Prétraitement des mammographies	40
4.4.1.1	Suppression du bruit	40
4.4.1.2	Rehaussement de contraste	42
4.4.2	Segmentation : Détection des masses par LPE	45
4.4.2.1	Application aux gradient morphologique	46
4.4.2.2	L'algorithme LPE	46
4.4.3	Descripteurs de forme	47
4.4.4	Classification par SVM	48
4.4.5.1	Évaluation des performances de classification	49
4.5	Implémentation numérique	51
	Conclusion générale et perspectives	52
	Morphologie mathématique	53
	Bibliographie	59

Table des figures

1.1	L'anatomie du sein chez la femme[3].	9
1.2	Exemple d'échographie mammaire[6].	10
1.3	Exemple d'imagerie IRM mammaire[8].	11
1.4	Exemple de mammographie.	12
1.5	Visualisation lever du bras pour oblique médiolatérale (MLO)[11]	13
1.6	Visualisation craniocaudal(CC)[12].	13
1.7	Les formes des opacités selon le BI-RADS. (a)ronde, (b)ovale, (c)lobulée, (d)irrégulière[15].	15
1.8	Les différents types de contour des opacités selon le BI-RADS. (a)circonscrit, (b)masqué, (c)microlobulé, (d)maldéfini, (e)spiculé,(f)distorsions architecturales[17].	16
1.9	Types de densité des masses selon le BI-RADS. (A)sein graisseux, (B)sein graisseux et hétérogène, (C)sein dense et hétérogène, (D)sein hyper dense et hétérogène[19].	17
2.1	Traitement d'image mammographique[24]	21
2.2	Exemple des artefacts d'une image mammographie.	23
2.3	En faisant monter l'eau dans les bassins 2 et 3, ceux-ci finissent par se rejoindre en la ligne de partage des eaux 2-3 (symbolisée par une flèche rouge). De la même façon, lorsque le niveau continue de monter, les bassins 1 et 2 se rejoignent en la ligne 1-2 (symbolisée par la flèche verte)[29].	25
2.4	L'image (a) est transformée en l'image (b). Dans cette image on retrouve les trois bassins versants numérotés de 1 à 3 ainsi que les pixels de ligne de partage des eaux étiquetés à 0 [31].	26
2.5	LPE contrôlée par marqueurs.(a) Les minimas sans contrainte, (b)processus de LPE avec contrainte des marqueurs.	27
3.1	Le séparateur idéal correspond intuitivement à l'hyperplan qui passe au milieu, c'est le séparateur de marge maximale.	30
3.2	Les "vecteurs de support" se trouvent à une distance égale à la marge d'un côté ou de l'autre de l'hyperplan de séparation[37].	31
3.3	Astuce à noyaux : projeter les données dans un espace de dimension beaucoup plus grande, où elles deviennent séparables linéairement.	34
3.4	Exemple de changement de l'espace de données.	35
4.1	Système de diagnostic assisté par ordinateur	38
4.2	Suppression des artefacts. (a) et (b) images originelle, (c) et (d) après suppression des artefacts.	41
4.3	Résultat de suppression du bruit.(a) et (b) les images sans artefacts, (c) et (d) application du filtre gaussien.	42

4.4	Amélioration du contraste.(a) et (b) les images filtrées, (c) et (d) rehaussement de contraste par expansion de la dynamique.	43
4.5	Application du filtre haut de forme. (a) et (b)Images rehaussées,(c) et (d)Filtre haut de forme blanc, (e) et (f)Application du filtre haut de forme noir.	44
4.6	Amélioration de contraste. (a) et (b)Filtre haut de forme noir, (c) et (d)Maximisation du contraste.	45
4.7	Détermination du gradient morphologique.(a) et (b)ouverture et fermeture géodésique, (c) et (d)gradient morphologique	46
4.8	L'extraction des masses.(a) et (b)le gradient morphologique. (c) et (d)DLPE contrôlé par marqueurs.	47
4.9	Extraction des masses.(a) mdb206 et (b) mdb083.	48
4.10	Calcul d'une matrice de confusion	49
4.11	matrice de confusion de notre modèle SVM	50
4.12	Interface graphique du CAOx proposé.	51
4.13	Éléments structurants plans isotropes élémentaires. L'origine de chaque élément structurant est son centre	53
4.14	Relation de voisinage pour un élément structurant carré à 8-connexités. . .	54
4.15	Erosion. (a) image en niveaux de gris originale (b) érosion avec un SE carré de taille 3 (c) érosion avec un SE carré de taille 6 (d) érosion avec un SE carré de taille 10	55
4.16	Dilatation. (a) image en niveaux de gris originale (b) dilatation avec un SE carré de taille 3 (c) dilatation avec un SE carré de taille 6 (d) dilatation avec un SE carré de taille 10.	56
4.17	Ouverture. (a) image en niveaux de gris originale (b) ouverture avec un SE carré de taille 3 (c) ouverture avec un SE carré de taille 6 (d) ouverture avec un SE carré de taille 10.	57
4.18	Fermeture. (a) image en niveaux de gris originale (b) fermeture avec un SE carré de taille 3 (c) fermeture avec un SE carré de taille 6 (d) fermeture avec un SE carré de taille 10.	57

Introduction générale

Le cancer du sein est le cancer le plus fréquent chez les femmes avec 450000 décès chaque année dans le monde. En Algérie, il représente près de 50 % des cancers gynécologiques. L'incidence du cancer du sein reste progressive en Algérie et il n'existe toujours pas des programmes de dépistage organisé. Du fait de son diagnostic tardif, il en résulte souvent un traitement lourd, mutilant et coûteux qui s'accompagne d'un taux de mortalité élevé. Cependant, si la tumeur a été prise en charge dès les premiers stades de son apparition, le taux de mortalité diminue considérablement.

Dans l'optique d'assurer le dépistage précoce d'une telle tumeur, les radiologues ont été amenés à augmenter la fréquence des mammographies particulièrement à partir de la cinquantaine. Il a été démontré que la mise en œuvre des campagnes de dépistage (tous les deux ans par exemple) s'avère très bénéfique et permet de diminuer le taux de décès d'environ 35%. Deux ou quatre mammographies sont nécessaires par patiente à raison d'une ou deux mammographies par sein. Par conséquent, le nombre de mammographies réalisées a augmenté de manière exponentielle. De ce fait, l'interprétation est devenue une tâche compliquée et dépendante de l'expertise du radiologue. Par ailleurs, une deuxième lecture peut améliorer d'environ 15% le taux de détection du cancer du sein. Au cours des dernières décennies, des techniques de calcul ont été mises au point dans le but de détecter automatiquement les structures associées aux tumeurs lors des examens de mammographie. La détection et la classification automatiques des masses dans les mammographies constituent toujours un défi de taille et jouent un rôle capital pour aider les radiologistes à statuer sur un diagnostic précis[1].

Plusieurs travaux de recherche ont été élaborés soit pour détecter automatiquement les lésions (micro-calcifications ou masses) ou fournir un second avis à propos de la lésion détectée à travers des systèmes de diagnostic assisté par ordinateur. Ceci a conduit à la conception de plusieurs méthodes de traitement d'images mammographiques toutes s'inspirant de la classification Breast Imaging Reporting and Data System (BI-RADS) de l'American College of Radiology (ACR) pour classifier les masses en deux catégories à savoir bénignes et malignes.

Dans cette même optique, notre objectif est de mettre au point un système d'aide automatique à la détection et la classification des masses (opacités) dans des images mammographiques. Ce système est basé sur l'utilisation en premier lieu d'une méthode de segmentation morphologique appelée la ligne de partage des eaux et en second lieu, les Machines à Vecteurs de Support (SVM) pour la classification.

Afin de décrire le travail effectué, ce mémoire est divisé en quatre chapitres. Le premier chapitre est consacré à la définition des notions de base sur l'anatomie du sein ainsi que la description de ses principaux composants. De plus, un aperçu sur les masses mammaires, leurs caractéristiques et leurs classifications sont présentées également dans ce chapitre. Le second chapitre est dédié à l'analyse des images mammographiques.

Introduction générale

En premier lieu nous décrivons les méthodes standards de pré-traitement utilisées en traitement d'images comme la suppression du bruit et l'amélioration du contraste. En second lieu, la méthode de la ligne de partage des eaux (LPE) est présentée afin de détecter et d'extraire les masses.

Le troisième chapitre est consacré à l'étude des Machines à Vecteurs de Support (SVM) pour la classification. Le chapitre quatre propose un système de diagnostic assisté par ordinateur pour la détection de masses dans les images de mammographie et leur classification en bénigne ou maligne.

Une conclusion et les perspectives envisagées sont données en dernier lieu.

A pink awareness ribbon is positioned behind the text, forming a loop that frames the words "Contexte" and "médicale".

Chapitre 1:
Contexte
médicale

Chapitre 1

Contexte Médicale

1.1 Introduction

Dans ce chapitre, nous présenterons brièvement l'anatomie du sein chez la femme, ce qui va nous permettre d'aborder la notion du cancer du sein et son impact sur la population féminine. Ensuite, nous exposerons la procédure du dépistage du cancer du sein, puis nous continuerons avec les principes généraux d'acquisition en imagerie du sein. Cette introduction permettra de comprendre l'articulation des diverses étapes qui composent le système de diagnostic que nous proposons.

1.2 L'anatomie du sien chez la femme

Un sein normal comprend des éléments mammaires glandulaires entourés de stroma adipeux et mammaire, qui sont entourés d'une structure fibreuse en nid d'abeille à ligaments de Cooper minces. Les éléments glandulaires sont composés de canaux lactifères menant du mamelon et de la branche aux canaux excréteurs, aux canaux inter-lobulaires et aux canaux terminaux conduisant aux acini producteurs de lait (Fig 1.1). Les conduits sont tapissés d'épithélium composé d'une couche myoépithéliale cellulaire externe et d'une couche cellulaire sécrétoire interne. Les canaux et le tissu glandulaire s'étendent vers l'arrière selon une distribution en éventail composée de 15 à 20 lobes drainant chacun des canaux lactifères, la majeure partie du tissu glandulaire se trouvant dans la partie supérieure de la poitrine près de l'aisselle[1]. Le tissu adipeux entoure le tissu glandulaire. La graisse rétro-glandulaire est située en arrière du tissu glandulaire et est décrite par le Dr Laszlo Tabar comme un «no man's land» dans lequel aucun tissu glandulaire ne doit être observé. Le muscle pectoral se trouve derrière la graisse au-dessus de la paroi thoracique[2].

La taille et le poids du sein varient durant la vie d'une femme. Ils sont principalement influencés par la quantité de tissu adipeux présent dans le sein.

Le développement du sein chez la femme se produit en deux étapes sous l'influence des hormones : à l'approche de la puberté, le volume du sein augmente en réponse à l'action de l'œstrogène sur les canaux galactophores, la deuxième phase survient sous l'influence de la production de progestérone quand débute l'ovulation.

La croissance du sein est essentiellement terminée environ un an et demi après la fin de

la période des premières menstruations, tandis que le développement des alvéoles continue en réponse à une stimulation hormonale.

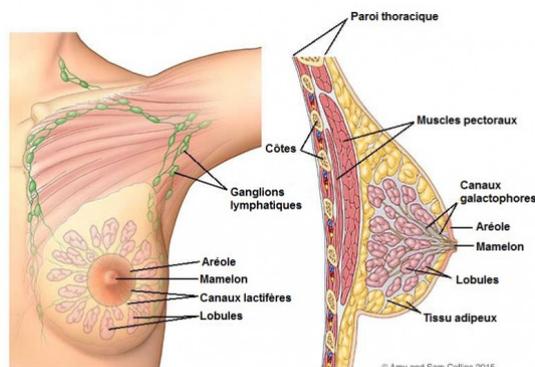


FIGURE 1.1 – L’anatomie du sein chez la femme[3].

1.3 Le cancer du sein

Le cancer est une maladie caractérisée par une prolifération cellulaire, ou tumeur maligne, anormalement importante formée à partir de la transformation par mutation ou instabilité génétique d’une cellule initialement normale.

Le cancer du sein est un type de cancer causé par le tissu mammaire, que ce soit par la paroi interne des canaux lactifères ou par les lobules qui alimentent les canaux en lait. Le cancer du sein est causé par une combinaison de multiples facteurs tels que l’hérédité, la composition des tissus, les substances cancérigènes, les niveaux d’immunité, les hormones[3].

Il est le résultat d’une multiplication de cellules anormales dans cet organe. Ces cellules sont souvent situées au niveau d’un canal galactophore d’un lobule. Elles peuvent également se retrouver au niveau des ganglions lymphatiques.

Les masses formées par cette multiplication excessive sont appelées tumeurs.

Il existe deux types de tumeurs cancéreuses du sein : les tumeurs non cancéreuses ou "bénigne" et les tumeurs cancéreuses "maligne".

A- Tumeurs bénignes

Lorsqu’une tumeur est diagnostiquée comme étant bénigne, les médecins la laissent généralement seule plutôt que de la retirer. Même si ces tumeurs ne sont généralement pas agressives envers les tissus environnants, elles peuvent parfois continuer à se développer, en appuyant sur les organes et en provoquant des douleurs ou d’autres problèmes. Dans ces situations, la tumeur est retirée, permettant à la douleur ou aux complications de s’apaiser.

B- Tumeurs malignes

Les tumeurs malignes sont cancéreuses et agressives car elles envahissent et endommagent

les tissus environnants. Lorsqu'une tumeur est soupçonnée d'être maligne, le médecin procédera à une biopsie afin de déterminer la gravité ou l'agressivité de la tumeur.

1.4 Dépistage du cancer du sein

Le dépistage du cancer consiste à réaliser des examens de surveillance d'une personne supposée en bonne santé, dans le but d'arriver à détecter des anomalies pouvant être des signes avant-coureurs d'un cancer, bien avant l'apparition des premiers symptômes. C'est une démarche qui s'avère très efficace en ce sens qu'elle permet de réduire de façon significative le taux de mortalité lié au cancer du sein en améliorant les chances de guérison. En effet, s'il est détecté tôt, le cancer du sein peut être guéri dans neuf cas sur dix, ce qui participe à une réduction de 15 à 21% du taux de mortalité lié à ce cancer.

1.4.1 L'échographie

Une échographie mammaire est un examen qui utilise des ultrasons pour produire des images de l'intérieur du sein. Cet examen est pratiqué par un radiologue. Une échographie est très utile pour voir la nature, liquide ou solide, des nodules palpés ou découverts sur une mammographie[5].

Elle permet de comparer ce que l'on sent avec les doigts (lors de la palpation) et ce que l'on voit sur l'écran. Elle permet aussi de guider un prélèvement. Une échographie mammaire est réalisable à n'importe quel moment du cycle menstruel. Il n'est pas utile d'être à jeun, ni de suivre un régime particulier. Aucune préparation n'est nécessaire.

L'échographie mammaire permet de voir s'il existe certaines anomalies dans le sein, mais elle ne permet pas toujours de déterminer avec certitude s'il s'agit ou non d'un cancer. Pour cela, il faut réaliser d'autres examens complémentaires. Dans certains cas, il n'est pas possible de donner d'emblée un diagnostic définitif. Le radiologue réalise ou conseille d'autres examens (IRM des seins ou prélèvements) pour établir le diagnostic.



FIGURE 1.2 – Exemple d'échographie mammaire[6].

1.4.2 L'imagerie par résonance magnétique (IRM)

L'imagerie par Résonance Magnétique, est un moyen d'investigation dans le dépistage et le suivi thérapeutique des cancers. Proposant une visualisation de coupes anatomiques extrêmement fines, elle permet de différencier les tissus pathologiques des tissus sains

et de réaliser une "cartographie" très précise et en trois dimensions des tumeurs. L'IRM utilise les modifications provoquées par un champ magnétique, envoyé par la machine, dans les noyaux des atomes d'hydrogène présents dans l'organisme. Cet examen est indolore et très fiable. L'imagerie par résonance magnétique n'utilise pas les rayons X [7]. Parmi toutes les modalités radiologiques, l'IRM est certainement la technique la plus performante dans la recherche des lésions du sein. Rarement, une prise de contraste diffuse du tissu fibroglandulaire peut limiter l'interprétation de l'examen, voire même masquer une lésion. La spécificité de l'IRM est un peu moins bonne. Des petits foyers de prise de contraste sont difficilement caractérisables. Certaines lésions bénignes comme les fibroadénomes peuvent avoir un comportement identique à celui des tumeurs malignes. Cette situation requiert très souvent une biopsie ou une exégèse. On peut rencontrer également des situations où l'on trouve à l'IRM une prise de contraste dont l'aspect est inquiétant, et dans le même temps, ne voir aucune lésion sur les mammographies et à l'échographie (même après réexamen). Il existe alors une indication à effectuer une biopsie sous IRM.

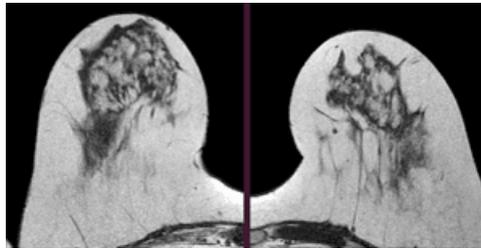


FIGURE 1.3 – Exemple d'imagerie IRM mammaire[8].

1.4.3 La mammographie

La mammographie a pour but de déceler au plus tôt des anomalies avant même qu'elles n'aient provoqué des symptômes cliniques. Elle peut permettre, ainsi, de détecter des cancers bien avant qu'ils ne soient palpables.

L'examen consiste à comprimer à tour de rôle les deux seins puis à les exposer à une faible dose de rayons X, en utilisant le mammographe[9].

La mammographie ne permet pas toujours de donner d'emblée un diagnostic définitif : elle permet de voir s'il existe une anomalie dans le sein, mais elle ne permet pas de déterminer avec certitude s'il s'agit ou non d'un cancer.

La douleur lors de l'examen liée à la compression du sein entre deux plaques est fréquente mais de faible intensité. Le premier risque consécutif à la mammographie de dépistage est le sur-diagnostic. De 20 à 49 % des femmes participantes à un dépistage mammographie régulier auront au moins un résultat faussement positif après 10 examens.

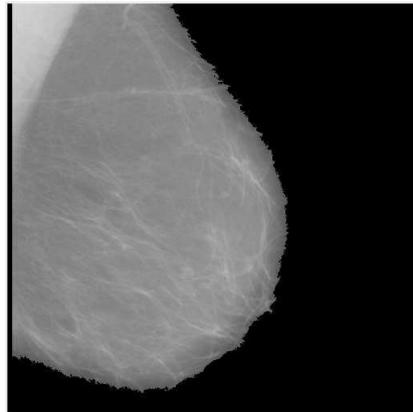


FIGURE 1.4 – Exemple de mammographie.

La mammographie numérique est une image électronique produite avant d'être sauvegardé sur ordinateur où elle subira une manipulation du degré de contraste dans les zones les plus dense notamment avant d'être visualisée. On distingue 2 positions de visualisation en mammographie numérique :

- lever du bras pour oblique médiolatérale (MLO) :

Lors de la visualisation du MLO, le patient est également amené à se pencher vers l'équipement pour une visualisation optimale des tissus. Le bras est tourné à 45 degré afin de démontrer la quantité maximale de tissu mammaire et de muscle pectoral. Parfois, l'angle est individualisé selon la taille de la poitrine avec une différence de 10 degré. L'autre sein du patient, non-imageur, est doucement pressé contre le corps et maintenu à l'écart[10](Figure 1.6).

- craniocaudal(CC) :

Pendant la vision CC, le patient est positionné de sorte que le mamelon se trouve approximativement au centre du détecteur. La patiente est obligée de se pencher vers l'appareil pour rapprocher le sein du détecteur. Le bas de la poitrine doit être soutenu et tiré vers le haut afin que les tissus les plus profonds et les plus bas soient inclus dans la vue, du côté en cours d'imagerie, est poussée vers le bas pour relâcher le muscle pectoral de manière à inclure le tissu mammaire dans le quadrant externe. La visualisation du muscle pectoral sur la vue en CC implique qu'aucun tissu le long de la paroi thoracique n'a été exclu(Figure 1.7).

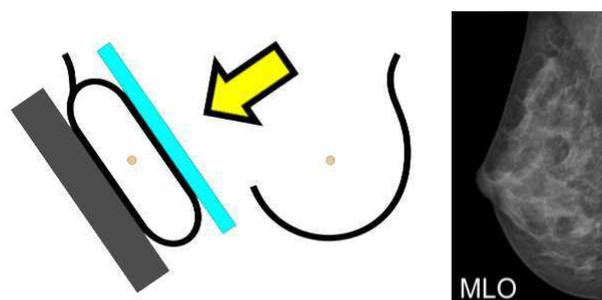


FIGURE 1.5 – Visualisation lever du bras pour oblique médiolatérale (MLO)[11]

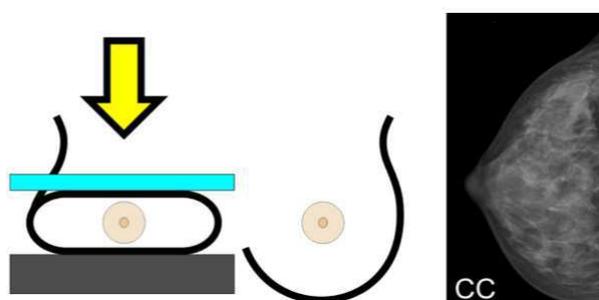


FIGURE 1.6 – Visualisation craniocaudal(CC)[12].

1.5 Diagnostic du cancer du sein assisté par ordinateur(DAOx)

Le diagnostic des radiologues à propos des lésions mammaires est une tâche subjective qui peut donc être entachée d'erreur dont les conséquences peuvent s'avérer mortelles pour les patientes. Plusieurs raisons comme la fatigue, le manque d'expérience en imagerie du sein et la quantité d'images à analyser chaque jour rendent cette tâche fastidieuse et justifient les erreurs de diagnostic. A l'inverse, un système automatique de diagnostic assisté par ordinateur(DAOx) conçu sur un modèle qui prend en compte des descripteurs pertinents de la lésion mammaire fera toujours preuve de robustesse, ce qui pourrait aider à réduire le nombre de biopsies inutiles. En effet, il a été prouvé qu'après une biopsie, seulement moins du tiers des images mammographiques suspectes sont réellement des mammographies de lésions cancéreuses [Meyer et al., 1990]. Aussi, la conception des systèmes de DAOx pourrait aider les radiologues à diminuer le nombre de biopsies et éviter à la patiente une situation de stress inutile.[13] Un système automatique de DAOx employé dans un contexte de dépistage ou de diagnostic servira donc de second avis au radiologue et n'a absolument pas vocation à se substituer à ce dernier. De façon générale, le fonctionnement d'un système de DAOx est une suite de plusieurs étapes toutes aussi importantes

les unes que les autres. Quand le système reçoit en entrée une image, il procède d'abord à un pré-traitement de celle-ci, c-à-d. élimination du bruit d'acquisition ou speckle et amélioration du contraste de l'image, afin d'obtenir la meilleure qualité d'image possible.

Cette étape est suivie de la détection des anomalies (p. ex. microcalcifications, opacité ou distorsions architecturales). Sur l'anomalie détectée sont ensuite extraits plusieurs attributs mathématique pouvant aider à décrire au mieux la nature de la lésion. Pour finir, les descripteurs extraits sont ensuite classifiés en s'appuyant sur des algorithmes adaptés à cette tâche, afin de décider de la nature de la masse.

1.5.1 L'importance des systèmes de diagnostic par ordinateur

Même si la lésion est très suspecte en mammographie ou échographie, le diagnostic préopératoire est aujourd'hui indispensable pour deux raisons :

- Il peut éviter de nombreuses chirurgies inutiles s'il ne s'agit pas d'un cancer.
- Il permet d'optimiser d'emblée la chirurgie s'il s'agit d'un cancer du sein : si ce cancer est de petite taille par exemple, une intervention sur un ganglion (ex : ganglion sentinelle) pourra être programmée et permettra à la patiente d'éviter un curage axillaire. Dans d'autres cas par exemple, le traitement pourrait être d'abord médical (néo-adjuvant) pour permettre une chirurgie ensuite moins lourde.

Dans de très rares cas, ce diagnostic pré-opératoire n'est pas possible en raison de la localisation des images radiologiques qui sont, par exemple, non accessibles à une biopsie. La chirurgie permettra de poser le diagnostic[14].

1.6 Description des masses

En imagerie du sein, on distingue deux grands types de lésions ; on a d'un côté les microcalcifications et de l'autre les masses. Étant donné que notre projet concerne les masses mammaires, nous ne ferons ici que la description de ces dernières. Par définition, une masse est une opacité importante occupant un espace et vue sur deux incidences différentes. Dans le cas où, la masse est observée uniquement sur une seule incidence, elle est appelée asymétrie jusqu'à ce que sa nature tridimensionnelle soit confirmée. Différents attributs permettent aux médecins de décrire les masses en vue de déterminer leur nature ; ce sont : la forme, le contour et la densité.

1.6.1 La forme

D'après la classification BI-RADS de l'ACR, les opacités mammaires peuvent avoir une forme ronde, ovale, lobulée ou irrégulière(Figure 1.8). (a) ronde : c'est l'aspect d'une masse sphérique, circulaire ou globuleuse.

(b) ovale : de telles masses présentent une forme d'ellipse.

(c) lobulée : la forme de la masse est légèrement ondulée.

(d) irrégulière : c'est le terme utilisé pour désigner les masses qui présentent une forme aléatoire et ne peuvent être classées dans les trois classes précédentes.

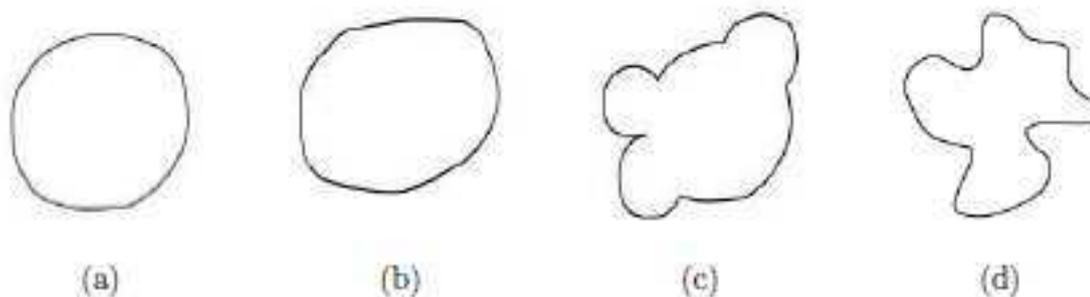


FIGURE 1.7 – Les formes des opacités selon le BI-RADS. (a)ronde, (b)ovale, (c)lobulée, (d)irrégulière[15].

1.6.2 Le contour

En dehors de leur forme, les opacités varient également au niveau de leur contour ; le BIRADS dénombre cinq classes de contour que sont les contours : circonscrit, masqué, microlobulé, maldéfini et spéculé. La (figure 1.9) présente un exemple de chacune de ces cinq catégories de contour.

(a) circonscrit : ce sont des contours bien définis où la frontière entre la masse et le fond de l'image est nette et qui correspondent en général à des lésions bénignes.

(b) masqué : ce sont pour la plupart les masses dont le contour est dissimulé par la superposition de tissus normaux, faisant penser à une masse circonscrite dont une partie du contour est cachée.

(c) micro-lobulé : c'est un contour qui contient de petites ondulations.

(d) mal-défini : c'est un contour qui peut laisser envisager la présence d'une infiltration.

(e) spiculé : c'est un contour comportant des structures filiformes qui rayonnent en s'éloignant du centre de la masse et qui suggèrent fortement une malignité.

Hormis les microcalcifications et les masses, il existe une troisième forme de signes radiologiques qui traduisent la présence d'un cancer et qui sont appelées les distorsions architecturales[16]. Elles ne comportent pas de centre dense, comme c'est le cas des opacités décrites plus haut.

(f) distorsions architecturales : ce sont des signes hautement suspects qui se traduisent par des structures linéaires ou des spécules qui convergent vers une même zone focale.

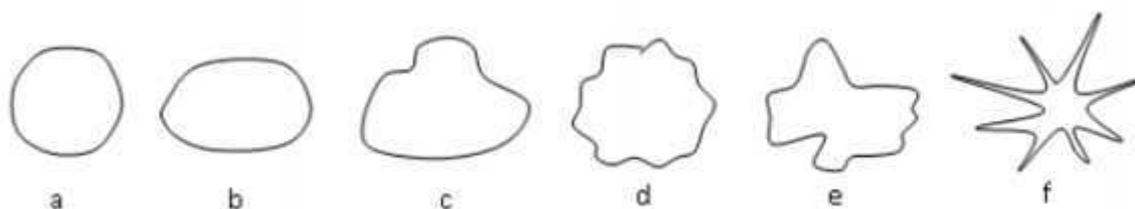


FIGURE 1.8 – Les différents types de contour des opacités selon le BI-RADS. (a) circonscrit, (b) masqué, (c) microlobulé, (d) maldéfini, (e) spiculé, (f) distorsions architecturales[17].

1.6.3 La densité

La composition du sein varie souvent d'une femme à une autre; l'un des aspects visibles de cette grande variabilité est la densité radiologique du tissu mammaire. Des études récentes ont suggéré que la densité mammaire est un facteur de risque important, en raison de ce que les pathologies peuvent être dissimulées à l'intérieur d'une région très dense.

Dans le but d'aider à une meilleure compréhension de la composition de la densité du sein, plusieurs travaux et modèles ont été proposés. fut l'un des premiers à aborder le sujet; il a classé de façon empirique la densité mammaire en quatre catégories – allant du sein grasseux au sein fibroglandulaire dense, ces travaux n'ont cependant pas fait l'unanimité dans la communauté scientifique en raison du fait que les mammographies utilisées dans cette étude sont issues du premier essai du programme de dépistage aléatoire et contrôlé et des conditions d'acquisition limitées des appareils de mammographie (c-à-d. positionnement et compression du sein) du début des années 70. Fortement influencés par les travaux de Wolfe, ont proposé un modèle basé sur le mélange de quatre briques qui constituent l'anatomie normale du sein; ce qui leur a permis par la suite de subdiviser les images en cinq catégories de risque en tenant compte de la distribution des briques du sein. L'avantage du modèle de Tabar et al., c'est qu'il est basé sur les corrélations de l'anatomie mammaire plutôt que sur une simple lecture des mammographies, comme l'a fait Wolfe. De plus, le modèle d'évaluation du risque mammaire en fonction de la densité que Tabar et al. ont proposé est plus discriminant que celui de Wolfe, permettant ainsi une précision reproductible de la classification[18]. Quant à [Boyd et al., 1995], ils ont conçu une méthode utilisant un outil de détection assistée par ordinateur pour estimer le pourcentage de région dense dans le sein; leur travaux ont servi à infirmer certaines conclusions de [Wolfe, 1976] à propos du risque de cancer lié aux quatre groupes proposés. Face aux différents modèles existants et dans le but de standardiser les comptes rendus des mammographies, l'ACR a mis au point une classification BI-RADS définissant quatre classes relatives à la composition du sein[Figure1.10].

(A) B1 : le sein est presque entièrement grasseux et homogène, radio transparent et facile à interpréter pour les radiologues (densité occupe moins de 25% de la glande mammaire). (B) B2 : le sein est grasseux et hétérogène, et on note la présence d'opacités fibroglandulaires dispersées dans le parenchyme (la densité occupe entre 25 et 50% de la glande mammaire).

(C) B3 : le sein est dense et hétérogène (la densité occupe entre 50 et 75% de la glande mammaire).

(D) B4 : le sein est hyper dense et homogène ; l'interprétation de la mammographie est donc très difficile car la densité peut dissimuler une masse (la densité occupe plus de 75% de la glande mammaire).

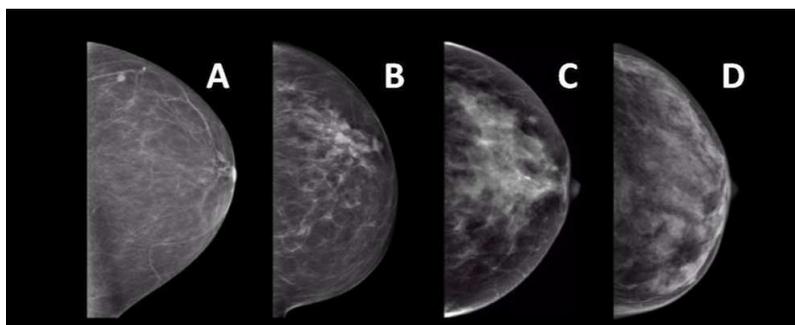


FIGURE 1.9 – Types de densité des masses selon le BI-RADS. (A)sein graisseux, (B)sein graisseux et hétérogène, (C)sein dense et hétérogène, (D)sein hyper dense et hétérogène[19].

1.7 Les bases de mammographie

Il existe plusieurs bases d'images utilisées par les chercheurs en imagerie du sein dans le but de concevoir des systèmes de DAOx, dont certaines sont publiques et d'autres sont privées[20]. Toutefois, toutes ne respectent pas les exigences que doivent satisfaire une base dédiée à la conception d'un système de DAOx. Dans un premier temps, nous allons évoquer brièvement les exigences relatives à la création d'une base de mammographies, puis nous finirons en détaillant quelques unes des bases disponibles tout en mettant l'accent sur leur forces et leurs faiblesses.

A. Les exigences : Dans des études menées entre 1996 et 1998, [a énuméré certaines exigences qui doivent être respectées dans la création de base de mammographies numériques. Certaines d'entre elles sont abordées ci-dessous.

B. La sélection des cas : Toute base de mammographies doit contenir divers cas avec des images de sein normal, de même que celles de toutes les pathologies rencontrées en imagerie du sein et ce quelque soit la densité mammaire. Les images de sein normal mais contenant des structures qui peuvent prêter à confusion doivent être maintenues dans la base, car elles sont importantes pour rendre les classifieurs plus robustes. Les cas doivent être collectés par des radiologues ayant une réelle expérience en mammographie ; chaque cas doit contenir quatre images obtenues à l'aide des incidences standards (MLO et CC), à moins que la femme n'ait subit auparavant une mastectomie du sein. Quant à l'acquisition des images, elle doit être réalisée dans les bonnes conditions, en veillant à ce que le positionnement de la femme et l'exposition aux rayons-X soient conformes à la procédure standard, de manière à éviter tout flou cinétique sur les images.

C. La vérité terrain : La preuve de la réalisation de la biopsie pour tous les cas évoqués doit être disponible. Par ailleurs, l'annotation doit inclure la vérité terrain concernant

la cytologie /histologie de tous les cas, et les limites du contour de la lésion obtenue par la segmentation manuelle, réalisée par les radiologues experts en imagerie du sein. **D. Les infomations assisté à chaque cas :** Les informations relatives à l'âge, aux antécédents personnels et familiaux de la patiente peuvent être incluses dans la mesure où elles peuvent s'avérer utiles pour améliorer les systèmes de DAOx. Quant à la classification de la densité du sein selon l'ACR et le BI-RADS, elles sont obligatoires.

E. L'organisme de la base et sa distribution : Il n'y a pas de format de fichier spécifique, toutefois, la plupart des images médicales sont sauvegardées au format DICOM (digital imaging and communications in medicine). Une partition de ces images en sous base d'apprentissage et de test est également conseillée en vue de faciliter la comparaison des différents systèmes de DAOe ou de DAOx. Enfin, les bases doivent être accessibles à tous et de préférence via internet, sans oublier qu'une maintenance permanente doit être obligatoirement assurée par les propriétaires de la base.

1.7.1 Mammographie Image Analysis Society Digital Mammogram Database(MIAS)

Mammographic Image Analysis Society (MIAS) [Suckling et al., 1994] est la plus ancienne base d'images publiquement disponible ; elle a été conçue au Royaume-Uni en 1994 et est encore largement utilisée dans l'état-de-l'art. La MIAS contient 161 cas pour un total de 322 images numérisées obtenues en incidence MLO, sur lesquelles sont présentes toutes les pathologies à savoir les lésions bénignes et malignes, mais également les images normales. Cette base dispose d'un nombre important de masses spiculées, ainsi que d'une information relative à la densité du sein ; toutefois, la classification des masses réalisée par les radiologues ayant évalué cette base ne respecte pas les standards de l'ACR. Cependant, avec l'augmentation de l'usage de la classification BI-RADS de l'ACR, certains auteurs se sont essayés à une classification des masses de la MIAS afin qu'elle puisse correspondre aux standards en vigueur en imagerie du sein.

Un autre défaut de la MIAS est l'annotation de ces images, qui consiste à indiquer le centre et le rayon de la région d'intérêt, c'est-à-dire la région où se situe la pathologie. Ce genre d'annotation est considéré comme insuffisant pour certaines études comme la segmentation des masses, où l'on souhaite que toutes les lésions circonscrites ou spiculées soient manuellement segmentées afin de faire une comparaison très précise avec les méthodes automatiques. Pour finir, l'autre inconvénient de la MIAS, c'est la résolution à laquelle les images sont numérisées et qui fait que cette base ne convient pas pour des expériences liées à la détection des micro-calcifications[21].

1.7.2 Digital Database For Screening Mammography(DDSM)

Constituée en 1999 aux Etats-Unis (USA), la Digital Database for Screening Mammography (DDSM) est la plus grande base publique de mammographies. Elle contient 2620 cas incluant deux images (incidences MLO et CC) de chaque sein, pour un total de 10480 images où se retrouvent toutes les pathologies, des images normales à celles contenant des masses bénignes et malignes. Pour chacune de ces images, les informations relatives à l'âge de la patiente, la densité du sein et une classification respectant le BI-RADS sont fournies.

A la différence de la MIAS, les annotations concernant les pathologies sont faites manuellement par les radiologues, qui encerclent de façon un peu grossière la région suspecte. Toutefois, des travaux de recherche ont souligné que la DDSM n'était pas adéquate pour la validation des algorithmes de segmentation, en raison du manque de précision de la segmentation manuelle[22].

1.7.3 Banco Web LAPIMO

Plus récente que les deux précédentes bases, la BancoWeb LAPIMO est une base créée en 2010 au Brésil. Elle dispose de 320 cas avec 1473 images (MLO et CC) contenant toutes les pathologies. Les antécédents de la patiente, de même qu'une classification BI-RADS sont fournis. En ce qui concerne les masses, seules certaines images sont annotées sous forme de ROI ; mais toutes les pathologies disposent d'une description textuelle[23].

1.8 Conclusion

Dans ce chapitre ont été abordées des notions de généralité en rapport avec le cancer du sein et les systèmes d'acquisition en imagerie du sein. Après avoir présenté de façon concise l'anatomie du sein, nous avons exposé les données statistiques qui prouvent que le cancer du sein est l'un des principaux problèmes de santé publique à l'échelle mondiale. L'objectif qui sous-tend la présentation de ces éléments, c'est de mettre en lumière les difficultés rencontrés par les radiologues dans l'analyse des mammographies. Dans le prochain chapitre nous allons proposer les étapes du traitement des images mammographique, qui permet d'améliorer les qualités des mammographies et donc l'obtention des meilleurs performances possible de segmentation et de classification .

A pink awareness ribbon is positioned behind the text, forming a loop that frames the words "Chapitre 2:" and "Traitement".

Chapitre 2:
Traitement
des
mammographies

Chapitre 2

Traitement des mammographies

2.1 Introduction

Les mammographies sont des images qui n'offrent pas un bon contraste entre les tissus mammaires sain et malin ; ceci s'explique par le fait que l'atténuation des rayons X entre ces tissus varie peu, surtout chez les jeunes femmes qui ont un parenchyme mammaire très dense . Le traitement d'image se résume comme suit :

- Le prétraitement consiste en diverses opérations visant à améliorer la qualité de l'image et à la faciliter de la segmentation. Ces opérations sont principalement le rehaussement du contraste, la modification des histogrammes et la réduction du bruit.
- L'analyse a pour but d'extraire les caractéristiques permettant de classifier les pixels de l'image.
- La segmentation est un traitement de bas niveau qui consiste à créer une partition de l'image en sous-ensembles appelés régions. La segmentation est alors obtenue par extraction des composantes connexes des pixels appartenant à la même classe. Ainsi une région sera constituée de pixels d'une même classe.

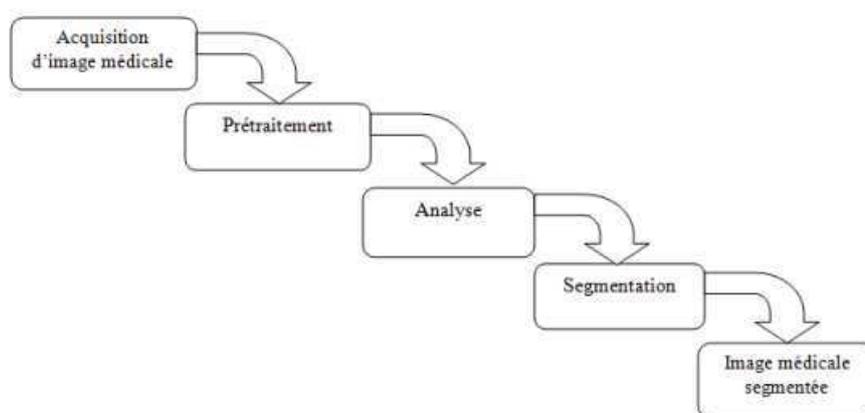


FIGURE 2.1 – Traitement d'image mammographique[24]

2.2 Prétraitement des mammographies

Les techniques standard de réduction du bruit et d'amélioration de contraste ont d'abord été appliquées aux mammographies, puis, on s'est rendu compte que, hormis l'amélioration apportée à l'image, ces approches contribuaient à amplifier de façon significative le bruit présent dans l'image. Des méthodes un peu plus spécialisées ont alors été proposées et ont globalement aidé à améliorer la qualité des mammographies, comme la relevé, qui a également insisté sur le fait que l'utilisation de ces techniques en milieu clinique peut réduire de moitié la dose de radiation délivrée durant l'acquisition des mammographies.

Cependant, le réel problème des algorithmes de prétraitement se situe dans le fait que certaines régions peuvent être traitées convenablement, tandis que d'autres le sont plus ou moins. Un manque de prétraitement peut engendrer la négligence de plusieurs détails concernant la lésion ; dans ce cas de figure, certaines tumeurs peuvent ne pas être détectées, causant ainsi des faux négatifs. A contrario, l'excès de rehaussement de contraste peut rajouter des détails ou artefacts à l'image, donnant ainsi au radiologue l'impression d'être en présence d'une tumeur, ce qui aboutit à des biopsies inutiles et une augmentation des faux positifs. A cela s'ajoute le fait que les approches locales et multi-échelles nécessitent la spécification de plusieurs paramètres (p.ex. la taille et le type de filtre, les seuils et les différentes résolutions) qui ont un grand impact sur la qualité des résultats obtenus[25].

Les techniques d'amélioration des images offrent une multitude de choix pour améliorer leur qualité visuelle. Un choix approprié de ces techniques est fortement influencé par l'importance des détails de l'image et par le type d'informations qu'on cherche à améliorer et/ou à supprimer. Ces techniques peuvent être divisées en deux catégories :

- Méthodes exploitant le domaine spatial, qui opèrent directement sur les pixels.
- Méthodes basées sur le domaine de fréquence, qui opèrent sur la transformée de Fourier d'une image.

Cette étape cherche l'amélioration de la qualité de l'image traitée. Elle opère, d'abord, d'une application d'un filtre passe-bas gaussien de faible taille afin d'augmenter la netteté des images puis un filtre chapeau était utile pour la modification des répartitions des valeurs des niveaux de gris dans l'image et donc pour améliorer le contraste. Elle applique par la suite un filtre morphologique pour diminuer les intensités sur l'image et donc un bon résultat du gradient.

2.2.1 Suppression du bruit

La réduction du bruit sur l'image est effectuée par le nettoyage d'image des artefacts de numérisation et l'application filtres gaussien afin de supprimer le bruit due au appareil d'acquisition(le mammogram dans notre cas).

* Suppression des artefacts

Une mammographie numérisée renferme généralement deux régions distinctives : la région exposée du sein et la région non exposée constituant le fond, ce fond peut contenir

Chapitre2 Traitement des mammographies

des artefacts que le système visuel humain peut facilement ignorer pendant l'interprétation par contre un système automatisé doit au préalable identifier et classifier ces artefacts qui engendrent des erreurs d'interprétation lors de l'analyse des images.

Le terme artefacts latin "arti facta" ou "effet de l'art", désigne une altération du résultat d'un examen radiologique selon certains procédés techniques utilisés (Figure 2.2). Les mammographies sont généralement marquées avec une certaine forme d'étiquette permanente d'identification contenant l'information concernant l'examen effectué.

Ces labels sont des indicateurs radiopaques exhibant la latéralité de la mammographie (R/L, Right/Left) ainsi que des indicateurs de projection MLO /CC (Figure 2.3 et 2.4).

Une raie de numérisation est un extremum local de haute intensité parallèle aux axes des abscisses de l'image mammographique.

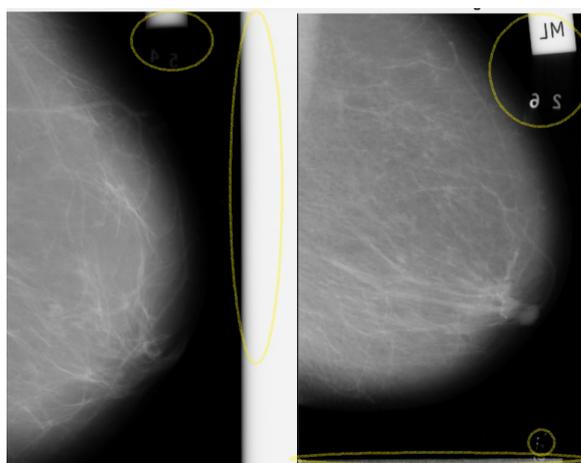


FIGURE 2.2 – Exemple des artefacts d'une image mammographique.

* Filtre gaussien

Un filtre gaussien est un filtre linéaire passe bas, donné par discrétisation de la fonction gaussienne :

$$G(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

2.1

La taille du filtre gaussien est gouvernée par σ qui doit être proportionnel à l'écart type du bruit (s'il est connu !!). En général un filtre gaussien avec $\sigma < 1$ est utilisé pour réduire le bruit, et si $\sigma > 1$ c'est dans le but de fabriquer une image qu'on va utiliser pour faire un masque flou personnalisé. Il faut noter que plus σ est grand, plus le flou appliqué à l'image sera important..

2.2.2 Amélioration de contraste

Le contraste d'une image peut être défini comme :

- L'écart-type des variations des niveaux de gris. Il s'exprime par l'équation :

$$C = 1/M * N \sum_{i=0}^M -1 \sum_j^N -1(I(i, j) - moy)^2$$

2.2

Avec, M et N sont les nombres respectifs de ligne et de colonne dans l'image I et moy est la moyenne des valeurs des niveaux de gris existants sur cette image. Après une étude comparatif des différentes méthodes de modification du contraste et d'amélioration des qualités des images mammographiques ,on peut juger que l'application de l'une des fonctions de transformation sur n'importe quel type d'image doit être argumentée par la préservation de ses caractéristiques. En effet, certaines transformations peuvent affecter des caractéristiques intéressantes, comme le cas du faible contraste et visibilité du contour du sein d'une image mammographique qui peut être affectée par une application d'une transformation exponentielle[26].

2.3 Segmentation des images mammographie

La segmentation est un problème omni-présent en traitement d'images et en vision par ordinateur dont le but est d'isoler un objet du fond de l'image. Dans le cadre de la conception d'un système de DAOx, la segmentation joue un rôle central en ce sens que les autres étapes à savoir la description de la masse, la détection des structures d'intérêt comme les spicules et la classification des différentes lésions en dépendent fortement[27]. Ainsi, plus la segmentation est précise, meilleure sera la description de la nature du contour, de la texture et de la forme de la masse, ce qui aura pour conséquence une bonne classification de la tumeur. La particularité de la segmentation réside dans le fait qu'il n'existe pas une méthode adaptée à toutes les applications. A ce titre, elle a fait l'objet de plusieurs publications, dans lesquelles les chercheurs ont tenté d'apporter des solutions diverses et variées qui tiennent souvent compte de la nature des images traitées.

De façon générale, les méthodes de segmentation sont regroupées en deux catégories que sont les approches supervisées et non-supervisées. Les approches non-supervisées réalisent une partition de l'image en plusieurs régions distinctes qui satisfont des propriétés liées au niveau de gris des pixels et à la texture. Ces approches sont également subdivisées en trois groupes que sont les méthodes basées régions, celles basées contour et celles basées clustering. Pour leur part, les approches supervisées c-à-d nécessitant une intervention humaine – sont fondées sur l'utilisation d'une information a priori sur la nature de l'objet à extraire et éventuellement le fond de l'image.

2.3.1 2.3.1 Segmentation par la ligne de partage des eaux(LPE)

Une idée pour segmenter les images est de déterminer les lignes le long desquelles les niveaux de gris varient rapidement. En faisant une analogie géographique, cela se rapproche de la notion de ligne de crête du module du gradient de l'image. Ici, l'image est vue comme une zone géographique dans laquelle les lignes de crêtes ou ligne de partage des eaux sont assimilées aux contours de l'image. Pour cela, on définit son complémentaire : les bassins versants. Un bassin versant est une zone géographique d'où une goutte d'eau, suivant la ligne de plus grande pente, arrivera dans ce minimum[28]. On associe d'ailleurs un minimum à un bassin versant.

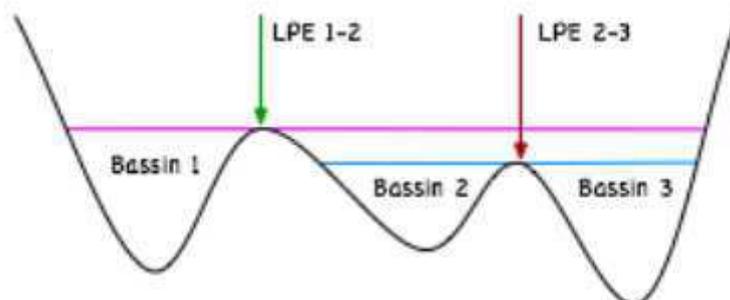


FIGURE 2.3 – En faisant monter l’eau dans les bassins 2 et 3, ceux-ci finissent par se rejoindre en la ligne de partage des eaux 2-3 (symbolisée par une flèche rouge). De la même façon, lorsque le niveau continue de monter, les bassins 1 et 2 se rejoignent en la ligne 1-2 (symbolisée par la flèche verte)[29].

Cette technique nous donne de façon efficace des contours fermés et squelettisés; ces contours fermés forment donc une partition de l’image. des reliefs (si l’on observe l’image en trois dimensions, le niveau de gris étant la hauteur) et la recherche de la ligne de partage des eaux nous donnera les frontières entre chaque région(La figure 2.6 illustre sommairement ce principe).

Une façon simple de trouver la ligne de partage des eaux est de partir des points les plus bas du relief et de les inonder peu à peu. On voit alors se former les bassins versants. Lorsque deux bassins versants se rejoignent, en général sur une ligne de crête, le point de jonction appartient à la ligne de partage des eaux.

* Méthode générale de la LPE

Nous considérons une image comme un tableau à deux dimensions dont chaque case porte une valeur de niveau de gris. Ce niveau de gris sera un entier (int). Nous commençons par ordonner l’ensemble des pixels (les cases du tableau) par niveau de gris croissant. Cela va nous permettre de trouver les pixels qui sont des minima locaux de l’image. Nous devons en effet commencer l’inondation par les pixels dont le niveau de gris est le plus bas.

De nombreux problème se posent dans un espace discret, notamment le fait que, si la ligne de crête est un plateau, la ligne de partage des eaux dépend fortement du balayage choisi pour analyser l’image. Une simple rotation pourrait nous fournir un résultat différent ce qui n’est guère acceptable. De plus, aborder le problème de l’inondation de manière simpliste nous conduirait à un algorithme trop gourmand en temps de calcul.

Nous allons donc utiliser un algorithme à pile FIFO(First IN First OUT) Les pixels à examiner seront placés dans une pile, laquelle sera vidée à un autre moment du processus d’inondation pour découvrir la LPE[30]. Le processus se décompose en plusieurs étapes :

- élaboration d’un tableau trié des pixels par ordre croissant de niveau de gris.
- parcours des pixels d’iso-niveau, pour les marquer et éventuellement les placer dans la queue de traitement en fonction de leur voisinage.

Chapitre 2 Traitement des mammographies

– parcours de la pile de pixels pour décider s'ils appartiennent à la ligne de partage des eaux et leur attribuer une valeur ou attribuer une valeur à leurs voisins.

– deuxième parcours des pixels d'iso-niveau pour les placer dans la pile de traitement ou placer leurs voisins dans la pile de traitement.

Ce n'est qu'à l'issue de ces étapes que les pixels de l'image auront obtenu une valeur qui correspondra à :

– une appartenance à la ligne de partage des eaux caractérisée par un niveau égal à zéro.

– une appartenance à un bassin versant caractérisée par un niveau supérieur à zéro.

Tous les pixels appartenant à un même bassin versant posséderont le même niveau, c.à.d le numéro du bassin. La figure suivante donne un aperçu des résultats complets de traitement.

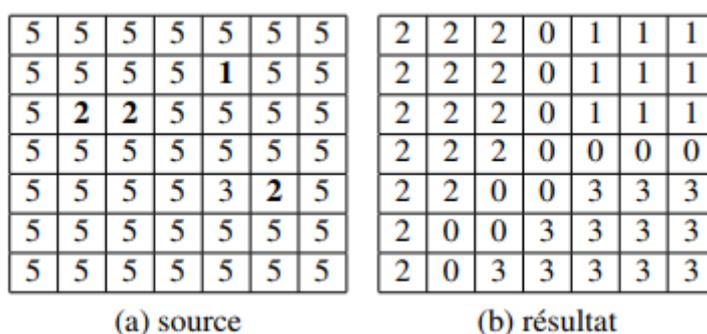


FIGURE 2.4 – L'image (a) est transformée en l'image (b). Dans cette image on retrouve les trois bassins versants numérotés de 1 à 3 ainsi que les pixels de ligne de partage des eaux étiquetés à 0 [31].

* LPE contrôlée par marqueurs

Le principe de la modification d'homotopie du gradient est d'imposer les marqueurs des régions à segmenter comme minima du gradient en supprimant tous les autres minima indésirables qui sont à l'origine de toute sur-segmentation (Fig 2.5). On inonde ensuite ce gradient à partir de tous les marqueurs. Une et une seule ligne de partage des eaux est alors présente entre chaque marqueur, et elle a tendance à se situer sur le contour des objets à segmenter, qui ont déjà été pré-détectés par le gradient.

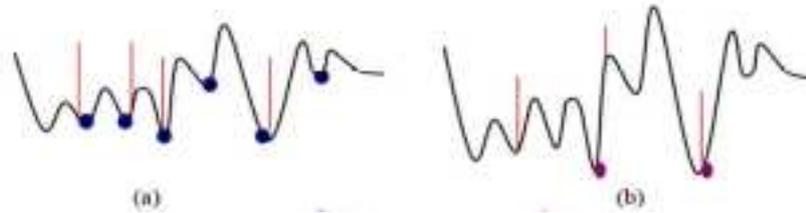


FIGURE 2.5 – LPE contrôlée par marqueurs. (a) Les minimas sans contrainte, (b) processus de LPE avec contrainte des marqueurs.

2.4 Conclusion

L'étape de prétraitement dans un processus de reconnaissance de formes constitue le point de départ de toute autre analyse de haut niveau regroupant les opérations de filtrage, d'amélioration de contraste et de segmentation. Nous avons présenté dans ce chapitre quelques méthodes de prétraitement, que nous avons utilisées dans notre projet.



Chapitre 3:
Classification
par
SVM

Chapitre 3

Classification par SVM

3.1 Introduction

Support Vector Machine SVM également appelé « séparateurs à vaste marge ou les machines à vecteurs de support » est une nouvelle approche de la classification de modèles supervisés qui a été appliquée avec succès à un large éventail de problèmes de reconnaissance de modèles. Elle constitue également un algorithme pour l'apprentissage des règles de classification et de régression à partir de données. Le modèle SVM est le plus approprié pour travailler avec précision et efficacité avec des espaces de fonctions de haute dimensionnalité. Le SVM est basé sur des bases mathématiques solides et donne des résultats simples et des algorithmes très puissants.

L'algorithme SVM a été développé dans les années 90 par le russe Vladimir Vapnik [35]. Initialement, les SVM ont été développés comme un algorithme de classification binaire supervisée. Il s'avère particulièrement efficace de par le fait qu'il peut traiter des problèmes mettant en jeu de grands nombres de descripteurs, qu'il assure une solution unique (pas de problèmes de minimum local comme pour les réseaux de neurones) et il a fourni de bons résultats sur des problèmes réels.[34].

3.2 Support vector machine(SVM)

Le SVM a été appliqué à de très nombreux domaines (bioinformatique, recherche d'information, vision par ordinateur, finance....)[35].

Le principe de base du SVM consiste de ramener le problème de la discrimination à celui, linéaire, de la recherche d'un hyperplan optimal. Deux idées ou astuces permettent d'atteindre cet objectif :

- La première consiste à définir l'hyperplan comme solution d'un problème d'optimisation sous contraintes dont la fonction objectif ne s'exprime qu'à l'aide de produits scalaires entre vecteurs et dans lequel le nombre de contraintes "actives" ou vecteurs supports contrôle la complexité du modèle.
- Le passage à la recherche de surfaces séparatrices non linéaires est obtenu par l'introduction d'une fonction noyau (kernel) dans le produit scalaire induisant implicitement une transformation non linéaire des données vers un espace intermédiaire (feature space) de plus grande dimension. D'où l'appellation couramment rencontrée de machine à noyau ou kernel machine.

3.2.1 Principe de fonctionnement

soit un ensemble d'exemples d'apprentissage (x_i, y_i) où les x_i sont les instances de données du monde réel et les y_i les étiquettes indiquant la classe à laquelle appartient l'instance. Pour le problème de reconnaissance de modèle à deux classes, $y_i = +1$ ou $y_i = -1$. Un exemple d'apprentissage (x_i, y_i) est appelé positif si $y_i = +1$ et négatif si $y_i = -1$. On cherche alors une « surface de séparation » $f : X \rightarrow \mathbb{R}$ tel que si $f(x) > 0$ alors x est affecté à la classe $+1$ et si $f(x) < 0$ alors x est affecté à la classe -1 .

Pour un problème de classification linéaire on suppose que les deux classes $(-1 \text{ et } +1)$ sont séparables par un hyperplan, la fonction f a donc la forme :

$$f(x) = \sum_{i=1}^n w_i x_i + b$$

3.1

où w est le vecteur orthogonal à l'hyperplan et b est le déplacement par rapport à l'origine. Pour juger la qualité d'un hyperplan en tant que séparateur on utilise la distance entre les exemples d'apprentissage et ce séparateur. Plus précisément, la « marge » d'un problème d'apprentissage est définie comme la distance entre le plus proche exemple d'apprentissage et l'hyperplan de séparation. Pour un hyperplan H on a [36](Figure 3.1) :

$$\text{Marge}(H) = \min d(x_i, H)$$

3.2

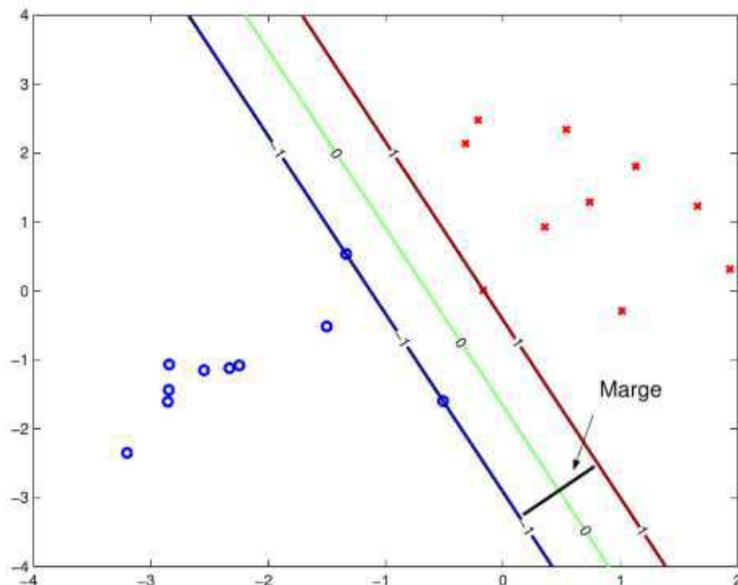


FIGURE 3.1 – Le séparateur idéal correspond intuitivement à l'hyperplan qui passe au milieu, c'est le séparateur de marge maximale.

Les SVM linéaires cherchent le séparateur (l'hyperplan de séparation) qui maximise la marge. On appelle cela "séparateur à vaste marge".

Les éléments de la classe 1 les plus proches de ce séparateur se trouvent à la même distance du séparateur que les éléments les plus proches de la classe 2 (cette distance est égale à la marge). Ces éléments, soit d'une classe soit de l'autre, s'appellent "vecteurs de support".

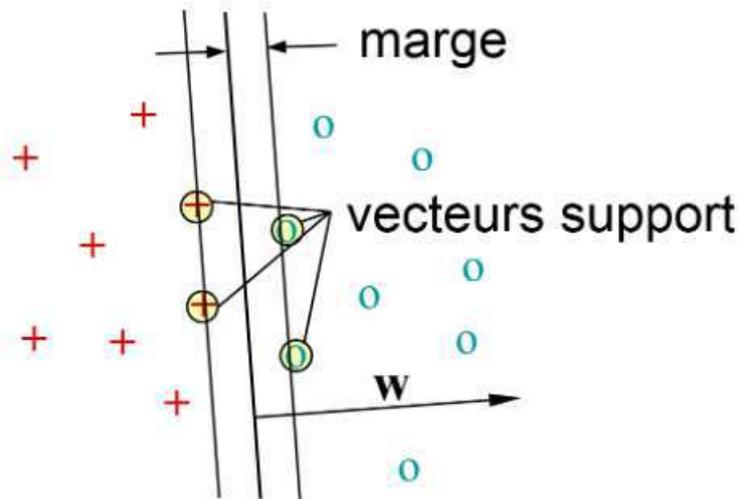


FIGURE 3.2 – Les "vecteurs de support" se trouvent à une distance égale à la marge d'un côté ou de l'autre de l'hyperplan de séparation[37].

Intuitivement, ce sont les vecteurs de support qui déterminent le séparateur (par l'intermédiaire de la fonction distance et de leur configuration géométrique). Une fois le séparateur $f(x)$ trouvé, la classification d'un nouvel exemple se fait par une simple décision à seuil zéro :

- $f(x) = 0$: l'élément se trouve sur la frontière de séparation, pas de décision.
- $f(x) > 0$: classe 1.
- $f(x) < 0$: classe 0.

3.3 Formalisme des SVM

Fondamentalement, un SVM cherchera simplement à trouver un hyperplan qui sépare les deux catégories d'un problème. Soit l'espace vectoriel E de dimension n , doté de la base (e_1, \dots, e_n) . L'équation caractéristique d'un hyperplan est de la forme :

$$w_1 \cdot e_1 + w_2 \cdot e_2 + \dots + w_n \cdot e_n = 0$$

3.3

Ainsi, si l'on se place dans \mathfrak{R}^n muni de la base (e_1, \dots, e_n) , pendant son entraînement le SVM calculera un hyperplan vectoriel d'équation $w_1 \cdot e_1 + w_2 \cdot e_2 + \dots + w_n \cdot e_n = 0$, ainsi qu'un scalaire (un nombre réel) b . C'est ce scalaire b qui va nous permettre de travailler

avec un hyperplan affine. Le vecteur $w = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$ est appelé vecteur de poids, le scalaire b

est

appelé biais.

Une fois l'entraînement terminé, pour classer une nouvelle entrée $x = \begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_n \end{pmatrix}$ le SVM

regardera le signe de :

$$f(x) = w_1.a_1 + w_2.a_2 + \dots + w_n.a_n + b = \sum_{i=1}^n w_i.a_i + b = w^t.a + b \tag{3.4}$$

Si $f(x)$ est positif ou nul, alors x est d'un côté de l'hyperplan affine et appartient à la première catégorie, sinon x appartient à la seconde catégorie.

3.4 Entraînement des SVM

En conséquence, étant donné un hyperplan de vecteur de poids w , et de biais b , nous pouvons calculer si un point x_k appartient à telle ou telle catégorie, grâce au signe de $f(x_k)$. Supposons que l'on assigne à tout point x_k un label l_k qui vaut 1 si x_k appartient à la première catégorie, et -1 si x_k appartient à la seconde catégorie. Alors, si le SVM est correctement entraîné, on a toujours $l_k f(x_k) \geq 0$, c'est-à-dire $l_k(w^t.x_k + b) \geq 0$. Le but d'un SVM, lors de l'entraînement, est donc de trouver un vecteur de poids w et un biais b tels que, pour tout point x_k de label l_k appartenant aux données d'entraînement $l_k(w^t.x_k + b) \geq 0$. Autrement dit, de trouver un hyperplan séparateur entre les deux catégories. On choisira l'hyperplan qui maximise la marge.

* Calcul de la marge

La distance d'un point x_k à l'hyperplan de vecteur poids w et de biais b est donnée par :

$$\frac{l_k(w^t.x_k + b)}{\|w\|} \tag{3.5}$$

où $\|w\|$ désigne la norme euclidienne de w . La marge (distance minimale de l'hyperplan à un des points d'entraînement) d'un hyperplan de paramètres (w, b) par rapport à un ensemble de points x_k est donc :

$$\min_k \frac{l_k(w^t.x_k + b)}{\|w\|} \tag{3.6}$$

*** Maximisation de la marge**

L'objectif est de trouver l'hyperplan qui permet de maximiser la marge. Autrement dit, chercher l'unique hyperplan dont les paramètres (w, b) sont donnés par la formule :

$$\operatorname{argmax}_{(w,b)} \frac{\min_k (l_k w^t \cdot x_k + b)}{\|w\|} \tag{3.7}$$

Il existe plusieurs couples (w, b) qui décrivent le même hyperplan. On décide de ne considérer que l'unique paramétrage (w, b) tel que les vecteurs support x_s vérifient $l_s(w^t \cdot x_s + b) = 1$ par conséquent $l_k(w^t \cdot x_k + b) \geq 1$, et l'égalité est atteinte si x_k est un vecteur support. Dit autrement, cette normalisation sur w et b permet de garantir que la marge citée dans l'équation 3.6 est alors de :

$$\frac{1}{\|w\|} \tag{3.8}$$

Le problème d'optimisation de la marge 3.7 se simplifie ainsi en :

$$\operatorname{argmax}_{(w,b)} \frac{1}{\|w\|} \tag{3.9}$$

tout en gardant d'hypothèses de normalisation, à savoir :

$$\forall_k l_k(w^t \cdot x_k + b) \geq 1 \tag{3.10}$$

le problème revient à :

$$\begin{cases} \text{maximiser} & \frac{1}{\|w\|} \\ \text{sous la contrainte} & \forall_k l_k(w^t \cdot x_k + b) \geq 1 \end{cases} \tag{3.11}$$

Que l'on peut reformuler pour des raisons pratiques de la façon suivante :

$$\begin{cases} \text{minimiser} & \frac{\|w\|^2}{1} \\ \text{sous la contrainte} & \forall_k l_k(w^t \cdot x_k + b) \geq 1 \end{cases} \tag{3.13}$$

Ce genre de problème est appelé problème d'optimisation quadratique, et il existe de nombreuses méthodes pour le résoudre. Dans le cas présent, on utilise la méthode des multiplicateurs de Lagrange. Cette résolution donnera une valeur optimale pour w . Pour retrouver b , il faut résoudre l'équation suivante :

$$\min_k (l_k(w^t \cdot x_k + b)) = 1 \tag{3.14}$$

3.5 SVM non linéaire et la fonction noyau

Dans ce paragraphe, le problème des données non linéairement séparables est présenté. Pour rappel, des données sont non linéairement séparables quand il n'existe pas d'hyperplan capable de séparer correctement les deux catégories. Ce qui, d'ailleurs, arrive quasiment tout le temps en pratique. Afin de remédier à ce problème, une solution consiste à transposer les données dans un autre espace (en général de plus grande dimension) dans lequel elles sont linéairement séparables (ou presque) et ensuite appliquer l'algorithme SVM sur les données transposées. On a donc une transformation d'un problème de séparation non linéaire dans l'espace de représentation en un problème de séparation linéaire dans un nouvel espace de plus grande dimension. Cette transformation non linéaire est réalisée via une fonction noyau [38].

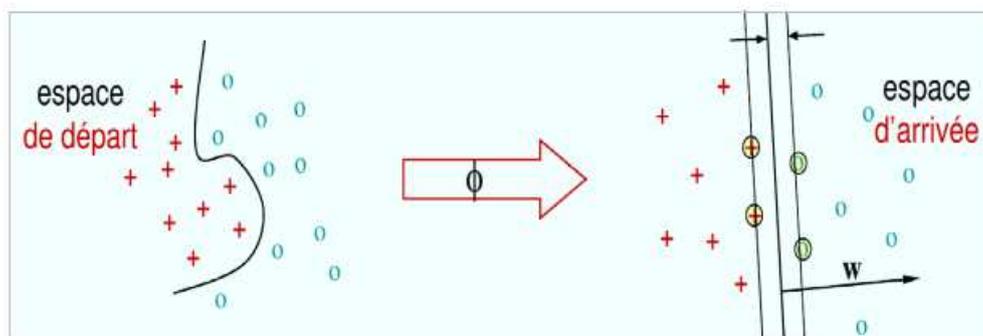


FIGURE 3.3 – Astuce à noyaux : projeter les données dans un espace de dimension beaucoup plus grande, où elles deviennent séparables linéairement.

En pratique, quelques familles de fonctions noyau paramétrables sont connues et il revient à l'utilisateur de SVM d'effectuer des tests pour déterminer celle qui convient le mieux pour son application. On peut citer les exemples de noyaux suivants : polynomiale, gaussien, sigmoïde et laplacien.

Exemple de séparation non linéaire

On a donc une transformation d'un problème de séparation non linéaire dans l'espace de représentation en un problème de séparation linéaire dans un espace de re-description de plus grande dimension. Cette transformation non linéaire est réalisée via une fonction noyau.

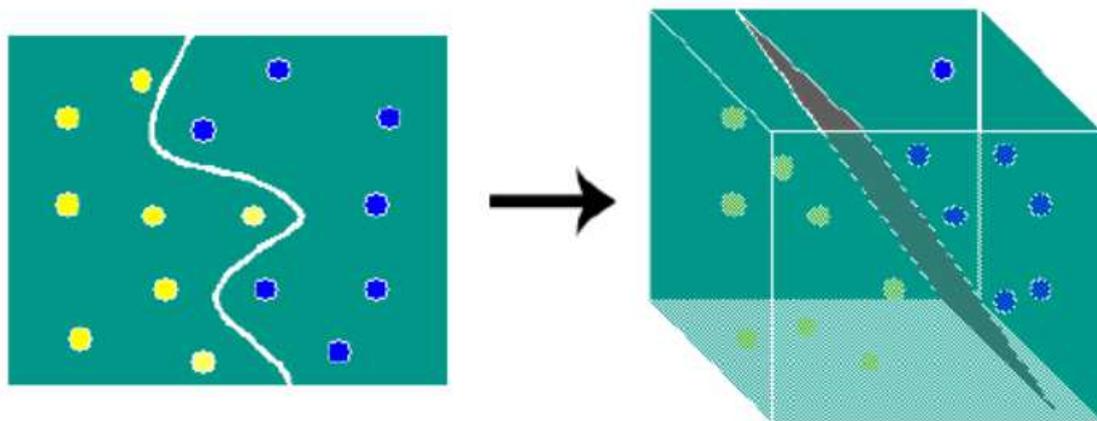


FIGURE 3.4 – Exemple de changement de l'espace de données.

3.6 Conclusion

Ce chapitre nous a permis de bien expliquer la classification supervisée en mentionnant la classification par machine à support vecteur .
Les machines à vecteurs de support constituent une approche intéressante pour la classification des données.



Chapitre 4:
Conception du
système d'aide
au diagnostic
assisté par
Ordinateur

Chapitre 4

Conception du système d'aide au diagnostic assisté par ordinateur

4.1 Introduction

Après avoir réalisé le prétraitement des images mammographiques suivie ensuite de la segmentation des masses et de l'extraction des spécules dans les chapitres précédents, nous abordons dans cette partie la dernière phase de la conception de notre système. Il s'agit de l'extraction des descripteurs qui doivent servir à la caractérisation des masses et à une prise de décision de notre système à propos de la nature de la masse.

Nous proposons une validation qualitative et quantitative des résultats obtenus. Cette validation se base sur l'exploitation du résultat d'une méthode de segmentation qui a été réservée dans les chapitres précédents.

L'extraction des descripteurs est une étape fondamentale dans tout système de traitement d'images mammographiques, car ce sont eux qui vont traduire l'aspect pathologique et la spécificité des lésions par des quantités facilement utilisables par les algorithmes de classification. Toutefois, les systèmes conçus sur de telles approches souffrent de plusieurs limitations qui justifient leur absence en milieu médical. Parmi ces limitations, nous pouvons évoquer le fait que les performances obtenues varient en fonction des descripteurs utilisés et du jeu de données ayant servi pour la validation, de plus, la combinaison de certains descripteurs peut dégrader les performances de ces systèmes, ce qui oblige à sélectionner de façon minutieuse les descripteurs les plus pertinents.

4.2 Problématique

A cause de l'augmentation du nombre de mammographies réalisées ces dernières années il reste difficile pour les radiologues experts de fournir des analyses précises et uniformes. D'où l'importance des outils d'aide au diagnostic développés durant ces vingt dernières années. Un système d'aide à la prise de décision (CAOx) dédiée à l'imagerie médicale est un programme numérique qui consiste à assister le spécialiste dans l'interprétation et le classement des images médicales. Son rôle est de détecter et de classer une ou plusieurs tumeurs dans une image comme étant tumeurs malignes ou bénignes.

L'objectif de ce projet est de mettre au point un système d'aide automatique à la détection et la classification des masses (opacités) dans des images mammographiques en utilisant la ligne des partages des eaux et les SVM.

4.3 Base de données

Nous avons utilisées[36] la base de données Mini-MIAS qui est une version réduite de la base d'image originaale MIAS. Les images sont de type MLO (incidence latérale oblique), la résolution spatiale de chaque image est de 1024 x 1024 pixels. Les différentes pathologies mammaires issues de la base de données MIAS qui font l'objet de notre étude représentent principalement des masses de type rond, des spécules et des distorsions architecturale bénignes et malignes .

Notre base de données contient 58 images réparties comme suit :

- (31) images de masses malignes.
- (27) images de masses bénignes.

4.4 Système d'aide au diagnostic assisté par ordinateur proposé

La conception du système d'aide au diagnostic proposé nécessite quatre étapes (figure 4.1) :

- 1- Le prétraitement des images mammographiques.
- 2- La détection des opacités (segmentation).
- 3- Extraction des caractéristiques (descripteurs).
- 4- La classification des masses en tant que malignes ou bénigne.

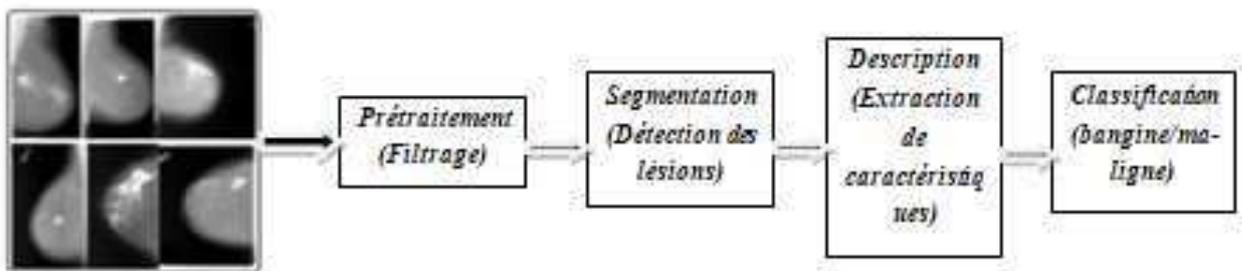
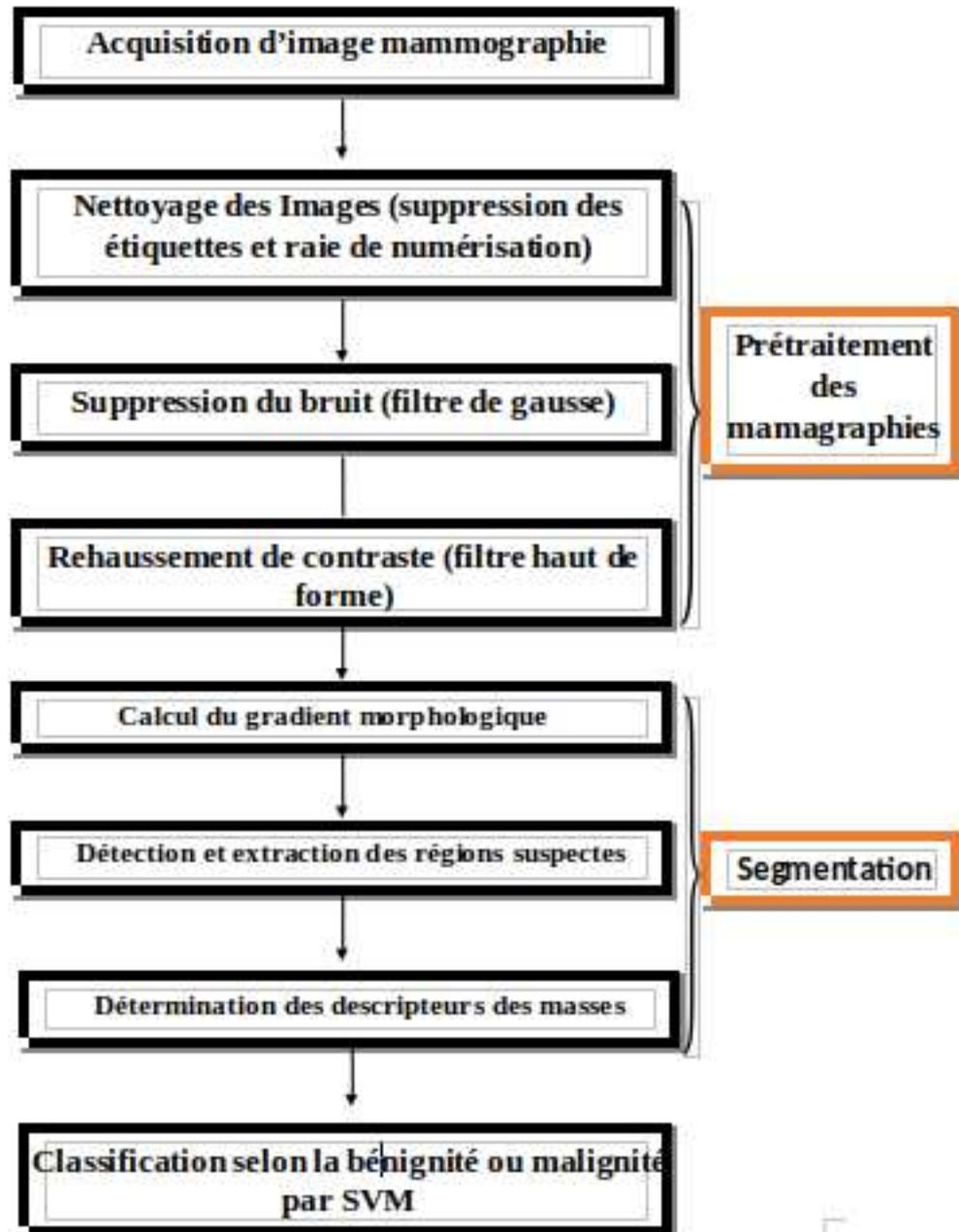


FIGURE 4.1 – Système de diagnostic assisté par ordinateur

Dans ce qui suit, nous allons détaillé ces différentes étapes comme indiqué dans l'algorithme 4.2.



L'algorithme 4.2 : Les différentes étapes du traitement des mammographies.

4.4.1 Prétraitement des mammographies

Le prétraitement regroupe l'ensemble des processus visant à améliorer les caractéristiques d'une image. Nous avons choisi les deux images (a) mdb206 et (b) mdb083 pour illustrer les résultats de notre traitement.

4.4.1.1 Suppression du bruit

Il s'agit essentiellement d'enlever le bruit (parasite) afin d'améliorer la qualité des images.

A- suppression des raies et étiquettes de numérisation

La détection et la suppression de raies de numérisation sont effectuées à l'aide de deux opérations : seuillage et suppression des petites régions. Nous avons réalisé un seuillage automatique sur l'image mammographique originale, par la suite toutes les régions qui ne font pas partie du sein sont supprimées. La figure 4.2 illustre des exemples pris de la base miniMias.

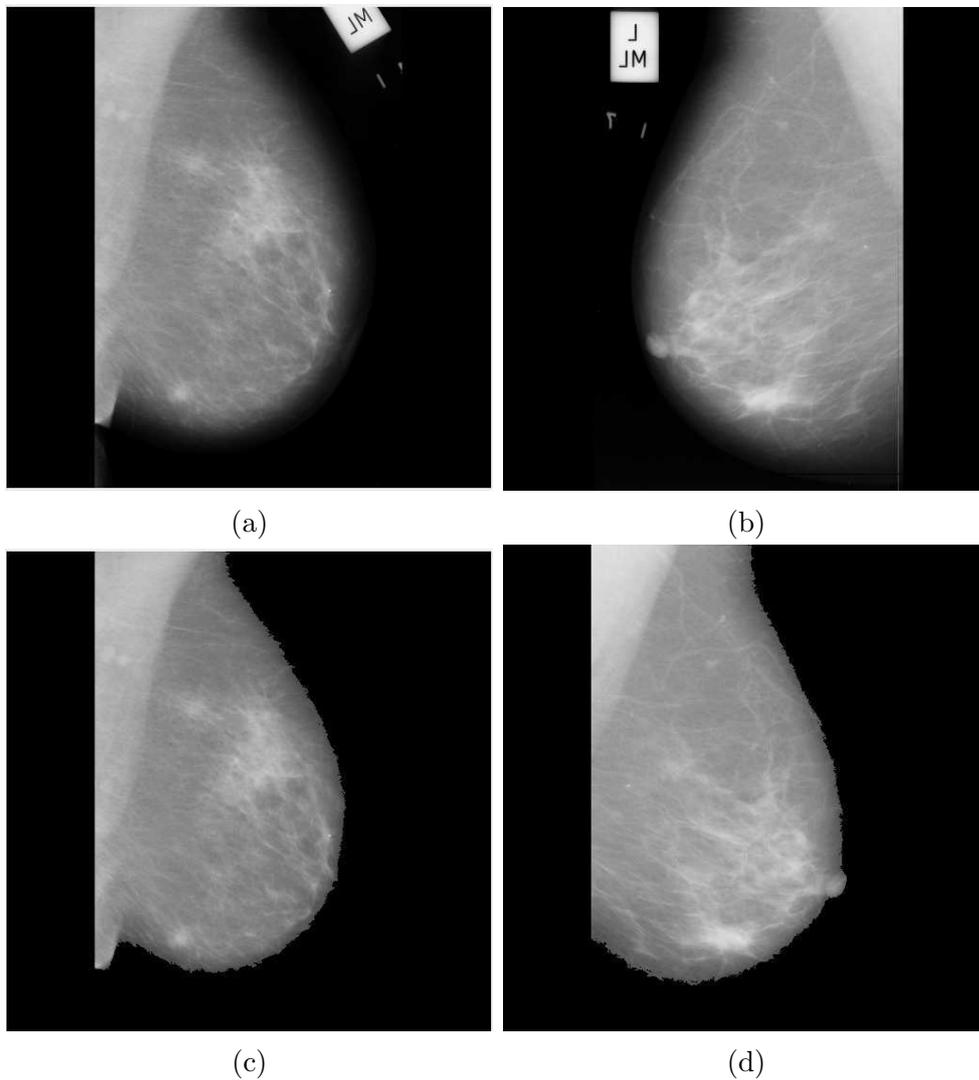


FIGURE 4.2 – Suppression des artefacts. (a) et (b) images originelle, (c) et (d) après suppression des artefacts.

B- Filtre gaussien

Nous avons utilisé un filtre gaussien ($\sigma = 0.7$) pour supprimer le bruit. Les résultats obtenus sont présentés dans la (figure 4.3).

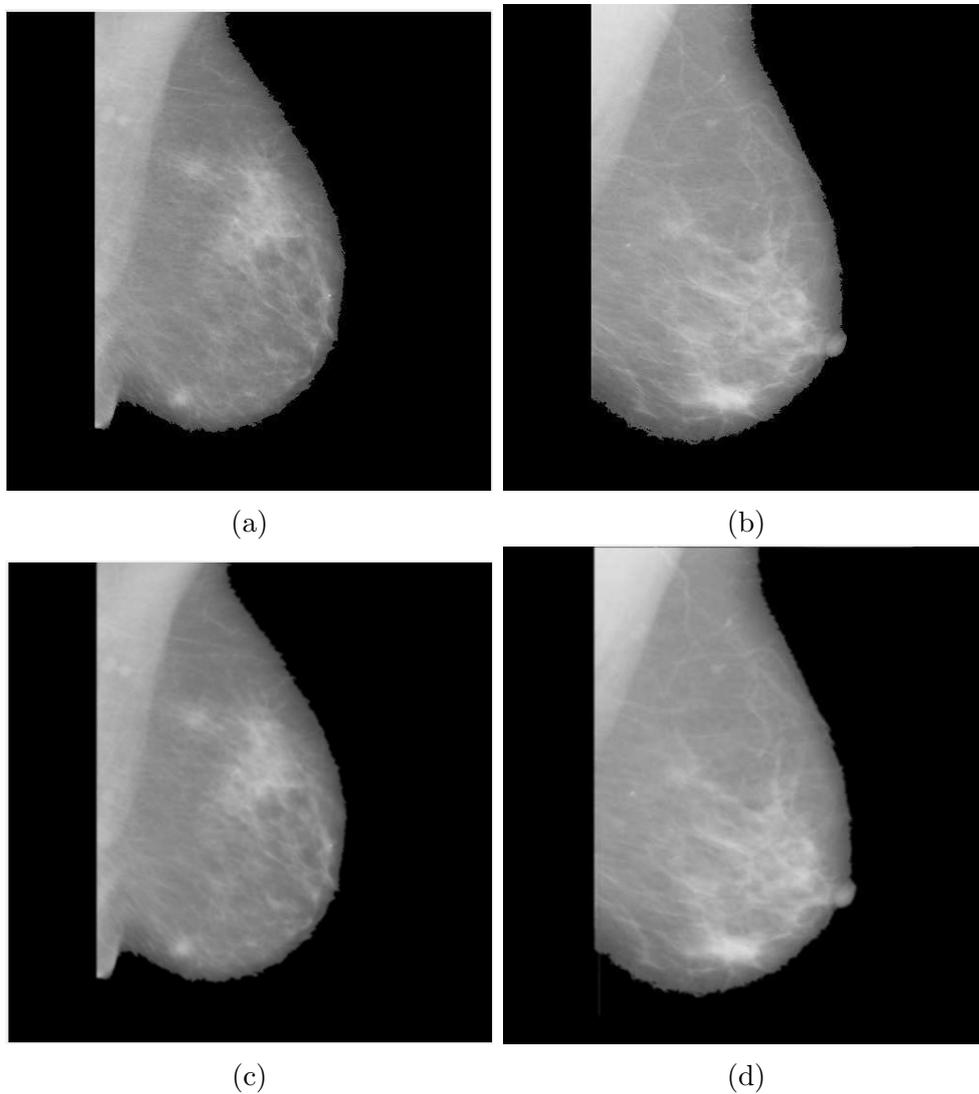


FIGURE 4.3 – Résultat de suppression du bruit.(a) et (b) les images sans artefacts, (c) et (d) application du filtre gaussien.

4.4.1.2 Rehaussement de contraste

Après le filtrage de l'image, un rehaussement de contraste est effectué pour mettre en évidence toute les taches de hautes intensité, autrement toutes les régions susceptibles d'être des masses.

A- Ajustement de contraste par expansion de dynamique

Dans un premier temps, nous avons ajustées le contraste en utilisant matlab(Figure 4.4) :

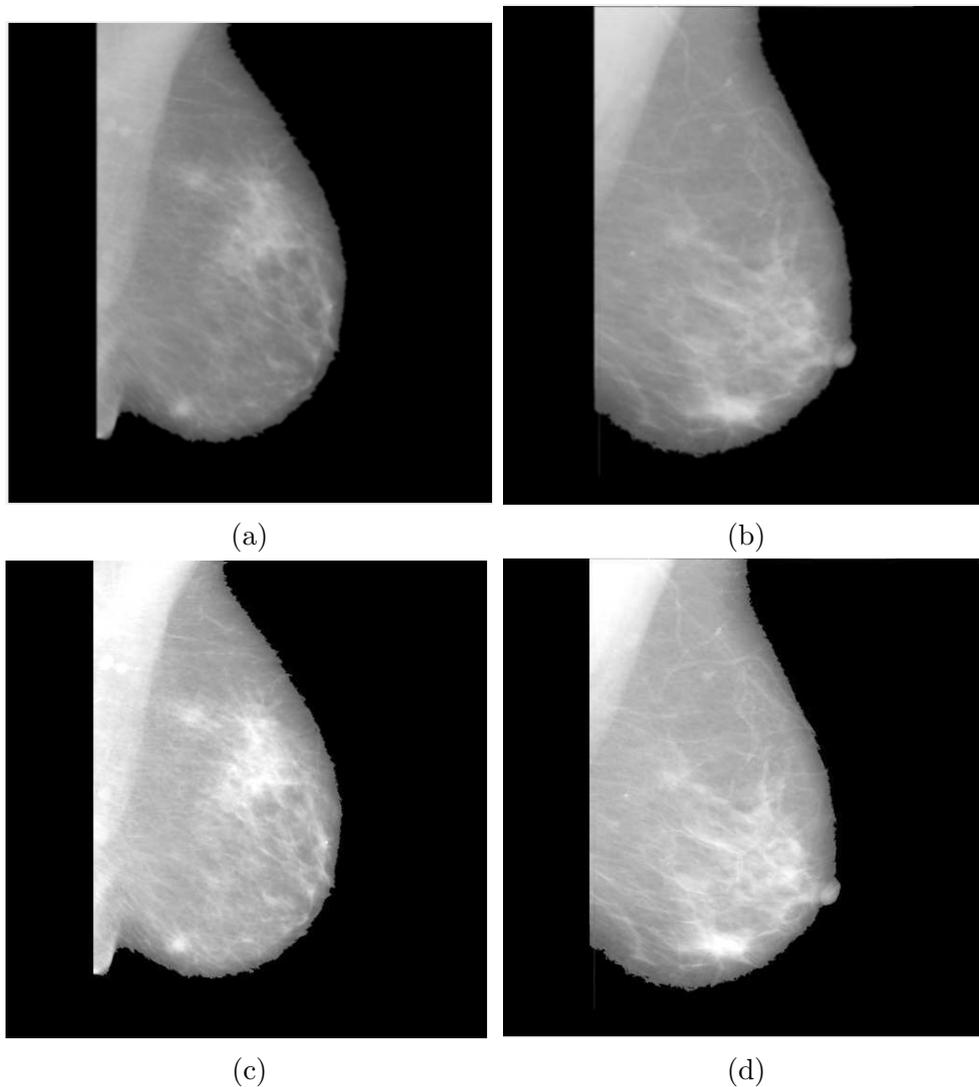


FIGURE 4.4 – Amélioration du contraste.(a) et (b) les images filtrées, (c) et (d) rehaussement de contraste par expansion de la dynamique.

Pour améliorer encore plus le contraste, on a décidées d'appliquer le filtre haut de forme :

B- Filtre haut de forme

L'algorithme proposé dans cette étape consiste à améliorer le contraste de l'image en utilisant deux filtre de type morphologiques :

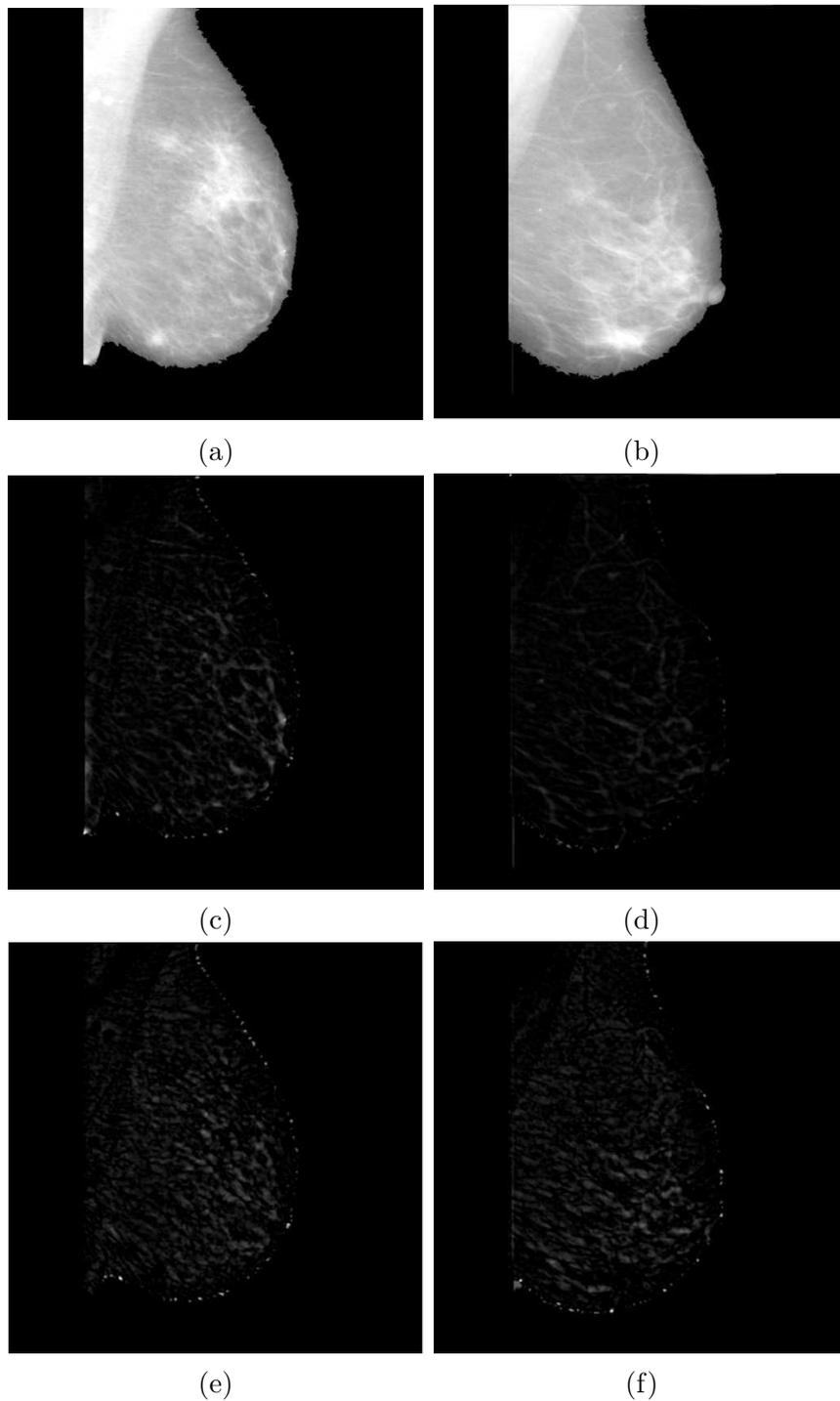


FIGURE 4.5 – Application du filtre haut de forme. (a) et (b) Images rehaussées, (c) et (d) Filtre haut de forme blanc, (e) et (f) Application du filtre haut de forme noir.

Pour maximiser le contraste entre les objets et les espaces qui les séparent, nous avons ajouté l'image top-hat à l'image d'origine, puis soustrayons celle-ci du résultat (Figure 4.6).

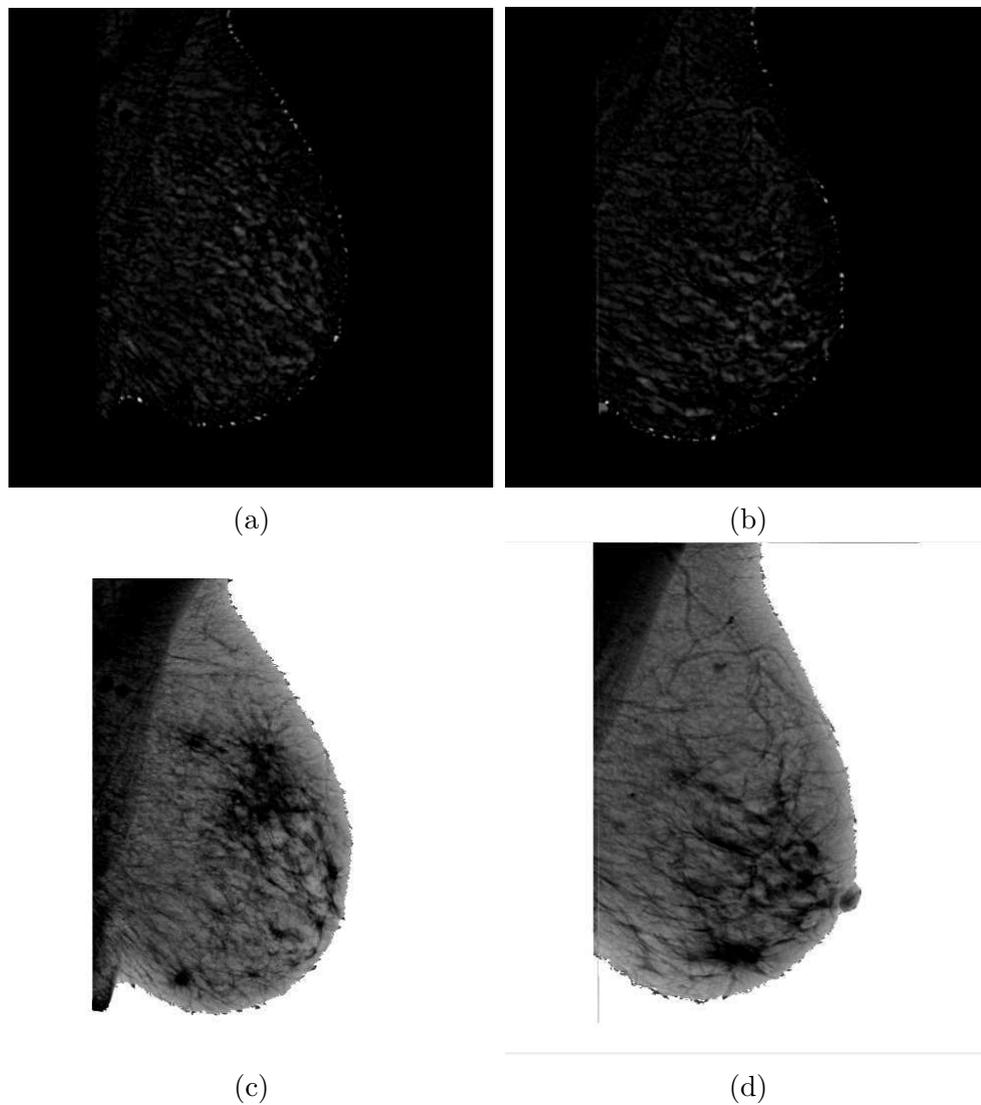


FIGURE 4.6 – Amélioration de contraste. (a) et (b) Filtre haut de forme noir, (c) et (d) Maximisation du contraste.

4.4.2 Segmentation : Détection des masses par LPE

L'étape de segmentation vise à identifier les régions du sein avec une plus grande possibilité d'être des masses. Dans ce travail, la segmentation est effectuée par LPE contrôlée par marqueurs.

4.4.2.1 Gradient morphologique

L'utilisation directe de la LPE sur le gradient de l'image va engendrer une sur-segmentation. Pour éviter ce problème, on a précédées cette étape par une ouverture-fermeture géodésique qui permet d'améliorer l'image en mettant en surbrillance les opacités.

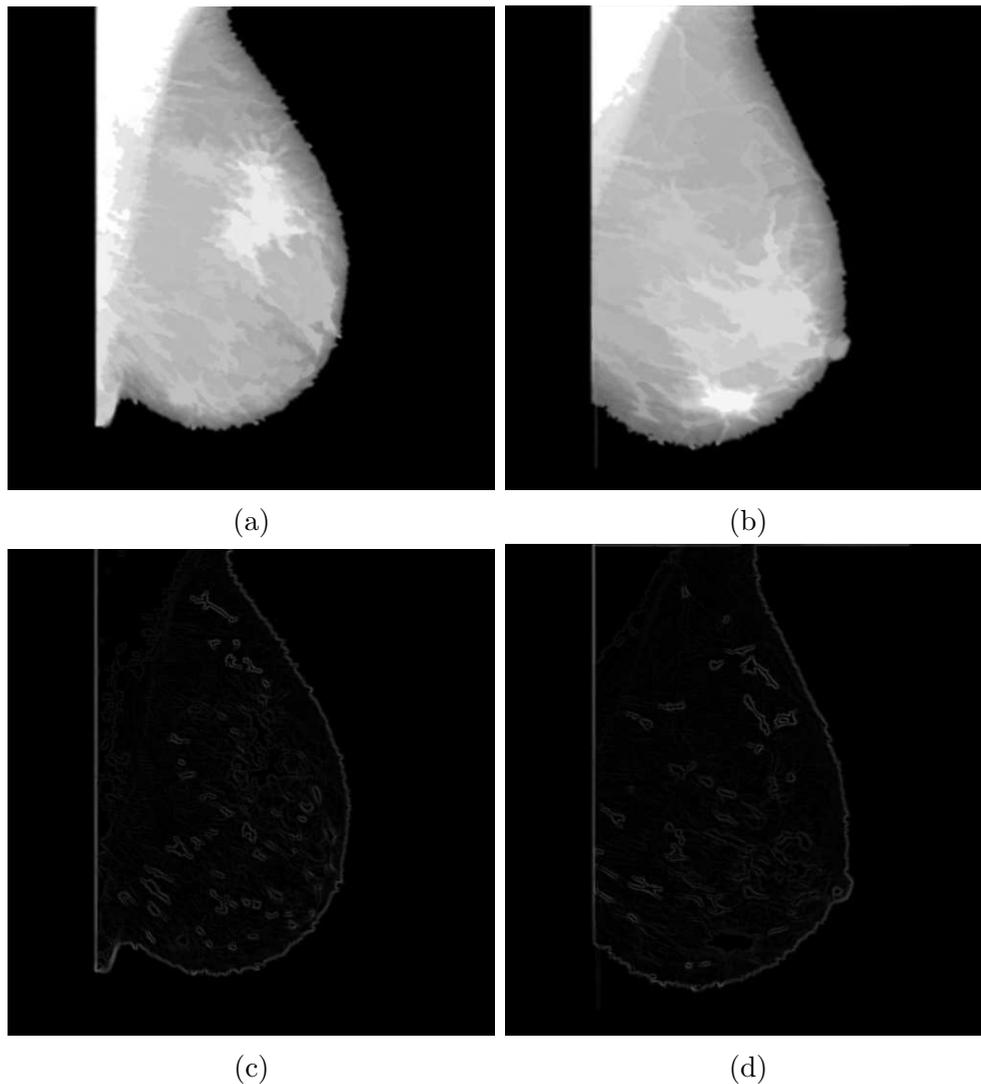


FIGURE 4.7 – Détermination du gradient morphologique. (a) et (b) ouverture et fermeture géodésique, (c) et (d) gradient morphologique

4.4.2.2 L'algorithme LPE

A cette étape nous possédons toutes les données nécessaires pour calculer la LPE afin de détecter et d'extraire les masses suspectes.

La figure 4.8 illustre les résultats de segmentation par LPE obtenus sur les images mdbxx et mdbyy. Ils montrent bien l'efficacité de l'algorithme à extraire les masses suspectes qu'elles soient rondes, ovales, de fort ou faible contraste, de contour régulier ou non.

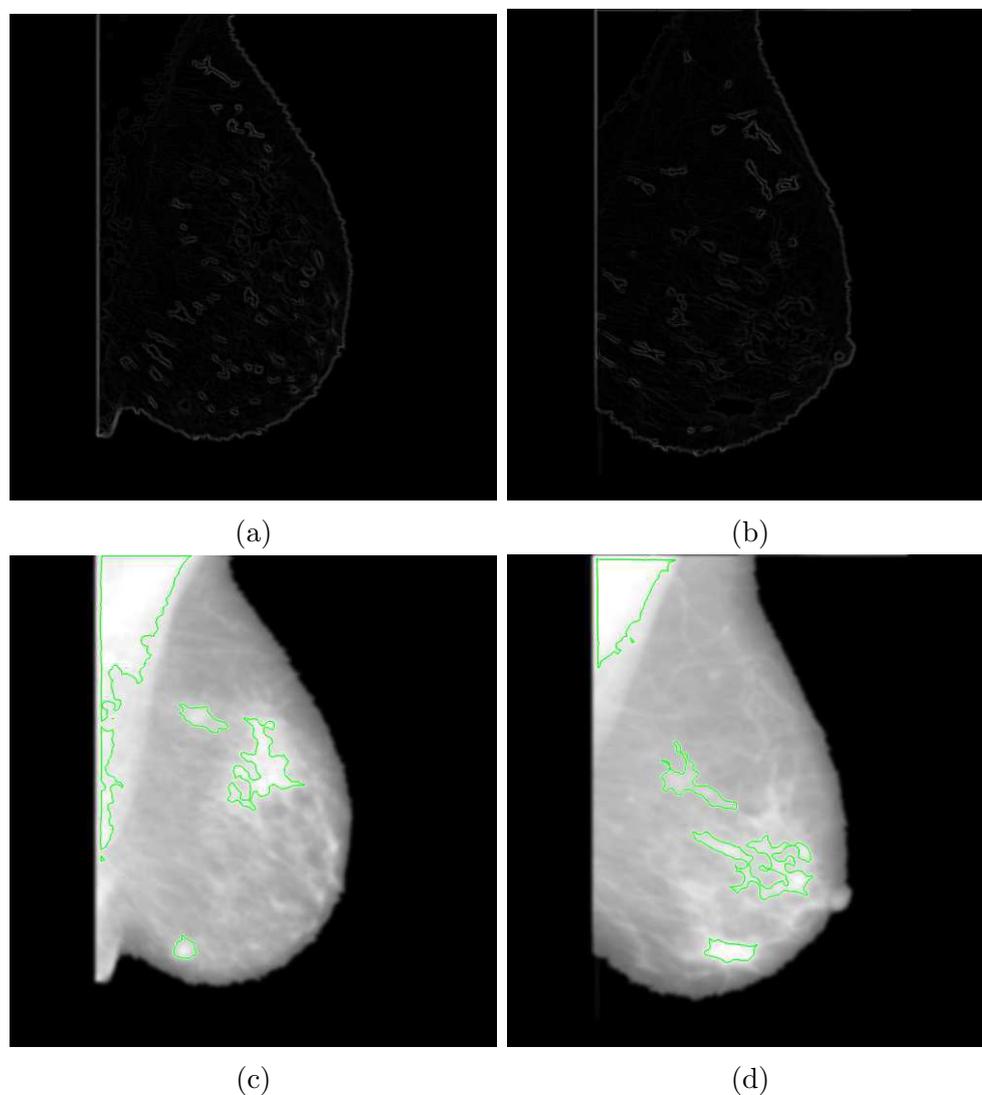


FIGURE 4.8 – L'extraction des masses. (a) et (b) le gradient morphologique. (c) et (d) LPE contrôlé par marqueurs.

4.4.3 Descripteurs de forme

L'analyse des images mammographiques permet non seulement de détecter les lésions mammaires mais aussi de préciser leur nature en termes de malignité ou bénignité en exploitant les caractéristiques extraites de la masse qui sont en général de types texturaux, photométriques et morphologiques. Dans cette étude, on s'est intéressé aux caractéristiques (descripteurs) morphologiques d'aspects géométriques tels que : la surface et le périmètre.

Chapitre 4 Conception du système d'aide au diagnostic assisté par ordinateur

A l'issue de la caractérisation, on dispose d'une description unique de chaque structure déterminée par l'algorithme LPE. Nous avons utilisés 12 descripteurs parmi lesquels :

- Area : Le nombre de pixels dans la région.
- Périmètre : La longueur de la limite de l'objet.
- Extent : la proportion de pixels dans le cadre de sélection qui se trouvent également dans la région.
- L'excentricité : est le rapport entre la distance entre les foyers de l'ellipse et la longueur de son axe principal.
- Equivdiametre : Diamètre d'un cercle de même surface que la région, retourné sous forme de scalaire.

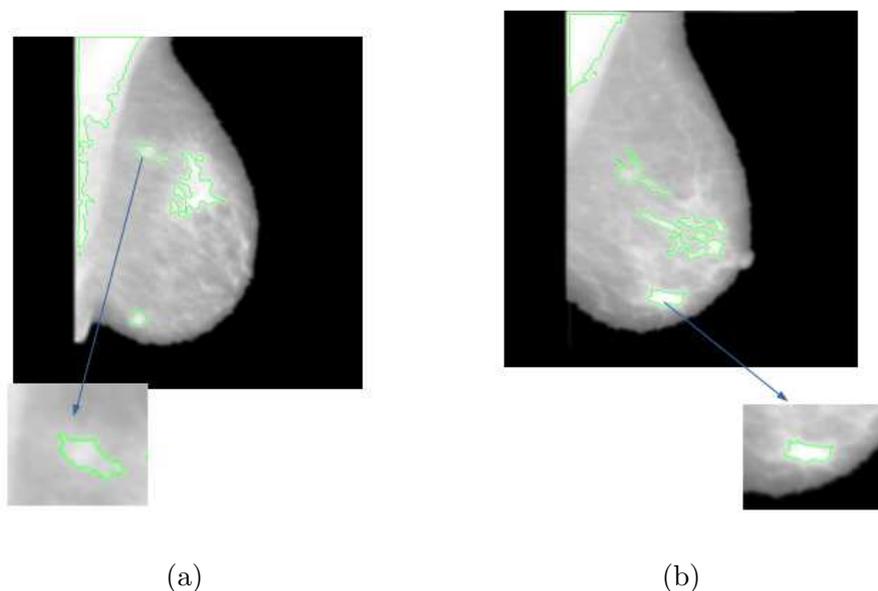


FIGURE 4.9 – Extraction des masses.(a) mdb206 et (b) mdb083.

L'image résultante de l'étape de segmentation dans la phase précédente montre bien le contour de la masse par rapport à l'arrière-plan (Figure 4.9). Nous avons laissé le choix au radiologue de choisir celle qu'il trouve la plus suspecte en la sélectionnant pour qu'ensuite l'algorithme puisse la classer comme étant maligne ou bénigne. Notre CAOx permet d'assister le radiologue à conforter sa décision dans le cas d'un doute par exemple. Pour rappel notre base de validation est formée par une matrice de 12x11.

4.4.4 Classification par SVM

Nous avons utilisé le classifieur SVM linéaire deux classes (voir chapitre 3). Avant l'étape de classification, la bases de données est divisée aléatoirement en base d'apprentissage et de test. 80% des données sont utilisées dans la base d'apprentissage de taille 12*47(nombre de descripteurs x nombre d'images) et 20% pour le test de taille 12*11.

4.4.5.1 Évaluation des performances de classification

Nous présentons dans ce paragraphe les résultats obtenus dans la phase de validation. On utilise généralement la matrice de confusion, aussi appelée tableau de contingence pour évaluer les performances du système.

A- Matrice de confusion

Une matrice de confusion est un résumé des résultats de prédictions sur un problème de classification. Les résultats sont ainsi comparés avec les valeurs réelles. Ceci permet non seulement de savoir quelles sont les erreurs commises, mais surtout le type d'erreurs commises.

		Classe réelle	
		-	+
Classe prédite	-	True Negatives <i>(vrais négatifs)</i>	False Negatives <i>(faux négatifs)</i>
	+	False Positives <i>(faux positifs)</i>	True Positives <i>(vrais positifs)</i>

FIGURE 4.10 – Calcul d'une matrice de confusion

Vrai positif (VP) : est le nombre de patient malade qui est classé malade.

Faux négatif (FN) : est le nombre de malade qui sont classé non malade.

Faux positif (FP) : est le nombre de patient non malade qui sont classé malade.

Vrai négatif (VN) : est le nombre de patient non malade qui sont classé non malade.

Dans une matrice de confusion, les lignes et les colonnes sont respectivement vraies et prédites. VP et VN représentent les décisions correctes prises par le classificateur, alors que FN et Les FP sont des erreurs de classification.

Voilà les résultats du calcul de matrice de confusion de notre modèle :

	Bénigne	Maligne
Bénigne	5	1
Maligne	1	4

FIGURE 4.11 – matrice de confusion de notre modèle SVM

Remarque :

On remarque d'après les résultats obtenus dans la matrice de confusion que le système a estimé 5 masses comme étant bénignes (VP) et 4 autres comme étant malignes (VN). Donc on peut dire que notre système a bien appris les exemples présentés.

B- Évaluation des performances :

*** Précision**

Le critère de performance le plus utilisé est le taux de classification correct, connu sous le nom de précision.

Il est possible de calculer cette précision comme suit :

$$\text{Précision} = ((VP + VN) / (VP + VN + FP + FN)) * 100$$

$$\text{Précision} = (5 + 4 / 11) * 100 = 81.82\%$$

*** sensibilité**

La sensibilité est la proportion de VP sur le nombre total de cas positifs. Mathématiquement, la sensibilité peut être calculée comme suit :

$$\text{Sensitivité} = VP / (VP + FN) = (4 / 5) * 100 = 80\%$$

*** spécificité**

est la proportion de VN sur le nombre total de cas négatifs. La spécificité peut être calculée comme suit :

Pour notre cas :

$$\text{Spécificité} = VN / (VN + FP) = (5 / 6) * 100 = 83.33\%$$

Chapitre 4 Conception du système d'aide au diagnostic assisté par ordinateur

Les résultats que nous avons obtenus sont satisfaisants. Ils montrent bien l'efficacité de notre système CAOx proposé. Les seuls cas où le système s'est trompé peuvent être expliqués par le fait que les structures dans ces images sont très petites et par conséquent leurs descripteurs correspondants peuvent être similaires.

4.5 Implémentation numérique

Afin de faciliter le système d'aide au diagnostic assisté par ordinateur, nous avons mis en œuvre une interface graphique avec laquelle l'utilisateur peut faire du prétraitement, de la segmentation ou de la classification.

Notre CAOx permet d'assister le radiologue à conforter sa décision dans le cas d'un doute par exemple. Il pourra également être utilisé par les étudiants en radiologie ou de médecins non expérimentés afin d'apprendre à analyser les mammographies.

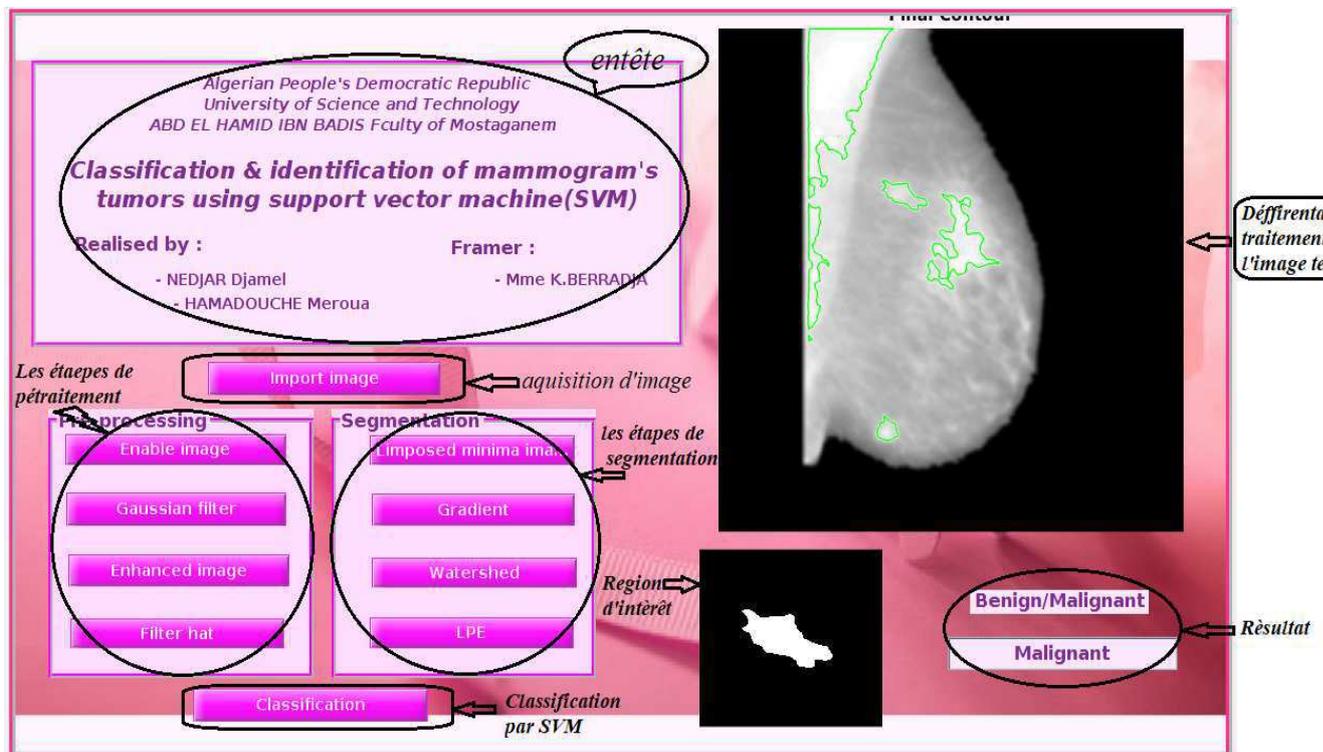


FIGURE 4.12 – Interface graphique du CAOx proposé.

Conclusion générale et perspectives

Dans ce travail, on s'est intéressé à l'étude des systèmes automatiques de diagnostic assistés par ordinateur en vue du diagnostic du cancer du sein, notamment à la description et à la classification des masses mammaires.

Les chercheurs dans le domaine de l'image se focalisent sur différentes étapes de la chaîne du diagnostic assisté par ordinateur afin d'améliorer le résultat de classification des masses en malignes/bénignes. Dans ce contexte, une étude concernant les outils ainsi que l'utilité du prétraitement des images mammographies est réalisée. La partie de pré-traitement des masses été fondé sur un ensemble d'opérations morphologiques, afin d'extraire la région du sein et le fond nettoyé qui a permet l'amélioration des résultats ultérieurs de segmentation des lésions mammaires.

Par la suite, nous avons prouvées la fiabilité de la segmentation morphologique des masses par la ligne de partage des eaux contrôlé par marqueurs. on se basant sur des connaissances de la morphologie mathématique.

Dans le problème de classifier les tumeurs selon ses bénignité ou malignité, était nécessaire de se reposer sur le choix d'un modèle pour l'aide au diagnostic basé sur le machine SVM(support vector machine)qui constitue une approche intéressante pour la modélisation de données.

Les résultats obtenus sont très intéressants, même s'ils sont encore loin des performances des meilleures méthodes proposées dans la littérature.

Ce mémoire a permis de dégager de nombreuses perspectives que nous les résumons dans ces quelques lignes :

1. Les attributs morphologiques que nous avons utilisés restent très simple, donc il serait possible de les enrichir par d'autres attributs de plus haut niveau : les attributs de texture, de photométrie, . . .
2. La méthode de machine à support de vecteur (SVM) choisit permet juste la classification de deux classes, il est possible d'ajouter une troisième classe où elle présente les sein normaux qui ne contient pas de tumeur. C'est le type des machines SVM se qu'on appel "multi-classes".
3. On pourrait tester notre système sur une autre base de données plus puissante tel que DDSM.
4. On pourrait utiliser aussi les réseaux de neurones comme classifieur pour une étude comparative avec les machines SVMs.

Morphologie mathématique

Introduction

L'origine de la morphologie mathématique remonte à une étude de la géométrie des grains effectuée au milieu des années soixante en France au laboratoire de l'école des Mines de Paris. G.Matheron a considéré la frontière d'un grain et a constaté que la frontière n'appartient ni au fond ni à la forme. Donc les grains sans contour et les grains avec contour sont indiscernables. J.Serra a appliqué ce formalisme mathématiques aux images. Il a introduit le formalisme ensembliste pour l'analyse des images binaires. La morphologie mathématique est aujourd'hui l'une des disciplines de traitement d'image. Son champ d'application s'est ensuite étendu et son utilisation est devenue rapidement très courante dans les applications de vision connues à ce jour. Ces applications touchent des domaines différents : multimédia, biologie, médecine...

A- Les opérateurs morphologiques

Avant de lister les opérateurs morphologiques "basiques", nous présentons dans le paragraphe suivant la notion d'un élément structurant (SE) en morphologie mathématique.

A.1- L'élément structurant

Un élément structurant (SE) est un ensemble appliqué à une image d'étude. Les SE "plans" consistent à un ensemble de point sans aucune valeur contrairement aux SE volumiques où des poids sont considérés. Les SE "plans" sont nommés ainsi car ces derniers ont uniquement deux dimensions dans le cas des images 2D.

Les opérateurs morphologiques fondamentaux requièrent la définition d'une origine pour chaque élément structurant. Cette origine permet le positionnement de l'élément structurant sur un point ou un pixel donné : un SE est à un point x signifie que son origine coïncide avec x . Un élément structurant est repéré donc par son origine[40]. Les éléments structurants plan et isotopes élémentaires pour des grilles-hexagonale et carrée- sont représentés par la figure 4.13.

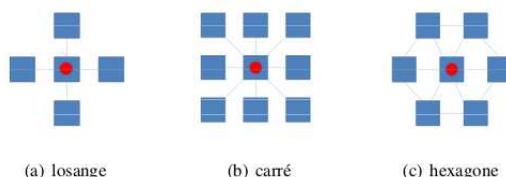


FIGURE 4.13 – Éléments structurants plans isotropes élémentaires. L'origine de chaque élément structurant est son centre

Un élément structurant définit une relation de voisinage et de connexité dans une image. La relation se fait du centre vers les voisins comme le montre la figure 4.14 qui présente la relation de voisinage d'un élément structurant carré à 8-connexités.

Dans ce présent travail, nous nous intéresserons uniquement à la morphologie qui utilise les SE plans. La forme et la taille de l'élément structurant doivent être adaptées aux propriétés géométriques des objets de l'image. Par exemple, des SE linéaires conviennent pour l'extraction des objets linéaires.

Nous allons détailler dans les sections qui suivent les opérateurs de base en morphologie.

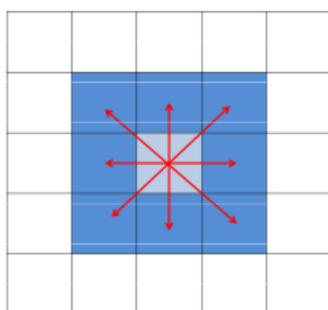


FIGURE 4.14 – Relation de voisinage pour un élément structurant carré à 8-connexités.

A.2- L'érosion

L'érosion ensembliste consiste à poser en chaque pixel x d'un objet X , la question : l'élément structurant B_x est-il contenu entièrement dans X ? L'ensemble érodé est constitué des points où la réponse à cette question est affirmative. L'ensemble traité représente soit les objets d'une image binaire ou le sous graphe d'une image en niveaux de gris. L'érosion d'un ensemble X par un élément structurant B est notée par $\varepsilon_B(X)$ et définie par l'ensemble des points, x , tel que B est inclus dans X quand son origine est placée sur x .

$$\varepsilon_B(X) = \{X | B_x \subset X\} \tag{A.1}$$

L'équation A.1 peut aussi être écrite sous forme d'intersections d'un ensemble de translations. Ces translations sont définies par l'élément structurant.

$$\varepsilon_B(X) = \bigcap_{b \in B} X - b \tag{A.2}$$

La définition précédente peut être appliquée sur des images binaires ainsi que sur des image en niveaux de gris : l'érosion d'une image f par un élément structurant B est notée $\varepsilon_B(f)$ et est définie comme étant le minimum des translations de f par les vecteurs b de B .

$$\varepsilon_B(f) = \bigwedge_{b \in B} f - b$$

A.3

D'où la valeur érodée d'un pixel donné est la valeur minimum de l'image dans la fenêtre définie par l'élément structurant quand son origine est placée sur x :

$$\varepsilon_B(f)(x) = \min_{b \in B} f(x + b)$$

A.4

L'érosion réduit les « pics » de niveaux de gris et élargit les « vallées » : elle tend donc à homogénéiser l'image, à l'assombrir et à étaler le bord des objets les plus sombres comme le montre la figure 4.15.

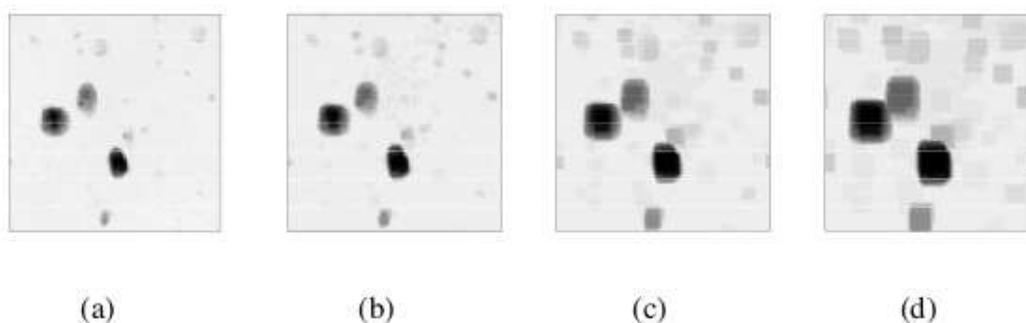


FIGURE 4.15 – Erosion. (a) image en niveaux de gris originale (b) érosion avec un SE carré de taille 3 (c) érosion avec un SE carré de taille 6 (d) érosion avec un SE carré de taille 10

A.3- La dilatation

La dilatation est l'opérateur adjoint de l'érosion. Elle consiste à poser en chaque pixel x d'un objet X , la question : "l'élément structurant B_x intersecte-t-il l'ensemble X ?" l'ensemble dilaté est constitué de points où la réponse à cette question est affirmative. La dilatation d'un ensemble X par un élément structurant B est notée $\delta_B(X)$, et est définie par l'ensemble de points x tel que B intersecte X quand son origine coïncide avec x :

$$\delta_B(X) = \{x | B_x \cap X \neq \emptyset\}$$

A.5

Dans l'équation A.5, B désigne l'élément structurant symétrique de B . Notons que la dilatation d'un seul point donne comme sortie l'élément structurant centré sur son origine. L'équation A.5 peut aussi s'écrire sous forme d'unions de translations. Ces translations sont définies par l'élément structurant.

$$\delta_B(X) = \bigcup_{b \in B} X_b$$

A.6

La définition précédente peut être appliquée sur des images binaires ainsi que sur des images en niveaux de gris : la dilatation d'une image f par un élément structurant B est notée $\delta_B(f)$ et est définie par le maximum de la translation de f par les vecteurs b de B .

$$\delta_B(f) = \bigvee_{b \in B} f_b$$

A.7

D'où la valeur dilatée d'un pixel donné est le maximum de l'image dans la fenêtre définie par l'élément structurant quand son origine est placée sur x :

$$[\delta_B(f)](x) = \max_{b \in B} f(x + b)$$

A.8

Cette transformation comble les « vallées » et épaissit les « pics » : elle homogénéise l'image, l'éclaircit et tend à faire disparaître les objets sombres.

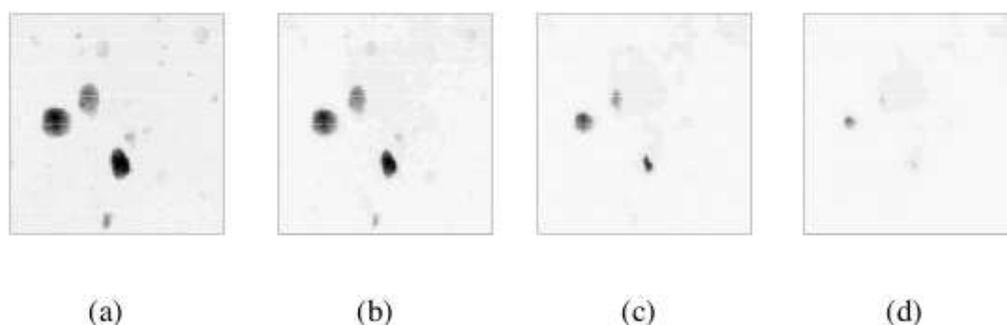


FIGURE 4.16 – Dilatation. (a) image en niveaux de gris originale (b) dilatation avec un SE carré de taille 3 (c) dilatation avec un SE carré de taille 6 (d) dilatation avec un SE carré de taille 10.

A.4- L'ouverture morphologique

L'ouverture d'une image f par un élément structurant B est notée $\gamma_B(f)$ et définie par l'érosion de f par B suivie par une dilatation par l'élément structurant symétrique B .

$$\gamma_B(f) = \delta_B \circ \varepsilon_B(f)$$

A.9

L'ouverture supprime les pics mais préserve les vallées, elle homogénéise l'image mais préserve les objets sombres.

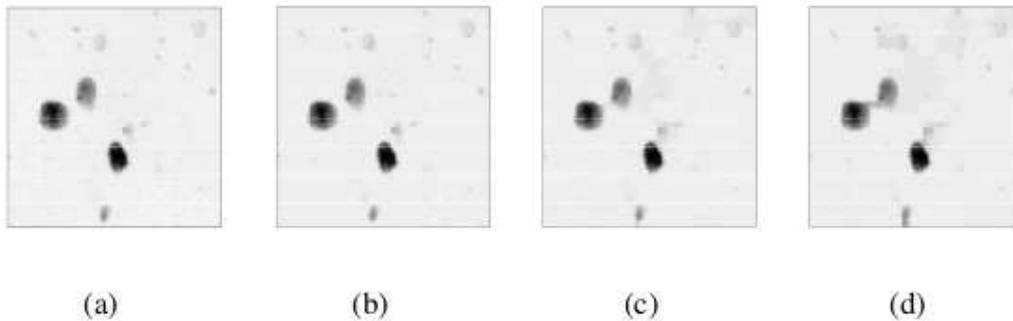


FIGURE 4.17 – Ouverture. (a) image en niveaux de gris originale (b) ouverture avec un SE carré de taille 3 (c) ouverture avec un SE carré de taille 6 (d) ouverture avec un SE carré de taille 10.

A.5- La fermeture morphologique

La fermeture d'une image f par un élément structurant B est notée $\varphi_B(f)$ et est définie par la dilatation de f par un élément structurant B , suivie d'une érosion par l'élément structurant symétrique B :

$$\varphi_B(f) = \varepsilon_B \circ \delta_B(f)$$

A.10

La fermeture comble les vallées, elle homogénéise et éclaircit l'image comme le montre la figure 2.6.

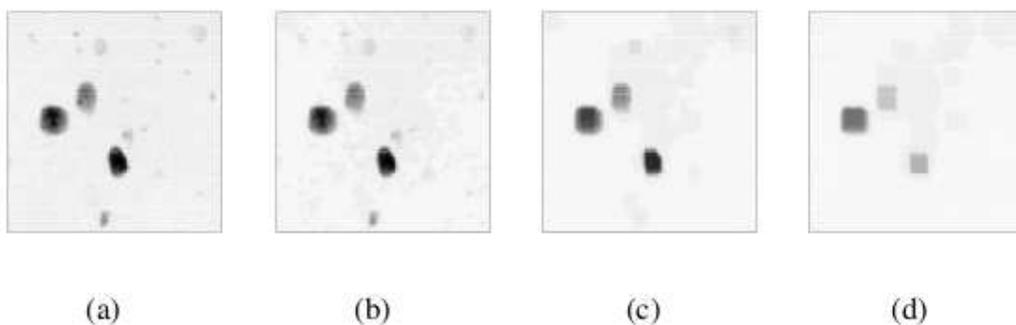


FIGURE 4.18 – Fermeture. (a) image en niveaux de gris originale (b) fermeture avec un SE carré de taille 3 (c) fermeture avec un SE carré de taille 6 (d) fermeture avec un SE carré de taille 10.

A.6- Le gradient morphologique

Une hypothèse courante dans l'analyse d'images consiste à considérer les objets de l'image comme des régions homogènes en niveaux de gris. Cela induit que sur la frontière entre deux objets, il existe une variation du niveau de gris. Les opérateurs du gradient visent à mettre en évidence ces variations. Plusieurs gradients ont été proposés pour l'analyse d'images. Dès lors, il n'y a pas de définition claire de l'opérateur gradient pour les données discrètes. Les gradients morphologiques sont des opérateurs qui mettent l'accent sur les variations locales de l'image selon le voisinage défini par l'élément structurant. Trois combinaisons peuvent être utilisées :

- différence arithmétique entre la dilatation et l'érosion : c'est le gradient morphologique basique proposé par Beucher. Il est noté par ρ :

$$\rho_B = \delta_B - \varepsilon_B$$

- différence arithmétique entre la dilatation et l'image originale (demi gradient supérieur) :
 $\delta_B - I$

- différence arithmétique entre l'image originale et son érodée (demi gradient inférieur) :
 $I - \varepsilon_B$

Nous avons présenté les opérateurs morphologiques basiques. Le calcul de l'érosion et la dilatation sert à trouver le gradient que nous allons avoir besoin dans le calcul de la ligne des partage des eaux. L'ouverture et la fermeture serviront pour des travaux ultérieurs.

Bibliographie

- [1] Cheikhrouhou, K. Djemal, and H. Maaref. Protuberance selection descriptor for breast cancer diagnosis. Third European Workshop on Visual Information Processing (EUVIP'11), Paris, France, pp280-285, July, 2011.
- [2] [En ligne] <https://www.biron.com/fr/centre-du-savoir/petit-guide-biron/cancer-du-sein/>
- [3] [En ligne] <https://www.biron.com/fr/centre-du-savoir/petit-guide-biron/cancer-du-sein/>.
- [4] [En ligne] : <http://www.vulgaris-medical.com/encyclopedie-medicale/sein-anatomie-physiologie>. [Accès le 12 06 2013].
- [5] Balu-Maestro C, Chapellier C (2001) Diagnostic échographique des lésions mammaires. Encycl Med Chir Gynécologie 810 G-17 : 7
- [6] [En ligne]<https://radiologie-la-defense.fr/centre-paris/sein/echographie-mammaire/>.
- [7] BI-RADS. Breast Imaging reporting and Data System. Atlas d'imagerie du sein – Ultrasons. Première édition française basée sur la première édition américaine.
- [8] [En ligne]<https://imageriemedicale.fr/examens/irm/mammaire/> .
- [9] BI-RADS. Breast Imaging reporting and Data System. Atlas d'imagerie du sein – Mammographie. Deuxième édition française basée sur la quatrième édition américaine.
- [10] BI-RADS. Breast Imaging reporting and Data System. Atlas d'imagerie du sein – Mammographie. Deuxième édition française basée sur la quatrième édition américaine.
- [11] [En ligne] <http://www.breastcare.at/brustkrebsvorsorge/mammographie>.
- [12] [En ligne] <http://www.breastcare.at/brustkrebsvorsorge/mammographie>. [13] [Allen et al., 2003] Allen, B. H., Oxley, M. E., and Collins, M. J. (2003). A universal segmentation platform for computer-aided detection. In *Digital Mammography*, pages 164–168. Springer.
- [14] [Baker et al., 2003] Baker, J. A., Rosen, E. L., Lo, J. Y., Gimenez, E. I., Walsh, R., and Soo, M. S. (2003). Computer-aided detection (cad) in screening mammography : sensitivity of commercial cad systems for detecting architectural distortion.

Bibliographie

American Journal of Roentgenology, 181(4) :1083–1088.

[15] Ikeda DM, Hylton NM, Kinkel K et al. Development, standardization and testing of a lexicon for reporting contrast-enhanced breast magnetic resonance (MR) imaging studies. J Magn Reson Imaging 2001 ;13 :896-902.

[16] G. Liasis et S. Petroudi, «Estimation of the Skin-Air Interface in mammograms Using level Sets,» IEEE, pp. 239-242, 2011.

[17] Ikeda DM, Hylton N, Kinkel K et al. Development, standardization and testing of a lexicon for reporting contrast-enhanced breast magnetic resonance (MR) imaging studies. 85th Annual Meeting of the Radiological Society of North America (RSNA), Chicago, USA, December 1999

[18] [Alto et al., 2005] Alto, H., Rangayyan, R. M., and Desautels, J. L. (2005). Content-based retrieval and analysis of mammographic masses. Journal of Electronic Imaging, 14(2) :023016– 023016.

[19] [Bailar, 1976] Bailar, J. C. (1976). Mammography : a contrary view. Annals of Internal Medicine, 84(1) :77–84.

[20] D’Osri, C., Bassett, L., Berg, W. et al. (2003). Breast imaging reporting and data system : Acr bi-rads, Technical report, American College of Radiology.

[21] [En ligne] [http ://peipa.essex.ac.uk/info/mias.html](http://peipa.essex.ac.uk/info/mias.html).

[22] D’Osri, C., Bassett, L., Berg, W. et al. (2003). Breast imaging reporting and data system : Acr bi-rads, Technical report, American College of Radiology.

[23] Monsees BS. The Mammography Quality Standards Act : an overview of the regulations and guidance. Radiol Clin North Am 2000 ; 38 :759–772 [Crossref] [Medline] [Google Scholar].

[24] H. Hwang, B. Choi, H. Lee, H. Yoon, Y. Nam, and H. Choi, Multiresolution wavelettransformed image analysis of histological sections of breast carcinomas, Cellular Oncology, vol. 27, 2005.

[25] [Baker et al., 2003] Baker, J. A., Rosen, E. L., Lo, J. Y., Gimenez, E. I., Walsh, R., and Soo, M. S. (2003). Computer-aided detection (cad) in screening mammography : sensitivity of commercial cad systems for detecting architectural distortion. American Journal of Roentgenology, 181(4) :1083–1088.

[26] M. LEHAMEL, Segmentation d’images texturées á partir des attributs fractals, Ph.D. dissertation, Université Mouloud Mammeri de Tizi-Ouzou, 2010.

Bibliographie

- [27] R. D. Yapa et Koichi Harada, «Breast Skin-Line Estimation and Breast Segmentation in Mammograms using Fast marching Method,» International Journal of Biological and Life Sciences, vol. 3, %11, pp. 54-62, 2007.
- [28] M.C. Christ , R. Parvathi, "Segmentation of Medical Image Using Clustering and Watershed Algorithms", American Journal of Applied Sciences, Vol.8, No.12, 2011, 1349-1352, ISSN 1546-9239.
- [29] S.Beucher, "The Watershed Transformation Applied to Image segmentation", Center De Morphologic Mathematique , Ecole Desmines Paris, 77305 Fontainebleau Cedex.
- [30] R. Frigate, and E. Silva, "Mathematical Morphology Application to Features Extraction in Digital Images", Pecora 17 –The Future of Land Imaging.Going Operational, 2008, November 18 – 20, Colorado.
- [31] H. I. Ali, "Digital Images Edge Detection Using Mathematical Morphology Operations", Iraq Journal of Science , Vol. 51, No.1, 2010.
- [32] John, G. H., Kohavi, R. et Pflieger, K. (1994). Irrelevant features and the subset selection problem. In MACHINE LEARNING : PROCEEDINGS OF THE ELEVENTH INTERNATIONAL, pages 121-129. Morgan Kaufmann.
- [33] Kira, K. et Rendell, L. A. (1992). The feature selection problem : Traditional methods and a new algorithm. In AAAI, pages 129-134, Cambridge, MA, USA. AAAI Press and MIT Press.
- [34] Joachims T, Scholkopf B, Burges C and Smola A, "Making large-scale SVM Learning Practical, Advances in Kernel Methods-Support Vector Learning", Cambridge, MA, USA, 1999.
- [35] Crammer Koby and Yoram Singer, "On the Algorithmic Implementation of Multi-class Kernel- based Vector Machines", Journal of Machine Learning Research, MIT Press, Cambridge, MA, USA, Vol.2, pages 265-292, 2001.
- [36] Mehmet Fatih Akay. Support vector machines combined with feature selection for breast cancer diagnosis. Expert systems with applications, 36(2) : 2009.
- [37] Cortes, C. and V. Vapnik (1995). Support vector networks. Machine Learning 20.
- [38] Mohamadally Hasan, Fomani Boris : " SVM machine a vecteurs de support ou separateur a vaste marge ". BD Web, ISTY3, Versailles St Quentin, France, janvier 2006.
- [39] S. Beucher. Segmentation d'images et morphologie mathématique. PhD thesis, Centre de Morphologie Mathématique, École des Mines de Paris, 1991.

Bibliographie

[40] R. Encficiaud. Algorithmes multidimensionnels et multispectraux en Morphologie Mathématique : Approche par méta-programmation. PhD thesis, Centre de Morphologie Mathématique, École des Mines de Paris, 2007.