



Fusion collaborative de classifieurs : application à la détection de pathologies en mammographie

Asma Touil

► To cite this version:

Asma Touil. Fusion collaborative de classifieurs : application à la détection de pathologies en mammographie. Traitement du signal et de l'image [eess.SP]. Ecole nationale supérieure Mines-Télécom Atlantique; Institut supérieur d'informatique et des techniques de communication (Hammam Sousse, Tunisie), 2021. Français. NNT : 2021IMTA0263 . tel-03510707

HAL Id: tel-03510707

<https://theses.hal.science/tel-03510707>

Submitted on 4 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE DE DOCTORAT DE

L'ÉCOLE NATIONALE SUPERIEURE MINES-TELECOM ATLANTIQUE
BRETAGNE PAYS DE LA LOIRE - IMT ATLANTIQUE

EN COTUTELLE AVEC

L'INSTITUT SUPERIEUR D'INFORMATIQUE ET DES TECHNOLOGIES
DE COMMUNICATION DE HAMMAM SOUSSE

ECOLE DOCTORALE N° 601
*Mathématiques et Sciences et Technologies
de l'Information et de la Communication*
Spécialité : *Signal, image, vision*

Par

Asma TOUIL

Fusion collaborative de classifieurs : Application à la détection de pathologies en mammographie

Thèse présentée et soutenue à L'Institut Supérieur de l'informatique et des technologies de la communication de *Hammam Sousse*, le 29/11/2021

Unité de recherche : Latim (IMT Atlantique – France) et LATiS (Eniso, Sousse – Tunisie)

Thèse N° : 2021IMTA0263

Rapporteurs avant soutenance :

El-Hadi Zahzah	Maitre de conférences (HDR), Université de La Rochelle (France)
Asma Ben Abdallah	Maitre de conférences (HDR), ISIMM Monastir (Tunisie)

Composition du Jury :

Président :	Ali Khenchaf	Professeur, ENSTA Bretagne (France)
Examineurs :	El-Hadi Zahzah	Maitre de conférences (HDR), Université de La Rochelle (France)
	Asma Ben Abdallah	Maitre de conférences (HDR), ISIMM Monastir (Tunisie)
	Imed Riadh Farah	Professeur, ISAMM Manouba (Tunisie)
Dir. de thèse :	Basel Solaiman	Professeur, IMT Atlantique, Brest (France)
Co-dir. de thèse :	Mohamed Ali Mahjoub	Professeur, Ecole Nationale d'Ingénieurs de Sousse (Tunisie)

Invité(s)

Pierre-Henri Conze	Maitre de conférences, IMT Atlantique – Brest (France)
Karim Kalti	Maître assistant, Faculté des Sciences de Monastir (Tunisie)
Jean-Paul Haton	Professeur, Université de Henry Poincaré – Nancy 1 (France)
Isabelle Thomas-Kergastel	Chef service de radiologie, CHRU de Brest (France)

DÉDICACES

Du profond de mon cœur, je dédie ce modeste travail

A l'âme éternelle de celui qui m'a toujours indiqué la bonne voie en me rappelant que la volonté fait toujours les grands hommes - A mon très cher père *Hassine TOUIL*. Je pourrais écrire un million de pages mais je ne pourrais jamais dire à quel point je t'aime, et je te regrette tous les jours. Je te dois ce que je suis aujourd'hui et ce que je serai demain et je ferai toujours de mon mieux pour rester ta fierté et ne jamais te décevoir.

A ma source inépuisable de tendresse, de patience et de sacrifice - A ma très chère mère, *Jamila SOUSSIA*. Quoi que je fasse ou que je dise, je ne saurai point te remercier comme il se doit. Ton affectation me couvre, ta bienveillance me guide et ta présence à mes côtés a toujours été ma source de force. En ce jour mémorable, pour moi ainsi que pour toi, reçoit ce travail en signe de ma vive reconnaissance et mon profond estime.

A celui avec avec qui nous partageons cette vie - A mon très cher oncle *Hassen*. Autant de phrases et d'expressions aussi éloquentes soit-elles ne sauraient exprimer ma gratitude et ma reconnaissance envers tes valeurs nobles. Merci pour le soutien permanent venu de toi, Merci pour les sacrifices que tu n'as cessé de m'apporter.

A ceux qui m'ont partagé mes moments d'enfance - A ma fierté, mes frères et sœur *Hamdi, Imen* et *Ahmed*. Puissent nos liens fraternels se consolider et se pérenniser encore plus. Je vous dois tout l'amour, avec tous mes vœux de voir réussir dans vos vies.

A celui qui nous a partagé la vie depuis une dizaine d'années - A mon beau frère *Mohamed Ali*. Aucun langage ne saurait exprimer mon respect et ma considération pour ton soutien et tes encouragements. Je te souhaite une vie comblée de bonheur et de réussite.

A mes trois anges si adorables et charmantes - A mes nièces *Elèe, Eya* et *Israa*. Vous avez apporté beaucoup de bonheur à notre famille. Je vous aime et je vous souhaite un avenir radieux et plein de bonheur et de réussite.

A mes chères sœurs *Rania, Hela, Fatma* et *Wissal*. En témoignage de l'amitié qui nous unit et des souvenirs de tous les moments que nous avons passés ensemble, je vous dédie ce travail et je vous souhaite une vie pleine de santé et de bonheur.

Asma

REMERCIEMENTS

Merci est un hommage au passé, une grâce présente, et une porte sur le futur.

Un travail de recherche ne se fait pas seul, mais bien entouré. Un certain nombre de personnes m'ont permis de réaliser mon travail dans des conditions idéales. Je ne vois pas de remerciement plus profond que celui provenant du cœur, validé par l'esprit et exprimé par la main.

Mes remerciements vont, en premier lieu, à mon cher directeur de thèse, Monsieur le professeur *Basel Solaiman* pour l'aide compétente qu'il m'a apporté, pour sa patience et son encouragement. Vous m'avez toujours réservé le meilleur accueil, malgré toutes vos obligations professionnelles. Vous étiez toujours à l'écoute de mes nombreuses questions, et intéressé à l'avancée de mes travaux. Les nombreuses discussions que nous avons eues ainsi que vos conseils sont pour beaucoup dans le résultat final de ce travail. J'ai eu le grand plaisir de travailler sous votre direction, et ai trouvé auprès de vous le conseiller et le guide qui m'a reçu en toute circonstance avec sourire et bienveillance. Votre parcours professionnel, votre compétence incontestable, votre charisme et vos qualités humaines font de vous un grand professeur. Votre capacité d'analyse et votre enthousiasme m'ont montré que le monde de la recherche pouvait être un univers passionnant. Je veux vraiment vous remercier car j'ai eu beaucoup de chance de vous avoir comme directeur de thèse.

Mes remerciements les plus sincères vont également à mon directeur de thèse, Monsieur *Mohamed Ali Mahjoub*, Professeur à l'école nationale des ingénieurs de Sousse, à l'attention et le soutien qu'il a porté à mon travail de doctorante. Votre gentillesse, vos qualités humaines, votre modestie n'ont rien d'égal que votre compétence. Je saisis cette occasion pour vous exprimer ma profonde gratitude tout en vous témoignant mon grand respect.

Mes remerciements les plus sincères et ma profonde gratitude s'adressent à mon encadrant de thèse, Monsieur *Karim Kalti*, Maître Assistant à la faculté des sciences de Monastir, avec qui j'ai eu le plaisir d'initier mes premiers pas dans le domaine de la recherche scientifique. Votre énergie et votre confiance ont été des éléments moteurs pour moi. Votre sens élevé de devoir, votre entière disponibilité, vos conseils avisés et votre écoute ont été prépondérants pour la bonne réussite de cette thèse qui n'aurait jamais pu être abouti sans vous. J'ai eu l'honneur et la chance de bénéficier de vos connaissances et compétences et de travailler avec vous, depuis mon master. Je vous remercie, mon cher Maître pour l'intérêt, de tout instant, dont vous avez fait preuve envers mes travaux de

recherche. Je vous remercie pour m'avoir montré ce qu'était le monde de la recherche et pour tout ce que vous m'avez appris, autant professionnel que personnel.

Mes remerciement les plus particulières s'adressent aussi à *Pierre-Henri Conze*, Maitre de conférences à l'IMT Atlantique de Brest, d'avoir encadrer cette thèse. Je vous suis également reconnaissante pour le temps conséquent que vous m'aviez accordé, vos qualités pédagogiques et scientifiques, votre franchise et votre sympathie. Vous étiez toujours montré à l'écoute et disponible tout au long de ces années de thèse. J'ai beaucoup appris avec cette collaboration et je vous adresse ma gratitude pour tout cela.

Je voudrais exprimer toute ma reconnaissance à Monsieur *Imed Riadh Farah*, Professeur à l'Institut Supérieur des Arts Multimédia de la Manouba, pour l'honneur qu'il m'a accordé en présidant mon jury de thèse.

Toute ma gratitude s'adresse aussi à Monsieur *Ali Khenchef*, Professeur à ENSTA Bretagne qui m'a fait l'honneur d'examiner ce rapport et de participer ma soutenance de thèse.

Mes remerciements s'adressent également à Madame *Asma Ben Abdallah*, Maitre de conférences à l'Institut supérieur d'Informatique et de Mathématiques de Monastir et Monsieur *El-Hadi Zahzah*, Maitre de conférences à l'université de La Rochelle qui se sont acquittés de la délicate tâche de rapporteur sur ce travail.

Mention spéciale à Monsieur *Jean-Paul Haton*, Professeur à l'université de Henri Poincaré qui s'est rendu disponible pour assister à cette soutenance. Je vous remercie pour l'honneur que vous m'avez accordé en acceptant de participer mon jury de thèse.

Mention spéciale à Madame *Isabelle Kergastel Thomas*, Chef de service de radiodiagnostic et imagerie médicale au Centre Hospitalier Universitaire de Brest (CHRU de Brest), pour son aide dans la phase de validation et pour l'effort qu'elle ait fourni pour valider la base d'images avec des vérités de terrain. Je souhaite vous remercier du temps et de l'attention que vous avez bien voulu me consacrer malgré vos nombreuses charges.

Je souhaiterai aussi adresser ma gratitude à Madame *Najoua Essoukri Ben Amara*, Professeur à l'Ecole Nationale des Ingénieurs de Sousse (Eniso) et directrice de notre laboratoire de recherche LaTiS pour sa gentillesse, ses encouragements et ses qualités humaines. Qu'il me soit permis de vous présenter, Madame, à travers ce travail le témoignage de mon grand respect et l'expression de ma profonde reconnaissance.

Je ne saurais terminer sans remercier Madame, *Corinne Le Lann*, l'assistante du département ITI de l'IMT Atlantique et Madame *Najla Ragoubi*, l'assistante du laboratoire LaTiS à l'Ecole Nationale des Ingénieurs de Sousse, pour leurs sympathies et leurs efficacités dans l'organisation et la résolution des problèmes administratifs, mais aussi pour les moments agréables qu'elles nous ont créé tout au long de ce travail de recherche.

En plus de toutes ces personnes liées à ce travail, je tiens à remercier chaleureusement toute personne qui a contribué de près ou de loin à la réalisation de ce travail de recherche.

RÉSUMÉ

Titre : Combinaison collaborative de classifieurs : Application à la détection de pathologies en mammographie.

Mots clés : Cancer du sein, Mammographie, Détection des microcalcifications, Superpixels, Morphologie mathématique, Similarité structurelle, Croissance de régions conditionnelle, Combinaison collaborative, Propagation des connaissances.

Résumé : Les systèmes d'aides à la décision dans le domaine médical ont prouvé leur capacité à apporter une aide aux experts au cours de leur diagnostic. Cette thèse s'inscrit dans ce contexte et s'intéresse à l'interprétation d'images mammographiques pour la détection des microcalcifications (MCs). Cette problématique s'avère complexe compte-tenu de la diversité anatomique du sein et de ses lésions mammaires. Partant d'un schéma de prétraitement, nous proposons, dans ce travail de recherche, une première méthode pour la détection des MCs. Basée sur la morphologie mathématique et l'indice de similarité structurelle SSIM, cette méthode offre des résultats fiables et se trouve capable de gérer la diversité des MCs en termes de formes et de contrastes. Dans l'idée d'étudier l'effet d'une analyse locale à base des connaissances *a priori*, nous avons proposé une deuxième méthode de détection. Exploitant le concept de croissance de région conditionnelle, elle part des maximas régionaux et intègre les connaissances *a priori* sous la forme de contraintes. La fiabilité des résultats obtenus montre l'utilité de ces connaissances intégrés. Bien que les deux méthodes proposées aient présenté des résultats fiables, elles se trouvent face à des difficultés liées à la diversité des images mammographiques ainsi que la diversité des morphologies des MCs. Partant de ce constat, et dans l'idée d'améliorer davantage la qualité des résultats obtenus, nous avons conçu la troisième et principale contribution dans ce travail de recherche. Il s'agit d'une approche de combinaison collaborative qui vise à détecter les MCs tout en minimisant le nombre de détection faux négatif (FN). L'idée de base est d'opérer sur l'image entière divisée en régions homogènes (des superpixels) utilisées pour identifier des zones suspectes et des objets candidats à partir des résultats des méthodes de détection appelées *détecteurs*. Les deux méthodes proposées ainsi que deux autres issues de la littérature sont utilisées comme détecteurs. Un processus de classification collaborative s'applique par la suite et permet de faire collaborer les résultats des détecteurs et de mettre en cause leurs décisions initiales en fonction des informations et connaissances partagées. Le raffinement appliqué permet de mettre à jour les étiquettes des objets candidats issus de chaque détecteur à l'aide d'une analyse locale et contextuelle basée sur des connaissances *a priori* sur les liens entre les superpixels et les MCs. D'une itération à une autre, ce processus réduit le désaccord entre les différents détecteurs et estime les termes de fiabilité locaux pour chaque superpixel. Le résultat final est obtenu par combinaison conjonctive des nouvelles décisions obtenues par le processus collaboratif. Les résultats expérimentaux montrent les apports obtenus en termes d'amélioration des performances de détection des MCs par rapport aux détecteurs existants ainsi qu'aux opérateurs de fusion ordinaires.

ABSTRACT

Title : Collaborative combination of classifiers : Application to microcalcification detection in mammograms

Keywords : Breast cancer, Mammography, Microcalcification detection, Superpixels, Mathematical morphology, Structural similarity index (SSIM), Conditional region growing, Collaborative combination, Knowledge propagation.

Abstract : CAD systems have proven their ability to assist experts in their medical diagnosis. This thesis is related to the field of mammographic image analysis and aims at detecting microcalcifications (MCs). This issue is quite complex due to the anatomical diversity of breast tissues as well as lesions. Starting from a pre-processing scheme, we propose in this study a first method of MC detection. Based on mathematical morphology and the SSIM structural similarity index, it offers reliable results and is able to deal with MC diversities in terms of shapes and contrasts. In order to examine the efficiency of a local knowledge-based analysis, we have proposed a second detection method. Based on the conditional region growing concept, it starts from a selected regional maxima and integrates *prior* knowledge as constraints for the growing process. The reliability of the obtained results shows the usefulness of this integrated knowledge. Despite the reliability, the two proposed methods, they faced with difficulties related to mammographic image and MC diversities. Based on this finding, and with the idea of further improving the quality of the obtained results, we proposed the third and main contribution in this research work. It deals with a collaborative combination approach that aims at detecting MCs while minimizing the false negative detections (FN). The main concept is to operate on the entire image divided into homogeneous regions (superpixels) which are used to identify both suspicious areas and candidate objects from the results of the used detection methods called *detectors*. The two proposed methods as well as two other existing methods are used as detectors. A collaborative classification process is, then, applied and makes the different MC detectors collaborate to refine their initial results on the basis of the shared information and knowledge. The applied refinement consists of iteratively updating the candidate object labels of each detector following a local and contextual analyses based on *prior* knowledge about the links between superpixels and MCs. This process iteratively reduces the disagreement between different detectors and estimates local reliability terms for each superpixel. The final result is obtained by a conjunctive combination of the new detector decisions reached by the collaborative process. The experimental results show the benefits gained in terms of improving MC detection performances compared to existing detectors as well as standard fusion operators.

TABLE DES MATIÈRES

Dédicaces	i
Remerciements	iii
Table des matières	x
Table des figures	xv
Liste des tableaux	xx
1 Contexte médical	7
1.1 Introduction	8
1.2 Étude anatomique	8
1.2.1 Anatomie du sein	8
1.2.2 Cancer du sein	9
1.3 Imagerie médicale	10
1.3.1 Mammographie	10
1.3.2 Interprétation des mammographies	11
1.3.3 Les anomalies détectables à partir d'une mammographie numérique	13
1.4 Calcifications mammaires	15
1.4.1 Définition selon le lexique médical	15
1.4.2 Diagnostic des calcifications mammaires	16
1.5 Microcalcifications: Caractéristiques d'apparition	18
1.5.1 Nombre et regroupement	18
1.5.2 Distribution	19
1.5.3 Morphologie	19
1.6 Bases de mammographies numériques	25
1.6.1 Exigences de création	25
1.6.2 Bases de mammographies	26
1.6.3 Base de mammographies utilisée	28
1.6.4 Génération de la vérité de terrain	29
1.7 Conclusion	29
2 Traitement d'images pour la détection des microcalcifications	32
2.1 Introduction	33
2.2 Problématique de détection des MCs	33
2.2.1 Intérêts de détection	33
2.2.2 Enjeux de détection	34
2.3 Détection des microcalcifications dans la littérature	34

2.3.1	Segmentation fréquentielle	36
2.3.2	Segmentation basée morphologie	38
2.3.3	Segmentation basée régions	40
2.3.4	Segmentation basée partitionnement	41
2.3.5	Segmentation à base de filtres	44
2.3.6	Étude comparative	46
2.4	Prétraitement des images mammographiques	48
2.4.1	Qualité et nécessité de prétraitement	48
2.4.2	Techniques de prétraitement standard	48
2.4.3	Prétraitement en mammographie: Synthèse et discussion	53
2.5	Approche proposée	54
2.5.1	Analyse critique	54
2.5.2	Approche proposée	56
2.6	Conclusion	59
3	Première contribution: Détection basée sur la morphologie mathématique	61
3.1	Introduction	62
3.2	Motivation	62
3.3	Morphologie mathématique et détection des MCs	63
3.3.1	Concepts de base de la morphologie mathématique	63
3.3.2	Limites	64
3.4	Méthode de détection proposée	65
3.4.1	Amélioration de la mammographie	66
3.4.2	Génération de la carte de suspicion	72
3.5	Méthodologie de validation	78
3.5.1	Métriques d'évaluation des performances	78
3.5.2	Méthode ('Zh') retenue pour comparaison	80
3.6	Évaluation qualitative des résultats	81
3.6.1	Fiabilité de détection	82
3.6.2	Fiabilité de la segmentation	86
3.6.3	Fiabilité par rapport à la méthode de comparaison	88
3.7	Évaluation quantitative des résultats	90
3.7.1	Fiabilité des résultats	91
3.7.2	Fiabilité par rapport à la méthode 'Zh'	91
3.8	Discussion	93
3.9	Conclusion	94
4	Deuxième contribution: Détection basée sur une croissance de région conditionnelle	97
4.1	Introduction	98
4.2	Croissance de région	98
4.2.1	Principe	98
4.2.2	Motivation pour la détection des MCs	99
4.3	Description globale de la méthode proposée	99
4.3.1	Croissance de région classique et MCs - Limites	99
4.3.2	Démarche de la méthode proposée	100
4.4	Sélection des germes de départ	102

4.4.1	Maximas régionaux comme germes initiaux	102
4.4.2	Sélection des maximas germes de départ	104
4.4.3	Description du processus de croissance	108
4.5	Contraintes définies pour la croissance de région conditionnelle	109
4.5.1	Contraintes spatiales	110
4.5.2	Contraintes de délinéation	111
4.6	Résultats et discussion	116
4.6.1	Méthodologie de validation et méthode de comparaison	116
4.6.2	Fiabilité de détection par rapport à la méthode de comparaison	116
4.6.3	Fiabilité de segmentation	118
4.6.4	Discussion	121
4.7	Conclusion	123
5	Combinaison collaborative pour la détection des microcalcifications	125
5.1	Introduction	126
5.2	Approche proposée	126
5.2.1	Motivations	126
5.2.2	Approche collaborative	127
5.2.3	Description générale	128
5.2.4	Espaces de représentation pixelique	129
5.3	Détection des zones suspectes	130
5.3.1	Principe et motivation	130
5.3.2	Représentation pixelique des résultats	131
5.4	Identification des objets candidats	133
5.4.1	Problèmes d'identification	133
5.4.2	Schéma d'identification proposé	136
5.5	Processus de classification collaborative	141
5.5.1	Description globale	141
5.5.2	Estimation des degrés de confiance	145
5.5.3	Ré-étiquetage des objets candidats	148
5.5.4	Convergence et représentation des résultats	151
5.6	Résultats expérimentaux et discussion	152
5.6.1	Détecteurs utilisés	153
5.6.2	Critères d'évaluation	155
5.6.3	Évaluation des performances	156
5.7	Conclusion	159
Bibliographie		I
A Annexe A : Schéma de prétraitement adopté		XV
A.1	Sélection des techniques de prétraitement	XV
A.2	Démarche d'évaluation	XVI
A.2.1	Critères d'évaluation	XVI
A.2.2	Images de test	XVI
A.3	Évaluation relative à la qualité d'apparition des microcalcifications	XVIII
A.3.1	Critères d'évaluation	XVIII
A.3.2	Protocoles d'évaluation	XVIII
A.3.3	Résultats d'évaluation	XIX

A.4	Synthèse et discussions	XXVII
A.5	Evaluation relative au lien entre les superpixels et les microcalcifications .	XXIX
A.5.1	Schémas de génération des superpixels	XXIX
A.5.2	Critères d'évaluation	XXX
A.5.3	Synthèse et discussions	XXX
A.6	Génération des superpixels: schéma adopté	XXXII

TABLE DES FIGURES

1.1	Structure anatomique du sein	9
1.2	Une lobe d'une glande mammaire	9
1.3	Les types du cancer du sein en fonction (a) de l'emplacement d'apparition et (b) du niveau de propagation	10
1.4	Les incidences standard les plus courantes pour un examen mammographique : Incidence CC et MLO (en bleu). Les incidences standard complémentaires: Médio Latérale à 90° (ML), Inféro Médial à Supéro Latéral oblique (ISO), de dessous (FB), Latéro Médial Oblique (LMO), Latéro Médial à 90° (LM) et Supéro Inférieur Oblique (ISO).	11
1.5	Les deux incidences de base pour un examen mammographique: (a) Incidence MLO et (b) incidence CC	11
1.6	Types de densités selon BI-RADS: (a) ACR1, (b) ACR2, (c) ACR3 et (d) ACR4	13
1.7	Exemple d'une asymétrie marquée sur les incidences MLO (a) et CC (b)	14
1.8	Exemple d'une distorsion architecturale sur une image mammographique avec un zoom avant	14
1.9	Exemple de macrocalcifications (a) et microcalcifications (b) sur une image mammographique	14
1.10	Exemple d'une masse sur une image mammographique	15
1.11	Schéma anatomique de l'UDTL	17
1.12	Exemples de microcalcifications malignes (a) et bénignes (b)	18
1.13	Distributions possibles des MCs	19
1.14	Les différents types de morphologies distingués par la classification de BI-RADS	19
1.15	Exemples de mammographies numériques de la base : (a) Mini-MIAS, (b) DDSM et (c) INBreast	27
1.16	Exemple d'une image mammographique (a) et son masque de vérité de terrain créé (b)	29
2.1	Projection en 3D (a) et 2D (b) du filtre spatial de Vivona <i>et al.</i> [1]	45
2.2	Illustration de la variation du contraste sur une image mammographique entre les MCs : (a) ROI d'une image mammographique sans (a) et avec superposition de la vérité de terrain (contours rouges en (b))	48
2.3	Les sous-fenêtres du filtre de Nagao	52
2.4	Les niveaux d'interprétation exploités dans l'approche proposée	57

3.1	Exemple de quatre éléments structurants associés à un pixel central (cercle rouge)	63
3.2	Exemple des résultats des opérations morphologiques élémentaires sur une image de synthèse	64
3.3	Schéma global de la méthode de détection des MCs proposée	66
3.4	Illustration d'un exemple de MCs de différentes tailles	67
3.5	Les éléments structurants utilisés par notre première contribution	68
3.6	Génération de la carte F_1 à partir d'une image mammographique	71
3.7	Le schéma de la combinaison adoptée	74
3.8	Schéma descriptif d'une boîte à moustache (Box-plot)	77
3.9	Représentation des différents types de pixels distingués pour l'évaluation d'une segmentation	79
3.10	Exemple d'une image mammographique (a) de la base INBreast utilisée dans l'évaluation, avec un zoom sur la ROI (b) où figure les MCs et une superposition des contours de la vérité de terrain (c)	81
3.11	Les cartes morphologiques générées suite à l'application des opérations morphologiques par la première étape de la contribution proposée	82
3.11	Les cartes morphologiques générées suite à l'application des opérations morphologiques par la première étape de la contribution proposée (cont.)	83
3.12	Illustration des résultats intermédiaires pour la génération de la carte de suspicion	84
3.13	Génération de la carte des MCs potentielles	86
3.14	Zoom sur les MCs détectées à partir de trois différentes zones sur une ROI	87
3.15	Illustration des exemples de MCs détectées à partir des mammographies originales (colonne de gauche) par la méthode 'Zh' [2] (colonne du milieu) et la méthode proposée (colonne de droite)	88
3.15	Illustration des exemples de MCs détectées à partir des mammographies originales (colonne de gauche) par la méthode 'Zh' [2] (colonne du milieu) et la méthode proposée (colonne de droite)(cont.)	89
3.15	Illustration des exemples de MCs détectées à partir des mammographies originales (colonne de gauche) par la méthode 'Zh' [2] (colonne du milieu) et la méthode proposée (colonne de droite)(cont.)	90
3.16	Représentation des différents types de pixels distingués pour l'évaluation d'une segmentation	94
4.1	Illustration de deux exemples de foyers de MCs avec des contrastes variables	100
4.2	Exemple d'une ROI avec superposition des maxima régionaux (points marqués en bleu) avec les contours vérité de terrain des MCs (contours rouge fermés)	101
4.3	Schéma global de la deuxième contribution proposée	101
4.4	Exemple d'une distribution statistique (a) des intensités des pixels d'une MC dans une fenêtre de taille 15×15 (b)	102
4.5	Exemple de superpixels SLIC (a) et LPE (b) générés à partir d'une image mammographique (contours bleu) avec superposition des contours GT des MCs (contours rouge)	105
4.6	Exemple de maxima régionaux sélectionnés (jaune) à partir de tous les maxima initialement générés (vert) avec la superposition des contours des superpixels LPE (bleu) et GT des MCs (rouge)	106

4.7	Organigramme du processus de la croissance de région conditionnelle proposée	109
4.8	Illustration d'un exemple des fenêtres d'analyses définies autour d'un germe de départ	111
4.9	Exemple des contours internes (b) et externes (c) d'une région R_c	113
4.10	Illustration des exemples de MCs détectées à partir de trois différentes mammographies (une mammographie par colonne) par la méthode ' M ' de comparaison (deuxième ligne) et la méthode ' CRC ' proposée (troisième ligne). Les flèches en rouge pointent sur les MCs qui n'ont pas été identifiées par une méthode donnée	117
4.11	Zoom sur les MCs détectées sur une sélection de trois zones différentes de deux ROIs	119
5.1	Exemple de résultats de deux méthodes de détection de MCs avec superposition des contours GT (contours rouge). Les pixels blancs sont supposés appartenir à des MCs par une technique donnée	127
5.2	Organigramme global de l'approche proposée	128
5.3	Exemple d'une carte thématique. Les pixels blancs sont supposés appartenir à des MCs. Les pixels noir sont considérés comme des pixels de fond de l'image (pixels de tissus mammaire sain)	129
5.4	Exemple d'une carte de suspicion. Plus les valeurs des pixels sont proches de '1' (couleur rouge), plus le pixel correspondant est suspicieux et est digne d'intérêt	130
5.5	Exemple de superpixels SP1 générés (contours bleu) avec superposition des contours GT (contours rouges)	131
5.6	Schéma global de la phase d'identification des zones suspectes	131
5.7	Conversion des cartes thématique et de suspicion	132
5.8	Schéma de prétraitement adopté par l'approche proposée	133
5.9	Illustration d'un exemple de zones suspectes de différentes tailles et formes identifiées à partir d'une carte thématique d'un même détecteur	134
5.10	Les différents scénarios possibles lors de la comparaison des zones suspectes issues de deux détecteurs différents	135
5.11	Schéma du déroulement de la phase d'identification des objets candidats	136
5.12	Illustration des exemples d'objets candidats identifiés à partir de trois différentes zones suspectes	138
5.13	Les différents scénarios possibles lors de l'unification des listes des objets	139
5.14	Illustration d'un exemple de détermination de l'objet O à considérer dans la liste globale des objets à partir de deux objets présentant un chevauchement	139
5.15	Illustration d'un exemple de détermination de l'objet O à considérer dans la liste globale des objets à partir de deux objets différents	140
5.16	Illustration d'un exemple de classification initiale des objets candidats représentés par leurs positions géométriques sur l'image mammographique (centre de gravité): Les objets associés aux classes C_{CT} , C_{CP} and C_I sont respectivement en rouge, vert et bleu	141
5.17	Organigramme global du processus de classification collaborative proposé	142
5.18	Illustration d'un exemple de graphe généré à partir d'un ensemble d'objets candidats. A ce stade, les objets de la classe d'incertitude (cercles bleu) ne sont pas considérés comme étant des nœuds du graphe	143
5.19	Schéma de déroulement de la phase de classification collaborative	144

5.20	Exemple d'un groupe de MCs qui apparait sur le même superpixel SP1 . .	146
5.21	Illustration d'un exemple de raffinement intra-superpixel	148
5.22	Exemple d'un raffinement inter-superpixel	150
5.23	Illustration des cartes pixeliques (b) générées par les différents détecteurs à partir une ROI d'une image mammographique en NG (a)	155
5.24	Illustration des différents types possibles des objets identifiés	155
5.25	Pourcentages des taux de TP obtenus suite à l'application de l'approche proposée	156
5.26	Boîtes à moustaches des taux des TPs obtenus suite à l'application de l'opérateur de fusion conjonctif avant et après l'application du processus collaboratif proposé	158
A.1	Echantillons d'images mammographiques pour l'évaluation des techniques de prétraitement appartenant à différentes catégories	XVI
A.2	Les ROIs sur les images mammographiques sélectionnées	XVII
A.3	Degrés d'importance des critères de validation des techniques de prétraite- ment	XVIII
A.4	Symboles utilisés pour évaluer la validation des critères	XIX
A.5	Résultats des différents prétraitements appliqués sur les échantillons des images considérées	XX
A.5	Résultats des différents prétraitements appliqués sur les échantillons des images considérées (cont.)	XXI
A.6	Résultats des différents prétraitements appliqués sur les ROIs des échan- tillons des images considérées	XXII
A.6	Résultats des différents prétraitements appliqués sur les ROIs des échan- tillons des images considérées (cont.)	XXIII
A.6	Résultats des différents prétraitements appliqués sur les ROIs des échan- tillons des images considérées (cont.)	XXIV
A.6	Résultats des différents prétraitements appliqués sur les ROIs des échan- tillons des images considérées (cont.)	XXV
A.7	Schéma adopté pour la génération des superpixels	XXXII

LISTE DES TABLEAUX

1.1	Interprétations et recommandations associées aux catégories distinguées par BI-RADS	12
1.2	Comparaison des caractéristiques des micro et macro-calcifications	16
1.3	Calcifications mammaires en fonction des emplacements d'apparition	17
1.4	Projection des différentes descriptions morphologiques des MCs sur une image mammographique	24
1.5	Synthèse des caractéristiques des images dans les bases de mammographies numériques	27
1.6	Synthèse des caractéristiques des MCs dans la sélection des images utilisées pour l'évaluation	28
2.1	Comparaison des techniques explorées dans la littérature pour la détection des MCs	47
2.2	Approche proposée: Niveaux d'interprétation et connaissances sémantiques	58
3.1	Description statistique de la répartition des valeurs de la sensibilité (SE) par la contribution proposée	91
3.2	Caractéristiques des MCs dans une sélection des mammographies numériques	92
3.3	Valeurs des mesures SI et SE retenues par la méthode proposée (' D_M ') et la méthode 'Zh' pour une sélection de mammographies numériques. Les meilleurs résultats sont en gras	93
4.1	Description statistique de la répartition des valeurs de la sensibilité (SE) par la méthode de comparaison (' M ') et la méthode proposée (' CRC ') sur l'ensemble des images issues de la base INBreast	118
4.2	Les valeurs Min, Max, Moyenne et Ecart-type des mesures d'évaluation. Les meilleurs résultats sont en gras	120
4.3	Pourcentages d'amélioration des métriques d'évaluation pour les différentes familles d'images mammographiques distinguées. Les couleurs rouge (respectivement bleu et vert) correspondent à OF (respectivement EF avec les familles F2 et F4)	121
5.1	Comparaison des détecteurs utilisés	154
5.2	Comparaison des taux moyens de TPs obtenus par l'approche proposée par rapport à chaque détecteur	157
5.3	Pourcentages des taux de TPs améliorés en comparaison avec chaque détecteur utilisé	157

5.4	Taux moyens de TP obtenus en comparaison avec les opérateurs d'union et d'intersection	158
5.5	Pourcentages d'amélioration des taux de TP en comparaison par rapport aux opérateurs union et intersection ainsi que le meilleur détecteur	159
A.1	Évaluation des résultats des prétraitements appliquées sur les images mammographiques	XXVIII
A.2	Synthèse des résultats de générations des superpixels SP1 à partir des différents schémas	XXXI

INTRODUCTION GÉNÉRALE

Problématique de la détection des microcalcifications

Une image est une source d'information composée d'un ensemble de pixels. Ces derniers sont représentés chacun par un vecteur d'attributs qui traduit les propriétés sémantiques des objets de cette image en propriétés numériques.

Dans le contexte médical, l'imagerie médicale est devenue un outil indispensable pour l'analyse et le diagnostic des pathologies pour les radiologues. À titre d'exemple, l'imagerie mammographique, à laquelle s'intéresse notre sujet, est utilisée pour tout examen clinique relatif à l'étude du cancer du sein qui est de plus en plus fréquent chez la femme.

En mammographie, les systèmes d'aide à la décision, (Computer-aided diagnosis, CADs), apportent aux cliniciens des outils informatiques dont l'objectif est de faciliter la prise de décision. Ces derniers concernent en particulier plusieurs structures pathologiques telles que les masses et les microcalcifications (MCs). En routine, les radiologues s'intéressent plus à des systèmes capables de détecter les microcalcifications. En effet, sur le plan du diagnostic du cancer du sein, ces lésions constituent le premier signe précurseur de la pathologie et leur détection précoce augmente l'espérance de vie chez les femmes affectées [3]. Sur le plan anatomique, ces lésions apparaissent regroupées dans des foyers avec une très grande variété de formes et de distributions. Cette diversité anatomique est liée à des phénomènes qui sont à l'origine de leur production. Elle se traduit par une diversité de caractéristiques d'apparition, spatiales et morphologiques, sur les images mammographiques. Par ailleurs, les microcalcifications peuvent être indissociables même par l'œil humain des autres tissus mammaires voisins à cause de leur contraste très faible par rapport à ces derniers [4]. La détection des microcalcifications constitue donc, non seulement un réel besoin mais aussi un grand défi.

Ce sujet de thèse s'inscrit dans le cadre de l'interprétation des images mammographiques dont l'objectif est de détecter les foyers de microcalcifications ainsi que de fournir une source d'information exploitable pour l'expert au cours de son diagnostic.

Au regard de l'utilité de l'interprétation des microcalcifications à partir des images mammographiques, différents travaux de recherche ont eu pour objectif de proposer des systèmes permettant d'identifier les ensembles de pixels adjacents pouvant faire référence à des microcalcifications potentielles. Nous pouvons distinguer deux familles d'approches exploitées par ces systèmes : les approches globales et les approches locales. Les approches globales opèrent sur l'image mammographique entière tandis que les approches locales opèrent sur une région d'intérêt (ROI) sélectionnée par le radiologue.

Les différents travaux s'appuient sur des modes de raisonnement spécifiques, qui explorent nécessairement une ou plusieurs propriétés des microcalcifications, afin de détecter les groupements de pixels adjacents qui les constituent. En revanche, ces travaux souffrent souvent de fortes erreurs de détections telles que les erreurs de type Faux Négatif et Faux Positif.

L'apparition de ces deux types d'erreurs est essentiellement dûe à :

- l'absence d'intégration des informations sémantiques contextuelles dans le raisonnement appliqué : seule l'information numérique, sensible à la qualité du système d'acquisition, est exploitée ;
- l'utilisation d'un seul niveau d'analyse pour la détection des microcalcifications, qui est le niveau pixélique : les microcalcifications sont décrites en tant que groupements de pixels adjacents vérifiant les critères des pixels suspects définis par la technique utilisée ;
- l'utilisation d'une technique de détection, seule, pour la résolution d'un tel problème complexe.

Contributions

Dans ce travail de recherche, nous proposons un système dédié à la fois à l'identification et la segmentation des microcalcifications à partir des images mammographiques. En effet, la forme et la morphologie d'un foyer de microcalcifications jouent un rôle important dans l'estimation de son degré de malignité. La contribution principale proposée est une approche qui fait collaborer les résultats de différentes méthodes de détection afin de générer un résultat plus fiable. Pour y parvenir, nous avons procédé à l'application d'un schéma de prétraitement dont l'objectif est d'améliorer la qualité des images ainsi que d'apparition des microcalcifications. Par la suite, nous avons mis en place deux méthodes de détection basées sur deux modes de raisonnements différents. Bien que ces méthodes aient montré une fiabilité acceptable, elles n'étaient toujours pas suffisantes et elles ont montré une sensibilité à la diversité des microcalcifications. Par ailleurs, elles ont fourni

des résultats différents en termes d'erreurs de détection. C'est dans ce contexte qu'est apparu l'intérêt d'une approche qui évalue les résultats de différentes méthodes afin de retenir celles qui sont les plus pertinentes et générer, ainsi, un résultat plus fiable. C'est ce qu'on appelle des approches collaboratives.

L'idée sous-jacente consiste en un raffinement itératif et automatique des décisions initiales prises par les méthodes de détection utilisées sur la base des informations partagées entre elles, ce qui permet de réduire leurs désaccords initiaux. Le raffinement appliqué est conduit à deux niveaux : un niveau local et un niveau contextuel. Il permet d'estimer des degrés de fiabilité locaux sur la base d'une connaissance préalable des objets recherchés (microcalcification) et de l'image traitée.

Chaque niveau d'analyse offre une représentation propre de la granularité de l'information spatiale étudiée dans l'image. Il applique un traitement indépendant basé sur des informations numériques et d'autres sémantiques afin de produire des connaissances sur l'objet sémantique étudié relativement à la représentation utilisé pour la description des microcalcifications.

Le choix de cette stratégie basée sur la propagation et l'échange d'informations est justifié par le fait que nous cherchons à simuler le raisonnement du radiologue au cours de son diagnostic. En effet, ce dernier, afin d'interpréter les microcalcifications sur une image mammographique, commence par localiser les petites régions ayant des formes plus ou moins circulaires et plus brillantes par rapport à leur voisinage local. Par la suite, le radiologue émet des hypothèses sur le type de ces premières détections en les situant par rapport à leur positionnement local et global sur l'image, tout en tenant compte de son expertise du domaine.

Structuration du manuscrit

Ce manuscrit est structuré en cinq chapitres suivis par une conclusion générale et des perspectives d'amélioration.

Le premier chapitre présente le cadre d'étude dans ce travail de recherche, à savoir le cancer du sein et l'ensemble des lésions qu'il peut produire. Dans sa première partie, il présente une description anatomique du sein suivie par une autre du lexique médical utilisé pour décrire ce problème de santé souvent diagnostiqué à l'aide d'une mammographie. Ensuite, la présentation de l'ensemble des lésions détectables à partir d'une image mammographique justifie l'intérêt que nous avons porté pour les microcalcifications. Une étude détaillée des caractéristiques de ces lésions est aussi abordée dans ce chapitre. Elle a permis de mettre en lumière les difficultés ainsi que les intérêts de proposer un système

de détection fiable afin d’apporter une aide aux radiologues lors de l’interprétation. Dans sa deuxième partie, ce chapitre aborde les différentes bases d’images qui existent dans la littérature afin de valider notre choix. Il décrit également la démarche que nous avons suivie pour concevoir la vérité-terrain utilisée pour la validation de notre système.

Le deuxième chapitre présente un état de l’art sur les méthodes mises en œuvre pour traiter la problématique de détection des microcalcifications à partir des images mammographiques. Cela nous permet d’évoquer les avantages et les inconvénients de chacune de ces méthodes et de mettre l’accent sur l’importance d’une étape de prétraitement pour améliorer la qualité de ces images. A partir des constatations établies suite aux études effectuées, nous arrivons à constater que les difficultés rencontrées sont causées par la diversité des caractéristiques des microcalcifications ainsi que l’absence des connaissances sémantiques par la démarche de détection suivie. Ceci nous conduit à présenter l’idée générale de notre contribution principale de ce travail qui consiste en une combinaison collaborative des résultats de plusieurs méthodes de détection afin de converger vers une décision collaborative optimale.

Le troisième chapitre décrit la première contribution de la thèse. Elle consiste à proposer une nouvelle méthode de détection basée sur la morphologie mathématique et les indices de similarité structurelle qui vise à traiter le problème de faible contraste des microcalcifications. Les résultats évalués ont prouvé sa capacité à gérer la diversité des microcalcifications en termes de contrastes et formes. Une étude comparative avec d’autres méthodes proposées dans l’état de l’art a également prouvé le gain offert en termes de réduction du nombre des faux négatifs ainsi que l’amélioration de la précision des contours. En revanche, la fiabilité de cette méthode n’empêche pas certaines difficultés liées à la forte similarité des microcalcifications à leurs tissus de superposition avec une grande conformité aux éléments structurants.

Le quatrième chapitre s’intéresse à la présentation de la deuxième contribution menée dans ce travail de recherche. Après avoir étudié l’intérêt d’une analyse locale à partir d’un point de départ, pour la détection des microcalcifications, nous mettons l’accent sur l’importance de l’intégration de connaissances *a priori* pour une meilleure segmentation. Nous proposons, par la suite, une méthode qui intègre des critères géométriques et d’autres relatifs au gradient afin de faire converger un point de départ vers les meilleurs contours. L’évaluation de cette contribution est menée selon une méthodologie similaire à celle de la première contribution. La pertinence en termes de segmentation offerte par cette contribution se trouve accompagnée par des fausses détections liées à la diversité des types de tissus de superposition des microcalcifications.

Dans le cinquième chapitre, les conclusions tirées des chapitres précédents nous amènent à conclure de l'absence d'une méthode capable, à elle seule, de gérer la diversité des microcalcifications sur les images mammographiques. Cela implique la génération des fausses détections dans les décisions fournies. La contribution mise en œuvre dans ce chapitre repose sur cette base. Il s'agit d'une combinaison collaborative de différentes méthodes de détection dans l'idée de tirer partie de leurs avantages et de converger vers une décision commune plus pertinente que celle fournie par une méthode individuelle. La dernière partie de ce chapitre est consacrée à l'évaluation et la comparaison de l'approche proposée. Les expérimentations prouvent la pertinence de cette approche en termes de qualité des résultats et améliorations apportées par rapport à certaines méthodes de détection dont les deux contributions précédentes font partie.

Ce manuscrit est clôturé par une conclusion générale qui résume l'ensemble des contributions et propose certaines perspectives d'amélioration.

CONTEXTE MÉDICAL

Sommaire

1.1	Introduction	8
1.2	Étude anatomique	8
1.2.1	Anatomie du sein	8
1.2.2	Cancer du sein	9
1.3	Imagerie médicale	10
1.3.1	Mammographie	10
1.3.2	Interprétation des mammographies	11
1.3.3	Les anomalies détectables à partir d'une mammographie numérique	13
1.4	Calcifications mammaires	15
1.4.1	Définition selon le lexique médical	15
1.4.2	Diagnostic des calcifications mammaires	16
1.5	Microcalcifications: Caractéristiques d'apparition	18
1.5.1	Nombre et regroupement	18
1.5.2	Distribution	19
1.5.3	Morphologie	19
1.6	Bases de mammographies numériques	25
1.6.1	Exigences de création	25
1.6.2	Bases de mammographies	26
1.6.3	Base de mammographies utilisée	28
1.6.4	Génération de la vérité de terrain	29
1.7	Conclusion	29

1.1 Introduction

L'imagerie médicale regroupe des moyens d'acquisition d'images du corps humain en se basant sur plusieurs domaines physiques auxquels sont associés des techniques d'imagerie. Depuis leur apparition, ces techniques ont conduit à un progrès de la médecine en permettant d'observer l'intérieur du corps humain et donc d'améliorer le diagnostic et le suivi thérapeutique.

L'imagerie mammographique, à laquelle s'intéresse notre travail, est la modalité médicale de référence utilisée pour l'analyse et le diagnostic du cancer du sein. L'interprétation des images mammographiques retenues est la tâche des experts en radiologie. Ces derniers ont recours à leur expertise de domaine et à leur savoir médical pour reconnaître et localiser les anomalies puis envisager l'ensemble des actes thérapeutiques à suivre.

Au regard de l'augmentation de la fréquence d'apparition du cancer du sein ces dernières décennies et donc du nombre de mammographies réalisées, les travaux de recherches ont mené au développement de systèmes d'aide à la décision (CAD). L'objectif de ces systèmes est d'apporter un deuxième avis aux radiologues afin d'améliorer la précision de leur diagnostic, sans toutefois les remplacer. Néanmoins, la pertinence de ce deuxième avis est fortement dépendante de la qualité des images acquises et de la maîtrise de la démarche d'interprétation suivie par le radiologue lors du diagnostic.

Dans ce chapitre, nous décrivons le cadre général de notre travail, à savoir le cancer du sein et son diagnostic à partir des images mammographiques. Nous commençons par une brève description de l'anatomie du sein chez la femme et du lexique utilisé dans le contexte du cancer du sein. A cette description s'ajoute la présentation des différentes lésions détectables à partir des images mammographiques. Dans une deuxième partie, nous nous focalisons sur la description d'un type particulier de lésions, autour duquel tourne notre sujet de thèse : les microcalcifications. Nous étudions ainsi, les caractéristiques anatomiques et histologiques de ces dernières dans l'objectif de maîtriser leur structures et de mettre en lumière les intérêts ainsi les difficultés de leur détection et donc les problématiques liées à la conception d'un système de détection fiable. Nous clôturons ce chapitre par une présentation de la base de mammographies numériques considérée pour notre partie expérimentale.

1.2 Étude anatomique

1.2.1 Anatomie du sein

Le sein est un organe globuleux qui occupe la partie antéro-supérieure du thorax. Il est situé par dessus du muscle pectoral qui garantit son maintien [5]. Il est principalement

constitué d'une glande mammaire, de tissu conjonctif (de soutien) et de tissu adipeux (gras), le tout est recouvert par la peau (Figure 1.1). La glande mammaire est divisée en 15 à 20 lobes (Figure 1.2), composée chacune de 20 à 40 lobules. Les lobules constituent le centre de production du lait, et sont liés à des canaux centraux (canaux galactophores) dont le rôle est d'acheminer le lait produit vers le mamelon.

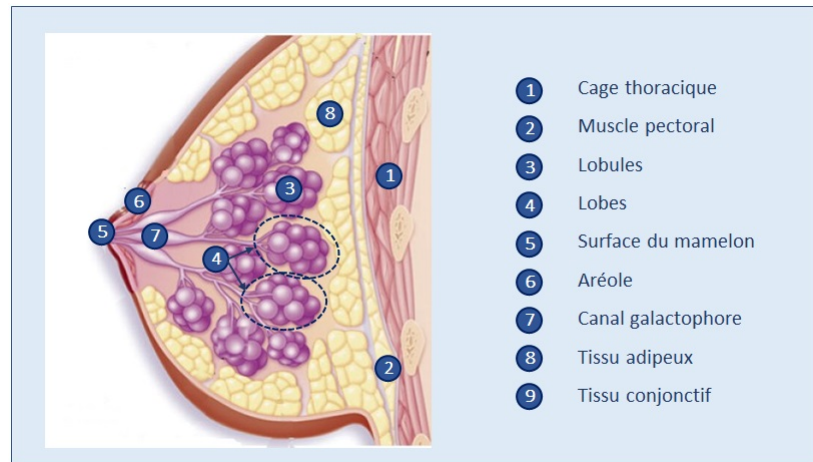


Figure 1.1 – Structure anatomique du sein

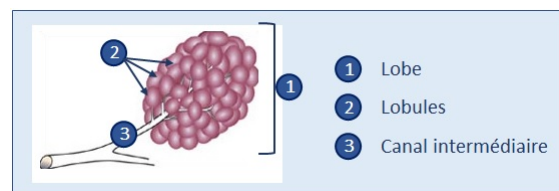


Figure 1.2 – Une lobe d'une glande mammaire

1.2.2 Cancer du sein

Le cancer du sein est une pathologie très fréquente à travers le monde qui cause le plus grand nombre de décès chez les femmes [6]. Les statistiques indiquent que le taux d'incidence a augmenté au cours des trois dernières décennies. Par contre, le diagnostic précoce et le progrès des traitements appliqués ont permis de réduire le taux de mortalité au cours de la même période [7]. Le taux de survie estimé est aussi amélioré. Il est de cinq ans après le diagnostic et il varie entre 80% et 90%, selon l'âge et le type de cancer [8].

Anatomiquement parlant, le cancer du sein est le résultat d'une multiplication anormale de cellules dans le sein qui engendre la formation de masses appelées lésions. Le plus souvent, les cellules concernées par cette multiplication excessive sont celles situées au niveau des canaux galactophores ou des lobules (Figure 1.3a). On parle alors, de cancer canalaire ou cancer lobulaire.

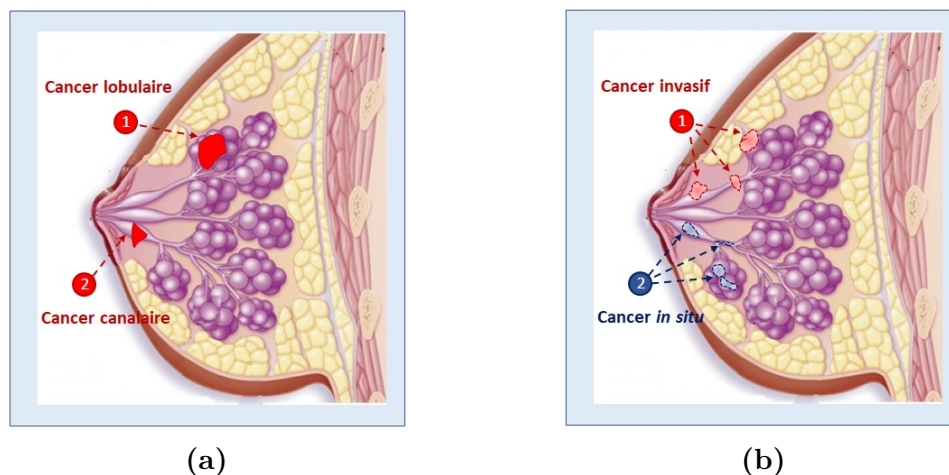


Figure 1.3 – Les types du cancer du sein en fonction (a) de l’emplacement d’apparition et (b) du niveau de propagation

Le stade d’évolution est, en outre, un autre facteur pour la distinction du cancer du sein. On parle de cancer *in situ* si les cellules cancéreuses sont encore dans leur emplacement d’origine. On parle, par ailleurs, de cancer *infiltrant/invasif* si les cellules cancéreuses sont propagées vers les tissus voisins ou dans le reste du corps (Figure 1.3b). Une telle propagation peut prendre des mois voir même des années [9].

1.3 Imagerie médicale

L’imagerie médicale est le domaine qui a le plus contribué au progrès médical depuis une vingtaine d’années, en s’imposant comme un pivot de développement futur de la médecine. Elle repose sur un ensemble de modalités qui permettent de "voir" l’intérieur du corps humain, la structure ou la fonction des organes de manière non-invasive.

1.3.1 Mammographie

Les modalités d’imagerie médicale exploitent des principes initialement réservés pour des domaines et concepts purement physiques tels que les rayons X, les ultrasons, la résonance magnétique nucléaire, etc. La mammographie, modalité à laquelle nous nous intéressons dans le cadre de ce travail, est un examen radiologique qui consiste à radiographier le sein pour visualiser et étudier la glande mammaire. Elle permet de faire un diagnostic précoce ou un examen de dépistage du cancer du sein. Cet examen s’effectue avec un appareil à rayons X, le mammographe, dédié uniquement à cet usage. Il génère des images numériques de hautes résolutions dites images mammographiques avec différents types d’incidences selon l’angle d’acquisition. On distingue deux types d’incidences : des incidences de base et des incidences complémentaires (Figure. 1.4). Les incidences de

base sont : l'incidence de face, dite incidence Carnio-Caudal (CC) et l'incidence oblique externe, dite Médio Latérale Oblique (MLO). Ces deux incidences sont les plus utilisées pour un examen de dépistage ou de diagnostic du cancer du sein (Figure 1.5).

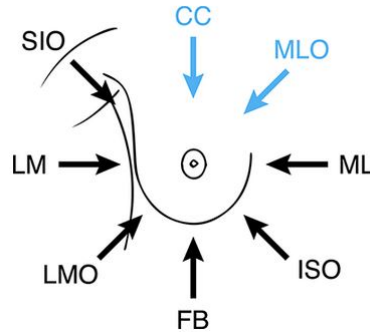


Figure 1.4 – Les incidences standard les plus courantes pour un examen mammographique : Incidence CC et MLO (en bleu). Les incidences standard complémentaires : Médio Latérale à 90° (ML), Inféro Médial à Supéro Latéral oblique (ISO), de dessous (FB), Latéro Médial Oblique (LMO), Latéro Médial à 90° (LM) et Supéro Inférieur Oblique (ISO).

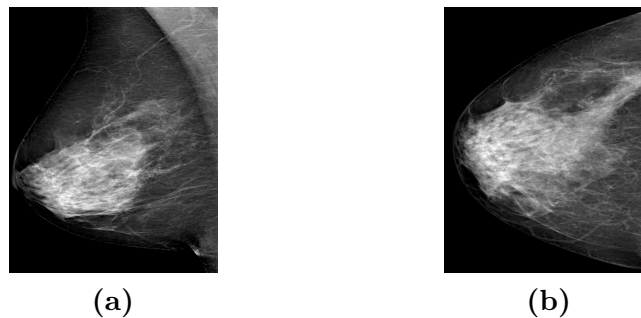


Figure 1.5 – Les deux incidences de base pour un examen mammographique : (a) Incidence MLO et (b) incidence CC

1.3.2 Interprétation des mammographies

1.3.2.1 Anomalies et risque de cancer

Une fois l'examen de mammographie réalisé, un expert du domaine (radiologue) examine les mammographies numériques générées et passe à la recherche de présence de changements et/ou d'anomalies. Suite au résultat de son analyse, le radiologue décide et prescrit les actes à suivre. Dans l'idée d'utiliser un lexique de description standard et de faciliter les lectures multiples, plusieurs classifications ont été élaborées. Les deux classifications les plus répandues et les plus utilisées sont celles de Breast Imaging Reporting and Data System (BI-RADS) [10] et de Le Gal [11]. La classification de Le Gal a été introduite en 1984. Toutefois, cette classification est simple et ne concerne qu'un seul type d'anomalies qui sont les microcalcifications. Elle s'est avérée donc insuffisante pour assurer un lexique de diagnostic standard en cas d'apparition d'autres signes pathologiques.

La classification de BI-RADS a été conçue par l'*American College of Radiology* (ACR) en 2003. Elle présente l'avantage d'établir une décision commune en fonction de n'importe quel signe pathologique retrouvé. Elle distingue six différentes catégories et associe à chacune un ensemble de recommandations avec un risque de malignité. Le tableau 1.1 donne un aperçu sur l'ensemble des catégories distinguées avec la conduite à tenir.

Tableau 1.1 – Interprétations et recommandations associées aux catégories distinguées par BI-RADS

Catégorie	Interprétation	Recommandation	Risque
<i>BI-RADS 0</i>	Évaluation incomplète	Examens complémentaires	
<i>BI-RADS I</i>	Mammographie normale	Dépistage régulier	0%
<i>BI-RADS II</i>	Région(s) normale(s)	Dépistage régulier	0%
<i>BI-RADS III</i>	Région(s) probablement bénigne(s)	Suivi sur 6 mois	<2%
<i>BI-RADS IV</i>	Région(s) probablement maligne(s)	Biopsie	>2 et <95%
<i>BI-RADS V</i>	Région(s) maligne(s)	Biopsie	>95%

1.3.2.2 Densité mammaire et risque de cancer

Un certain nombre de systèmes d'évaluation de risques ont été développés pour estimer la probabilité que les femmes développent un cancer du sein. Les études menées sur ce thème, ont montré que la densité mammaire peut être un indicateur important pour la présence et le développement du cancer du sein [12–16]. Elles ont aussi montré que la précision des approches d'interprétation des mammographies numériques est fortement influencée par le degré et la répartition spatiale de la région dense sur ces images [17]. La densité mammaire est définie en tant que l'entité radiologique appréciant la quantité des structures mammaires denses (radio-opaques) caractérisant le tissu fibroglandulaire par rapport aux structures adipeuses (radio-transparentes) caractérisant le tissu graisseux. Les systèmes développés produisent quatre différentes classifications [18,19] : Wolfe (1976) [20,21], Tabàr (1982) [22,23], Boyd (1995) [24,25], et BI-RADS (1993) [26,27]. Bien que ces classifications partent de différents principes de raisonnement pour caractériser la densité mammaire, plusieurs travaux ont pu déterminer une corrélation fiable entre elles [18]. La classification la plus utilisée par les radiologues est celle de BI-RADS. Elle exprime la densité par des notes qualitative ainsi que quantitative, ce qui permet une répartition des seins en quatre types (Figure 1.6) :

- *Graisseux homogène* (ACR1) : Le sein est clair et entièrement graisseux. Le pourcentage du tissu fibroglandulaire est inférieur à 25%. Avec ce type de seins, l'interprétation des mammographies numériques générées est facile ;

- *Graisseux hétérogène* (ACR2) : Le sein est clair avec des opacités dispersées. Le pourcentage de tissu fibroglandulaire varie entre 25% et 50% ;
- *Dense hétérogène* (ACR3) : Le sein est dense et hétérogène. La partie fibroglandulaire qu'il présente occupe entre 50% et 75% de la glande mammaire. La présence de ce pourcentage de densité rend difficile l'interprétation des mammographies numériques, vu sa capacité de masquer des masses de petites tailles ;
- *Dense homogène* (ACR4) : Le sein extrêmement dense et homogène. Il présente un pourcentage de tissu fibroglandulaire supérieur à 75%. Ce taux est capable de non seulement augmenter la complexité d'interprétation mais aussi de diminuer la sensibilité de la mammographie numérique générée.

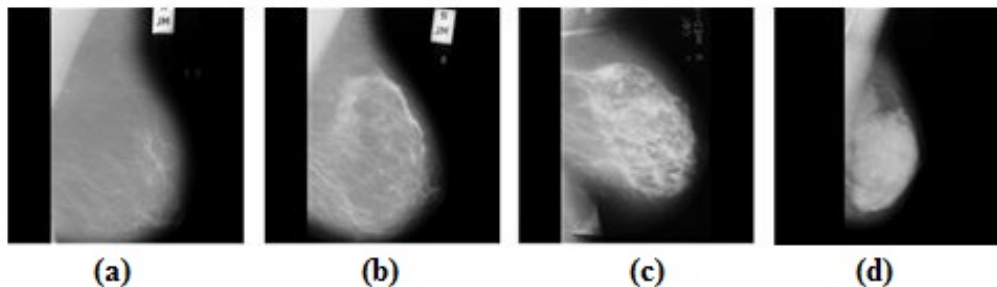


Figure 1.6 – Types de densités selon BI-RADS : (a) ACR1, (b) ACR2, (c) ACR3 et (d) ACR4

1.3.3 Les anomalies détectables à partir d'une mammographie numérique

L'utilité majeure de la mammographie se trouve dans sa capacité à permettre la détection d'anomalies pouvant être des signes avant-coureur du cancer, bien avant que des premiers symptômes se manifestent ou que les femmes soient en mesure de les palper. Les signes d'anomalies détectables par un radiologue à partir d'une image mammographique sont classés, selon la classification de BI-RADS [10] en quatre différentes catégories :

- *L'asymétrie mammaire* : Elle est définie comme une asymétrie de répartition des densités et des structures anatomiques dans une paire d'images mammographiques. Elle apparaît entre les densités observées sur les vues, droite et gauche, aussi bien en incidence MLO qu'en incidence CC et est généralement liée à l'existence d'une lésion cancéreuse (Figure 1.7) ;

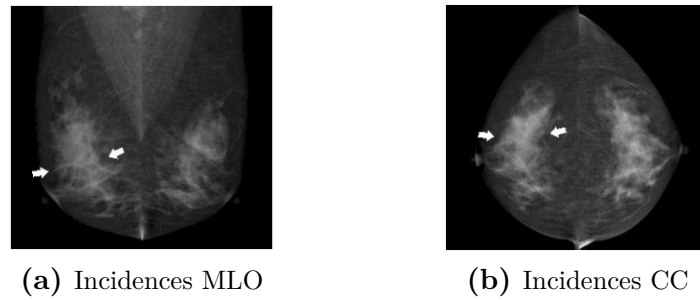


Figure 1.7 – Exemple d’une asymétrie marquée sur les incidences MLO (a) et CC (b)

- *Les distorsions architecturales* : Elles sont définies comme une déformation de l’architecture normale du tissu mammaire. Une distorsion architecturale inclut généralement des spéculations (Figure 1.8) rayonnant d’un point focal, et la rétraction ou la distorsion du bord du sein si elle se situe à une position suffisamment proche ;

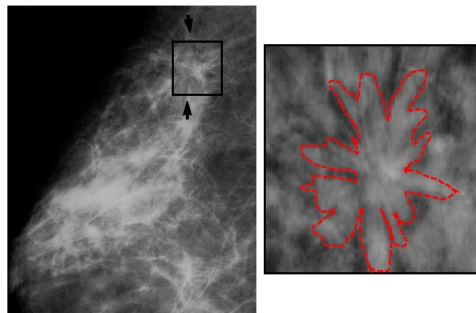


Figure 1.8 – Exemple d’une distorsion architecturale sur une image mammographique avec un zoom avant

- *Calcifications* : Elles sont définies comme des dépôts de calcium de petite dimension (Figure 1.9). Elles sont distinguées en microcalcifications ou macrocalcifications en fonction de leur nombre, taille ainsi que la densité d’apparition par rapport au tissus de superposition [28]. Une calcification mammaire, pour les radiologues, est un signe radiologique très important qui doit être soigneusement traité puisqu’il prédit un future risque [29] ;



Figure 1.9 – Exemple de macrocalcifications (a) et microcalcifications (b) sur une image mammographique

- *Les masses* : Elles sont définies comme des lésions occupant un espace sur le sein (Figure 1.10). Une masse apparait sur les deux incidences (MLO et CC). Sa visibilité diffère en fonction du type du tissu de superposition [30]. Sa caractérisation est basée sur certains critères tels que la forme (ronde, ovale, lobulaire et irrégulière), la densité ainsi que le type de contour (circonscrit, effacé, micro-lobulé, mal défini et spiculé) [31].

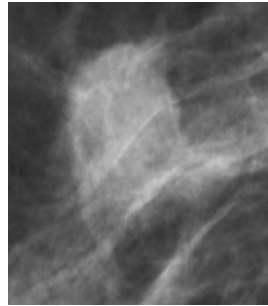


Figure 1.10 – Exemple d’une masse sur une image mammographique

Dans ce travail de recherche, l’interprétation des images mammographiques se focalise sur la détection d’un signe pathologique particulier : les calcifications mammaires. Ce signe précoce doit être soigneusement étudié pour prévoir la possibilité de production du cancer du sein et décider des actes à suivre [32]. Une détection fiable nécessite une analyse approfondie des propriétés histologiques et anatomiques des calcifications. Celle-ci sera l’objectif des sections suivantes.

1.4 Calcifications mammaires

1.4.1 Définition selon le lexique médical

Pour les experts en radiologie, les calcifications sont des « *entités amorphes, radiopaques (denses) et cellulaires produites par un dépôt de calcium dans les canaux galactophores* ». Elles sont généralement associées à des modifications dans le tissu mammaire et n’ont aucun lien avec la quantité de calcium absorbée par l’alimentation. En fonction de la taille et la visibilité sur les images mammographiques, les radiologues distinguent les calcifications mammaires en macrocalcifications et microcalcifications. Les macrocalcifications sont de nature généralement bénigne. Par contre, les microcalcifications peuvent être un signe avant-coureur du cancer du sein. Le tableau 1.2 présente une comparaison des descriptions pour ces deux types de calcifications.

Tableau 1.2 – Comparaison des caractéristiques des micro et macro-calcifications

	Macrocalcifications	Microcalcifications
<i>Taille</i>	Supérieure à 1mm	Inférieure à 1mm (généralement entre 0.1 et 0.5mm)
<i>Fréquence</i>	Fréquentes chez les femmes âgées de plus que 50 ans	
<i>Malignité</i>	Malignes dans des cas très particuliers	Malignes dans 30 à 50% des cas
<i>Apparition</i>		Apparaissent sous plusieurs formes et distributions
<i>Visibilité</i>	Clairement visibles où elles présentent un contraste très élevé, un contour net et bien défini	Parfois indistinguables par rapport au tissu voisin où elles présentent un contraste faible et un contour flou

La présence des calcifications mammaires n'est généralement accompagnée d'aucun symptôme et n'est révélée que par une mammographie. La surveillance et le traitement thérapeutique à suivre en cas de présence, s'il est requis, est question du diagnostic effectué. En effet, pour les radiologues, les calcifications mammaires peuvent toutefois être un signe avant-coureur de cancer du sein. En revanche, elles peuvent, dans quelques configurations, être associées à un processus néoplasique qui n'est pas nécessairement cancéreux.

1.4.2 Diagnostic des calcifications mammaires

Le diagnostic d'un ensemble de calcifications mammaires nécessite une étude histologique et une étude de caractéristiques. L'étude histologique concerne les emplacements d'apparition ainsi que les phénomènes de production. Tandis que l'étude de caractéristiques concerne les propriétés d'apparition.

1.4.2.1 Phénomènes de production

La production des calcifications mammaires peut être due soit à :

- Des accumulations et stagnations des sécrétions mammaires entre les lobules et les canaux galactophores. On peut avoir à ce niveau une compression ou une modification de la structure des canaux galactophores ou encore une déformation de l'Unité Terminale Ducto-Lobulaire (UDTL). L'UDTL est constituée de l'ensemble de lobules avec leur canaux galactophores terminaux inter et intra lobulaire (Figure 1.11) ;
- Dépôts calcaires dans des zones de nécrose qui produit une perte d'une zone cutanée plus ou moins étendue.

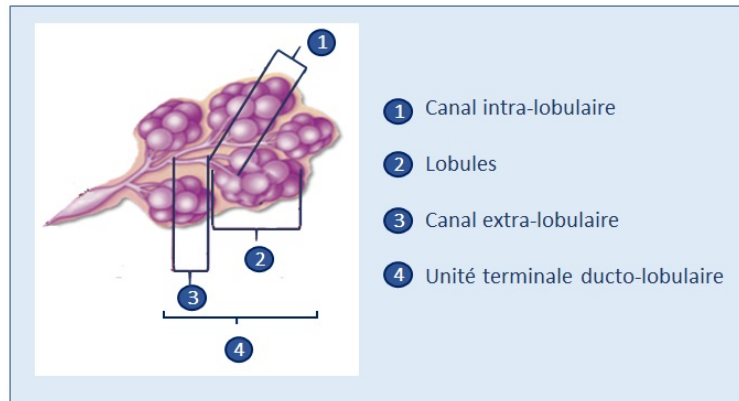


Figure 1.11 – Schéma anatomique de l'UDTL

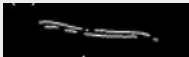

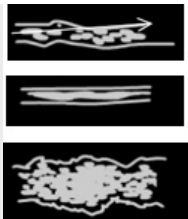
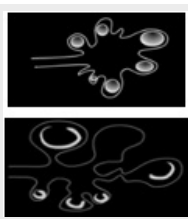
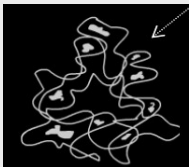
1.4.2.2 Emplacements de production

Les calcifications mammaires peuvent apparaître à plusieurs niveaux dans la glande mammaire :

- Au niveau du tissu de soutien (tissu conjonctif) : Elles sont dues à des altérations qui peuvent accompagner les affectations tumorales de l'arbre galactophorique ;
- Au contact avec les tissus constituant les canaux galactophores (épithélium galactophorique) : Elles sont dues à des proliférations des cellules qui génèrent des déséquilibres métaboliques responsables des nécroses cellulaires ;
- Au niveau des canaux galactophores : Elles sont dues à des sécrétions mammaires.

En termes d'emplacement d'apparition, les calcifications mammaires peuvent être classifiées en cinq différentes classes (Tableau 1.3).

Tableau 1.3 – Calcifications mammaires en fonction des emplacements d'apparition

Types	Vasculaire	Glandes sébacées	Canalaire	Lobulaire	Strome
Exemple					
Endroit d'apparition	Vaisseaux sanguins	Tissus sous-cutanés	Canaux galactophores	Lobules	Tissu conjonctif

Notre intérêt dans ce travail de recherche, est accordé à la détection des microcalcifications vu le risque de malignité qu'elles présentent par rapport aux macrocalcifications. En effet, les macrocalcifications sont souvent bénignes et ne demandent que très rarement

de suivi ou de traitement. En outre, leur détection est une tâche très facile du fait qu'elles sont clairement visibles sur les images mammographiques. Par contre, les microcalcifications s'avèrent compliquées à repérer dans le tissu mammaire du fait de leur taille très faible avec la sensibilité des emplacements d'apparition. La section suivante détaille les caractéristiques d'apparition des microcalcifications, sur une image mammographique, et sur la base desquelles les experts associent des degrés de suspicion.

1.5 Microcalcifications : Caractéristiques d'apparition

Une microcalcification (MC) est un dépôt de calcium composé de plusieurs substances radio-opaques qui se traduisent par des petits points clairs de formes variantes [3]. Elles apparaissent souvent rassemblées en groupes dits de foyers et éparpillées sur un espace plus ou moins grand dans une image mammographique. La taille très petite et les densités locales élevées sont les éléments de base qui distinguent les microcalcifications (notées MCs) des autres types de lésions. Par contre, elles peuvent être facilement confondues avec des pixels bruits ou des pixels des tissus localement voisins. Une caractérisation fiable de ces lésions permet de décider leur degré de suspicion mais doit inclure la morphologie, la distribution et si possible le nombre, l'évolutivité et la taille de la zone considérée [33–35].

1.5.1 Nombre et regroupement

Le nombre et la densité des MCs présentes sur une image mammographique sont des facteurs importants dans le diagnostic effectué. En effet, plus le nombre de MCs est important, plus le risque d'avoir un cancer du sein est important. Elles sont marquées nombreuses si leur nombre dépasse la valeur 10 (selon la classification de Le Gal [11]). Sauf que, selon les experts, il n'existe pas un seuil au-dessous duquel le risque du cancer est égal à zéro.

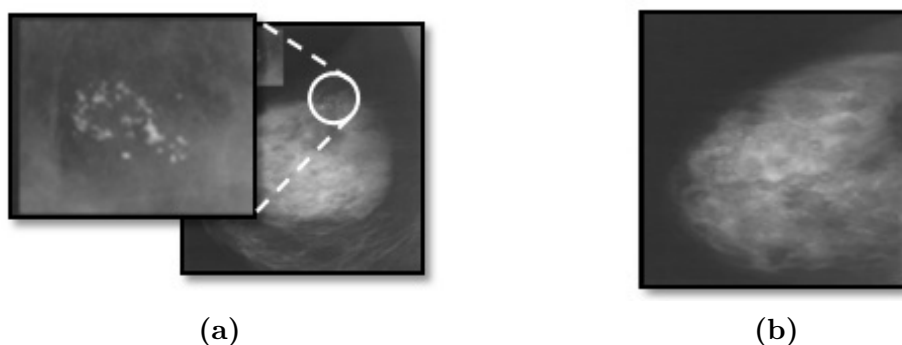


Figure 1.12 – Exemples de microcalcifications malignes (a) et bénignes (b)

1.5.2 Distribution

La distribution des MCs (Figure 1.13) décrit leur répartition géométrique sur une image mammographique. Le processus pathologique à l'origine de leur production va conditionner cette répartition qui constitue un élément indispensable dans l'association d'un degré de malignité. A titre d'exemple, un foyer de MCs est dit malin s'il regroupe un grand nombre de MCs qui sont situées étroitement les unes des autres. Tandis qu'un foyer bénin englobe un petit nombre de MCs éparpillées [3]. Ainsi, une analyse robuste des MCs doit prendre en considération la forme en plus du nombre et des distances géométriques entre les différentes calcifications.

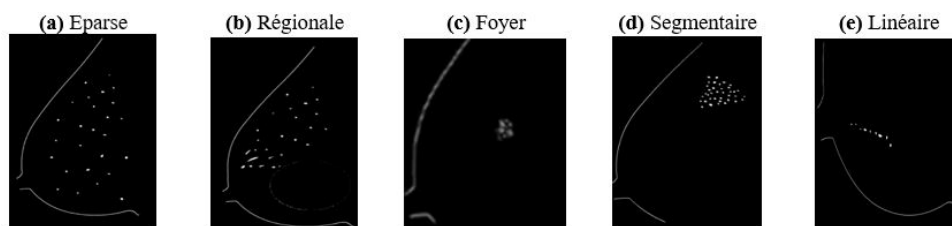


Figure 1.13 – Distributions possibles des MCs

1.5.3 Morphologie

La morphologie des MCs concerne leurs formes d'apparition possibles sur une image mammographique et est généralement liée à l'emplacement de formation et d'accumulation. La figure 1.14 résume les différents types de MCs, distinguées par la classification BI-RADS et issues de l'analyse de leurs morphologies.

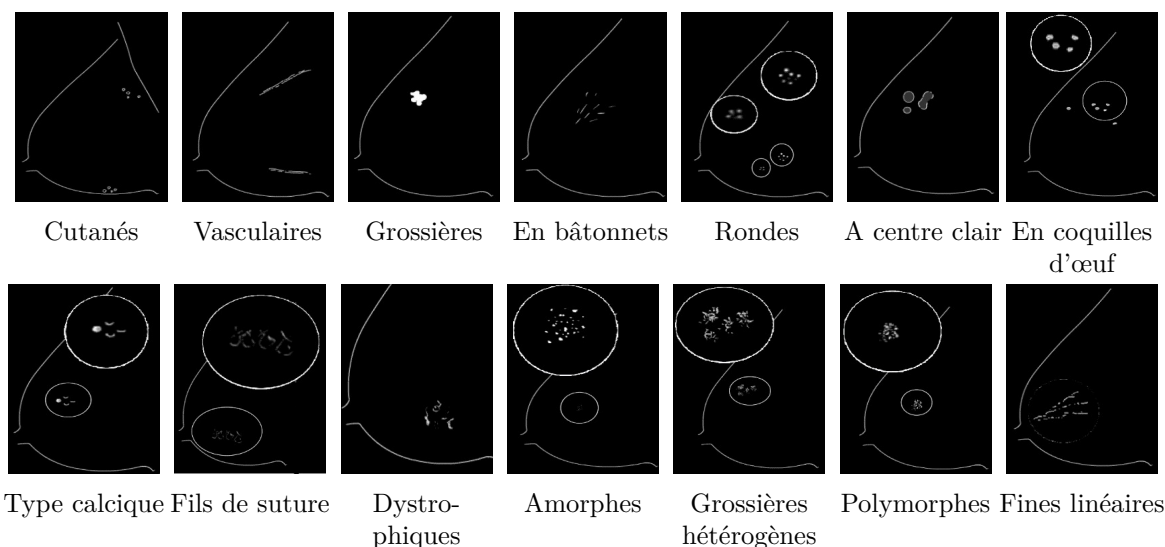
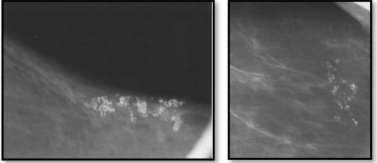
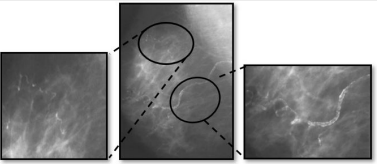
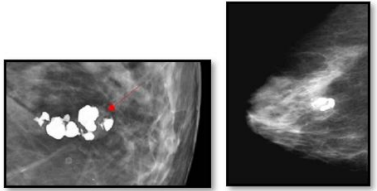
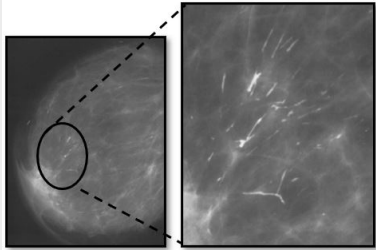
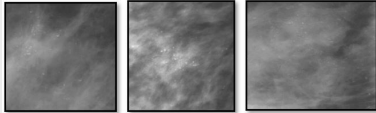
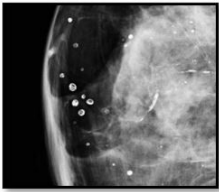
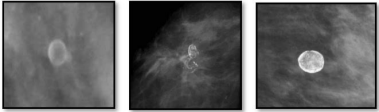


Figure 1.14 – Les différents types de morphologies distingués par la classification de BI-RADS

Le tableau 1.4 détaille les propriétés anatomiques des différents types de MCs précédemment présentés et donne des exemples pour chacun.

	Description anatomique et mammographique	Exemples
<i>Cutanées</i>	<ul style="list-style-type: none"> • Calcifications avec des formes rondes ou annulaires, • Des incidences tangentielles à la peau sont généralement demandées pour une meilleure visualisation de son origine cutané. 	
<i>Vasculaires</i>	<ul style="list-style-type: none"> • Calcifications en rails ou linéaires, • Constituent une observation normale après la ménopause, • Sont associées aux structures tubulaires, • Apparaissent comme des lignes hyperdenses. 	
<i>Grossières</i>	<ul style="list-style-type: none"> • Sont aussi appelées : 'Popcorn-like', • Sont considérées comme des calcifications bénignes dues à un phénomène de fibroadénome, • Apparaissent avec des intensités assez élevées et de grande taille (supérieure à 2/3mm). 	

En bâtonnets	<ul style="list-style-type: none"> • Sont aussi appelées des 'calcifications sécrétoires', • Sont associées à une ectasie canalaire sécrétante, • Sont produites dans les canaux galactophores en direction vers le mamelon, • Apparaissent souvent alignées et occasionnellement branchées, généralement pleines, mais peuvent parfois être à centre clair, • Diamètre supérieur à 1mm. 	
Rondes	<ul style="list-style-type: none"> • Reflètent des dépôts calciques dans les acini lobulaires (lobes), • Présentent des formes rondes ou ovales avec des contours bien définis et sont souvent regroupées en 'foyers', • Présentent des tailles variables (elles sont dites <i>ponctiformes</i> si la taille est $< 0.5\text{mm}$). 	
A centre clair	<ul style="list-style-type: none"> • Correspondent à des calcifications de cytotéatonecrose, ou des débris calciques canaux,aires, • Sont généralement rondes ou ovales, • Ont une paroi plus épaisse que celle des calcifications en coquille d'oeuf, • Présentent des tailles variables (du mm vers cm). 	

<i>En coquille d'oeuf</i>	<ul style="list-style-type: none"> • Dites aussi 'pariétales', • Correspondent à des calcifications de cytotéatonecrose et des calcifications des parois des kystes, • Sont très fines ayant l'apparence d'un dépôt calcique sur une surface d'une sphère. 
<i>A type de lait calcique</i>	<ul style="list-style-type: none"> • Correspondent à des sédimentations déclives des sécrétions dans des formes kystiques, • Apparaissent différemment sur les incidences mammaires : <ul style="list-style-type: none"> • Dépôts amorphes avec des limites floues sur les incidences de face, • Nettes semi-lunaires en croissant, curvilignes à concavité supérieure ou linéaires dessinant une partie déclive des kystes sur les incidences de profil. 
<i>Fils de suture</i>	<ul style="list-style-type: none"> • Correspondent à des dépôts calciques sur du matériel de suture, • Elles sont fréquentes dans le sein irradié, • Présentent des formes linéaires suivant le trajet des fils de sutures. 

<i>Dystrophiques</i>	<ul style="list-style-type: none"> • Apparaissent dans des zones irradiées ou après un traumatisme mammaire, • Apparaissent sous plusieurs formes (irrégulières, grossières) avec des tailles variantes (supérieure à 5mm), • Des calcifications à centre clair ou en coquille d'oeuf sont fréquemment associées. 	
<i>Amorphes</i>	<ul style="list-style-type: none"> • Dites aussi : 'indistinctes ou poussiéreuses', • Calcifications très fines, • Peuvent être éparses, groupées en foyer ou ayant une distribution linéaire ou segmentaire. 	
<i>Grossières hétérogènes</i>	<ul style="list-style-type: none"> • Sont irrégulières et variables en taille et en forme (taille > 0.5mm), • Se distinguent des calcifications dystrophiques par leur taille plus petite, • Apparaissent : <ul style="list-style-type: none"> • Sous forme de regroupements bilatéraux multiples, • Avec une distribution segmentaire ou linéaire, • En foyers. 	

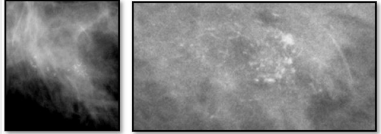
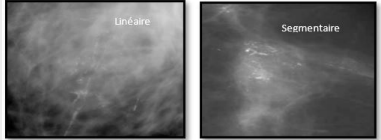
<i>Polymorphes</i>	<ul style="list-style-type: none"> • Présentent des contours bien distincts, • Plus visibles que les calcifications amorphes, • Présentent des tailles ($<0.5\text{mm}$) et des formes variables. 	
<i>Fines linéaires</i>	<ul style="list-style-type: none"> • Dites aussi : 'fines linéaires ramifiées', • Correspondent à des calcifications de type canalaire, • Se présentent sous formes très fines ou curvilignes irrégulières et parfois discontinues, • Peuvent présenter une distribution segmentaire, • Présentent une taille $<0.5\text{mm}$. 	

Tableau 1.4 – Projection des différentes descriptions morphologiques des MCs sur une image mammographique

1.6 Bases de mammographies numériques

1.6.1 Exigences de création

Dans l'objectif d'étudier la fiabilité de leurs systèmes d'aide à la décision, les chercheurs utilisent des bases d'images dites bases d'évaluation. Une base d'évaluation, afin de pouvoir être utilisée dans la validation de tels systèmes, doit répondre à certaines exigences relatives à sa création, dont quelques-unes sont communes et d'autres sont spécifiques au contexte d'application. Dans le contexte de détection et du diagnostic du cancer du sein, les bases utilisées sont souvent des bases de mammographies numériques et qui doivent garantir :

- une exhaustivité des cas qu'elles présentent ;
- un respect des standards de représentation des données médicales ;
- une disponibilité des vérités terrain associées.

1.6.1.1 Exhaustivité des cas

L'exhaustivité des cas exige que les mammographies numériques dans la base présentent tous les cas possibles à savoir des cas normaux et d'autres pathologiques. Les cas pathologiques doivent à leur tour être capables de décrire l'ensemble de toutes les anomalies possibles quelque soit le type de densité mammaire. En outre, les incidences de base pour un examen de mammographie (MLO et CC) sont aussi demandées pour chaque cas.

1.6.1.2 Représentation des données

Une base de mammographies est un ensemble d'images numériques avec un fichier descriptif associé. La représentation des données concerne les informations communiquées, dans ce fichier associé, et qui peuvent s'avérer utiles pour l'amélioration des images et/ou du diagnostic. Elles concernent essentiellement les paramètres d'acquisition et les autres informations relatives aux propriétés de densité mammaire. Les images sont, généralement, sauvegardées au format DICOM (Digital Imaging and Communications in Medicine) et elles représentent le résultat de la projection de l'organe mammaire tridimensionnel sur un espace bi-dimensionnel.

1.6.1.3 Vérité terrain

La vérité terrain est l'élément indispensable pour n'importe quelle base. C'est une information fournie par les radiologues experts. Elle peut être différente d'une base à une autre et d'un système à un autre. Dans le contexte des mammographies numériques, elle peut passer d'une simple localisation d'une lésion, par désignation de son centre de

gravité, à une délinéation exacte. L'objectif de base du système conçu exige le type de vérité terrain requis.

1.6.2 Bases de mammographies

Un nombre important de bases de mammographies numériques sont exploitées par les chercheurs afin de concevoir leurs systèmes d'aide à la décision. Les bases les plus répandues et les plus utilisées ces dernières années sont : la base Mini-MIAS [36], la base DDSM [37] et la base INBreast [38].

1.6.2.1 Mammographic Image Analysis Society Digital Mammogram Database

Mini-MIAS [36] est la base publique de mammographies numériques la plus ancienne. Elle a été fournie en 1994 par la société savante anglaise MIAS (*Mammographic Image Analysis Society*). Elle comporte en total 322 images obtenues en incidence MLO pour 161 cas dont certaines sont pathologiques et d'autres normaux. Les pathologies représentées dans cette base décrivent également l'ensemble de toutes les lésions possibles. Malgré son ancienneté, cette base continue d'être utilisée dans la littérature jusqu'à présent.

Les défauts majeurs de la base Mini-MIAS résident dans la faible résolution de ses images et l'absence de vérités terrains. En effet, la résolution des images est de $50\mu\text{m}$ et ne convient pas à la détection des MCs [38]. D'autre part, les annotations des lésions présentes sont faites pour les masses uniquement et elles n'indiquent que les centres et les rayons des régions d'intérêts au sein desquelles elles se situent.

1.6.2.2 Digital Database for Screening Mammography

Digital Database for Screening Mammography (DDSM) [37] est la plus grande base de mammographies numériques publique. Elle comporte en total 2620 cas soient 10480 images en incidences MLO et CC. Elle a été élaborée aux Etats-Unis (USA) en 1999.

Contrairement à la base Mini-MIAS, la résolution des images dans cette base varie entre $42.5\mu\text{m}$ et $200\mu\text{m}$ qui permet la détection des MCs. En revanche, les annotations qu'elle présente, constituent des délinéations manuelles un peu grossières des lésions de type masse mais pas des MCs.

1.6.2.3 INBreast Database

A notre connaissance, INBreast [38] est la base de mammographies numériques la plus récente. Elle date de 2010 et comporte un total de 115 cas (410 images d'incidence MLO et CC) dont 90 proviennent de femmes ayant les deux seins (4 images par cas) et 25 cas de patientes mastectomisées (2 images par cas). Les images de cette base présentent des

résolutions qui varient entre 50 et 70 μ m et sont acquises dans un centre du sein, situé dans un hôpital universitaire en Portugal. Les divers types de lésions sont inclus dans cette base ainsi que les différentes classes de densité. L'annotation par des contours précis réalisés par des spécialistes est également fournie pour un seul type de lésions qui sont les masses.

La figure 1.15 illustre une sélection des images des trois bases que nous venons de présenter. Le tableau 1.5 résume les différentes caractéristiques qu'elles présentent.

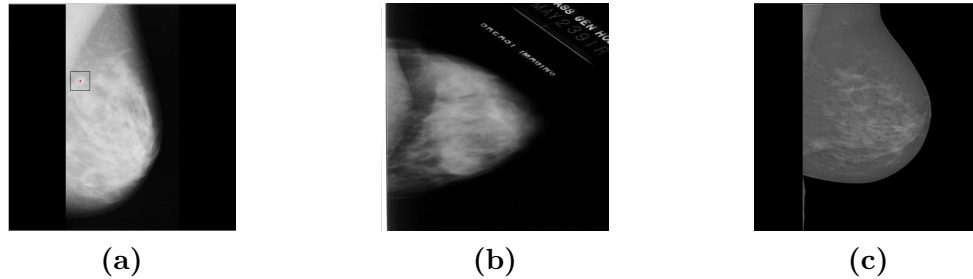


Figure 1.15 – Exemples de mammographies numériques de la base : (a) Mini-MIAS, (b) DDSM et (c) INBreast

Tableau 1.5 – Synthèse des caractéristiques des images dans les bases de mammographies numériques

Information	Mini-MIAS	DDSM	INBreast
Origine	Royaume-Uni	Etats-Unis	Portugal
Année	1994	1999	2010
Nombre de cas	320	2620	115
Types des incidences	MLO	MLO et CC	MLO et CC
Types d'images	PGM	LJPEG	DICOM
Résolution	50 μ m	42.5-200 μ m	50-70 μ m
Résolution	8 bits/pixel	8 et 16 bits/pixel	14 bits/pixel
Types des lésions	Tous	Tous	Tous
Vérité terrain	Centre de gravité (Masses)	Contours grossiers (Masses)	Contours précis (Masses)
Classification BI-RADS	Non	Oui	Oui
Densité	ACR	ACR	ACR
Accès	Public	Public	Public

1.6.3 Base de mammographies utilisée

Suite à la synthèse présentée ci-dessus, nous constatons que toutes les bases de mammographies présentes et disponibles à ce jour ont comme point commun l'absence de vérités terrain pour les lésions de type MCs. Seules des annotations, parfois grossières et parfois précises, sont fournies pour les lésions de type masse avec les bases DDSM et INBreast. En outre, ces deux dernières, présentent un nombre suffisant de cas avec des MCs, en les comparant à ceux dans la base Mini-MIAS (25 cas uniquement). L'évaluation d'un système est ainsi faisable à l'aide de ces deux bases à condition de concevoir des vérités de terrain. En revanche, la qualité des images de la base DDSM est considérée moyenne par rapport à celle de la base INBreast. En effet, ces images montrent plusieurs types d'artefacts, dus à des problèmes techniques provoquant une dégradation de la qualité (mauvaise apparition des tissus, présence d'artefacts, ...).

Dans ce travail, les différentes évaluations ont été menées sur les images de la base INBreast. Cette base comporte 90 cas avec des lésions de types calcifications mammaires (macro et micro-calcifications). A partir de ces dernières, nous avons effectué une sélection de 50 images avec uniquement des MCs. Nous n'avons pas considéré les images avec des macrocalcifications ni celles présentant un faible nombre de MCs (<5 MCs) claires avec des contours nets. Les divers types de MCs possibles au sein du tissu mammaire apparaissent dans cette sélection. Le tableau 1.6 présente une synthèse des critères sur la base et à partir desquels nous avons classés les MCs présentes dans les images traitées. Bien que le nombre des images utilisées pour l'évaluation soit petit, le nombre de MCs qu'elles présentent est très important. En effet, le nombre de MCs par image varie entre 5 et 1261 MCs. Le nombre moyen est de l'ordre de 125 MC par image. Le nombre total de MCs, sur l'ensemble de toutes les images considérées est de l'ordre de 6281 MCs.

Tableau 1.6 – Synthèse des caractéristiques des MCs dans la sélection des images utilisées pour l'évaluation

Contraste	Taille	Tissue(s) de superposition
Très faible	Très faible	Tissu nodulaire
Moyen	Variable	Tissu linéaire
Variable	Variable	Tissu homogène dense
Élevé	Variable	Variable

1.6.4 Génération de la vérité de terrain

Du fait que notre attention soit focalisée sur l'identification et la bonne délimitation des MCs, une vérité terrain avec des contours précis est nécessaire. Avec l'absence d'une segmentation manuelle précise dans toutes les bases existantes, nous nous sommes trouvés face à l'obligation d'en créer une. Pour concevoir cette vérité de terrain, nous avons commencé par réaliser des segmentations manuelles sur chaque image pour dessiner les contours exacts autour de chaque MC. Le résultat final est un masque binaire, par image, qui regroupe l'ensemble de toutes les MCs y présentes. L'ensemble des 50 masques générés est, dans un deuxième temps, analysé et validé par Dr. Isabelle KARGASTEL THOMAS, experte radiologiste en imagerie mammographique au Centre Hospitalier Universitaire de Brest (CHRU de Brest). La figure 1.16 suivante montre l'exemple d'un masque généré à partir d'une image mammographique avec un zoom sur la région qui regroupe les MCs (rectangle de contour rouge).

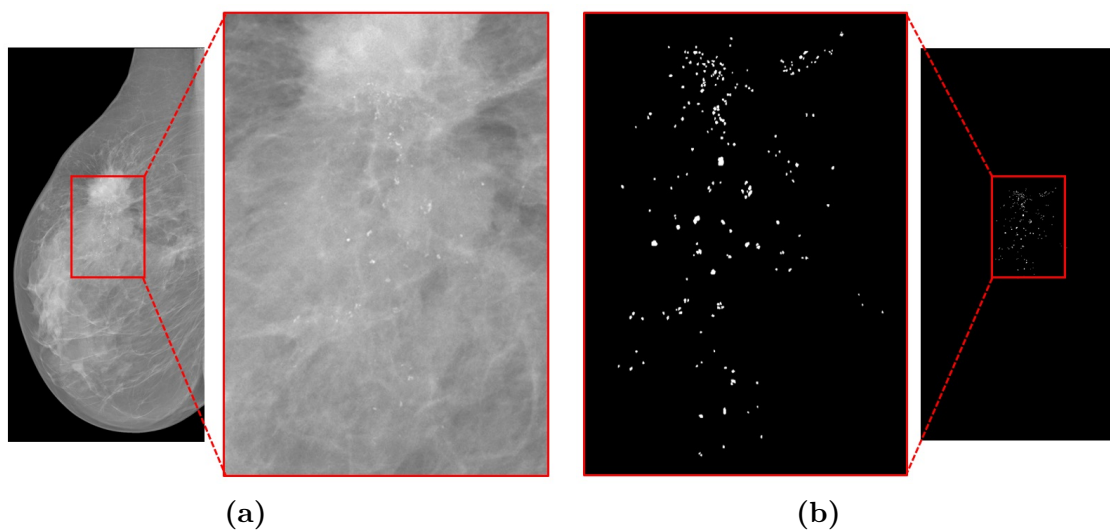


Figure 1.16 – Exemple d'une image mammographique (a) et son masque de vérité de terrain créé (b)

1.7 Conclusion

Ce premier chapitre a été consacré à la présentation du cadre d'étude de ce travail de recherche, à savoir le cancer du sein et les lésions cancéreuses qu'il peut produire. Après avoir présenté l'anatomie du sein, nous avons exposé les différentes lésions cancéreuses détectables à partir des images mammographiques. Par la suite, nous avons étudié les propriétés anatomiques et d'apparition des microcalcifications, lésions auxquelles nous nous sommes intéressés dans ce travail de recherche. Cette étude avait pour but de mettre en avant les problématiques relatives à la bonne détection et délimitation des microcalcifi-

cations à partir des images mammographiques et qui sont dues essentiellement à la variété des caractéristiques anatomiques et d'apparition de ces dernières. Nous avons clôturé ce chapitre par une synthèse des différentes bases d'évaluation existantes. Nous avons également présenté notre base conçue et expliqué le processus suivi pour l'obtention des vérités de terrain.

La problématique de la bonne détection des microcalcifications sera détaillée dans le chapitre suivant où nous allons présenter une étude approfondie des solutions proposées dans la littérature pour la résoudre. Nous allons également discuter des avantages et des inconvénients de ces méthodes pour enfin proposer et justifier le schéma de la démarche proposée.

TRAITEMENT D'IMAGES POUR LA DÉTECTION DES MICROCALCIFICATIONS

Sommaire

2.1	Introduction	33
2.2	Problématique de détection des MCs	33
2.2.1	Intérêts de détection	33
2.2.2	Enjeux de détection	34
2.3	Détection des microcalcifications dans la littérature	34
2.3.1	Segmentation fréquentielle	36
2.3.2	Segmentation basée morphologie	38
2.3.3	Segmentation basée régions	40
2.3.4	Segmentation basée partitionnement	41
2.3.5	Segmentation à base de filtres	44
2.3.6	Étude comparative	46
2.4	Prétraitement des images mammographiques	48
2.4.1	Qualité et nécessité de prétraitement	48
2.4.2	Techniques de prétraitement standard	48
2.4.3	Prétraitement en mammographie: Synthèse et discussion	53
2.5	Approche proposée	54
2.5.1	Analyse critique	54
2.5.2	Approche proposée	56
2.6	Conclusion	59

2.1 Introduction

L'image mammographique est l'outil indispensable pour le diagnostic du cancer du sein en se basant sur une recherche de constituants anatomiques et d'anomalies. L'efficacité de ce diagnostic est cruciale pour les radiologues et elle dépend de deux facteurs : la précocité et la précision. En effet, plus le diagnostic est précoce, plus la probabilité de survie est élevée [39]. En outre, plus la délimitation des anomalies est précise, plus la probabilité de réussite des traitements est élevée [40].

Actuellement, les travaux de recherche se dirigent vers le développement de systèmes d'aide au diagnostic. Ces systèmes interprètent l'image dans le but de distinguer les régions saines des régions présentant des anomalies potentielles et donc de faciliter le travail fastidieux de l'expert. Pour aboutir à cet objectif, ces systèmes se basent sur des techniques de traitement d'images ainsi que sur des connaissances de domaine afin de prendre une décision.

Le travail de recherche mené dans le cadre de cette thèse et son application visent à concevoir un système d'interprétation d'images mammographiques en simulant le raisonnement d'un expert dans la recherche des MCs. Pour ce faire, nous commençons ce chapitre par l'explication de l'intérêt ainsi que des problématiques relatives à l'interprétation d'images mammographiques dans l'objectif de détecter les MCs. Nous poursuivons par la présentation des techniques et approches proposées dans la littérature pour la résolution de cette problématique avant de discuter de leurs avantages et inconvénients. Nous passons aussi en revues les techniques de prétraitement utilisées. Suite à cette analyse, nous présentons dans la dernière partie de ce chapitre l'idée générale de l'approche proposée.

2.2 Problématique de détection des MCs

2.2.1 Intérêts de détection

A travers les éléments discutés dans le chapitre précédent, nous nous rendons compte de l'importance des systèmes d'aides à la décision en tant que piste pour l'amélioration de l'efficacité du diagnostic. En revanche, la plupart des travaux menés dans cet axe se sont intéressés à la recherche et l'interprétation des lésions de type masse [30, 41] ou distorsion architecturale [35, 42]. Ils ont accordé très peu d'intérêt à la recherche des MCs malgré leur utilité sur le plan du diagnostic du cancer du sein. En particulier, 55% des maladies mammaires avec lésions tumorales s'accompagnent par la présence de MCs apparaissant

sous la forme de petites 'tâches' minuscules, regroupées en foyers. Les recherches effectuées ont, aussi, démontré qu'environ 50% de cancers non-palpables et 93% du carcinome canalaire *in situ* sont diagnostiqués par l'apparition des MCs sur une mammographie [43].

2.2.2 Enjeux de détection

Les MCs figurent parmi les principaux signes indirects précoces du cancer du sein visibles à la mammographie. Le diagnostic et la détection de ces lésions constituent non seulement un besoin, mais également un défi pour les radiologues [44]. En effet, ces tâches de très petite dimension peuvent être localisées ou largement diffusées le long des canaux mammaires, dispersées ou groupées dans tout le tissu mammaire. La variété anatomique des tissus mammaires justifie la variété de caractéristiques d'apparition de ce type de lésions généralement liées à l'emplacement de formation.

De plus, la densité mammaire est l'un des facteurs provoquant la sensibilité de l'examen mammographique et donc de la mammographie numérique générée. En effet, un sein avec un pourcentage élevé de tissu fibroglandulaire (ACR3 ou ACR4) augmente la complexité d'interprétation de la mammographie et peut dissimuler la présence de MCs et/ou masses de petite taille [45].

L'identification des MCs est donc une tâche difficile, non seulement en raison de la grande variété de composition mammaire et de l'anatomie mammaire très texturée, mais aussi en raison du faible contraste inhérent aux mammographies et de la taille très petite de ces lésions, que l'on peut confondre avec des pixels bruits ou des pixels des tissus voisins.

2.3 Détection des microcalcifications dans la littérature

L'identification de la présence des lésions cancéreuses est la tâche pour laquelle les radiologues analysent et interprètent les mammographies numériques. La première solution adoptée, par les radiologues, pour la réduction des diagnostics erronés consiste à multiplier le nombre de lecture d'une même mammographie. En revanche, les conclusions d'études récentes, ont montré que cette première piste apparaît inefficace dans le contexte du diagnostic du cancer du sein [46].

Avec l'apparition des CADs, non seulement, les erreurs de détection et de diagnostic ont été efficacement réduites, mais aussi la charge de travail des radiologues avec

l'augmentation du nombre des examens effectués [47]. L'étape qui cherche à identifier la présence des lésions cancéreuses à partir d'une mammographie par ces systèmes s'appelle *détection*. Elle regroupe l'ensemble de techniques de traitement d'images dont l'objectif est de séparer le fond de l'image des objets d'intérêts qu'elle présente. Cette étape joue un rôle central dans la fiabilité du système conçu dans le sens qu'elle permet de localiser et segmenter l'ensemble des MCs et donc de faciliter leur interprétation et classification.

Cette problématique a fait l'objectif de plusieurs travaux de recherche qui se sont basés sur diverses méthodes de segmentation. La classification de ces méthodes peut être réalisée selon deux critères : le type et le niveau de l'analyse.

En fonction du type d'analyse, les méthodes proposées sont regroupées en deux catégories : les méthodes supervisées et les méthodes non-supervisées. La principale différence réside dans le fait que l'analyse supervisée se fait sur la base d'algorithmes entraînés sur un ensemble de données étiquetées et qui sont appelés des classifieurs. L'objectif de ces derniers est d'apprendre une fonction qui reproduit le mieux de la relation entre les entrées et les sorties à partir d'un échantillon de données et des résultats souhaités. En revanche, les algorithmes d'analyse non supervisée tentent à extraire les structures cachées des données non initialement étiquetées. Les méthodes non supervisées sont fondées sur des concepts de segmentation standards, tels que la morphologie mathématique, contours actifs, partitionnement, ..., pour la détection des MCs. Chacun de ces concepts présente un point fort en rapport avec des caractéristiques des MCs. Contrairement aux méthodes non-supervisées, les méthodes supervisées traitent le problème de détection des MCs en utilisant des techniques d'apprentissage automatique [43, 48–54] et profond [55–63]. Elles sont basées sur des classifieurs entraînés et un ensemble de descripteurs pour caractériser les MCs [64].

Dans ce travail de recherche, nous allons nous intéresser aux méthodes de détection non-supervisées que nous présenterons dans les sections suivantes un état d'art. Bien que les méthodes supervisées constituent une grande avancée ces dernières années, celles-ci nécessitent une grande quantité de données pour estimer le modèle robuste à utiliser. D'autre part, face à la grande diversité en termes de formes et de tailles des MCs, un nombre important de données annotées sera requis. Par conséquent, la qualité de la segmentation des contours fournis par ces méthodes n'est pas garantie et ne permet, donc, pas d'apporter une aide dans l'étude de la nature d'une MC. Bien évidemment, la localisation seule est, de son côté, insuffisante pour un diagnostic fiable.

En fonction du niveau d'analyse, les méthodes existantes sont regroupées en deux familles : les méthodes globales [8, 60, 65–68] et les méthodes locales [3, 53, 69]. Les méthodes globales utilisent l'image mammographique entière comme une région de travail pour la

recherche des MCs tandis que les méthodes locales se focalisent sur des régions d'intérêts (ROI) extraites à partir de l'image. Quelque soit leur niveau d'analyse, les méthodes existantes s'appuient sur des raisonnements spécifiques, qui explorent nécessairement une ou plusieurs propriétés des MCs, afin de détecter les groupements de pixels adjacents les constituants. Les points communs entre ces deux familles d'approches sont :

- L'exploitation des techniques de détection qui opèrent sur le plan de représentation pixelique pour la détection des MCs ;
- Le suivi d'une recherche basée sur la caractérisation des MCs à partir des valeurs numériques de l'image mammographique.

Dans ce cadre, on peut distinguer cinq types de techniques explorées dans la littérature pour la détection des MCs :

- Des techniques de segmentation fréquentielle ;
- Des techniques de segmentation basées morphologie ;
- Des techniques de segmentation basée régions ;
- Des techniques de segmentation basée partitionnement ;
- Des techniques de segmentation basée filtrage.

Une présentation de leurs principes et quelques travaux associés, sont présentées dans les sections suivantes.

2.3.1 Segmentation fréquentielle

La mammographie numérique est une image fortement texturée. La segmentation basée sur une analyse multi-échelle à base d'ondelettes a montré une grande efficacité dans le contexte de détection des MCs [68]. Les travaux de recherche qui ont exploité la notion d'ondelettes, l'ont intégré aussi bien dans l'étape de détection que dans une étape de pré ou de post-détection. Citons à titre d'exemple les travaux de Mencattini *et al.* [70] et Rizzi *et al.* [71] où les transformations d'ondelettes ont été utilisées pour réduire le bruit dans les images avant de procéder à la segmentation. Dans le même ordre d'idées, Balakumaran *et al.* [72] ont appliqué les ondelettes de dyadic avant d'utiliser l'algorithme Fuzzy Shell Clustering (FSC) pour la segmentation des MCs. FSC est un algorithme de segmentation itératif introduit par N. Davé en 1990 [73]. Il permet d'identifier les objets avec des structures nodulaires à partir d'une image. L'utilisation des ondelettes, dans ce travail, avait l'objectif de déterminer les zones probablement suspectes ou les régions d'intérêts qui peuvent présenter des MCs.

Avec un raisonnement différent, l'utilisation des ondelettes a constitué, dans certains autres travaux, une étape intermédiaire dans le schéma de détection complet. A titre

d'exemple, Malar *et al.* [74] ont appliqué les ondelettes de Haar afin d'extraire des caractéristiques de texture à utiliser dans des étapes de segmentation et classification suivantes.

Une autre piste consistait à faire intervenir les transformations d'ondelettes dans plusieurs étapes du schéma de détection. Citons l'exemple de Hamad *et al.* [75] qui ont proposé une méthode de détection des MCs à base de la transformation en ondelettes discrètes (DWT). Cette technique s'effectue en trois principales étapes : le prétraitement, l'extraction et le raffinement. La première étape rehausse les ROIs sélectionnées par des experts. Par la suite, six niveaux différents de transformations en ondelettes (1D) sont appliqués sur ces régions. A partir des résultats obtenus, le meilleur niveau de transformation est identifié (le dernier niveau à partir duquel les MCs disparaissent) et l'image est reconstruite. Par la suite, les auteurs appliquent, à nouveau, une transformation en ondelettes 2D avec le même niveau optimal retenu. Une fois la décomposition effectuée, les coefficients désignant des pixels du fond sont remis à zéro et l'image est reconstruite. Cette nouvelle image présente une zone noir désignant le fond avec des petites zones blanches désignant les MCs. Hamad *et al.* dans cette reconstruction se sont basés sur le fait que les MCs sont les zones denses qui projettent des fréquences très élevées par rapport au reste de l'image. Cependant, sur les zones les plus denses du sein, les annotations ainsi que les vaisseaux sanguins présentent eux mêmes des fréquences très élevées vues qu'elles se caractérisent par des valeurs d'intensités très élevées (proches de celles des MCs). Ces zones peuvent donc persister avec les MCs suite à l'application de la transformation en ondelettes et l'image résultat devient donc une image bruitée qui contient des fausses MCs. Une étape de raffinement est ainsi appliquée et est basée sur des opérations morphologiques.

En tant qu'étape principale pour la segmentation des MCs, les travaux décrits dans [76] et [77] sont fondés sur les transformations en ondelettes discrètes (DWT). Avec un raisonnement similaire, Lakshmanan *et al.* dans [78] et [79], Vigneshwaran *et al.* dans [80] ainsi que Guo *et al.* dans [81] ont utilisé les *ContourLet* (CT) pour la segmentation des MCs. La fiabilité des résultats obtenus a montré l'efficacité des analyses multi-échelles dans la recherche des MCs [81]. En contre partie, la partie expérimentale, présentée dans [80] a montré une supériorité, en termes de fiabilité des résultats, des transformations en CT par rapport à DWT. Dans le même contexte, Alarcon-Aquino *et al.* [82] en 2009 ont exploité les transformées d'ondelettes complexes (*Complex Wavelet Transform* : DT-CWT) et non pas la transformée en ondelettes classique. L'utilisation de cette dernière est, d'une part, argumentée par sa capacité de produire des informations relatives aux phases des données, chose qui n'est pas possible par la transformée en ondelettes classique, aussi bien que sa capacité de traiter, contrairement à DWT, les cas d'invariance de décalage et de directionnalité [82]. Dans l'objectif de sélectionner les premières zones de MCs potentielles, les auteurs ont appliqué un seuillage adaptatif sur l'image retenue suite à l'application des

ondelettes. Conformément au raisonnement de Hamad *et al.* [75], une opération morphologique est ensuite appliquée dans le but d'améliorer les contours des objets dans l'image résultat. La reconstruction de l'image mammographique, par l'application de la transformation inverse de CWT résulte en l'image mammographique d'origine regroupant les objets correspondant aux MCs.

2.3.2 Segmentation basée morphologie

La morphologie mathématique est une théorie non linéaire fondée sur des concepts mathématiques simples issus de la théorie des ensembles [83]. Elle permet l'analyse de structures géométriques et est très largement utilisée en traitement d'images servant à résoudre des problèmes aussi bien de filtrage que de segmentation et classification. Appliquée sur une image, son objectif principal est d'identifier des objets à partir des formes prédéfinies appelées éléments structurants. Un élément structurant est un masque, de taille et de forme quelconques, qui décrit le voisinage d'un pixel central à utiliser dans le traitement. Il est accompagné par une opération morphologique qui est appliquée sur l'image afin de préserver et d'améliorer l'apparence des objets ayant un profil similaire à celui de l'élément structurant défini. L'*érosion* et la *dilatation* sont les opérations de base de la morphologie mathématique. Pratiquement toutes les autres opérations (*tophat*, *ouverture*, *fermeture*, ...) sont déduites à partir des compositions de celles-ci.

Désignant par I une image en niveaux de gris et par E un élément structurant, l'*érosion*, \otimes , respectivement la *dilatation*, \oplus , de l'image I par E sont respectivement exprimées par les équations (3.1) et (3.2) :

$$(I \otimes E)(P) = \min\{I(p_i), p_i \in E_P\} \quad (2.1)$$

$$(I \oplus E)(P) = \max\{I(p_i), p_i \in E_P\} \quad (2.2)$$

avec :

- "max" et "min" désigne respectivement le *maximum* et le *minimum* ;
- E_P réfère à l'élément structurant de centre le pixel P .

On trouve la morphologie mathématique à la base d'une diversité de travaux de recherche pour la détection des MCs. L'idée de base part du fait que ces lésions constituent de petites tâches, dans la mammographie, de forme proche du circulaire et peu contrastées par rapport aux tissus sains environnants [84]. Un élément structurant de forme circulaire et de taille spécifique pourrait alors être capable d'ajuster l'apparition de ces lésions à partir des mammographies [85] et donc faciliter la détection. Le travail de Zhang *et al.* [2] est basé sur la combinaison des résultats de deux opérations morphologiques pour détecter

les MCs à partir d'une région d'intérêt sélectionnée par un radiologue. Les opérations morphologiques utilisées associent un effet flou aux régions homogènes mais elles permettent de retourner celles dont le profile correspond à celui des éléments structurants utilisés. Dans l'objectif de réduire la sensibilité à la taille variable des MCs, les auteurs ont utilisé deux types d'éléments structurants. Les résultats obtenus montrent une bonne détection des MCs même avec des tailles variables. D'autres pistes ont été suivies dans le même objectif et ont consisté à exploiter la notion des maxima étendus avec les opérations morphologiques pour la délinéation des contours des MCs tels que dans [86]. La comparaison de certains travaux qui ont suivi une démarche similaire est détaillée dans [87].

Autre que les opérations morphologiques, on trouve la segmentation par *Ligne de partage des eaux (LPE)* [88] qui désigne une famille de méthodes de segmentation d'images issues de la morphologie mathématique. Elle considère une image de niveaux de gris comme un relief topographique dont on cherche à simuler son inondation. Betal *et al.* [89], à la fin des années 90, ont proposé une approche fondée sur la morphologie mathématique suivie par l'algorithme LPE pour la délinéation des contours exacts des MCs. L'opération morphologique *Tophat* par reconstruction est initialement appliquée, ce qui permet la distinction des MCs du tissu mammaire sain. L'application de l'algorithme LPE, en deuxième lieu, fournit une méthode objective pour la segmentation des MCs initialement identifiées. Les mêmes concepts sont exploités en 2016 par Ciecholewski *et al.* [69] sur des régions d'intérêts de taille 512×512 . Ce travail fait appel à d'autres transformations de gradient de l'image subissant une segmentation LPE, ce qui permet de renforcer la qualité d'apparition des contours sur l'image.

Un raisonnement similaire a été mené par Bougioukos *et al.* [90] qui ont utilisé la technique *Frequency Histogram of Connected Elements (FHCE)* pour la segmentation des MCs à partir des ROIs initialement sélectionnées par les radiologues. La technique FHCE [91] exploite des opérateurs morphologiques pour chercher les éléments connexes à partir d'une image. Elle se base sur l'hypothèse qu'un pixel a une grande probabilité d'être dans la même classe que ses voisins. Pour ce faire, elle utilise un masque de parcours, un opérateur morphologique et un seuil de séparation déduit à partir d'un algorithme de clustering. La technique FHCE appliquée divise la ROI en deux classes. La première classe correspond aux différents pixels connectés désignant des MCs et la deuxième correspond aux pixels du fond.

Dans divers autres travaux, la morphologie mathématique n'était pas le traitement principal pour la détection des MCs, mais un outil de base pour l'amélioration de la qualité de leur apparition. Nous pouvons citer à titre d'exemple, les travaux de Quintanilla *et al.* dans [92] et [93] et les travaux de Vivona *et al.* dans [1]. Dans le même ordre d'idées, les

opérations morphologiques ont été exploitées comme une étape de posttraitement après la détection. A titre d'exemple, Suhail *et al.* [94] ont utilisé l'opération morphologique Tophat avec un seuillage adaptatif dans le but de réduire les fausses détections issues du résultat d'une étape de détection initialement appliquée. Une fausse détection était définie comme étant l'ensemble de pixels adjacents appartenant à un tissu mammaire sain mais considérés comme pixels référant à une MC.

2.3.3 Segmentation basée régions

Les méthodes de segmentation basées régions sont à la base d'une analyse locale du voisinage d'un point de départ. Le principe est de faire évoluer ce point de départ vers les contours précis de la région décrivant l'objet ou la forme recherchée à partir d'une image donnée. En plus de l'analyse spatiale pour la prise de décision, ces méthodes partagent la notion de points de départ. En revanche, elles se distinguent par la démarche suivie pour l'évolution de ces points. En fonction de cette démarche, nous pouvons distinguer deux familles de méthodes : Les méthodes à base de *Croissance de Régions* et les méthodes à base de *Contours actifs* où les points de départ associés sont respectivement appelés *germes de départ* ou *seeds* et *courbes déformables*.

Un germe de départ est un ou plusieurs pixel(s) adjacents sur l'image représentant l'objet recherché. Le principe des méthodes basées sur la croissance de régions consiste essentiellement à étendre un germe de départ selon des critères d'homogénéité, dans l'idée de le faire croître de manière itérative vers les contours exactes de l'objet qu'on cherche à segmenter. Cette extension est réalisée par l'agrégation de pixels voisins vérifiant certains critères d'homogénéité par rapport au germe de départ et la région segmentée à chaque itération.

Une courbe déformable est une courbe à deux dimensions formée par une série de points mobiles et placée dans une zone d'intérêt ou autour d'un objet dans l'image. Le principe des méthodes basées sur les contours actifs consiste à faire évoluer de manière itérative la courbe vers les contours exactes de l'objet désiré. Cette évolution est fondée sur la minimisation d'une fonction d'énergie exprimée sur la base d'une énergie interne et d'une énergie externe. L'énergie interne correspond à la morphologie et aux caractéristiques de la courbe tandis que l'énergie externe provient de l'analyse du gradient dans l'image.

Ces deux familles de méthodes ont été largement utilisées pour la détection des MCs du fait qu'elles offrent une segmentation itérative basée sur une analyse locale et permettent d'intégrer des critères adaptatifs. En 2010, Malek *et al.* [95] ont proposé un algorithme de détection de MCs à base de l'algorithme de *croissance de régions* standard (RG) [96].

Il correspond à l'un des rares travaux qui ont traité l'image entière pour la recherche des MCs. L'objectif de base était l'initialisation des germes de départ fiables du fait qu'ils ont un impact sur la dégradation de la qualité des résultats. Cette initialisation est basée sur la recherche et l'analyse des plateaux maxima à partir de la totalité de l'image mammographique. En partant de ces germes, ils appliquent l'algorithme de croissance de région standard pour estimer les zones susceptibles de correspondre à des MCs. Des opérations morphologiques sont, par la suite, appliquées dans le but de raffiner la qualité des contours des régions segmentées. Dans le même ordre d'idée, Duarte *et al.* [3] ont utilisé l'algorithme des contours actifs géodésiques [97] associé à un filtre anisotropique [98] pour la délimitation des MCs. Les contours actifs géodésiques constituent une extension de l'algorithme des contours actifs classique [99] avec l'avantage de pouvoir détecter plusieurs objets simultanément.

Une autre piste a consisté en l'utilisation de l'algorithme de croissance de région en association avec l'algorithme des contours actifs dans les travaux de Liu *et al.* [53] et Arikidis *et al.* [66] dans le but d'améliorer la précision des contours issus de la segmentation des MCs. Toutefois les approches proposées sont supervisées, elles commencent dans un premier temps par détecter et correctement segmenter les zones susceptibles de correspondre à des MCs dans le but de concevoir la base d'apprentissage. La segmentation de ces MCs débute par une initialisation des germes de départ par des radiologues. A partir de ces germes, une première délimitation des MCs est obtenue par l'application de l'algorithme de croissance de région. L'application des contours actifs sur les régions initialement segmentées a pour but de raffiner la qualité des contours obtenus.

2.3.4 Segmentation basée partitionnement

Un partitionnement, ou souvent appelé *clustering*, d'une image est une segmentation automatique de ses pixels dans l'objectif de la diviser en différentes zones homogènes pour ensuite en créer des groupes dits *clusters* ou *classes*. Ces groupes doivent être les plus homogènes individuellement mais les plus hétérogènes entre eux.

Les techniques de seuillages, qui offrent un partitionnement binaire d'une image, ont été largement utilisées pour la détection des MCs. La majorité a été fondée sur une technique de seuillage spatialement adaptative [100], [101], [102], [94]. Dans [103], les auteurs ont discuté l'utilisation de plusieurs types de seuillages utilisés pour la détection des MCs.

Une autre piste consistait à combiner d'autres techniques avec le seuillage adaptatif pour la détection des MCs. A titre d'exemple, Veni *et al.* [4] ont proposé une nouvelle approche fondée sur la méthode de détection de coins (SUSAN) et une technique de seuillage

adaptative du contraste. SUSAN *Smallest Univalve Segment Assimilating Nucleus* est un détecteur des points d'intérêts établi par Smith Prady en 1997. Il est considéré aussi comme un détecteur de contours. Comme l'indique son nom, il opère par une recherche des petites zones USAN contenant des pixels similaires au pixel en cours de traitement. Pour la détection, il utilise un masque circulaire de centre le pixel r_0 à analyser. Le résultat est une image divisée en petites régions contenant chacune un ou plusieurs objets susceptibles de correspondre à des MCs. Comme étape suivante, les auteurs ont appliqué sur chacune de ces régions, caractérisées par une taille maximale estimée à partir des informations cliniques, deux filtres spatiaux (filtre de contraste et filtre de distance) et un filtre de comptage du nombre des MCs (filtre de classe) afin de distinguer les 'vraies' MCs des fausses détections. Le filtre de contraste (Equation. (2.3)) compare le contraste estimé sur l'objet (dans la région au cours du traitement) et le seuil du contraste calculé sur la même région et qui est égal à la moyenne de ses 'mid-five' valeurs de NG. La taille des régions/blocs utilisés est généralement de 32x32 pixels. Les 'mid-five' valeurs sont les cinq niveaux de gris moyens dans la plage des valeurs d'intensité du bloc analysé.

$$Contrast_{Obj} = moy(NG_{Obj}) - moy(NG_{Back}) \quad (2.3)$$

Le filtre de distance est un filtre qui juge qu'un objet donné est, ou non, une MC en fonction de la distance minimale qui le sépare de son plus proche objet voisin. Il utilise une fenêtre de recherche de 40×40 pixels ($0.2\text{cm} \times 0.2\text{cm}$). Si un objet ne partage pas sa fenêtre de recherche avec un autre objet, il sera considéré comme une fausse MC et il sera effacé. Les deux autres filtres (filtre de contraste et filtre de classe) sont basés sur des critères définis par Kalleri *et al.*. Ces critères ne considèrent comme MCs qu'un groupe de minimum trois objets séparés par des distances inférieures à 0.2 cm et situés dans un emplacement de 1cm^2 .

En plus des techniques de partitionnement binaires à base de seuillage, K-Means est l'algorithme de clustering le plus connu. Il est introduit par J. McQueen *et al.* [104] en 1967. Fondé sur un processus itératif, il partitionne les pixels d'une image en K *clusters* défini chacun par l'ensemble des pixels qui le caractérise et son centroïde. Ce processus cherche à minimiser la somme des distances entre chaque pixel, représenté par un vecteur d'attributs x , et le centroïde m_j de son *cluster* C_j .

L'algorithme K-Means a été mis en œuvre dans [105] pour segmenter les MCs à partir d'une ROI d'une mammographie numérique. Il a été accompagné par une opération morphologique de prétraitement et un algorithme de classification supervisé pour la détermination de la classe des MCs segmentées (bénin ou malin).

Dans l'idée de remplacer l'affectation dure à un seul cluster, par le K-Means, l'algo-

rithme *Fuzzy C-Means* (FCM) [106] à été défini par Bezdek *et al.* en 1984. En particulier, ils ont introduit la notion d'ensembles flous dans la définition des *clusters* générés, où ils associent une valeur $\mu_{ij} \in [0, 1]$, dite un degré d'appartenance, à chaque pixel x_i relativement au cluster C_j . Le cluster auquel un pixel est associé est celui dont le degré d'appartenance associé est le plus élevé. Cet algorithme génère les clusters par un processus itératif en minimisant une fonction objective J basée sur un critère quadratique représentant la distance pondérée séparant chaque pixel x_i ($i = 1..N$ où N est le nombre des pixels dans l'image) du centre c_j du $j^{\text{ème}}$ cluster ($j = 1..K$).

Les travaux de recherche qui ont exploité des algorithmes de partitionnement précédemment décrits ou une de leurs variantes pour la détection MCs sont diverses. L'approche proposée par Quintanilla *et al.* en 2011 [92], à titre d'exemple, repose sur l'algorithme de segmentation possibiliste floue *Possibilistic Fuzzy C-Means* (PFCM) pour détecter les MCs. L'algorithme PFCM a été proposé par Pal *et al.* en 2005 [107]. Il présente une version hybride des algorithmes FCM et PCM [108]. C'est un processus itératif qui cherche à optimiser une fonction objective pour la construction des classes, à partir d'une image. Contrairement à FCM, avec PFCM, les pixels de l'image sur laquelle il est appliqué possèdent deux différentes cartes qui sont mis à jour au cours de chaque itération :

- Une carte de degrés d'appartenance (μ_{ij}) qui exprime le degré d'appartenance d'un pixel donné aux classes résultats de la segmentation. Les valeurs de cette carte sont estimées en fonction des écarts d_{ij} (appelés aussi distances) des niveaux de gris des pixels de l'image x_i par rapport aux centres des classes C_j ;
- Une carte de degrés de typicalité (t_{ij}) qui exprime le degré de *typicalité* de chaque pixel par rapport à sa classe d'appartenance. Les valeurs de cette carte sont estimées en fonction de la dispersion des niveaux de gris des classes construites, ainsi que de l'écart des niveaux de gris des pixels à traiter par rapport à celui des autres pixels de la classe courante.

La fonction objective de l'algorithme PFCM utilisée est définie comme suit :

$$J_{PFCM} = \sum_{j=1}^C \sum_{i=1}^N (\alpha \mu_{ij}^m + \beta \mu_{ij}^n) * d_{ij}^2 + \sum_{j=1}^C \gamma_j \sum_{i=1}^N (1 - t_{ij})^n \quad (2.4)$$

Les paramètres α , β , γ , m et n sont des variables de pondérations à définir.

Dans les travaux de Quintanilla *et al.*, l'algorithme PFCM est appliqué sur une région d'intérêt (ROI) prétraitée, sélectionnée par des radiologues, pour chercher les pixels MCs. Il divise cette région en deux classes : Une classe *objet* et une classe *fond*. La classe *objet* regroupe les pixels qui décrivent le tissu mammaire sur cette région tandis que la classe

fond regroupe les pixels du tissu de soutien du tissu mammaire présent. Les pixels de la classe *fond* possèdent des valeurs de niveaux de gris plus faibles que celles des pixels de la classe *objet*. Pour identifier les pixels MCs, les auteurs proposent d'appliquer un seuil fixe (égal à 0.2) sur la carte des degrés de typicalité des pixels de la classe *objet* obtenue. L'hypothèse de restriction de la recherche des MCs dans la classe *objet* se base sur le fait que les pixels de ces dernières présentent des valeurs d'intensités un peu plus élevées par rapport aux pixels des tissus mammaires de superposition. Pour eux, les pixels MCs sont les pixels intrus dans la classe *objet* qui représente le tissu mammaire du fait qu'ils présentent des valeurs de niveaux de gris *écartés* par rapport à la majorité des pixels dans cette classe. D'autres travaux des mêmes auteurs publiés en 2012 [93] et 2013 [109] ont montré la supériorité de l'algorithme PFCM par rapport à FCM et K-Means dans la recherche des MCs. Cette supériorité apparaît dans le taux d'identification, la précision des contours segmentés ainsi que la dépendance à une étape de prétraitement. En effet, les résultats obtenus ont montré une performance fortement liée à un prétraitement avec l'application des algorithmes K-Means et FCM contrairement à PFCM [110].

2.3.5 Segmentation à base de filtres

En traitement d'image, le terme filtrage désigne une opération dont l'objectif est de supprimer au maximum le bruit sur une image tout en préservant ses informations pertinentes. La pertinence des informations contenues dans une image varie en fonction de l'objectif d'interprétation. Ces informations sont généralement localisées au niveau des pixels contours qui, en dessinant des courbes fermées, décrivent les formes et les structures de base des objets recherchés. Les méthodes d'identification des objets d'intérêts à partir d'une image en utilisant des filtres, se distinguent en deux groupes. Le premier groupe identifie les extremums de la dérivée première (les maximums locaux de l'intensité du gradient). Le second groupe identifie les annulations de la dérivée seconde (annulations du Laplacien ou d'une expression différentielle non linéaire). L'utilisation des dérivées secondes est plus commode du fait qu'elle explore les deux directions dans l'image. C'est sur cette base que certains travaux de recherche ont exploré des filtres basés sur la dérivée seconde appelés aussi des filtres 'passe-haut' pour la recherche et l'identification des MCs à partir des images mammographiques. A titre d'exemple, le filtre *Laplacien de la Gaussienne* (LoG) a été utilisé dans les travaux de Cheng *et al.* [111] ainsi que Vivona *et al.* [1] et Basile *et al.* [44]. Ce filtre est défini comme la combinaison du *filtre Gaussien*, permettant la suppression du bruit, et *Laplacien* permettant la détection des contours (Equation. (2.5)). Il est calculé pour chaque pixel de coordonnées x et y et se base sur la mesure de variance σ .

$$-\nabla^2 H_\sigma(x, y) = -\frac{1}{\pi\sigma^2} \left(\frac{x^2 + y^2 - 2\sigma^2}{2\sigma^4} \right) e^{\left(\frac{x^2 + y^2}{\sigma^2} \right)} \quad (2.5)$$

La reconnaissance des MCs par ce filtre dépend de la valeur de σ utilisée qui influe directement sur la taille maximale des objets qu'on peut identifier à partir d'une image [1]. En effet, ce paramètre contrôle l'effet du lissage appliqué sur les images. Une augmentation de sa valeur renforce le lissage appliqué ainsi que le risque de perdre la capacité d'extraire certains détails. Ainsi, afin de surmonter cette dépendance et d'optimiser l'efficacité de la reconnaissance, Vivona *et al.* ont mis en œuvre un filtre spatial avec une forme qui permet la détection de petites et grandes MCs. Cette forme est caractérisée par trois régions (Figure. 2.1) et est conçue avec la somme des poids égaux à zéro afin d'obtenir un effet sur la dérivé :

- Une première région centrale, circulaire (de rayon R_1), dont les pixels ont une intensité positive et une surface égale à la plus petite taille d'une seule calcification, avec une forme de la partie positive du Laplacien de la Gaussienne ;
- Une deuxième région, également circulaire, adjacente à la première région et externe, dont les pixels ont une intensité nulle. Cette région s'étend jusqu'à une surface égale à la taille maximale d'une seule calcification détectable (distance R_2) ;
- Une troisième région, la plus externe et assez étroite, des deux valeurs négatives obtenues à partir de la même fonction logarithmique et renormalisées de sorte que la somme des intensités de ses pixels (négatives) est égale à la somme des intensités des pixels de la première région (positives).

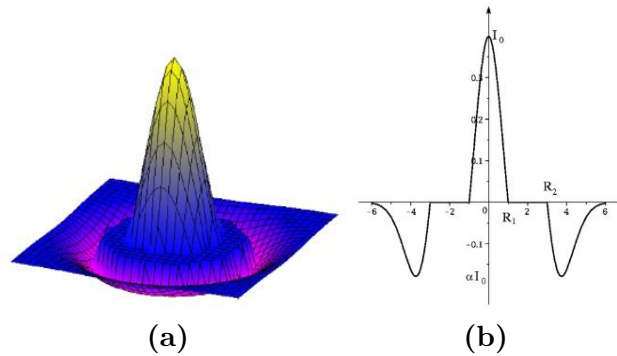


Figure 2.1 – Projection en 3D (a) et 2D (b) du filtre spatial de Vivona *et al.* [1]

Supposons $\Delta R = R_2 - R_1$, l'expression mathématique du filtre $F(x, y)$ basée sur les trois régions, précédemment décrites, est définie par l'équation (2.6) suivante :

$$F(x, y) = \begin{cases} LoG(x, y) & \text{Si } x^2 + y^2 < R_1^2 \\ 0 & \text{Si } R_1^2 \leq x^2 + y^2 \leq R_2^2 \\ \alpha LoG[(x - \Delta R), (y - \Delta R)] & \text{Si } x^2 + y^2 > R_2^2 \end{cases} \quad (2.6)$$

où α est un facteur de normalisation choisi de telle sorte que le volume I_{S_1} du cercle de rayon R_1 soit égal à celui I_{S_2} de la surface extérieure du cercle de rayon R_2 .

En analysant ce filtre et le résultat de son application sur une position (x, y) d'une image, on constate que la MC doit prendre une valeur positive dans (x, y) (c.à.d. doit se situer dans la partie positive du filtre) [1], tandis que le fond doit chevaucher avec la partie négative. Le filtre développé ici garanti également des chevauchements partiels entre les MCs et la partie positive du filtre. En fait, pour une MC caractérisée par des dimensions linéaires entre R_1 et R_2 , la partie positive qui se chevauche avec le filtre garde la partie des dimensions linéaires R_1 .

Plus récemment, Alsheh *et al.* ont combiné la *Différence de Gaussiennes* (DoG) avec l'*Histogramme de gradients orientés* (HoG) dans [8] pour la recherche des MCs à partir des images mammographiques. Le filtre DoG (Equation. (2.7)) est un filtre passe-bande qui s'exprime par une différence entre deux gaussiennes de variances respectivement σ_1 et σ_2 avec ($\sigma_1 > \sigma_2$). Pour les auteurs, et afin de limiter les effets du bruit et de maximiser l'amélioration des MCs, les valeurs de σ_1 et σ_2 sont fixées respectivement au rayon de la plus petite taille possible d'une MC (qui est de 0.05 mm) et à sa taille moyenne théorique (qui est de 0.15 mm).

$$DoG_{\sigma_1, \sigma_2}(x, y) = G_{\sigma_1} - G_{\sigma_2} \quad (2.7)$$

avec

$$G_{\sigma} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$$

Dans le même ordre d'idée, Melendez *et al.* dans [112] ont introduit un filtre de Butterworth (BBP) dans le domaine de Fourier. Le Butterworth filtre est un filtre linéaire, décrit par Stephen Butterworth en 1930 [113]. Il a une capacité de maintenir une réponse d'amplitude monotone dans la bande passante. Ce qui résulte en une suppression des pixels du tissu fibroglandulaire et une amélioration des pixels contours des MCs. Pour améliorer encore le contraste des bords détectés et réduire le bruit du fond, les auteurs ont appliqué un filtre médian avec une correction gamma. Les MCs segmentées sur l'image résultent de l'application d'opérations morphologiques sur l'image précédemment obtenue. Les résultats de l'application de ce filtre ont été comparés avec ceux obtenus avec les filtres DoG et LoG et ont montré sa supériorité aussi bien dans l'identification que la segmentation [112].

2.3.6 Étude comparative

Les travaux que nous venons de présenter dans les sections précédentes ont tenté de proposer des approches fiables pour détecter les MCs à partir des mammographies numériques. La diversité qu'ils présentent, en termes de principes de raisonnement ou type de techniques exploitées, justifie la complexité de la problématique de détection des MCs. En guise de récapitulation, et dans l'objectif de pouvoir correctement concevoir

notre approche, nous allons discuter dans le tableau suivant l'ensemble des avantages et inconvénients de chaque type de techniques.

Tableau 2.1 – Comparaison des techniques explorées dans la littérature pour la détection des MCs

Technique utilisée	Principe, Avantages et Inconvénients
Segmentation fréquentielle	Caractérisation des MCs en tant que des 'régions' de hautes fréquences <ul style="list-style-type: none"> Détection des MCs avec des formes exactes ; Fiabilité modérée par des fausses détections du fait qu'elles présentent une sensibilité aux bruits et aux types de tissus de superposition.
Segmentation basée morphologie	Détection basée sur une analyse spatiale à partir des initialisations <ul style="list-style-type: none"> Détection basée sur des connaissances <i>à priori</i> sur les caractéristiques des MCs (formes, tailles, ...) ; Fiabilité relative au type de tissu de superposition ainsi aux paramètres d'initialisations.
Segmentation basée région	Détection basée sur une analyse spatiale itérative à partir de points de départ <ul style="list-style-type: none"> Analyse spatiale très fine et contrôlée pour une prise de décision ; Nécessité des initialisations des points de départ pour démarrer la recherche ; Fiabilité dépend des points de départ et des conditions d'arrêt ainsi que la densité du sein qui affecte le contraste des contours des MCs.
Segmentation basée partitionnement	Détection basée sur la minimisation des fonctions objectives <ul style="list-style-type: none"> Modélisation et intégration des connaissances <i>à priori</i> par la fonction à optimiser ; Sensibilité à la variation du contraste des MCs par rapport aux tissus de superposition.
Segmentation basée filtrage	<ul style="list-style-type: none"> Réduction de la sensibilité au contraste variable des MCs ; Fiabilité relative au couplage entre les paramètres utilisés et les tailles des MC ; Sensibilité au bruit.

Tout en essayant de se positionner par rapport à l'existant, nous remarquons que

la majorité des travaux ont recours à une étape de prétraitement de l'image avant de procéder à la détection. Nous essayons d'étudier cette notion dans la section suivante.

Avant d'introduire l'idée générale de notre approche proposée dans le cadre de ce travail de thèse, nous présentons dans la section suivante un ensemble de prétraitement utilisés dans la littérature, dans le contexte de l'analyse de mammographies.

2.4 Prétraitement des images mammographiques

2.4.1 Qualité et nécessité de prétraitement

La qualité des images mammographiques est un élément fondamental pour le diagnostic qui est sensible à la densité du sein ainsi que la fiabilité du système d'acquisition (mammographe). En effet, les tissus mammaires apparaissent sur cette image en plusieurs formes et contrastes vue la variété de leurs natures anatomiques. Cette variété peut être aperçue entre deux types de tissus mammaires qu'entre un tissu mammaire et une lésion cancéreuse présente. Nous citons à titre d'exemple le cas des MCs qui apparaissent avec des contrastes généralement faibles par rapport à leurs voisinages immédiats. Un contraste similaire ou parfois plus élevé peut apparaître sur la même image entre deux tissus mammaires de différentes natures. Cette variation dépend du type de densité du sein ainsi du type de tissu de superposition des MCs (Figure. 2.2).

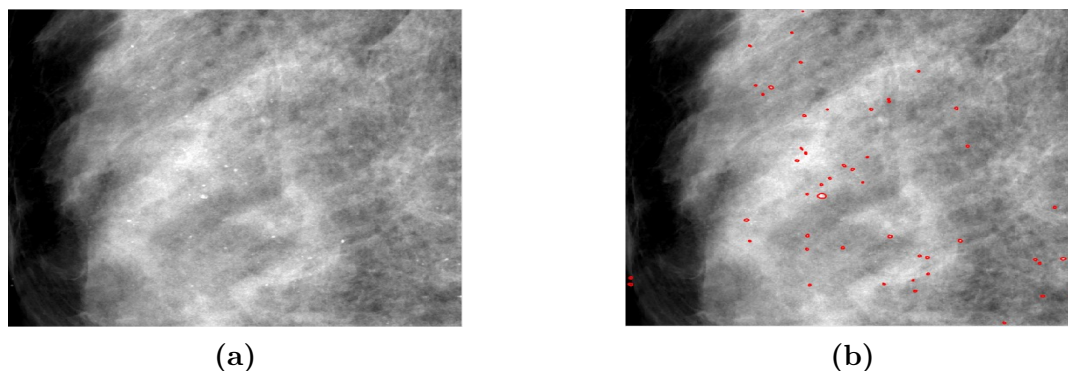


Figure 2.2 – Illustration de la variation du contraste sur une image mammographique entre les MCs : (a) ROI d'une image mammographique sans (a) et avec superposition de la vérité de terrain (contours rouges en (b))

2.4.2 Techniques de prétraitement standard

L'objectif de base des techniques de prétraitements est d'améliorer la qualité d'une image donnée. Ça reste toujours un problème subjectif. Cependant, il est connu que l'œil humain est sensible aux forts contrastes. De manière générale, les techniques d'amélioration d'images tentent donc de changer le contraste et/ou la luminosité afin d'accroître la

séparabilité des régions présentant des objets différents. Dans la littérature, nous pouvons distinguer deux types de techniques de prétraitements :

- Des techniques agissant dans le domaine spatial : opérations pixeliques ;
- Des techniques agissant dans le domaine fréquentiel : opérations fréquentielles.

Les techniques de prétraitement agissant sur le domaine spatial opèrent au niveau pixelique spatial ou global (distribution des niveaux de gris) afin d'améliorer la qualité d'une image donnée. En contrepartie, les techniques agissant sur le domaine fréquentiel opèrent par des transformations d'ondelettes pour la modification et la reconstruction d'une image.

Les MCs, sur une image mammographique, apparaissent comme des composants de haute fréquence en plus des composants mammaires denses. L'application des techniques de prétraitement à base d'ondelettes permet d'améliorer l'apparition des composants de hautes fréquences mais elles présentent l'effet indésirable d'augmenter le bruit [114]. Par conséquent, ces techniques peuvent améliorer le taux de détection des MCs, en particulier dans les images mammographiques denses, mais peuvent aussi augmenter le taux des faux positifs vis-à-vis du processus de détection. Partant de ce constat, nous allons restreindre notre synthèse sur les techniques de prétraitements à base d'opérations pixeliques. Parmi ces techniques, on peut distinguer :

- Les techniques de changement de contraste et/ou de luminosité ;
- Les techniques de filtrage.

2.4.2.1 Changement de contraste et/ou de luminosité

Le contraste d'une image est une propriété intrinsèque qui permet de quantifier la capacité de discriminer entre deux régions distinctes. Il constitue une valeur C appartenant à l'intervalle $[0,1]$ et définie en fonction des intensités maximale I_{max} et I_{min} des pixels ou zones dont on veut calculer le contraste (Équation. (2.8)).

$$C = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \quad (2.8)$$

L'opération de changement de contraste d'une image, revient à effectuer une transformation sur les pixels de cette dernière ayant une même valeur d'intensité. Cette opération peut être :

- Linéaire avec une équation de la forme : $a.x + b$;
- Non-linéaire avec une équation de la forme : $a.x^n$.

où :

x réfère au niveau de gris d'un pixel de l'image,

a , b et n réfèrent à des paramètres de transformation non-nuls.

Dans les paragraphes qui suivent, nous allons présenter certaines techniques de changement de contraste utilisées dans la littérature, dans le cadre de la détection des MCs.

a. Correction Gamma

La correction Gamma est la technique de transformation non linéaire la plus répandue pour l'amélioration de la qualité des images mammographiques [115]. C'est une opération d'amplification logarithmique des niveaux de gris d'une image I donnée. Elle se base sur un paramètre Gamma (γ) dit aussi le facteur de contraste. Avec une telle transformation, les pixels de niveaux de gris moyens seront plus modifiés que les pixels de niveaux de gris sombres ou clairs. La formule générale de cette transformation est comme suit :

$$I' = \left(\frac{1}{\max(I)} \right)^{\left(\frac{1}{\gamma} \right)} \quad (2.9)$$

b. Égalisation d'histogramme

Pour l'amélioration de la qualité des images mammographiques, les techniques d'égalisation d'histogramme sont les techniques de prétraitement linéaires les plus utilisées [116]. L'histogramme d'une image est la courbe qui représente la répartition de ses valeurs de niveaux de gris. L'égalisation d'histogramme est une méthode d'ajustement du contraste visant à transformer l'histogramme initial en un histogramme 'uniformé'. Elle consiste à répartir les intensités des pixels d'une image sur l'ensemble de la plage des valeurs possibles afin d'accroître le contraste. Parmi les techniques les plus utilisées dans la littérature, nous pouvons citer :

- L'égalisation d'histogramme standard ;
- L'égalisation d'histogramme adaptative ;
- L'égalisation bi-histogramme avec préservation de luminosité (*Brightness Preserving Bi-histogram Equalization*).

L'égalisation d'histogramme standard est une technique de modélisation *a posteriori* qui projette chaque niveau de gris d'entrée $h(i)$, ($i = 1..N$) de l'image I de N pixels à un niveau de gris proportionnel à sa densité cumulée et donc la probabilité de chaque niveau de gris dans l'image résultante I' est uniformément distribuée. L'expression utilisée pour étaler les niveaux de gris d'une image est la suivante :

$$I' = \frac{256}{N^2} * \sum_i h(i) \quad (2.10)$$

L'égalisation d'histogramme adaptative est une variante de la technique d'égalisation d'histogramme standard. Son raisonnement se base sur le principe suivant : elle divise

l'image originale en plusieurs sous-images qui ne se chevauchent pas. L'histogramme de chacune de ces images est découpé, puis égalisé, afin de limiter l'amélioration de chaque pixel. Avec une telle égalisation, les détails de l'image apparaissent clairement par rapport à l'arrière-plan. En même temps, l'arrière-plan de l'image est également amélioré en tant que premier plan de l'image, ce qui conduit à une image en sortie avec un fort contraste [117].

La technique 'Brightness Preserving Bi-histogram Equalization' [118] divise l'histogramme d'une image, ayant L comme valeur d'intensité maximale, en deux parties à partir de l'intensité moyenne $((L-1)/2)$. Ces deux histogrammes sont par la suite égalisés indépendamment. Avec une telle égalisation, les valeurs d'intensités inférieures à $(L-1)/2$ seront étalées vers la valeur zéro. Les valeurs d'intensités supérieures à $(L-1)/2$ seront étalées vers la valeur $(L-1)$.

Les performances de ces techniques de prétraitement ont été évaluées par plusieurs travaux dans le contexte d'amélioration d'apparition des MCs. Nous pouvons citer les travaux de Moradmand et *al.* 2012 [101], Wu et *al.* 2013 [119], Kayode et *al.* 2015 [120], Akila et *al.* 2015 [121], ...

2.4.2.2 Changement de contraste et luminosité : Window/Level

Les modalités d'imagerie médicale tentent à produire des images avec une profondeur de 12 et 16 bits par pixel, ce qui produit $2^{12} - 1$ et $2^{16} - 1$ valeurs possibles de niveau de gris. Cependant, les écrans utilisés pour l'analyse de ces images ne supportent généralement qu'une profondeur de 8 bits. Window/Level est une opération spécialement utilisée pour les images médicales et connue sous le nom de *fenêtrage*. Elle a pour objectif d'ajuster à la fois le contraste et la luminosité dans une image afin d'améliorer sa qualité. Elle consiste à sélectionner, dans un premier temps, un intervalle $[x_{min}, x_{max}]$ de la plage des valeurs des niveaux de gris du récepteur (selon la profondeur de l'image). Par la suite, elle affiche les valeurs des pixels dans cet intervalle sur toute la luminosité qui s'étend du blanc au noir. L'intervalle $[x_{min}, x_{max}]$ est appelé *Window* et le point situé à son milieu est appelé *Level*. Le lien explicite du Window/Level avec le contraste (C) et la luminosité (L) est le suivant :

$$Level = \left(1 - \frac{L}{100}\right) \frac{x_{max} - x_{min}}{x_{min}} \quad (2.11)$$

$$Window = \left(1 - \frac{C}{100}\right) \frac{x_{max} - x_{min}}{x_{min}} \quad (2.12)$$

Avec ce changement :

- Tout pixel dans l'intervalle *Window* sera affecté à un changement de contraste ;
- Tout pixel avec une intensité inférieure (respectivement supérieure) à 'window' devient noir (respectivement blanc) (changement de luminosité).

2.4.2.3 Techniques de filtrage

Contrairement aux techniques de prétraitement basées sur la modification des histogrammes pour l'amélioration du contraste, les filtres spatiaux constituent une analyse spatiale des pixels de l'image. Ils cherchent, par cette analyse, à lisser les valeurs des niveaux de gris des pixels tout en améliorant la qualité d'apparition des objets présents sur l'image. Dans ce qui suit, nous allons passer en revue les trois filtres les plus utilisés pour l'amélioration des images mammographiques à savoir le filtre de Nagao, le filtre du masque flou et le filtre de diffusion anisotropique.

a. Filtre de Nagao

C'est un filtre qui a l'effet d'un filtre moyenneur avec conservation des contours. Il constitue une des solutions proposées pour remédier au problème d'affectation de la texture de l'image en adoucissant ces contours par un filtre moyenneur simple [5]. Pour calculer la nouvelle valeur à affecter à un pixel P donné dans l'image, ce filtre, considère, dans un premier temps, neuf 'sous-fenêtres' dans le voisinage de ce pixel. Par la suite, il vérifie l'homogénéité de chacune de ces sous-fenêtres. Celle qui apparaît la plus homogène, sera sélectionnée comme un support pour calculer la nouvelle valeur de niveau de gris à affecter au pixel P . Cette valeur est la moyenne des valeurs de niveaux de gris dans la sous-fenêtre gagnante. En d'autres termes, le filtre de Nagao restreint l'application du filtrage moyen à la sous-fenêtre considérée comme la plus homogène dans le voisinage d'un pixel P . Cette caractéristique est celle qui a encouragé certains travaux pour appliquer ce filtre sur les images mammographiques afin de chercher à améliorer la qualité du contraste des MCs ainsi qu'augmenter l'homogénéité des différents objets (composantes anatomiques) sur l'image.

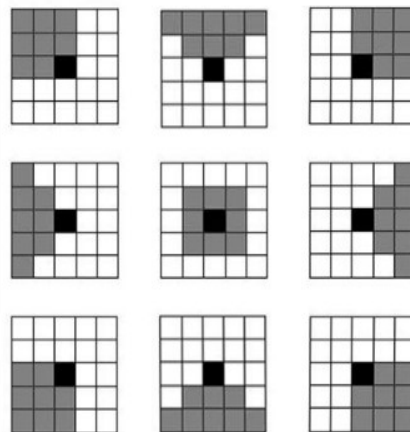


Figure 2.3 – Les sous-fenêtres du filtre de Nagao

b. Filtre de 'Masque flou' (Unsharp Masking)

C'est un filtre de rehaussement de contours avec augmentation de la netteté de l'image [122]. L'image résultat $UM(x,y)$ de ce filtre peut être exprimée comme l'addition de l'image originale $I(x,y)$ avec le résultat du filtre passe-haut $G(x,y)$ multiplié par un coefficient de contraste $K(x,y)$:

$$UM(x,y) = I(x,y) + K(x,y) \cdot G(x,y) \quad (2.13)$$

Le filtre passe-haut $G(x,y)$ revient à convoluer, dans un premier temps, l'image originale $I(x,y)$ par un filtre moyennneur. Le résultat de cette convolution est par la suite soustrait de l'image originale (Equation. (2.14)).

$$G(x,y) = I(x,y) - \frac{1}{mn} \sum_{i=1}^n \sum_{j=1}^m I(x_i, y_j) \quad (2.14)$$

Avec m et n sont le nombre de lignes respectivement colonnes de l'image I .

La capacité d'améliorer la qualité des contours ainsi que les composantes de hautes fréquences sur les images, par ce filtre, justifie son utilisation fréquente dans la littérature. Parmi les travaux, nous pouvons citer ceux de Wu *et al.* 2010 [122], Moradmand *et al.* 2012 [101] et Mohanalin *et al.* 2014 [123].

c. Diffusion anisotropique

La diffusion anisotropique est une diffusion non linéaire qui a pour objectif de filtrer une image tout en préservant ses contours. Elle est introduite par Pietro Perona et Jitendra Malik [98] en 1990 dans le but de remédier au problème d'atténuation des contours par une diffusion isotropique. Une diffusion isotropique consiste à un filtrage linéaire basé sur un noyau de convolution gaussien. Elle diffuse uniformément l'intensité des pixels de l'image autour de leurs voisinages. Contrairement à la diffusion isotropique, la diffusion anisotrope consiste à effectuer une forte diffusion dans les zones à faible gradient (zone homogène), et une faible diffusion dans les zones à fort gradient (contours). Cette caractéristique est la base de son utilisation pour les images mammographiques [101].

2.4.3 Prétraitement en mammographie : Synthèse et discussion

Le principe sous-jacent dans l'amélioration du contraste en mammographie est d'élargir la différence d'intensité entre les objets et le fond et de produire des représentations fiables des structures du tissu mammaire. Une méthode efficace pour améliorer la mammographie doit viser à améliorer la texture et les caractéristiques des objets recherchés.

Suite à la synthèse présentée plus haut, la classification des techniques de prétraitement peut être faite comme suit [101, 120] :

- Des techniques basées sur la modélisation de l'histogramme dont l'objectif principal est de modifier la forme de l'histogramme original à mesure que le contraste s'améliore ;
- Les techniques à base d'opérations morphologiques dont l'objectif est d'extraire des composantes d'image qui peuvent être utilisées pour calculer des caractéristiques relatives à la forme d'une région, ses limites et sa superficie. La valeur de chaque pixel dans l'image de sortie est basée sur une comparaison du pixel correspondant dans l'image d'entrée avec ses voisins ;
- Les techniques basées sur un masquage flou qui créent une version filtrée passe-bas de l'image originale. Les valeurs d'image qui en résultent sont multipliées par un facteur de pondération et soustraites de l'image originale. L'image finale préserve une grande partie des détails de l'image originale, mais les grandes structures présentent moins de contraste, ce qui réduit la plage dynamique requise pour l'afficher.

Malgré la variété des techniques de prétraitement utilisées, leur fiabilité par rapport à l'amélioration du contraste d'apparition des MCs reste toujours dépendante de la densité qui caractérise l'endroit de leur apparition [120].

Des travaux récents [48, 54] ont mentionné que la faible qualité d'apparition des MCs est due non seulement aux propriétés anatomiques de ces dernières, mais aussi à l'effet du bruit de type 'quantum' produit sur l'image à cause des interactions du corps humain avec les rayons X des modalités utilisées. Ce dernier peut être décrit par une distribution de Poisson et supprimé par l'application des filtres de débruitage, tel que celui de Frost.

2.5 Approche proposée

2.5.1 Analyse critique

Les approches proposées dans la littérature pour la détection des MCs souffrent souvent des deux principaux types d'erreurs de détections qui sont :

- Des erreurs de type faux négatif : groupement de pixels adjacents qui correspond à une 'vrai' MC mais considéré, à tort, comme un groupement de pixels normaux par l'approche utilisée ;
- Des erreurs de type faux positif : groupement de pixels adjacents qui correspond à un groupement de pixels normaux mais considéré, à tort, comme une MC par l'approche utilisée.

L'apparition de ces deux types d'erreurs est essentiellement due à trois limitations. La première limitation est l'absence de l'intégration des informations sémantiques contextuelles dans le raisonnement appliqué. En effet, les techniques utilisées pour la détection opèrent sur les valeurs numériques de l'image mammographique. Elles négligent, d'une part, la dépendance de la fiabilité de cette information numérique à celle du système d'acquisition. D'autre part, elles n'exploitent pas la richesse du niveau informationnel sémantique des images mammographiques dans le processus de prise de décision. En effet, une image mammographique traduit, avec les valeurs numériques associées à ses pixels, la sémantique de répartition des différents tissus mammaires qu'elle représente. Nous pouvons citer à titre d'exemple, que la présence d'une zone asymétrique entre les deux vues droite et gauche du même sein peut augmenter le facteur de risque pour un groupement de pixels mentionnés comme douteux. Un tel raisonnement, basé uniquement sur des valeurs numériques (les niveaux de gris), s'avère ainsi insuffisant pour assurer une détection fiable puisqu'il se base sur des informations incertaines dont la fiabilité dépend de la qualité du système d'acquisition utilisé et dont leur contenu informationnel sémantique contextuel n'est pas exploité.

La deuxième limitation est l'utilisation d'un seul niveau d'analyse pour la détection des MCs, qui est le niveau pixelique local. Ces MCs sont décrites en tant que groupements de pixels adjacents vérifiant les critères des pixels douteux définis par la technique utilisée. En revanche, l'image mammographique est une projection du sein mammaire tridimensionnel dans un espace à deux dimensions. La complexité et la similarité des structures anatomiques présentes sur cet organe multi-tissulaire engendrent une très grande similarité d'apparition 'locale' et parfois 'globale' entre ces tissus mammaires et certaines MCs.

La troisième limitation est l'utilisation d'une seule et même technique de détection pour la détection des MCs qui est en elle-même un problème complexe. En effet, les caractéristiques des MCs varient en fonction des endroits et des causes d'apparition et elles peuvent aussi être partageables avec d'autres pixels appartenant à d'autres types de tissus mammaires. Cette propriété pose le problème du bon choix des paramètres à utiliser (ex. choix des seuils et des conditions d'arrêts) avec les techniques de détection qui opèrent sur la totalité de l'image.

En conclusion, la restriction de la recherche des MCs sur une ROI permet de surmonter le problème de variabilité en terme de densité mammaire ainsi que du matériel utilisé pour la génération des images. En revanche, le processus de détection appliqué est un raisonnement de bas niveau qui ne prend en considération que les valeurs numériques des pixels. Ce qui met en exergue de nouveau le problème de fiabilité de ces valeurs numériques et la perte du niveau informationnel sémantique. En effet, selon les radiologues, la qualité des images générées est relative à celle du matériel utilisé [48].

2.5.2 Approche proposée

L'exploitation des techniques de segmentation pour la détection des MCs est un point commun pour les différentes approches proposées dans la littérature. Cependant, l'utilisation individuelle de ces techniques engendre parfois des problèmes de distinction entre les pixels des MCs et les pixels des tissus mammaires. La variété des caractéristiques des MCs et l'analyse numérique explorée, par ces dernières, constituent la principale cause de ces difficultés. En effet, le radiologue, au cours de son diagnostic, intègre et combine plusieurs informations du niveau sémantique. Afin d'interpréter les MCs, Il commence par détecter les "petites" régions ayant des formes plus ou moins circulaires et plus brillantes par rapport à leur voisinage local. Par la suite, il émet des hypothèses sur le type de ces premières détections en les situant par rapport à leur positionnements local, contextuel et global sur l'image tout en tenant compte de son expertise du domaine. Toutes ces informations, ensembles, ne peuvent pas être intégrées de façon conjointe dans une seule technique de segmentation et ne figurent pas dans le même espace de représentation. Partant de ce constat, et afin d'améliorer la qualité des prises de décision, il est envisageable d'explorer et de combiner plusieurs techniques de segmentation qui modélisent, chacune, l'une des caractéristiques des MCs relativement à un espace d'interprétation bien défini.

L'approche que nous proposons, dans ce travail de recherche consiste ainsi à proposer un système d'interprétation de scène qui exploite à la fois une représentation hiérarchique de plusieurs niveaux d'interprétation sémantique permettant la fusion de « connaissances » issue de ces niveaux. On désigne par le terme « connaissances », les informations issues des différents niveaux d'interprétations qui constituent bien des contenus informationnels sémantiques de l'image (connaissances de l'expert) représentés, dans des espaces bien définis et interprétés à un niveau d'interprétation donné. Le choix de cette structure hiérarchique pour les niveaux d'interprétation avec une propagation et un échange d'informations est justifié par le fait que nous cherchons à simuler, par notre approche proposée, le raisonnement du radiologue au cours de son diagnostic.

Chaque niveau d'interprétation utilisé dans notre approche offre une représentation propre de la granularité de l'information spatiale dans l'image. Il applique un traitement indépendant des autres niveaux tout en exploitant des informations numériques et d'autres sémantiques afin de produire des « connaissances » sur l'objet sémantique étudié relativement à la représentation qu'il a utilisé pour la description des MCs. Ces connaissances correspondent bien à des sources d'informations hétérogènes et complémentaires qui doivent collaborer ensemble afin de produire la décision finale.

En particulier, nous distinguons trois différents niveaux d'interprétation dans l'approche proposée : le niveau pixelique, le niveau primitive visuelle et le niveau des objets.

Le niveau pixelique constitue le plus bas niveau d'interprétation possible des MCs.

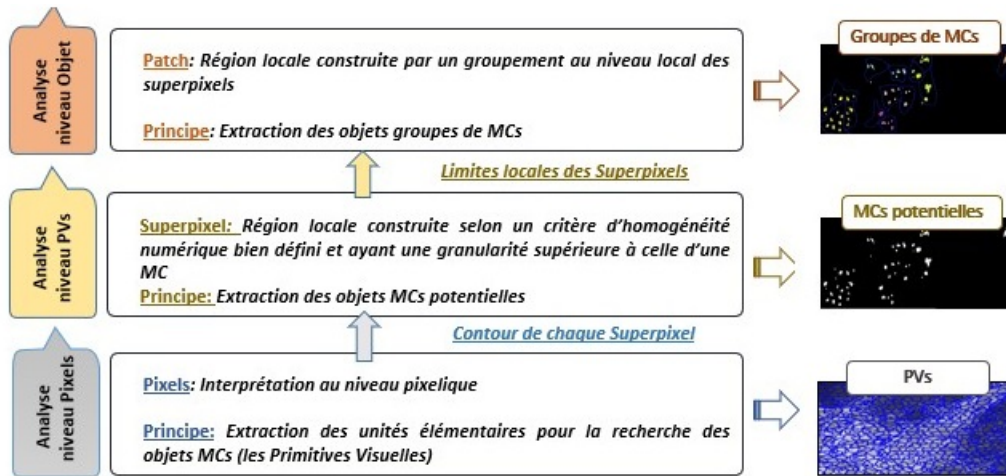


Figure 2.4 – Les niveaux d'interprétation exploités dans l'approche proposée

L'interprétation dans ce niveau tente à extraire des régions locales à partir de l'image ayant une granularité supérieure à celle d'une MC que l'on considère comme étant des Primitives Visuelles (PVs). Les connaissances sur les limites géométriques (contours) de ces dernières sont propagées vers le niveau d'interprétation suivant, celui des primitives visuelles.

L'interprétation au niveau des primitives visuelles se base sur un traitement numérique appliqué sur les pixels de ces régions et a pour objectif la détection des MCs potentielles. On appelle MC potentielle tout groupement de pixels adjacents vérifiant les critères des pixels de tissus suspects par le traitement appliqué. Les connaissances de ce niveau, concernent les limites locales des primitives visuelles comportant des MCs potentielles.

L'interprétation au niveau objet vise à identifier des régions de véritables labels sémantiques qui correspondent aux connaissances de l'expert. Dans notre contexte, un objet est formé par le regroupement de primitive visuelles adjacentes comportant des MCs potentielles (que l'on appelle des 'patches'). L'interprétation de ce niveau consiste à mettre en exergue les groupes de MCs. On désigne par groupe de MCs tout groupement de MCs potentielles dans un patch. Les connaissances de ce niveau concernent les limites géométriques des patches et sont propagées dans le niveau scène suivant.

Le passage d'un niveau d'interprétation à un autre exploite des connaissances sémantiques dont l'objectif de base est de traduire le 'savoir-faire' des radiologues pour une prise de décision. De ce fait, pour chaque niveau d'interprétation, un type particulier de connaissances sémantiques est intégré pour servir à la prise de décision. Le Tableau 2.2 illustre les différents niveaux d'interprétation avec les connaissances sémantiques correspondantes.

Tableau 2.2 – Approche proposée : Niveaux d'interprétation et connaissances sémantiques

Niveau d'interprétation	Connaissances sémantiques
Niveau des Pixels	Caractérisation pixelique de la densité anatomique mammaire
Niveau des Primitives Visuelles (PVs)	Caractéristiques morphologiques des MCs individuelles (formes, contrastes, ...)
Niveau des Objets	Caractérisation des groupes des MCs (distributions, formes, ...)

Notons que pour notre approche, la propagation des connaissances obtenues entre les niveaux d'interprétation est effectuée simultanément dans les deux sens : le sens ascendant et le sens descendant. Le sens ascendant concerne la propagation des résultats obtenus (les connaissances) d'un niveau antérieur au niveau supérieur. Il permet au niveau suivant de prendre une décision. En revanche, le sens descendant consiste à rétropropager le résultat d'un niveau supérieur vers les niveaux antérieurs. Il permet de raffiner le résultat déjà obtenu au niveau précédent en fonction de l'interprétation réalisée au niveau courant. A titre d'exemple, le résultat du niveau d'interprétation des objets pourra modifier celui précédemment obtenu (aux niveaux des primitives visuelles et pixelique). Les informations apprises à partir de chaque niveau d'interprétation constituent une sorte de « connaissances » traduisant le savoir descriptif sur les MCs dans ce dernier. L'hétérogénéité des espaces de représentation de ces dernières assure une hétérogénéité des connaissances obtenues qui doit être respectée par la fusion à appliquer. A ce stade, nous proposons de représenter ces connaissances dans un espace possibiliste. Ainsi, chaque niveau d'interprétation projette ses informations apprises sur une carte de suspicion prenant des valeurs dans l'intervalle $[0..1]$ et représentant les possibilités d'appartenance des pixels de l'image mammographique à la classe des pixels suspects. Cette représentation unique des différentes connaissances nous permettra de concevoir une interprétation homogène pour les différents niveaux d'interprétation sémantique sans perte des connaissances. Le choix de cette représentation possibiliste est justifié par la capacité de cette théorie à gérer l'incertitude et l'ambiguïté des informations. Une telle représentation nous permet aussi de concevoir un nouvel espace de représentation de l'objet à chercher (les MCs) indépendant de l'information numérique incertaine et du niveau de représentation sémantique utilisé. D'autre part, la sortie de l'approche proposée peut être conçue comme une source d'information permettant l'aide à la décision pour les radiologues (interprétée à l'aide d'un lexique médical et non seulement une décision brute). NivInterpretApproche2

La décision finale de notre approche est le résultat d'une combinaison collaborative des différentes connaissances apprises par les niveaux d'interprétations qui correspondent à des sources d'informations hétérogènes. Du fait que les connaissances obtenues par ces

sources soient bien complémentaires, en termes d'espaces de représentations et de type d'informations sémantiques modélisées, cette combinaison doit respecter les niveaux de représentation afin de pouvoir exploiter de façon optimale les connaissances des différents niveaux.

2.6 Conclusion

Dans ce chapitre, nous avons commencé par introduire l'intérêt ainsi que les difficultés relatives à l'interprétation d'une image mammographique pour détecter les MCs. Les solutions existantes dans la littérature pour la résolution de cette problématique, ont été par la suite présentées. Nous avons distingués cinq différentes techniques de segmentation explorées par ces méthodes : segmentation fréquentielle, basée morphologie, régions, partitionnement et filtrage. Nous avons poursuivi cette présentation avec une étude des avantages et inconvénients de ces travaux existants dans l'objectif de correctement concevoir notre propre contribution, décrite dans la dernière partie de chapitre.

PREMIÈRE CONTRIBUTION : DÉTECTION BASÉE SUR LA MORPHOLOGIE MATHÉMATIQUE

Sommaire

3.1	Introduction	62
3.2	Motivation	62
3.3	Morphologie mathématique et détection des MCs	63
3.3.1	Concepts de base de la morphologie mathématique	63
3.3.2	Limites	64
3.4	Méthode de détection proposée	65
3.4.1	Amélioration de la mammographie	66
3.4.2	Génération de la carte de suspicion	72
3.5	Méthodologie de validation	78
3.5.1	Métriques d'évaluation des performances	78
3.5.2	Méthode ('Zh') retenue pour comparaison	80
3.6	Évaluation qualitative des résultats	81
3.6.1	Fiabilité de détection	82
3.6.2	Fiabilité de la segmentation	86
3.6.3	Fiabilité par rapport à la méthode de comparaison	88
3.7	Évaluation quantitative des résultats	90
3.7.1	Fiabilité des résultats	91
3.7.2	Fiabilité par rapport à la méthode 'Zh'	91
3.8	Discussion	93
3.9	Conclusion	94

3.1 Introduction

Malgré qu'elle ait fait l'objet de plusieurs travaux, la problématique de détection des MCs reste toujours un problème complexe qui attire l'attention des chercheurs. Ces derniers s'orientent vers l'étude des caractéristiques de ce type de lésions pour concevoir le meilleur processus de détection. Dans le chapitre précédent, nous avons discuté les solutions proposées dans la littérature pour la résolution de cette problématique à partir des images mammographiques. Nous avons constaté que la fiabilité est souvent affectée par la variation du contraste des MCs.

Nous proposons d'analyser d'une façon fine le problème de détection des MCs dans le contexte de faible contraste. Une méthode, qui a fait l'objet d'une publication dans une revue scientifique [124] est proposée dans ce chapitre. Basée sur la morphologie mathématique, c'est une nouvelle méthode de détection automatique des MCs, à partir des images mammographiques entières. Le résultat du traitement morphologique appliqué change les aspects d'apparition de certaines structures anatomiques dans l'image y compris les MCs. Ce changement sera à la base de la conception d'une carte de suspicion assurant un nouvel espace de représentation et d'interprétation indépendant de l'incertitude menée dans l'espace des niveaux de gris classique. Cette carte est localement analysée afin d'estimer des seuils adaptatifs permettant d'identifier les zones susceptibles de correspondre à des MCs, appelées 'zones suspectes'.

3.2 Motivation

Les solutions proposées dans la littérature, pour la détection des MCs, font face à des difficultés qui sont essentiellement liées à l'adaptation des méthodes de segmentation à la mammographie. En effet, autre que la méthodologie utilisée, le choix des meilleurs paramètres pour ces méthodes est aussi un facteur important qui influe sur la pertinence des résultats. Dans le contexte des MCs, le contraste faible et la grande similarité des pixels de ces lésions comparé aux pixels des tissus voisins sont à l'origine de ces difficultés.

Pour réduire cette sensibilité, la plupart des méthodes développées procèdent par une étape de prétraitement pour améliorer la qualité des images mammographiques et l'aspect des MCs avant d'appliquer l'étape de détection. Elles utilisent généralement des techniques de modification du contraste [16, 67, 89], de filtrage adaptatif [51, 63, 125] et d'amélioration par une analyse fréquentielle [72, 80], *etc.* Cependant, suite à ces améliorations, les MCs peuvent être confondues avec les pixels des tissus voisins ou denses. Cela

peut induire la génération des fausses détection de type faux positif aussi bien faux négatif.

Le point fort des opérations morphologiques réside dans leur capacité à améliorer la qualité d'apparition des zones séparées par un faible contraste tout en respectant leurs formes géométriques [67]. Ces opérations locales n'auront aucun effet sur la densité mammaire globale dans l'image. Cette caractéristique nous amène à proposer une première méthode basée sur la morphologie mathématique pour la détection automatique des MCs, du fait que la majorité de ces lésions apparaissent avec un contraste local faible par rapport aux pixels voisins et sont caractérisées par une forme géométrique circulaire.

3.3 Morphologie mathématique et détection des MCs

3.3.1 Concepts de base de la morphologie mathématique

Comme nous l'avons précédemment mentionné, la morphologie mathématique est une théorie permettant l'analyse des structures géométriques et basée sur des concepts mathématiques simples issus de la théorie des ensembles [83]. Une opération morphologique, accompagnée par son élément structurant E , est capable de préserver et d'améliorer la qualité d'apparition des objets, sur l'image traitée, ayant un profil similaire à celui de l'élément structurant. La figure 3.2 présente un exemple de quatre éléments structurants de différentes tailles et formes.

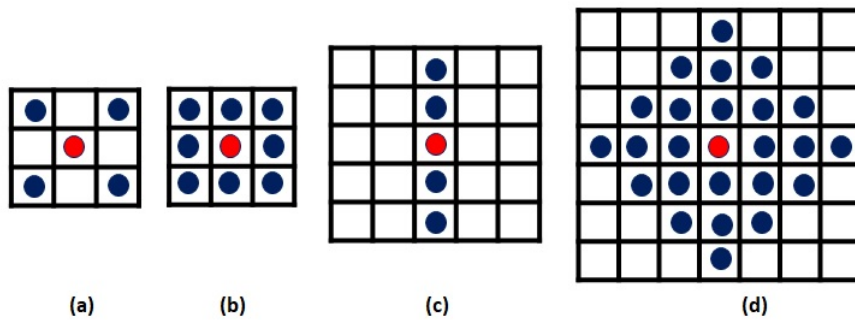


Figure 3.1 – Exemple de quatre éléments structurants associés à un pixel central (cercle rouge)

L'*érosion*, la *dilatation*, l'*ouverture* et la *fermeture* sont les opérations de base de la morphologie mathématique. Toute autre opération, peut être déduite à partir de certaines combinaisons de celles-ci.

Désignant par I une image en niveaux de gris et par E un élément structurant, l'*érosion*, \otimes , respectivement la *dilatation*, \oplus , de I par E sont exprimées par les équations (3.1) et (3.2) :

$$(I \otimes E)(P) = \min\{I(p_i), p_i \in E_P\} \quad (3.1)$$

$$(I \oplus E)(P) = \max \{I(p_i), p_i \in E_P\} \quad (3.2)$$

avec :

"max" et "min" désigne respectivement le *maximum* et le *minimum*. E_P désigne l'élément structurant de centre le pixel P .

De la même manière, les opérations d'*ouverture* (\otimes) et de *fermeture* (\oplus) sont respectivement définies par les équations (3.3) et (3.4) suivantes :

$$(I \circ E)(P) = (I \otimes E) \oplus E \quad (3.3)$$

$$(I \bullet E)(P) = (I \oplus E) \otimes E \quad (3.4)$$

La figure 3.2 illustre des résultats d'application de ces quatre opérations morphologiques élémentaires sur une image de synthèse. Les pixels noirs décrivent l'objet et les pixels blancs décrivent le fond. Seuls les pixels entourés par un contour rouge font partie du nouvel objet résultant.

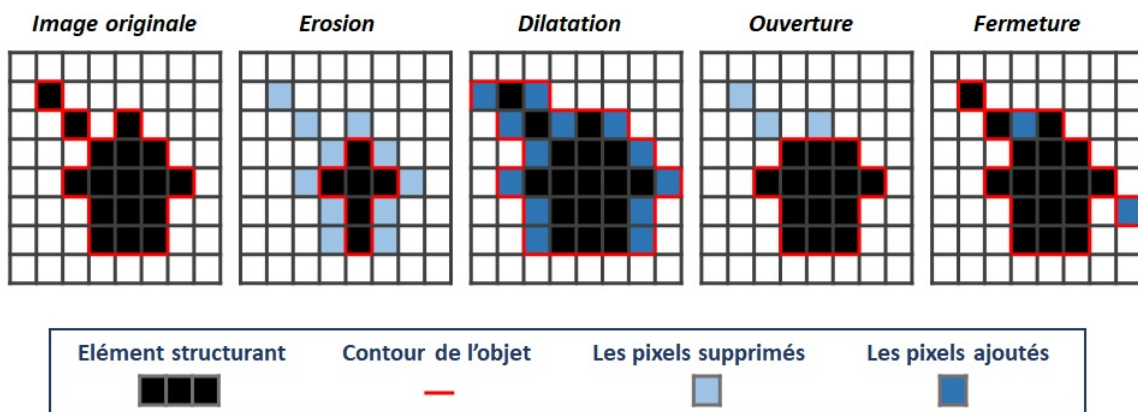


Figure 3.2 – Exemple des résultats des opérations morphologiques élémentaires sur une image de synthèse

3.3.2 Limites

La corrélation entre le profil des éléments structurants utilisés avec celui des objets à chercher à partir d'une image constitue l'élément clé de la fiabilité des méthodes à base de morphologie mathématique. En effet, les éléments structurants ont pour objectif l'amélioration des caractéristiques morphologiques des objets rendant ainsi plus facile le traitement à suivre, par la suite, pour leur détection. Dans le contexte des MCs, les travaux fondés sur la morphologie mathématique ont utilisé des éléments structurants de

forme circulaire et de taille $k \times k$ fixe. Ils étaient basés sur l'hypothèse que ces lésions apparaissent avec une forme circulaire et ne dépassent pas une taille limite bien définie. Cette hypothèse constitue l'une des causes d'échec de ces méthodes. En effet, l'estimation de la valeur de k est effectuée à l'égard de la résolution des images qui diffère d'une base de données à une autre. Par ailleurs, ceci exige un ajustement de la valeur utilisée avec tout changement de base. D'autre part, la diversité anatomique des MCs implique une diversité des tailles d'apparition possibles non seulement d'une mammographie à une autre, mais aussi au sein de la même mammographie. Le caractère multi-tissulaire et la variété des types de tissus de superposition sont à l'origine d'une telle diversité. Partant de ce constat, la détermination de la taille adéquate de l'élément structurant n'est pas triviale et aura un vrai impact sur la qualité de détection.

En outre, suite à l'application des opérations morphologiques, une étape de seuillage pour la détermination des zones suspectes qui constituent les MCs est toujours appliquée. La valeur estimée pour le seuillage est souvent statique et utilisée sur la totalité de la zone analysée. En revanche, l'utilisation d'un seuil statique augmente la sensibilité aux caractéristiques des mammographies (densité mammaire, types de tissus de superposition, ...). En effet, la diversité des tissus mammaires permet à différentes zones de la mammographie de partager les mêmes caractéristiques de forme et de taille des MCs. Cela peut entraîner un grand nombre de fausses détections à savoir des faux positifs, si la valeur du seuil est petite et des faux négatifs si la valeur du seuil est élevée.

Sur la base de toutes ces constatations, nous proposons une nouvelle méthode basée sur la morphologie mathématique pour la détection automatique des MCs. Elle cherche à concevoir une démarche de détection qui profite des avantages des opérations morphologiques et surpasse les points faibles des méthodes précédemment proposées. La description de cette méthode est détaillée dans la section suivante.

3.4 Méthode de détection proposée

L'objectif principal est l'amélioration de la fiabilité de détection des MCs, en termes du taux d'identification ainsi que de précision de segmentation. Les points clés de la méthode proposée concernent essentiellement :

- L'utilisation d'éléments structurants adéquats afin de réduire la sensibilité à la diversité des tailles des MCs ;
- La génération d'une carte de suspicion suite à l'application des opérations morphologiques sur les pixels de l'image mammographique ;

- L'estimation automatique de valeurs de seuils adaptatifs basée sur une analyse locale des degrés de suspicion à partir des régions homogènes en niveaux de gris.

L'organigramme de la méthode proposée est divisé en trois étapes principales (Figure. 3.3) : Une étape d'amélioration de la mammographie numérique, une étape de génération de la carte de suspicion et une étape de détermination des MCs.

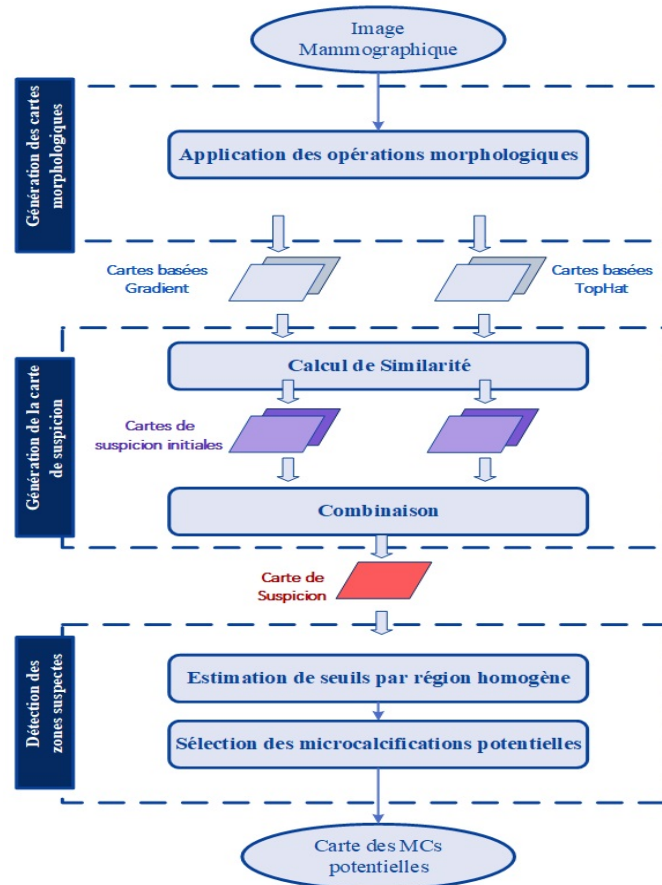


Figure 3.3 – Schéma global de la méthode de détection des MCs proposée

3.4.1 Amélioration de la mammographie

L'hypothèse de base dans l'analyse d'une image, pour la recherche d'un objet donné, consiste à le caractériser comme une région homogène en niveaux de gris avec une forme géométrique bien définie. Cela induit que les pixels sur la frontière de ce dernier présentent une variation de niveaux de gris par rapport aux pixels voisins. La mise en évidence de cette variation peut s'avérer utile pour améliorer la segmentation de ces objets. En revanche, dans le contexte des MCs, le caractère multi-tissulaire des mammographies ainsi la taille très petite de ces lésions sont toujours la cause de la confusion des pixels qui les

caractérisent avec d'autres pixels bruits dans l'image. Ce qui fait, en plus de la variation des niveaux de gris des pixels contours, que l'homogénéité des pixels de la région décrivant une MC doit aussi être incluse dans cette mise en évidence.

Dans l'approche proposée, la première étape se concentre sur deux objectifs principaux. Le premier concerne le rehaussement des contours fins dans l'image. Le deuxième concerne la mise en valeur des structures ayant des profils similaires à ceux des MCs. Pour ce faire, nous proposons d'utiliser une diversité d'éléments structurants et d'opérations morphologiques. L'idée principale, avec cette diversité, est de réduire la sensibilité à la variété de taille et de contraste des MCs.

3.4.1.1 Définition des éléments structurants

a. Motivation

Avec une opération morphologique, un objet spécifique sur une image est correctement détecté sauf si sa taille est inférieure ou égale à celle de l'élément structurant utilisé. La figure 3.4 montre un exemple de trois MCs individuelles de différentes tailles. Les contours discontinus, avec les couleurs respectivement jaune et bleu, font référence à deux éléments structurants, respectivement SE_1 et SE_2 , de tailles différentes. Les contours continus (rouge) désignent les contours vérité de terrain (GT) délimitant les MCs.

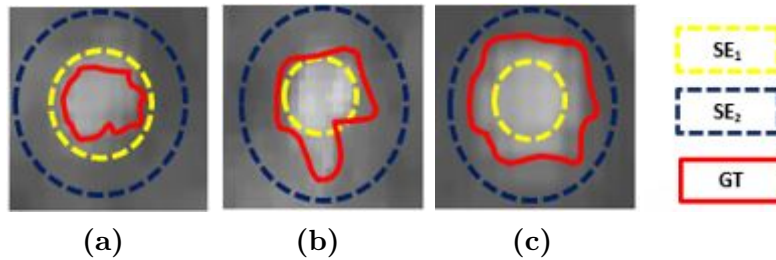


Figure 3.4 – Illustration d'un exemple de MCs de différentes tailles

Concentrons-nous sur la première MC (Figure. 3.4a), sa taille est inférieure à celle des deux éléments structurants (SE_1 et SE_2). Son contour sera ainsi mis en valeur et correctement détecté par l'utilisation de ces deux éléments. Par contre, les contours des deux autres MCs (Figure. 3.4b et Figure. 3.4c) apparaissent dans la zone de transition entre le premier (SE_1) et le deuxième (SE_2) élément structurant. Leurs contours seront correctement détectés par l'élément structurant le plus grand (SE_2) mais pas par le plus petit (SE_1).

En partant de ces constatations, nous pouvons conclure que l'utilisation d'un seul élément structurant peut ne pas être fiable pour la détection de toutes les MCs. En effet, il peut générer des fausses alarmes si sa taille est trop petite, voir masquer des vraies

alarmes (des vrais MCs) si sa taille est trop grande. C'est pour cette raison que nous avons proposé l'utilisation de plusieurs éléments structurants avec des tailles et de formes variables.

b. Choix des éléments structurants

Selon les radiologues, les MCs apparaissent avec un diamètre qui ne dépasse pas 1 mm. Du fait qu'il soit primordial d'améliorer l'aspect d'apparition des MCs, nous devons correctement choisir les tailles et les formes des éléments structurants en fonction des résolutions des mammographies numériques. La résolution des images de la base INBreast est de l'ordre de 70 microns/pixel. Ainsi, le nombre de pixels qui représente une MC est, généralement, compris entre 8 pixels (pour une taille de 0.5 mm) et 15 pixels (pour une taille de 1 mm).

Les éléments structurants que nous avons utilisés dans cette méthode sont illustrés dans la figure. 3.5. Leurs tailles sont de l'ordre de 3×3 (resp. 5×5) pour B_{11} et B_{21} (resp. B_{12} et B_{22}) et 5×5 et 7×7 pour B_3 afin d'assurer des masques plus grands que les MCs.

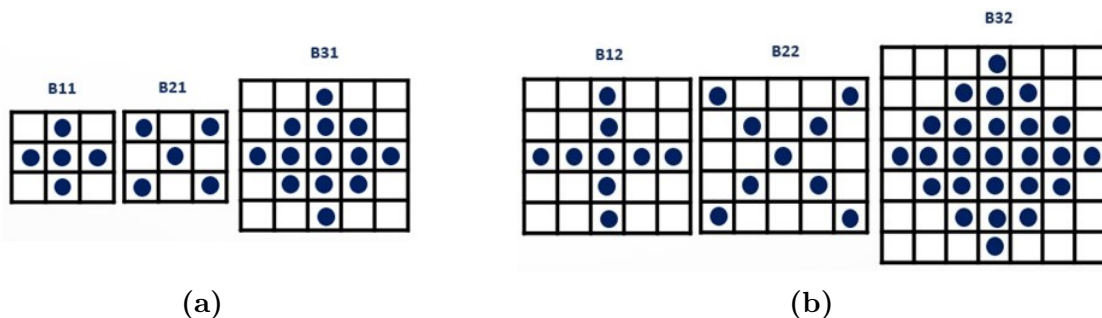


Figure 3.5 – Les éléments structurants utilisés par notre première contribution

Bien que les éléments structurants que nous avons définis soient de différentes tailles, ils présentent aussi des profils différents et seront donc utilisés pour des finalités complémentaires. Nous allons décrire, dans le paragraphe suivant, l'ensemble des opérations morphologiques à appliquer, avec leurs éléments structurants associés. Nous allons également discuter le principe et l'intérêt d'en combiner plusieurs.

3.4.1.2 Traitement morphologique appliqué

La particularité de la morphologie mathématique consiste à dégrader l'information contenue dans l'image qu'elle traite afin de ne préserver que l'information pertinente. Malgré cette dégradation, elle cherche à améliorer les détails relatifs aux objets recherchés dans l'image. La dégradation et la mise en évidence des détails de l'image sont toujours faites par les mêmes opérations qui tiennent compte de la géométrie ainsi que des connais-

sances *a priori* sur la forme des objets à rechercher.

Avec la morphologie mathématique, les transformations définies varient d'une opération 'basique' simple à une combinaison de deux ou plusieurs opérations. Elles partagent l'objectif de préserver au mieux l'information utile dans une image et de faire disparaître autant que possible le bruit. L'information utile préservée décrit les objets recherchés.

Outre les opérateurs morphologiques basiques, les résidus morphologiques sont définis comme une deuxième famille de transformations morphologiques. Ils font appel à des opérations morphologiques basiques combinées avec une différence. Les résidus morphologiques les plus connus sont : les *gradients morphologiques* et les *chapeaux haut de forme*.

a. Gradient morphologique

Les opérateurs basiques privilégient la notion de forme plutôt que l'information sur les intensités des pixels par le traitement appliqué. En revanche, le gradient morphologique est destiné à extraire les zones de variation d'intensité locale dans une image qui coïncident généralement avec les contours des objets. Il permet, ainsi, de mettre en évidence les contours locaux dans l'image traitée en tenant compte des profils des éléments structurants utilisés.

On peut distinguer trois différentes combinaisons utilisées pour la définition d'un gradient morphologique :

- Le gradient *symétrique* défini comme la différence entre les résultats de la dilata-tion \oplus et de l'érosion \otimes appliquées sur l'image originale I en utilisant l'élément structurant B ;

$$G_S(I) = (I \oplus B) - (I \otimes B) \quad (3.5)$$

- Le gradient *interne* défini comme la différence entre l'image originale I et l'image érodée \otimes par l'élément structurant B ;

$$G_S(I) = I - (I \otimes B) \quad (3.6)$$

- Le gradient *externe* défini comme la différence entre l'image dilatée \oplus et l'image originale I par le même élément structurant B ;

$$G_S(I) = (I \oplus B) - I \quad (3.7)$$

b. Chapeau haut de forme

Les transformations *chapeaux haut de forme* (dites aussi *TopHat*) sont définies comme le résultat de la différence entre l'image originale I et l'ouverture ou la fermeture de la même image par l'élément structurant B . Le premier type, appelé *chapeau haut de forme blanc*, permet de détecter les zones localement claires dans l'image répondant au profil de l'élément structurant utilisé. En revanche, le deuxième type est appelé *chapeau haut de forme noir* et s'intéresse à détecter les zones localement sombres dans l'image traitée.

$$TopHat_{Blanc}(I) = I - I \circ B \quad (3.8)$$

$$TopHat_{Noir}(I) = I - I \bullet B \quad (3.9)$$

Compte tenu de la correspondance entre les objectifs des opérations morphologiques discutées et la caractérisation des MCs, les opérations du gradient morphologique et en particulier le chapeau haut de forme blanc peuvent être très utiles pour la détection des MCs. En effet, ces lésions définissent des petites zones localement claires, de forme bien déterminée avec un contour souvent flou. L'application du gradient morphologique sur l'image permet de mettre en évidence les pixels contours. Cela induit la réduction du caractère flou entre les objets dans l'image, y compris les MCs, et leurs tissus de superposition. D'autre part, l'application de l'opération *TopHat* permet d'améliorer le contraste des régions claires ayant un profil proche de celui des éléments structurants (les MCs). Cela induit aussi le rehaussement du contraste entre les pixels caractérisant la région MC et ceux des tissus mammaires voisins.

3.4.1.3 Génération des cartes morphologiques

Dans la méthode proposée, eux opérations morphologiques avec les éléments structurants définis auparavant (Figure. 3.5) sont appliquées afin de générer des cartes morphologiques à utiliser dans le reste du traitement. Nous avons distingué un couple d'éléments structurants (C_{B1}, C_{B2}) . C_{B1} est le triplet qui regroupe les éléments structurants B_{11} , B_{21} et B_{31} tandis que C_{B2} est le triplet qui regroupe les éléments B_{12} , B_{22} et B_{32} (Figure 3.5). Les éléments décrits par ce couple présentent des formes similaires mais des tailles différentes. Cette différence permet à la fois de gérer la variété des tailles des MCs et d'accentuer davantage les structures rehaussées. En contre partie, les éléments qui définissent chaque triplet, dans ce couple, présentent des tailles et des formes différentes. Ils sont utilisés avec des opérations différentes, pour des finalités complémentaires.

Dans un premier temps, nous appliquons l'opération du gradient morphologique, sur

l'image, avec les éléments B_{11} et B_{21} de C_{B1} respectivement B_{12} et B_{22} de C_{B2} . Le résultat est un ensemble de quatre cartes de gradients morphologiques G_{11} , G_{21} , G_{12} et G_{22} décrites par les équations (3.10), (3.11), (3.12) et (3.13).

$$G_{11}(I) = (I \oplus B_{11}) - (I \otimes B_{11}) \quad (3.10)$$

$$G_{21}(I) = (I \oplus B_{21}) - (I \otimes B_{21}) \quad (3.11)$$

$$G_{12}(I) = (I \oplus B_{12}) - (I \otimes B_{12}) \quad (3.12)$$

$$G_{22}(I) = (I \oplus B_{22}) - (I \otimes B_{22}) \quad (3.13)$$

Par la suite, les deux premières cartes morphologiques F_1 et F_2 sont obtenues en faisant la somme de l'image originale avec les cartes de gradients G_{11} et G_{21} (Figure 3.6). Cette opération s'appelle *dualité* et elle permet :

- D'associer un effet flou aux régions homogènes et de réduire la fluctuation des niveaux de gris dans l'image traitée ;
- D'augmenter le contraste faible entre les régions de l'image traitée y compris les MCs et les tissus voisins.

$$F_1(I) = I + G_{11}(I) + G_{21}(I) \quad (3.14)$$

$$F_2(I) = I + G_{12}(I) + G_{22}(I) \quad (3.15)$$

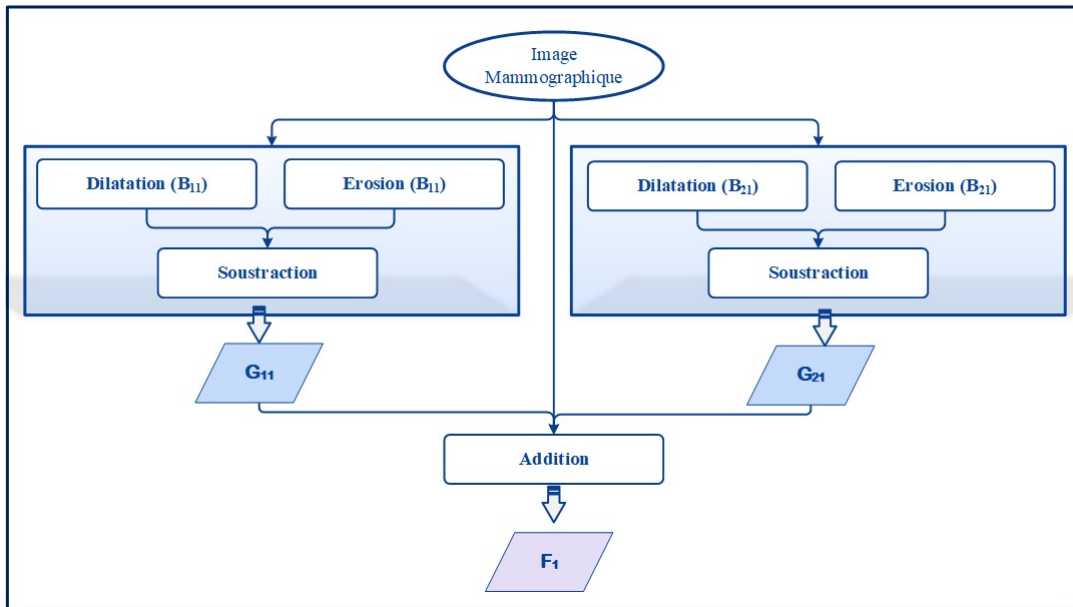


Figure 3.6 – Génération de la carte F_1 à partir d'une image mammographique

En parallèle avec les cartes F_1 et F_2 obtenues par l'application de la dualité, nous appliquons l'opération *TopHat* avec les éléments structurants B_{31} (Equation (3.17)) respec-

tivement B_{32} (Équation (3.16)) sur l'image originale. Cette opération permet de rehausser les zones localement claires dont la taille est inférieure à celle de l'élément structurant. Ces zones constituent les MCs ainsi que les régions dans l'image avec des profils similaires.

$$Top_1(I) = I - (I \circ B_{31}) \quad (3.16)$$

$$Top_2(I) = I - (I \circ B_{32}) \quad (3.17)$$

3.4.2 Génération de la carte de suspicion

Les cartes morphologiques précédemment obtenues (F_1, F_2, Top_1 et Top_2) sont utilisées, dans le cadre de la méthode proposée, pour fournir une source d'information alternative appelée carte de suspicion (SM). C'est une carte continue, de valeurs comprises entre 0 et 1, qui associe un degré de suspicion pour chaque pixel de l'image. Un pixel P avec un degré de suspicion proche de zéro correspond à un pixel du tissu mammaire sain. En revanche, un pixel avec un degré de suspicion élevé (proche de 1) décrit un pixel d'un tissu fortement suspect (MC ou autre lésion). Il est intéressant de noter que cette carte de suspicion proposée constitue une sorte de transformation de la représentation classique de l'image en niveaux de gris vers un nouvel espace de représentation semi-sémantique. Elle offre l'opportunité d'explorer des informations sémantiques, dans le processus de prise de décision, et de réduire l'impact des informations "incertaines" en niveaux de gris.

3.4.2.1 Motivation

L'utilisation des opérations morphologiques est dans l'idée d'altérer les niveaux de gris des pixels des zones suspectes sur l'image mammographique. La génération de la carte de suspicion repose essentiellement sur la quantification des changements apportés sur ces pixels. Elle peut être quantifiée par une simple soustraction de l'image altérée par rapport à l'image originale. Pour une évaluation plus sophistiquée, nous proposons d'utiliser l'indice de similarité structurelle SSIM (*Structural SIMilarity index*) [126] qui est une métrique permettant de quantifier la similarité entre deux images. Cet indice a été initialement conçu afin de mesurer la qualité visuelle d'une image transformée (souvent compressée) par rapport à une image originale de référence. L'évaluation de la qualité de l'image, comparée à une image objective de référence, est estimée en prenant en compte la dégradation de ses informations structurelles. Elle repose sur le fait que les paramètres de luminance (l), contraste (c) et structure (s) sont les plus adaptés pour extraire les informations structurelles d'une image donnée [127]. C'est cette combinaison de paramètres qui nous a motivé à choisir cet indice de similarité pour l'estimation de la carte de suspicion.

3.4.2.2 Indice de similarité structurelle (SSIM)

L'expression de l'indice SSIM, entre deux images I et J , est définie comme la combinaison pondérée de ses trois paramètres comparatifs (l , c , et s). (Équation (3.18)). Chacun de ces paramètres est calculé pour chaque pixel p de l'image en utilisant une fenêtre de taille $N \times N$ (souvent 8×8 ou 11×11) centrée sur le pixel considéré :

$$SSIM(I, J) = l(I, J)^\alpha \cdot c(I, J)^\beta \cdot s(I, J)^\gamma \quad (3.18)$$

$$l(I, J)^\alpha = \frac{2\mu_I\mu_J + k_1}{\mu_I^2 + \mu_J^2 + k_1} \quad (3.19)$$

$$c(I, J)^\beta = \frac{2\sigma_I\sigma_J + k_2}{\sigma_I^2 + \sigma_J^2 + k_2} \quad (3.20)$$

$$s(I, J)^\gamma = \frac{\sigma_{IJ} + k_3}{\sigma_I\sigma_J + k_3} \quad (3.21)$$

où :

- μ_I , μ_J et σ_I , σ_J désignent respectivement la moyenne locale et l'écart type des deux images I et J (par rapport à une fenêtre locale) ;
- σ_{IJ} fait référence à la covariance croisée entre la fenêtre locale des images I et J ;
- α , β , γ : sont trois paramètres de combinaison positif réel utilisés pour ajuster l'importance associée à chacun des termes ;
- k_1 , k_2 et k_3 : sont trois valeurs constantes permettant de stabiliser l'indice SSIM lorsque les moyennes et les variances deviennent faibles.

3.4.2.3 Génération des sources d'information : Cartes de suspicion initiales

L'application des opérations morphologiques basées sur les éléments structurants définis auparavant permet de générer quatre différentes cartes : les cartes F_1 et Top_1 en utilisant les éléments structurants dans C_{B1} et les cartes F_2 et Top_2 en utilisant les éléments structurants dans C_{B2} . En quantifiant leurs similarités, une par une, avec l'image originale, nous obtenons quatre cartes $SSIM_M$ (Équation (3.22)) de valeurs continues. Ces cartes sont considérées comme des cartes de suspicion initiales fournissant chacune une source d'information précieuse à explorer pour la génération de la carte de suspicion finale.

$$SSIM_M = SSIM(M, I), \text{ où } M \in \{F_1, F_2, Top_1, Top_2\} \quad (3.22)$$

Avant d'être transformées en cartes de suspicion, les cartes morphologiques sont normalisées en utilisant la méthode de mise à l'échelle min-max, donnée par l'équation (3.23)

ci-dessous :

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (3.23)$$

où x fait référence à la valeur originale et x' désigne la nouvelle valeur normalisée.

3.4.2.4 Fusion des sources d'information : Génération de la carte de suspicion

Une fois obtenues, les cartes sont combinées afin de générer une carte de suspicion finale sur laquelle une dernière étape de seuillage adaptatif est appliquée (Figure. 3.7).

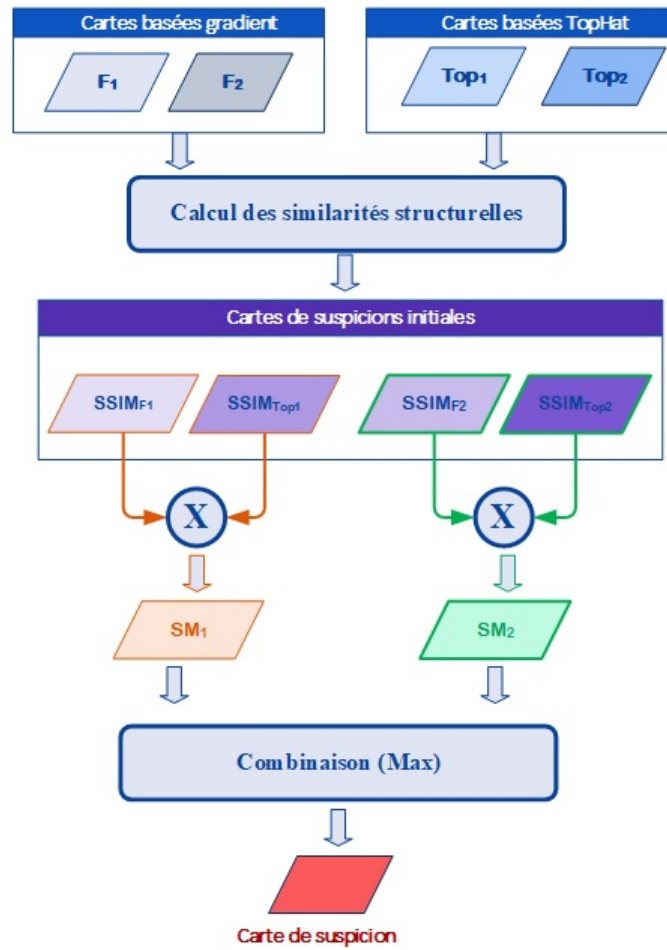


Figure 3.7 – Le schéma de la combinaison adoptée

Les profils des éléments structurants utilisés sont conformes aux tailles et formes des MCs. Ainsi, les quatre cartes de suspicion initiales présentent des degrés élevés dans les régions qui correspondent à des MCs ou d'autres régions ayant des profils similaires. Pour les transformer en une seule carte de suspicion, les quatre cartes initiales sont soumises à deux combinaisons successives. La première est appliquée sur les cartes issues d'un même triplet d'éléments structurants, à savoir $SSIM_{F1}$ et $SSIM_{Top1}$ de C_{B1} et $SSIM_{F2}$ et $SSIM_{Top2}$ de C_{B2} . Cette combinaison donne naissance à deux cartes SM_1 et SM_2 dé-

finies par les équations (3.24) et (3.25) en tant que résultat de l'application du produit probabiliste. Ce dernier figure comme un opérateur de combinaison conjonctif. Son utilisation repose sur l'idée que l'information gardée doit être affirmée par les différentes sources d'informations [128]. Cela correspond bien aux caractéristiques des cartes combinées. En effet, l'opération de dualité et *TopHat* cherchent, toutes les deux, à rehausser les régions homogènes avec un profil similaire à celui des MCs. Une région, à partir de ces cartes, ne sera retenue que si elle présente une concordance par les deux sources combinées.

$$SM_1 = SSIM_{F1} * SSIM_{Top1} \quad (3.24)$$

$$SM_2 = SSIM_{F2} * SSIM_{Top2} \quad (3.25)$$

La deuxième combinaison est fondée sur l'opérateur disjonctif et est appliquée pour combiner les cartes SM_1 et SM_2 . Dans ce travail, ces deux cartes cherchent à présenter des structures similaires en formes mais différentes en tailles. Ainsi, l'idée derrière le choix de cet opérateur est de préserver le maximum d'informations qui pourraient être explorées, ce qui permet d'apporter de nouvelles informations sur les régions que nous essayons de détecter (les MCs). Cet opérateur est généralement adapté à la fusion de sources d'informations non concordantes [129]. Il est basé sur le fait qu'elles ne sont pas totalement fiables et qu'elles peuvent apparaître avec un certain conflit entre elles. Le fruit de cette combinaison est la carte de suspicion finale SM exprimée par l'équation (3.26).

$$SM(P) = \max\{SM_1(P), SM_2(P)\} \quad (3.26)$$

3.4.2.5 Détection par seuillage adaptatif

Afin de détecter les régions de MCs potentielles, que nous appelons des zones suspectes, un seuil peut être appliqué à la carte de suspicion générée par l'étape précédente. Cependant, compte tenu du contraste faible des MCs par rapport au tissu de superposition et le caractère multi-tissulaire de la mammographie, il est préférable de suivre une analyse locale afin d'estimer des valeurs de seuils précises au lieu d'utiliser une seule valeur statique. Pour cela, nous proposons de diviser l'image originale en régions locales homogènes appelées superpixels. Un superpixel est défini comme un groupe de pixels connectés ayant des caractéristiques similaires en niveaux de gris [130]. Une telle division permet de réduire le caractère multi-tissulaire dans l'image du fait que chaque superpixel présente nécessairement le même type de tissu mammaire et par conséquent des pixels avec des niveaux de gris homogènes. D'autre part, elle offre la possibilité de bénéficier des avantages d'une analyse locale pour la recherche des MCs tout en préservant les connaissances globales sur l'image.

Dans ce travail, nous proposons l'utilisation de l'algorithme de référence SLIC [131] (*Simple Linear Iterative Clustering*) pour la génération des superpixels. Proposé par Achanta *et al.*, il décrit une version modifiée de l'algorithme de k -moyennes (K-Means). Sa contribution principale concerne la mesure de similarité utilisée qui regroupe à la fois la similarité en niveaux de gris et en proximité géométrique. Cet algorithme ainsi que la notion de superpixels ont été utilisés à plusieurs reprises dans ce travail de recherche avec des finalités différentes. Le dernier chapitre (chapitre. 5) présentera une description détaillée de l'algorithme SLIC et discute ses avantages qui nous ont conduit à le choisir.

La projection des contours des superpixels sur la carte de suspicion nous offre la possibilité d'exploiter l'impact des opérations morphologiques appliquées, traduit en degrés de suspicion, sur des zones localement homogènes référant à des pixels d'un même type de tissu mammaire.

Censés représenter le même type de tissu mammaire, les pixels du même superpixel doivent, logiquement, présenter des degrés de suspicion avec des valeurs similaires. Les pixels adjacents, dans un superpixel, qui pourraient être considérés comme des pixels de MCs, sont ceux qui ne remplissent pas le critère des pixels 'normaux'. En d'autres termes, ce sont les pixels *aberrants* (dits aussi *outliers*) dont les degrés de suspicion associés sont légèrement écartés par rapport à la majorité des pixels dans le superpixel considéré. Partant de cette idée, et dans l'objectif d'identifier ces pixels, nous avons eu recours à la technique de boîte à moustaches (*boxplot*) pour représenter la dispersion des valeurs des degrés de suspicion afin de pouvoir identifier les pixels aberrants. Boxplot (Figure. 3.8) est un graphique qui fournit un résumé visuel sur la dispersion des valeurs dans un ensemble de données. Parmi les éléments qu'il représente, la médiane (Q_2), le premier (Q_1) et le troisième (Q_3) quartile ainsi les deux extrémités inférieure et supérieure [132]. Il indique également l'ensemble des valeurs aberrantes.

La description de ces éléments est donnée comme suit :

- Q_1 , respectivement Q_3 , représente la plus petite valeur dans l'ensemble de données telle qu'au moins 25%, respectivement 75%, des données présentent des valeurs inférieures ou égales à Q_1 respectivement Q_3 ;
- L'écart inter quartile (IQR : *Inter Quartile Range*) est la distance entre les quartiles inférieur et supérieur (Q_1 et Q_3) ;
- Les extrémités inférieure et extérieure sont situées aux intervalles $[Q_1 - 1,5 \times IQR, Q_1]$ et $[Q_3, Q_3 + 1,5 \times IQR]$.

Une donnée aberrante est caractérisée par sa valeur distante du reste des données. Avec

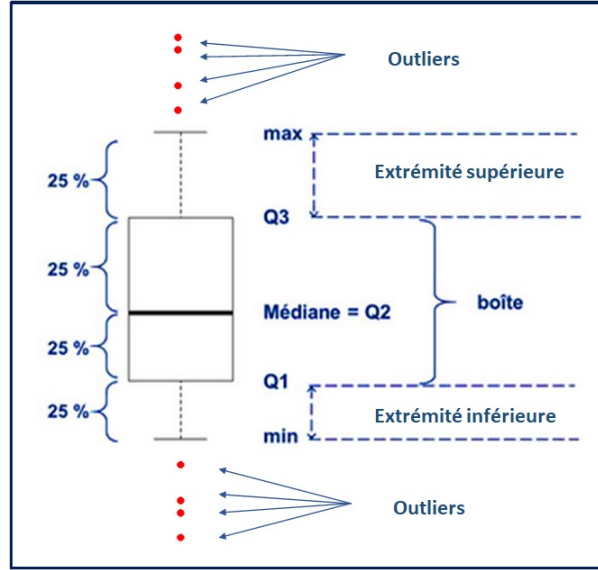


Figure 3.8 – Schéma descriptif d'une boîte à moustache (Box-plot)

les méthodes standards, l'identification de ces données est souvent basée sur le calcul de la moyenne et la variance sur les valeurs dans tout l'ensemble de données. Avec les boîtes à moustaches, l'identification est fondée sur le critère de John Wilder Tukey qui est moins sensible aux valeurs extrêmes [133]. Elle considère une donnée X_i comme aberrante si elle se trouve au-delà des extrémités inférieures ou supérieures (Équations (3.27) et (3.28)) de la boîte à moustaches.

$$X_i < Q_1 - 1.5 \times IQR \quad (3.27)$$

où

$$X_i > Q_3 + 1.5 \times IQR \quad (3.28)$$

Dans notre travail, l'ensemble de données que nous traitons correspond aux degrés de suspicion dans la carte SM définie pour chaque superpixel généré. Les pixels qui décrivent une MC apparaissent avec des valeurs de niveaux de gris proches, mais plus élevées, que leurs pixels voisins. Sur la carte de suspicion, ces pixels sont soumis à des modifications suite à l'application des opérations morphologiques. Par conséquent, ils présentent les degrés de suspicion plus élevés par rapport aux autres pixels. Ceci implique qu'un pixel aberrant P référant à un pixel d'une MC ne peut pas satisfaire la première condition prévue dans l'équation (3.27).

D'autre part, les superpixels analysés décrivent des tissus mammaires homogènes et présentent des valeurs de niveau de gris similaires. Ainsi, l'intervalle de variation des degrés de suspicion pour un superpixel donné est souvent faible. Cela induit que la valeur supérieure ($Q_3 + 1,5 \times IQR$) peut être supérieure ou égale au degré de suspicion le plus

élevé dans le superpixel traité. Les pixels P , qui correspondent à des pixels des MCs, pourraient alors présenter des degrés de suspicion inférieurs à cette dernière et ne pas satisfaire à la deuxième condition indiquée par l'équation (3.28). Malgré qu'une telle valeur s'avère efficace pour la réduction du taux de génération des faux positifs, elle augmente le risque de perte des vrais MCs et donc le taux de génération des faux négatifs. Du fait que notre objectif à ce niveau soit de réduire au maximum le nombre de faux négatifs, nous avons, donc, choisi d'utiliser la valeur minimale Q_3 dans l'intervalle défini par l'extrémité supérieure comme une valeur de seuil. L'efficacité de ce choix a été justifiée par la pertinence des résultats obtenus. Un pixel P est donc considéré comme un pixel aberrant si et seulement s'il présente un degré de suspicion $SM(P)$ supérieur à la valeur seuil estimée (Équation (3.29)) pour le superpixel auquel il appartient. Les pixels aberrants adjacents groupés constituent une zone suspecte dans la carte de détection pixelique finale, qui correspond à une MC potentielle.

$$SM(P) \geq Q_3 \quad (3.29)$$

3.5 Méthodologie de validation

Dans cette section, nous allons présenter la méthodologie suivie pour évaluer les performances de la méthode que nous venons de proposer. L'objectif principal de cette évaluation porte sur la fiabilité des résultats obtenus en termes d'identification et de segmentation des MCs.

Pour cette évaluation, nous avons utilisé l'ensemble des 50 images mammographiques retenues à partir de la base INBreast. Pour chaque image, le résultat du processus de détection appliqué, modélisé sous la forme d'une carte binaire, est comparé avec la vérité de terrain (GT) fournie par le radiologue (Chapitre. 1).

3.5.1 Métriques d'évaluation des performances

En vue d'évaluer les performances de la méthode proposée, nous proposons de comparer le masque des MCs potentielles obtenu (appelé masque candidat) avec le masque vérité de terrain (GT) généré et validé par l'expert en radiologie. Ces deux masques sont des cartes pixeliques avec des valeurs binaires (0 ou 1). Les pixels avec une valeur nulle (0) désignent les pixels du tissu mammaire sain dans les deux masques. En revanche, les pixels avec une valeur égale à 1 désignent les pixels caractérisant les vrais MCs dans le masque GT et supposés appartenir à des MCs dans le masque candidat. Cette superposition nous ramène à distinguer trois types de pixels (Figure. 3.9) :

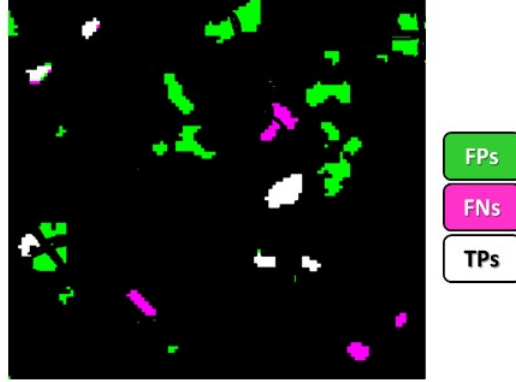


Figure 3.9 – Représentation des différents types de pixels distingués pour l'évaluation d'une segmentation

- Les pixels *Vrai Positif* (*True Positive* : TP) : Il s'agit des pixels MCs qui sont correctement détectés par la méthode proposée. En d'autres termes, il s'agit de l'ensemble des pixels MCs présentant un accord positif entre le masque GT et le masque candidat ;
- Les pixels *Faux Positif* (*False Positive* : FP) : Il s'agit de l'ensemble des pixels du fond identifiés, à tort, comme des pixels MCs par la méthode proposée ;
- Les pixels *Faux Négatif* (*False Negative* : FN) : Il s'agit des pixels MCs identifiés, à tort, comme des pixels du fond par la méthode proposée.

Étant donnée cette catégorisation, plusieurs mesures peuvent être définies et utilisées. Dans ce travail, le choix des mesures d'évaluation a été effectué compte tenu des objectifs de base annoncés. Elles concernent essentiellement deux critères d'évaluations. Le premier critère est lié à la fiabilité relative de l'identification. Tandis que le deuxième critère est lié à la qualité de la segmentation.

La sensibilité SE (*Sensitivity*) est la métrique la plus utilisée pour l'évaluation de la fiabilité relative à l'identification. Elle est parfois appelée le *taux de vrais positifs*. Elle représente la proportion des pixels correctement identifiés (TP) par rapport à ceux qui doivent être identifiés (TP+FN) (Équation (3.30)).

$$SE = \frac{TP}{TP + FN} \quad (3.30)$$

Pour l'étude de la fiabilité de segmentation, les trois métriques retenues sont l'indice de Dice (*Similarity Index* : SI), le taux de sous-segmentation (*Overlap Fraction* : OF) et le taux de sur-segmentation (*Extra Fraction* : EF). Elles comparent le résultat de la détection obtenue à la vérité de terrain fournie.

Soit S_C l'ensemble des pixels considérés comme des pixels de MCs par la méthode de

détection appliquée et S_G l'ensemble des vrais pixels de MCs à partir du masque GT qui nous sert de référence. Les métriques utilisées sont estimées sur la base de l'intersection et l'union des régions décrites par S_C et S_G de cardinal $|S_C|$ et $|S_G|$.

L'indice de Dice, connu aussi sous le nom d'indice de similarité, est défini comme le rapport des pixels d'intersection de S_C et S_G par rapport à ceux de l'union (Équation (3.31)). Plus la valeur de SI est proche de 1, meilleure est la qualité de superposition.

$$SI = \frac{2 * |S_C \cap S_G|}{|S_C| + |S_G|} \quad (3.31)$$

Le taux de sous-segmentation (OF) mesure la proportion des pixels MCs correctement détectés par la méthode appliquée par rapport à ceux du masque GT de référence (Équation (3.32)).

$$OF = \frac{|S_C \cap S_G|}{|S_G|} \quad (3.32)$$

Le taux de sur-segmentation (EF) mesure la proportion des pixels de fond considérés, à tort, comme des pixels MCs par la méthode appliquée par rapport aux pixels MCs du masque GT de référence (Équation (3.33)).

$$EF = \frac{|S_C \cap \overline{S_G}|}{|S_G|} \quad (3.33)$$

Plus la valeur de OF (respectivement EF) est proche de 1 (respectivement 0) plus la segmentation est considérée comme fiable.

3.5.2 Méthode ('Zh') retenue pour comparaison

Dans l'objectif d'évaluer les performances de cette contribution, une comparaison avec des travaux de la littérature est requise. En revanche, en raison de la variété des bases de mammographies utilisées par les travaux de recherche ainsi que les objectifs d'évaluations qu'ils présentent, il est difficile de faire une comparaison objective. Pour remédier à cette difficulté, nous avons implémenté dans ce travail de recherche la(les) méthode(s) retenue(s) pour la comparaison. De cette manière, nous pouvons garantir une évaluation équitable et objective.

Pour cette méthode proposée, nous avons choisi la méthode proposée par Zhang *et al.* [2]. Ce travail est basé sur la combinaison des résultats des opérations morphologiques pour détecter les MCs à partir d'une région d'intérêt sélectionnée par les radiologues. Le processus de recherche des MCs commence par améliorer le contraste de la région d'intérêt traitée par l'application de la correction Gamma. Par la suite, l'application des opérations morphologiques génère deux cartes morphologiques. Sur ces cartes, un seuillage statique avec une combinaison disjonctive sont appliqués pour créer la carte binaire représentant les MCs. Ce choix est justifié par le fait que la morphologie mathématique est le concept de base sur lequel les deux méthodes sont fondées.

3.6 Évaluation qualitative des résultats

Dans ce paragraphe, nous allons présenter et discuter l'ensemble des résultats obtenus à chaque étape du processus de détection appliqué par la méthode proposée. L'image mammographique utilisée pour la présentation des résultats est donnée par la figure 3.10.

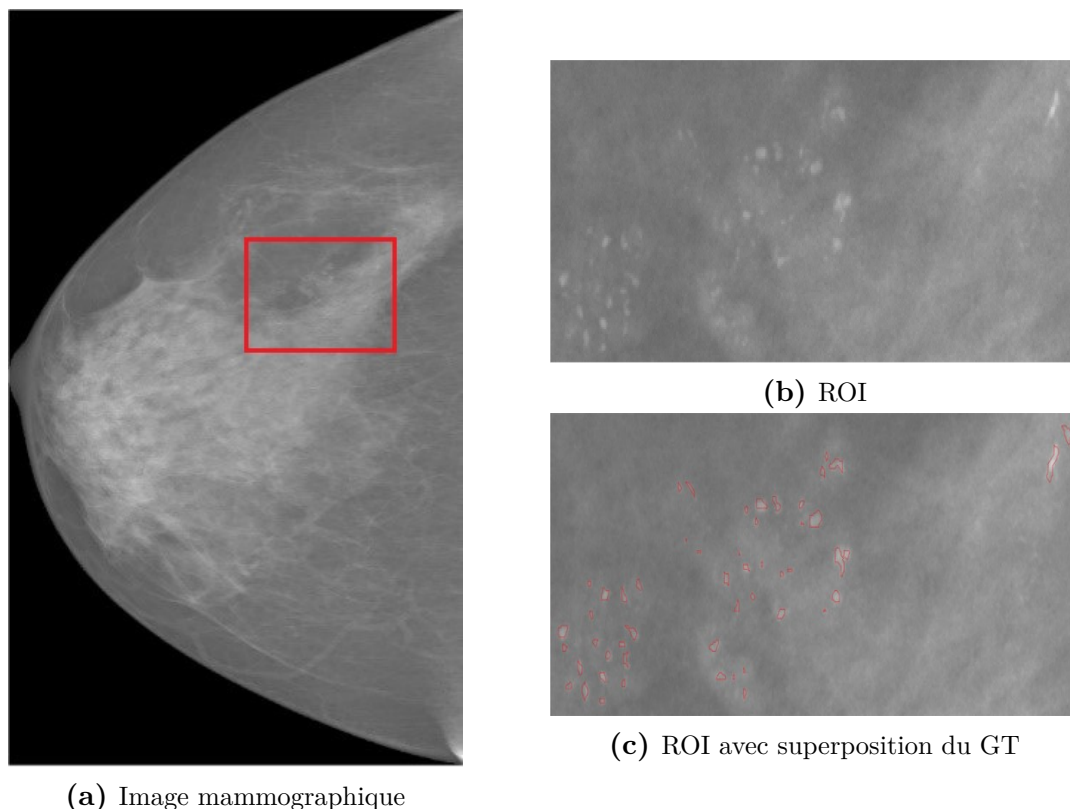


Figure 3.10 – Exemple d'une image mammographique (a) de la base INBreast utilisée dans l'évaluation, avec un zoom sur la ROI (b) où figure les MCs et une superposition des contours de la vérité de terrain (c)

Il est intéressant de mentionner que tous les traitements ont été appliqués sur l'image

mammographique entière. Par contre, les résultats exposés dans les figures, tout au long de la partie suivante, concernent seulement les régions d'intérêts (ROIs) là où figurent les MCs. Nous avons opté pour ce mode de visualisation afin de permettre une meilleure appréciation des MCs et des détails des images mammographiques.

L'image illustrée par la figure précédente montre la présence d'un foyer de MCs réparti en trois groupes qui comportent des MCs avec une diversité de caractéristiques à savoir la taille, la forme et le contraste. On constate également une grande diversité des tissus de superposition.

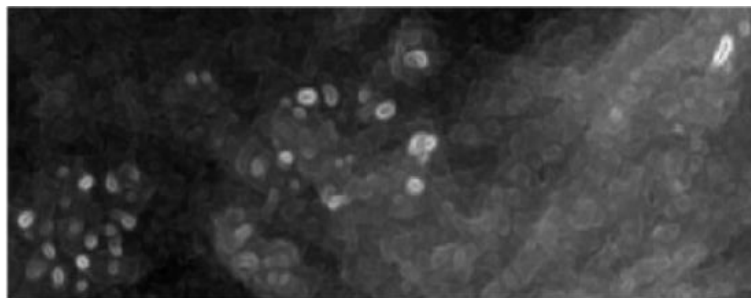
L'utilisation de cette image pour la présentation des résultats permet de :

- Justifier l'utilité des opérations effectuées par rapport aux objectifs désirés ;
- Étudier la sensibilité de la méthode proposée face à la diversité des caractéristiques des MCs (spécifiquement à la diversité du contrastes et de tailles).

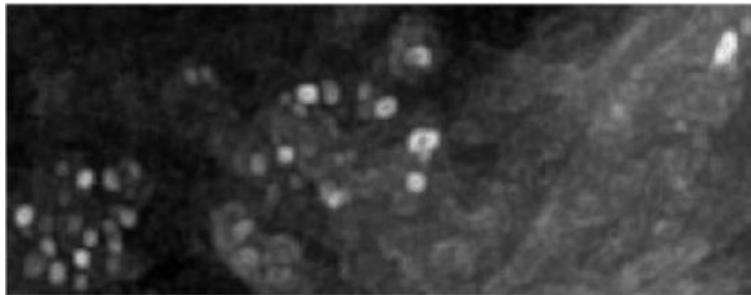
3.6.1 Fiabilité de détection

3.6.1.1 Évaluation de l'étape d'amélioration de la mammographie

Rappelons que l'objectif principal de cette première étape est de transformer l'image originale en niveaux de gris en cartes morphologiques qui rehaussent l'apparition des MCs. Un exemple de génération de ces cartes est illustré par la figure 3.11. Chaque ligne dans cette figure montre une des cartes morphologiques obtenues.



(a) La carte morphologique F_1



(b) La carte morphologique F_2

Figure 3.11 – Les cartes morphologiques générées suite à l'application des opérations morphologiques par la première étape de la contribution proposée

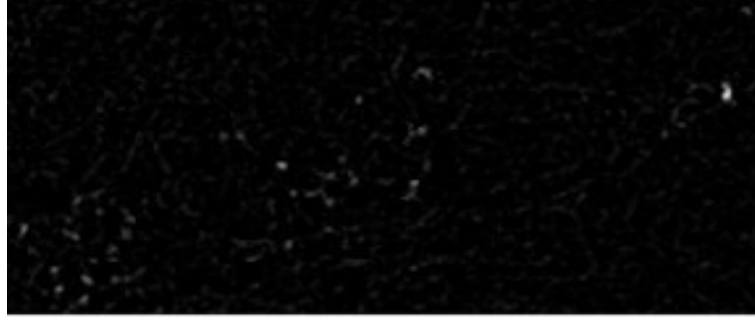
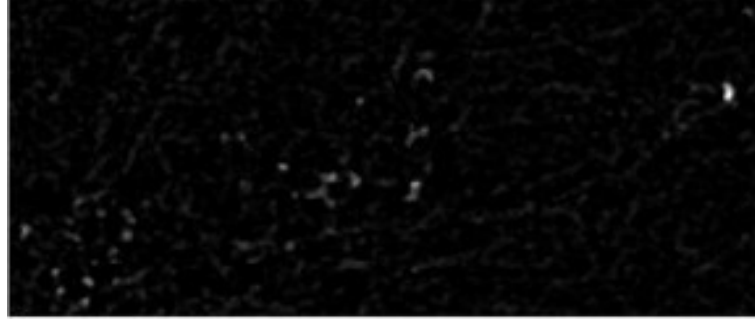
(c) La carte morphologique Top_1 (d) La carte morphologique Top_2

Figure 3.11 – Les cartes morphologiques générées suite à l’application des opérations morphologiques par la première étape de la contribution proposée (cont.)

Étant donné que l’on dispose de la vérité de terrain de cette image, à savoir la position et les contours des MCs, il est donc facile d’interpréter ces cartes afin de vérifier leur compatibilité avec les objectifs définis. Une analyse globale de ces cartes montre qu’elles avaient réussi à éclaircir les noyaux et les contours des MCs par rapport aux tissus de voisinage. En outre, l’analyse locale, de ces mêmes cartes, montre que malgré l’utilisation des opérations morphologiques redondantes et des éléments structurants de profils similaires, elles se révèlent complémentaires en termes de l’effet du rehaussement des pixels constituant les MCs. Cette complémentarité est observée à la fois sur les cartes issues des mêmes opérations morphologiques (F_1 , F_2 et Top_1 , Top_2) ainsi que sur celles issues du même triplet d’éléments structurants (F_1 , Top_1 et F_2 , Top_2). En effet, bien que les opérations de dualité aient agi conformément sur l’image mammographique, le changement des tailles des éléments structurants influe sur les contours rehaussés. De la même manière, nous pouvons justifier la complémentarité observée sur les cartes Top_1 et Top_2 au niveau des pixels noyaux des MCs (pixels proches du centre d’une MC).

Ceci nous conduit à conclure que les éléments structurants ainsi que les opérations morphologiques utilisés ont été bien choisis et ont réussi à satisfaire les objectifs souhaités. Cette première étape s’est ainsi avérée pertinente et capable de contribuer à l’amélioration de l’extraction des MCs.

3.6.1.2 Évaluation de l'étape de génération des cartes de suspicion

Cette étape transforme les cartes morphologiques, précédemment obtenues, en une carte de suspicion, sur laquelle un seuillage adaptatif sera appliqué afin de détecter les zones suspectes (MCs potentielles). Dans ce paragraphe, la démarche suivie pour la génération de la carte de suspicion SM est évaluée.

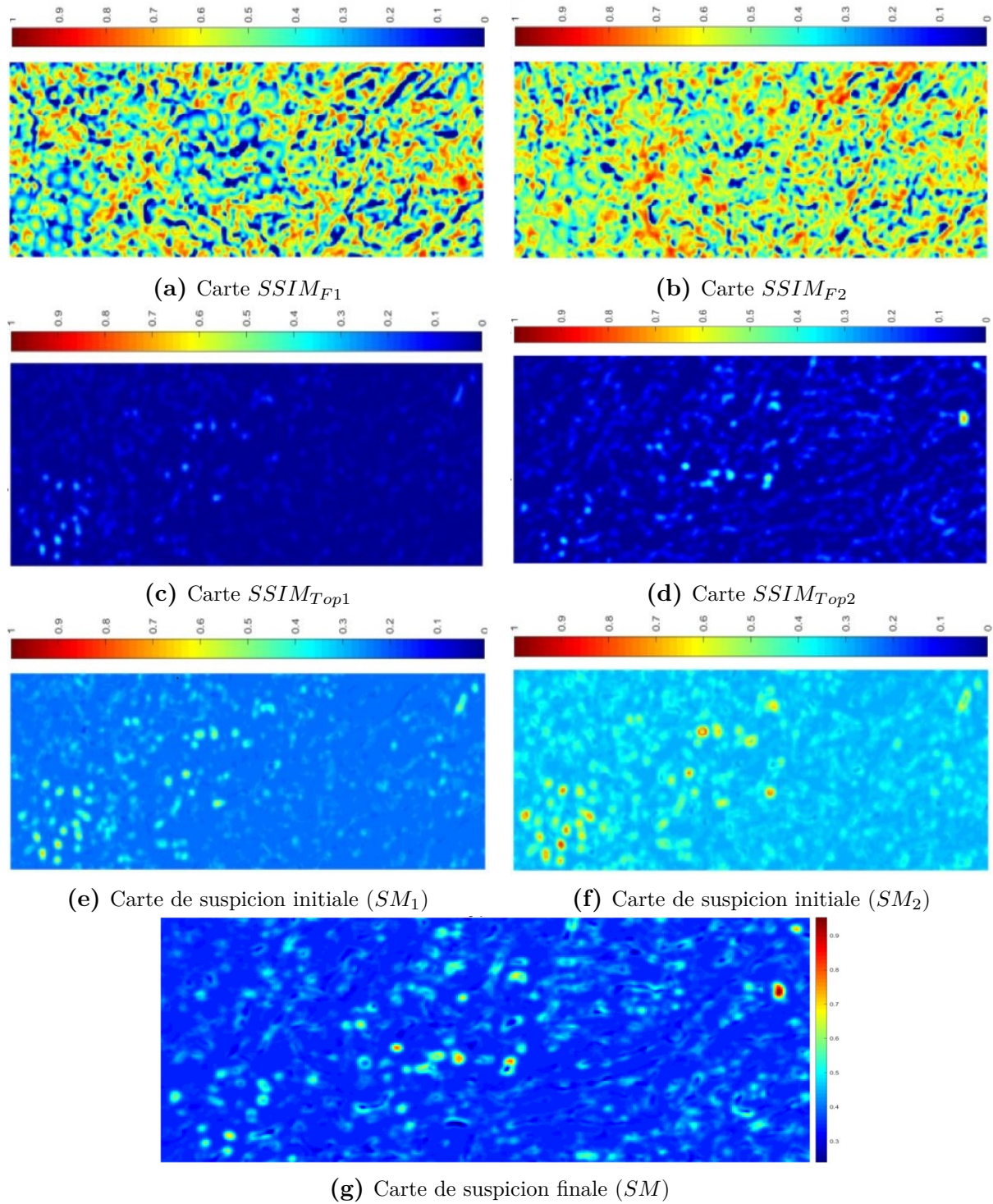


Figure 3.12 – Illustration des résultats intermédiaires pour la génération de la carte de suspicion

La figure 3.12 montre les différents résultats des transformations appliquées sur les cartes morphologiques permettant l'obtention de la carte de suspicion finale. La première et la deuxième ligne représentent les cartes $SSIM_M$ (avec $M \in \{F_1, F_2, Top_1 \text{ et } Top_2\}$) issues de la quantification des similarités structurelles, réalisée par l'indice de similarité SSIM, entre les cartes morphologiques correspondantes et l'image originale.

L'analyse de ces cartes montre que les pixels décrivant les structures linéaires fines sur l'image mammographique présentent des degrés de suspicion élevés, dans les cartes $SSIM_{F_1}$ et $SSIM_{F_2}$, par rapport à ceux des régions homogènes. Les degrés les plus élevés (avec des valeurs entre 0.8 et 1) sont retrouvés sur les pixels des structures très fines et présentant des pixels contours dans leurs voisinages immédiats. Ces pixels correspondent aussi bien aux régions décrivant des structures fines dans l'image que des MCs. Cette constatation n'est pas valable pour les cartes $SSIM_{Top_1}$ et $SSIM_{Top_2}$ où seulement les pixels des régions localement claires par rapport au voisinage immédiat et présentant des formes circulaires possèdent les degrés de suspicion les plus élevés par rapport aux autres pixels dans l'image.

Nous pouvons remarquer qu'avec cette nouvelle représentation, on n'interprète pas l'image sur la base des niveaux de gris de ses pixels mais plutôt des caractéristiques morphologiques de ses structures mammaires (forme, taille, ...).

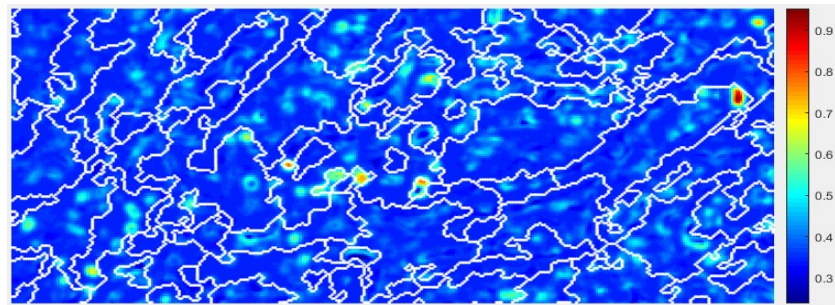
A partir de ces cartes, l'application de la combinaison permet d'obtenir les cartes SM_1 et SM_2 (troisième ligne de la figure 3.12) puis la carte de suspicion finale SM . La confusion entre les pixels des structures fines et ceux des MCs par les cartes $SSIM_{F_1}$ et $SSIM_{F_2}$ est corrigée par la combinaison conjonctive effectuée avec $SSIM_{Top_1}$ et $SSIM_{Top_2}$. En effet, nous pouvons remarquer, sur la troisième ligne, que les degrés de suspicion les plus élevés sont affectés aux pixels représentant les structures fines et localement claires dans l'image. Elles correspondent aux pixels ayant un accord dans les deux cartes. Les degrés de suspicion élevés initialement attribués aux pixels des structures linéaires, qui ne correspondent pas à des MCs, ont diminué de valeur. Par contre, les pixels des structures fines avec des profils similaires ont augmenté de valeurs de degrés de suspicion attribuées.

La combinaison finale des deux cartes SM_1 et SM_2 a augmenté davantage les valeurs des degrés attribués aux pixels des MCs et diminué celles des autres pixels.

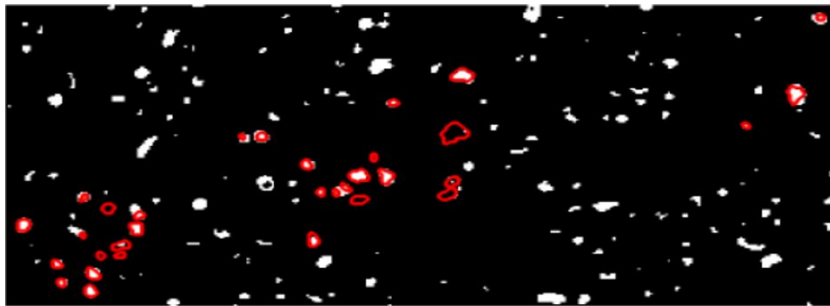
3.6.1.3 Évaluation de l'étape de détection des MCs potentielles

Cette dernière étape consiste à considérer la carte de suspicion SM , obtenue dans l'étape précédente, et à appliquer le seuillage adaptatif, décrit dans la section 3.4.2.5. L'objectif principal est de générer la carte de segmentation binaire représentant les zones sus-

pectes qui peuvent correspondre à des MCs. Cette carte est considérée comme le masque candidat à évaluer en le comparant avec le masque GT.



(a) *SM* avec superposition des contours des superpixels SLIC (contours blancs)



(b) Carte des MCs potentielles avec superposition du GT (contours rouges)

Figure 3.13 – Génération de la carte des MCs potentielles

L'analyse des résultats, illustrés dans la figure 3.13, montre que le masque binaire obtenu présente quelques fausses détections de type faux négatifs. En revanche, les pixels adjacents qui correspondent à des vrais MCs apparaissent sur la carte de suspicion avec des degrés élevés en les comparant aux autres pixels du tissu mammaire. Toutefois cette étape de seuillage peut échouer et générer quelques faux négatifs, la pertinence des résultats reste intacte avec la présence de la carte de suspicion qui nous permet de préserver toutes les informations disponibles sur les MCs ainsi que sur les tissus mammaires.

3.6.2 Fiabilité de la segmentation

Dans le but d'évaluer la pertinence de la segmentation des MCs par la contribution décrite dans ce chapitre, quelques résultats sont illustrés dans la Figure 3.14. L'image utilisée présente des MCs de type ACR4, réparties sur une grande partie du tissu homogène du sein avec des tailles variables et de faibles contrastes (Figure 3.14b).

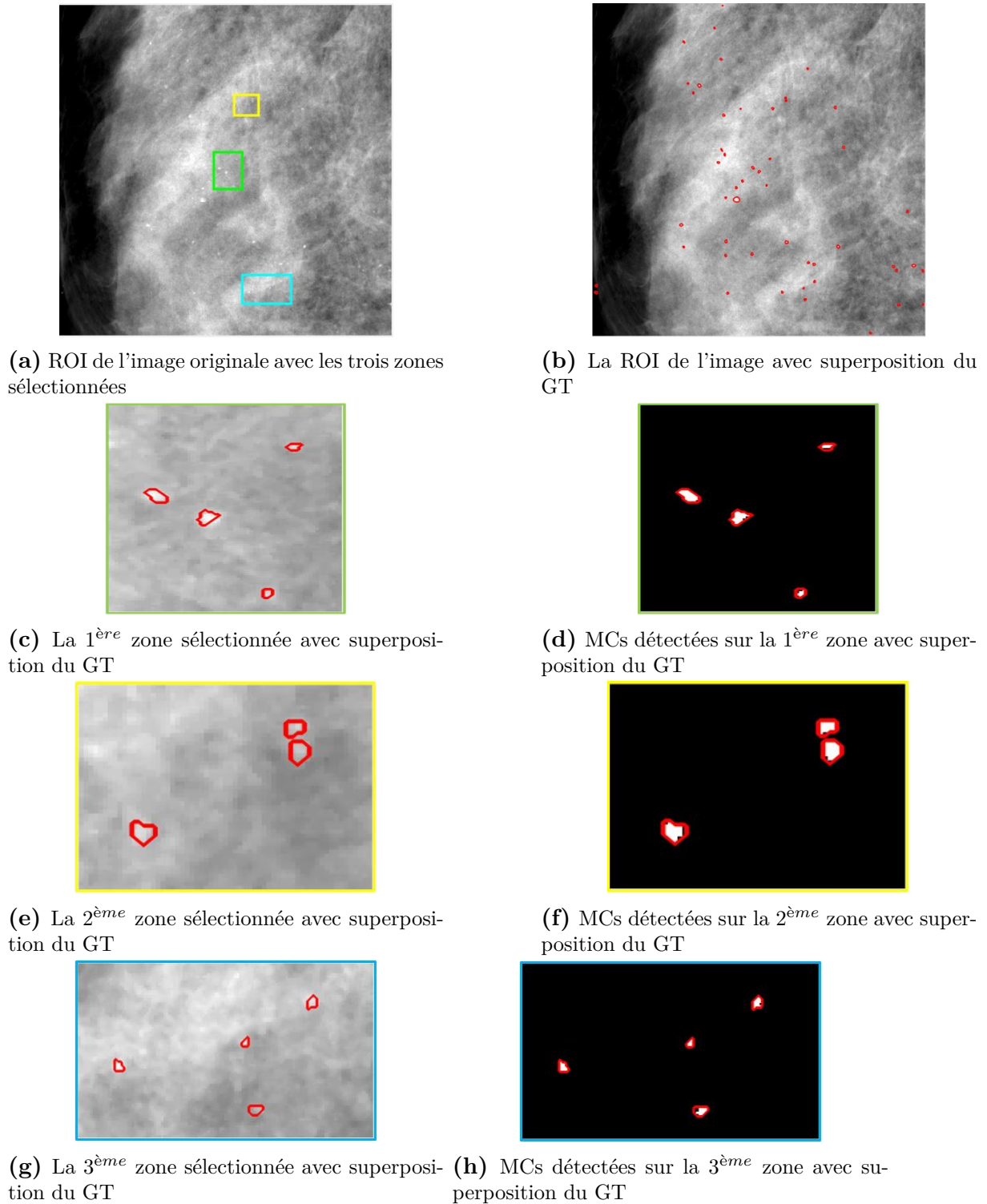


Figure 3.14 – Zoom sur les MCs détectées à partir de trois différentes zones sur une ROI

Les résultats montrés, sur cette figure, concernent trois petites zones rectangulaires (contours respectivement en jaune, vert et bleu sur l'image (Figure 3.14a)). Ces trois sélections sont positionnées sur la ROI dans l'image traitée, où se situent les MCs, et présentent divers types de MCs. Chaque ligne de cette figure montre les MCs détectées pour chacune des zones sélectionnées. Contrairement aux deux premières, la troisième

zone comporte deux types de tissus mammaires qui représentent des MCs de différents contrastes. L'analyse "visuelle" des résultats obtenus, sur l'ensemble des images d'évaluation, montre que la méthode proposée est capable de correctement segmenter les MCs de différentes tailles (zone avec le contour vert), peu contrastées (zone avec un contour jaune) et superposées sur différents types de tissus mammaires (zone avec le contour bleu ciel). Cette fiabilité a été aussi affirmée par la médecin radiologue avec laquelle nous avons travaillé.

3.6.3 Fiabilité par rapport à la méthode de comparaison

Les pixels d'une carte des MCs potentielles détectées sont distingués en trois types en fonction de leur superposition avec les pixels des vrais MCs (TP, FN et FP). Dans cette partie expérimentale, nous nous concentrons sur les pixels de types TP et FN, qui sont supposés représenter les pixels MCs correctement et erronément détectés par la méthode étudiée. Nous représentons par la figure 3.15 des cartes résultats de la superposition des GT avec les cartes MCs obtenues par la méthode proposée et la méthode retenue pour la comparaison [2] (notée 'Zh'). Les pixels marqués en blanc correspondent aux pixels TP. Les pixels marqués en violet correspondent aux pixels FN. En revanche, les pixels noirs regroupent les pixels du fond de l'image ainsi que les pixels considérés, à tort, comme des pixels MCs par la méthode proposée (entre autre ce sont les pixels FP).

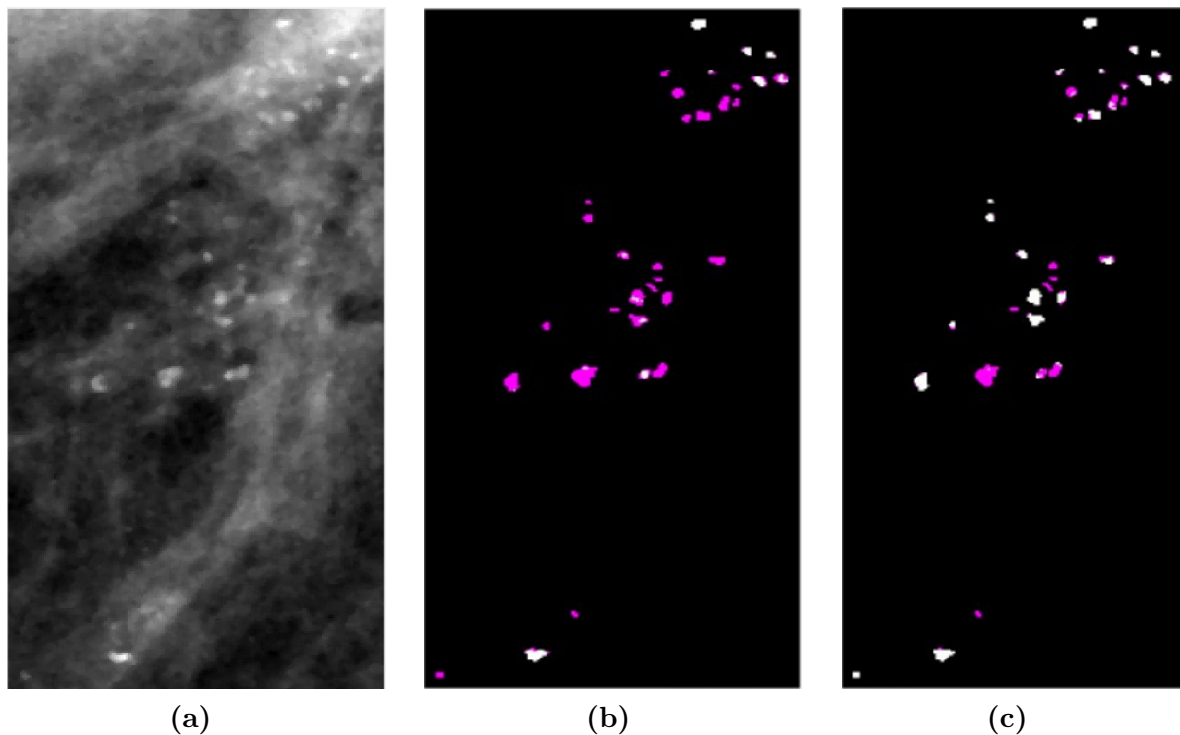


Figure 3.15 – Illustration des exemples de MCs détectées à partir des mammographies originales (colonne de gauche) par la méthode 'Zh' [2] (colonne du milieu) et la méthode proposée (colonne de droite)

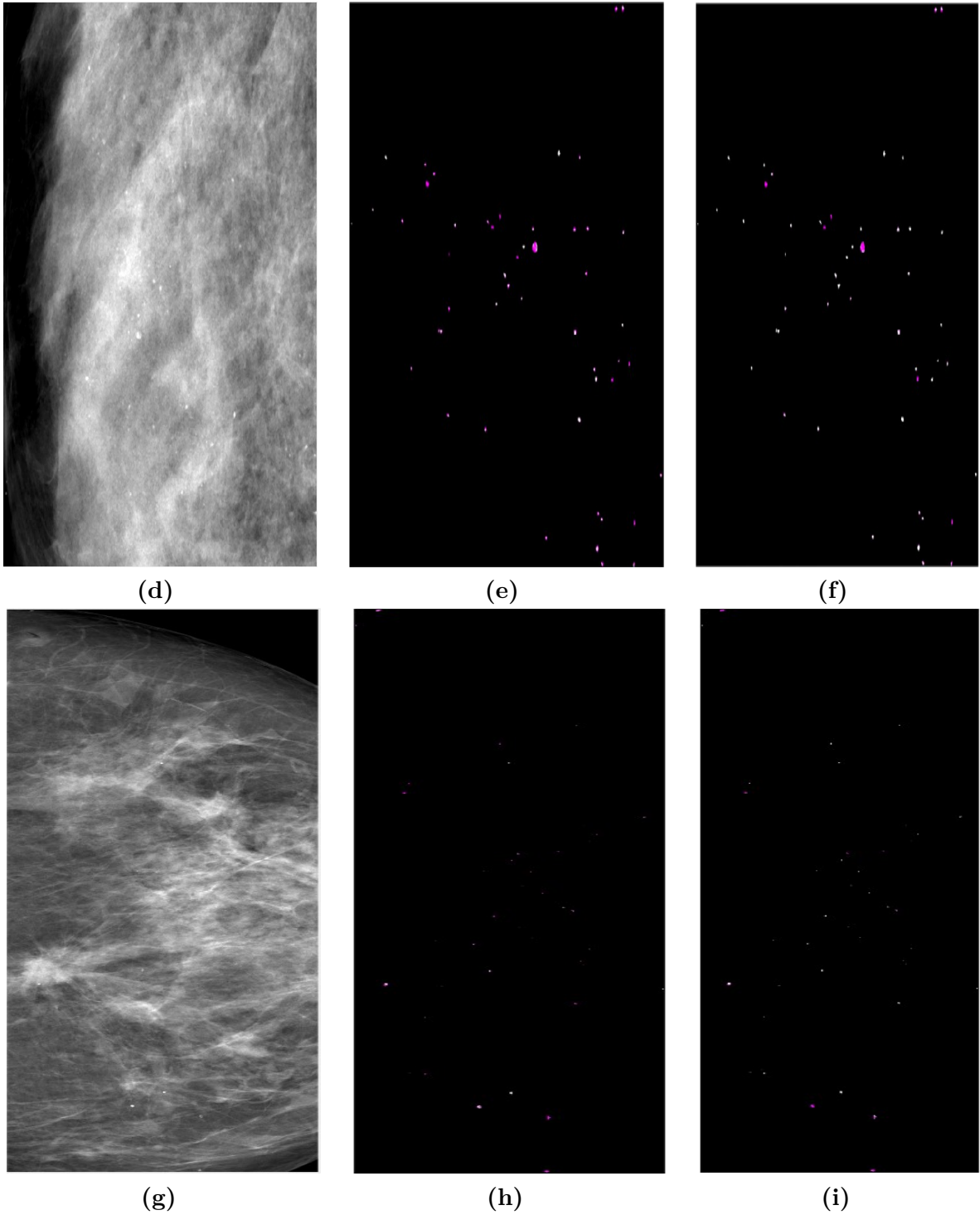


Figure 3.15 – Illustration des exemples de MCs détectées à partir des mammographies originales (colonne de gauche) par la méthode 'Zh' [2] (colonne du milieu) et la méthode proposée (colonne de droite)(cont.)

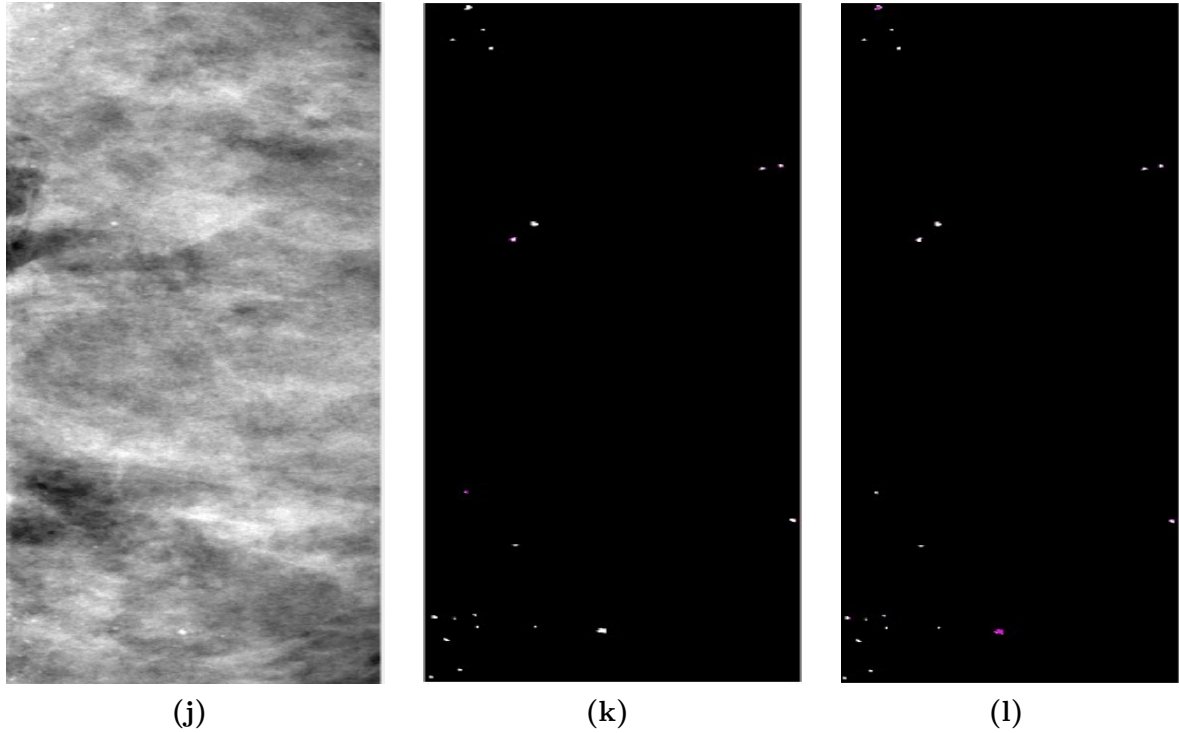


Figure 3.15 – Illustration des exemples de MCs détectées à partir des mammographies originales (colonne de gauche) par la méthode 'Zh' [2] (colonne du milieu) et la méthode proposée (colonne de droite)(cont.)

Chaque ligne concerne une image mammographique. La première colonne présente les ROI où les MCs apparaissent dans la mammographie originale. La deuxième colonne montre les masques obtenus par la méthode 'Zh'. La troisième colonne montre les masques obtenus par la méthode proposée.

L'analyse visuelle de ces résultats montre la précision de la méthode proposée par rapport à la méthode 'Zh' en termes des taux des MCs détectées (TP) tout en réduisant le nombre des MCs non détectées (FN).

3.7 Évaluation quantitative des résultats

Toute analyse qualitative doit être accompagnée par une analyse quantitative afin d'être justifiée. Pour ce faire, nous allons étudier, dans ce paragraphe, les valeurs des métriques d'évaluation utilisées, à savoir SE, SI, OF et EF obtenues. Cette étude ne concernera que les MCs détectées (les FPs ne seront pas considérés).

Cette évaluation porte sur deux volets. Le premier volet concerne l'étude de la pertinence de la méthode proposée. Le deuxième volet concerne sa supériorité par rapport à la méthode de comparaison retenue.

3.7.1 Fiabilité des résultats

Les valeurs de SE et SI sont exprimées en fonction des taux de vrai positif TP et faux négatif FN. Elles nous permettent ainsi d'étudier les pourcentages de détection. Une synthèse des résultats obtenus est présentée dans le tableau 3.1.

Tableau 3.1 – Description statistique de la répartition des valeurs de la sensibilité (SE) par la contribution proposée

	MIN	MAX	Moyenne \pm écart-type
SE	0.3333	1	0.6795 ± 0.1603
SI	0.2140	0.9636	0.7062 ± 0.1615
OF	0.1198	0.7295	0.4423 ± 0.1582
EF	0	0.1803	0.05316 ± 0.0526

La valeur moyenne et l'écart-type relativement à la même métrique SE sont respectivement de l'ordre de 0.68 et 0.16. En termes de similarité entre les pixels des vrais MCs et ceux considérés comme des pixels candidats par la méthode proposée, nous obtenons une valeur SI moyenne de l'ordre de 0.7 avec 0.16 comme écart-type. Ces valeurs montrent que la méthode proposée offre des résultats acceptables en termes de détection des MCs. En effet, nous nous sommes trouvés avec une identification (SE) de plus que 50% des MCs avec un seuil de 0.5.

Le même tableau présente, aussi, les valeurs retenues pour quantifier les pourcentages de sous et de sur-segmentation (EF et OF). L'analyse de ces dernières montre que notre méthode offre une segmentation acceptable en termes de sous-segmentation et précise en termes en sur-segmentation. En effet, les valeurs moyennes de OF et EF obtenues sont respectivement de l'ordre de 0.44 et 0.05 avec 0.16 et 0.05 d'écart-type.

3.7.2 Fiabilité par rapport à la méthode 'Zh'

Dans l'idée d'évaluer la pertinence de la méthode proposée par rapport à la méthode de référence 'Zh', nous avons effectué une sélection de 10 images dont 7 d'entre elles présentent des MCs avec des contrastes variables et des tailles différentes à partir de l'ensemble des images de test. Les images sélectionnées sont distinguées en trois familles notées (F_1 , F_2 et F_3). Les caractéristiques de contraste et de taille de chacune de ces mammographies sont indiquées dans le tableau 3.2.

Tableau 3.2 – Caractéristiques des MCs dans une sélection des mammographies numériques

Caractérisation des MCs			
	Contraste	Taille(s)	Tissu de superposition anatomique
<i>Image 1</i> (F_1)	Très faible	Très petites	Tissu nodulaire et radio-transparent
<i>Image 2</i> (F_1)	Moyen	Variables	Tissu nodulaire et radio-transparent
<i>Image 3</i> (F_1)	Variable	Variables	Tissu nodulaire et linéaire
<i>Image 4</i> (F_2)	Variable	Très petites	Tissu nodulaire
<i>Image 5</i> (F_1)	Très faible	Variables	Tissu nodulaire
<i>Image 6</i> (F_3)	Moyenne	Petites	Tissu linéaire
<i>Image 7</i> (F_2)	Elevé	Très petites	Tissu nodulaire
<i>Image 8</i> (F_3)	Moyenne	Variables	Tissu linéaire
<i>Image 9</i> (F_1)	Moyenne	Variables	Tissu linéaire
<i>Image 10</i> (F_3)	Faible	Petites	Tissu dense homogène (masse)

Afin d'évaluer la fiabilité de la méthode proposée, nous procédons par l'analyse des valeurs SI, présentées dans le tableau 3.1, obtenues en appliquant la méthode proposée et la méthode 'Zh'. L'analyse de ces dernières montre clairement la fiabilité de la méthode proposée en termes d'amélioration du taux de détection des vrais positifs et de réduction des faux négatifs par rapport à la méthode 'Zh'. En fait, les valeurs illustrées montrent que la méthode proposée augmente ou, du moins, maintient le même taux des TPs sur les mammographies avec des MCs à très faible contraste et de tailles différentes (images appartenant à F_1 et F_2 dans le tableau 3.2).

Cependant, la méthode proposée semble être légèrement moins fiable que la méthode 'Zh' pour les mammographies de la famille F_3 dans le tableau 3.2. Dans ces dernières, les microcalcifications se superposent aux tissus de structures linéaires à partir desquelles elles dérivent leurs caractéristiques. Les degrés de suspicion des pixels MCs sont alors très proches à ceux des pixels voisins ainsi qu'à la majorité des pixels du superpixel sur lequel ils apparaissent. Par conséquent, l'ensemble des pixels adjacents correspondant à des MCs dans ces superpixels ne pourrait pas remplir le critère des pixels aberrants. Ils peuvent donc apparaître comme des FNs dans la carte de MCs potentielle obtenue. Malgré cet échec, la méthode proposée reste instructive pour les radiologues puisque les degrés de suspicion peuvent les guider vers les régions de l'image sur lesquelles ils doivent se concentrer pendant l'interprétation.

Tableau 3.3 – Valeurs des mesures SI et SE retenues par la méthode proposée ($'D_M'$) et la méthode $'Zh'$ pour une sélection de mammographies numériques. Les meilleurs résultats sont en gras

	SE		EF	
	$'Zh'$	D_M	$'Zh'$	D_M
<i>Image 1</i>	0.47368	0.78947	0.0218	0.0075
<i>Image 2</i>	0.58333	0.8125	0.0170	0.0133
<i>Image 3</i>	0.33333	0.72549	0.0158	0.0427
<i>Image 4</i>	0.88571	0.88571	0.3523	0.1370
<i>Image 5</i>	0.62136	0.83495	0.141	0.0132
<i>Image 6</i>	0.95349	0.84884	0.0693	0.0559
<i>Image 7</i>	1	1	0.0588	0.0641
<i>Image 8</i>	0.61905	0.5	0.02708	0.0151
<i>Image 9</i>	0.03947	0.51316	0.0061	0.0052
<i>Image 10</i>	1	0.92857	0.40769	0.08461

D'autre part, une analyse plus approfondie des valeurs obtenues montre que cette première impression n'est pas toujours correcte et ne reflète pas vraiment une meilleure fiabilité. Le tableau 3.3 montre que cette supériorité est généralement couplée à un taux de sur-segmentation élevé pour la méthode $'Zh'$ prouvant ainsi que cette dernière est, réellement, incapable de distinguer ce type de MCs des tissus mammaires voisins.

3.8 Discussion

Revenons sur les valeurs des sous-segmentations (OF) et de sur-segmentations (EF) notées dans le tableau 3.1. Une délimitation fiable exige des valeurs proches de 0 pour EF et proches de 1 pour OF. La valeur moyenne pour EF est de l'ordre de 0.053. Cette valeur est considérée comme acceptable et elle justifie la capacité de notre méthode à ne pas étendre les contours des MCs vers des tissus mammaires voisins. Dans l'idée d'affirmer davantage cette constatation, nous pouvons envisager un enrichissement de la base de données utilisée pour l'évaluation par de nouvelles images de test.

Par ailleurs, la valeur moyenne de OF est de l'ordre de 0.44. Cette valeur semble être éloignée de la valeur maximale 1. Elle indique une faiblesse dans la détermination de tous les pixels de la zone caractérisant une MC. L'utilisation des éléments structurants avec différentes tailles peut être à l'origine de cet échec. En effet, l'application des opérations morphologiques sur une MC avec de profil proche de l'élément structurant le plus grand

implique sa division en deux parties qui sont le noyau et la frontière interne (Figure 3.16). Les pixels du noyau seront rehaussés deux fois (par l'élément structurant de petite ainsi de grande taille). Par contre, les pixels de la frontière interne seront rehaussés une seule fois (par l'élément structurant de grande taille). Ceci implique un changement important au niveau des pixels du noyau, et donc des degrés de suspicion très élevés dans la carte SM, par rapport aux pixels de frontière, qui auront des valeurs de degrés de suspicion proches des valeurs de la majorité des autres pixels dans le superpixel traité. Par conséquent, les pixels des frontières internes peuvent être considérés comme des pixels normaux par le seuillage appliqué.

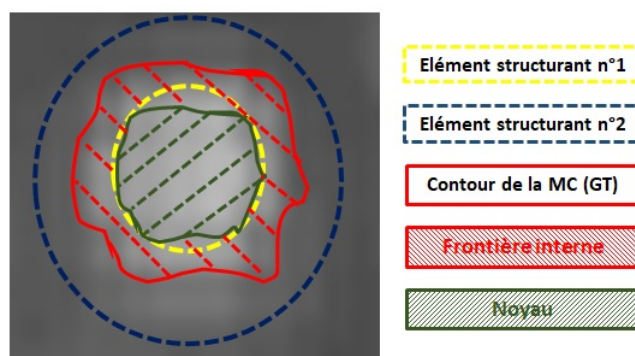


Figure 3.16 – Représentation des différents types de pixels distingués pour l'évaluation d'une segmentation

En partant de ces constatations, nous pouvons conclure, que l'amélioration du contraste et l'étude de la forme des MCs sont nécessaires, mais non suffisantes pour la détermination des contours précis des MCs. Une telle fiabilité nécessite une analyse locale plus approfondie. Cette conclusion sera à la base de notre deuxième contribution qui sera présentée et discutée dans le chapitre suivant.

3.9 Conclusion

Dans ce chapitre, nous avons présenté notre première contribution dans le cadre de ce travail de recherche, qui consiste en une nouvelle méthode de détection de MCs. Basée sur des opérations morphologiques et des indices de similarité structurale. Elle est composée de trois étapes successives : l'amélioration de la mammographie, la génération des cartes de suspicion et l'identification des MCs potentielles.

Parmi ses points forts, notre méthode combine l'indice de similarité et les opérations morphologiques d'une manière différente afin de traiter la diversité des caractéristiques des MCs. Elle exploite différents éléments structurants, la transformation des cartes morphologiques dans un nouvel espace de représentation (carte de suspicion), basé sur l'indice

SSIM et l'analyse locale des superpixels afin d'estimer des valeurs de seuil adaptatives à partir desquelles les pixels de MCs sont identifiés.

Les résultats obtenus sur les images de la base de données INBreast, ont été comparés à une méthode de référence fondée sur le même concept de morphologie mathématique. Les performances obtenues prouvent la capacité de la méthode proposée à traiter la diversité des formes et des contrastes des MCs. De plus, les comparaisons globales montrent des gains importants de la méthode proposée en termes de réduction des détections de type FNs et de sur-segmentation.

En revanche, un des problèmes majeurs de cette méthode concerne sa faiblesse face à deux types de MCs : Les MCs superposées à des tissus linéaires et les MCs ayant des tailles conformes à un seul élément structurant. Le point commun ici, est la non correspondance du profil des MCs avec celui des deux éléments structurants utilisés avec la présence d'une forte similarité aux tissus voisins.

Le chapitre suivant se base sur cette constatation et présente une autre méthode proposée, pour la détection des MCs. Cette dernière consiste en une analyse locale à base de connaissances *a priori* dans l'objectif de mieux identifier et segmenter les MCs.

DEUXIÈME CONTRIBUTION : DÉTECTION BASÉE SUR UNE CROISSANCE DE RÉGION CONDITIONNELLE

Sommaire

4.1	Introduction	98
4.2	Croissance de région	98
4.2.1	Principe	98
4.2.2	Motivation pour la détection des MCs	99
4.3	Description globale de la méthode proposée	99
4.3.1	Croissance de région classique et MCs - Limites	99
4.3.2	Démarche de la méthode proposée	100
4.4	Sélection des germes de départ	102
4.4.1	Maximas régionaux comme germes initiaux	102
4.4.2	Sélection des maximas germes de départ	104
4.4.3	Description du processus de croissance	108
4.5	Contraintes définies pour la croissance de région conditionnelle	109
4.5.1	Contraintes spatiales	110
4.5.2	Contraintes de délinéation	111
4.6	Résultats et discussion	116
4.6.1	Méthodologie de validation et méthode de comparaison	116
4.6.2	Fiabilité de détection par rapport à la méthode de comparaison	116
4.6.3	Fiabilité de segmentation	118
4.6.4	Discussion	121
4.7	Conclusion	123

4.1 Introduction

Avec la méthode proposée dans le chapitre précédent, nous avons justifié la fiabilité des opérations morphologiques dans la réduction de la sensibilité au faible contraste et à la variation des tailles des MCs. Par ailleurs, nous avons mis l'accent sur l'importance de l'analyse locale autour de chaque MC afin de pouvoir identifier ses contours exacts. La méthode présentée dans ce chapitre part de ce constat. Elle consiste en une nouvelle méthode de croissance de région conditionnelle (CRC) et a fait l'objet de deux publications scientifiques ([134] et [40]). Son objectif est de correctement identifier les contours des MCs en partant d'un ensemble de germes de départ. Les germes de départ sont déterminés sur la base de la détection de maxima régionaux et de l'analyse des superpixels. L'étape de croissance de région, quant à elle, est contrôlée par un ensemble de critères dérivés des connaissances *a priori* sur les MCs. Ces critères sont distingués en deux catégories : des critères géométriques et des critères de gradient.

4.2 Croissance de région

4.2.1 Principe

La méthode de croissance de région (*Region Growing*) est une méthode de segmentation par régions. Proposée par Meurle *et al.* en 1968 [135], elle consiste à faire évoluer progressivement des régions autour de leurs germes de départ (Algorithme 1).

Algorithme 1 : Algorithme de croissance de région classique

Entrées : I : Image , S : L'ensemble des germes de départ

Sorties : Image segmentée

pour chaque $s \in S$ **faire**

```

1    $R_c \leftarrow s$                                 /* Initialisation de la région à croître ( $R_c$ ) */
   pour chaque  $P_v \in PV_{R_c}$  faire
       si  $\text{HomCrit}(P_v, S)$  est Vrai alors
           /* Ajout du pixel  $P_v$ , à partir de la liste  $PV_{R_c}$ , à la région  $R_c$  */
2        $R_c \leftarrow R_c \cup P_v$ 
           /* Ajout des pixels voisins de  $P_v$  qui ne sont pas dans  $R_c$  à  $PV_{R_c}$  */
3        $PV_{R_c} \leftarrow PV_{R_c} \cup \text{Voisins}(P_v)$ 
       fin
   fin
fin

```

Cette croissance s'appuie sur certains critères d'homogénéité et d'adjacence. Le critère d'homogénéité (**HomCrit**) consiste en un seuillage souvent estimé sur la base d'une analyse de l'intensité locale [136] autour du germe de départ traité. La décision d'agrégation de

nouveaux pixels Pv à une région en cours d'évolution Rc , à partir d'une liste de pixels voisins candidats PV_{Rc} , se base sur la comparaison du degré de similarité entre le germe de départ et le pixel voisin candidat par rapport à un seuil estimé.

4.2.2 Motivation pour la détection des MCs

La majorité des travaux de recherche existants interprètent la présence des MCs, sur des images mammographiques, à partir des ROIs sélectionnées par les radiologues. Cette restriction a pour objectif de réduire la sensibilité aux choix des paramètres pour les techniques utilisées. Les techniques de segmentation les plus utilisées sont celles qui suivent un traitement itératif local telles que la croissance de région et les contours actifs. Le point fort de la croissance de région réside dans le déroulement de son processus de croissance qui fait croître une région, initialement définie par un germe de départ, vers les meilleurs pixels contours qui délimitent l'objet recherché. La prise de décision sur l'évolution ou l'arrêt d'une région à une itération t donnée est issue d'une analyse locale. Ce principe nous a encouragé à proposer une méthode basée sur un concept similaire pour la détection des MCs. En effet, une analyse locale autour de chaque MC nous permettra de surpasser les problèmes de variation de contraste, due à la diversité des tissus de superposition, même entre deux MCs spatialement proches sur la même mammographie.

4.3 Description globale de la méthode proposée

4.3.1 Croissance de région classique et MCs - Limites

Les méthodes à base de croissance de région exploitent le fait que les pixels d'une même région sont homogènes [81]. Les deux étapes clés pour ces méthodes sont le choix des germes de départ et la définition des critères d'homogénéité. Avec l'algorithme classique, le critère d'homogénéité résulte d'une analyse de l'histogramme de l'image traitée. Il s'agit d'une valeur numérique qui représente le seuil de variation maximale acceptée, en termes de niveaux de gris, entre un germe de départ et ses pixels voisins candidats (Pv).

La figure 4.1 illustre un exemple de deux ROIs présentant chacune un foyer de MCs. Le premier foyer (Figure 4.1a) montre des MCs avec différents contrastes. Le second foyer (Figure 4.1b) montre des MCs avec des contrastes très faibles par rapport au tissu de superposition qui est relativement homogène. Ils sont également plus faibles que les contrastes entre différents types de tissus mammaires issus de la même région. Dans ce cas de figure, le critère basé sur l'analyse du contraste local paraît insuffisant pour une bonne délinéation des contours des MCs recherchées. Le processus d'évolution, d'une région à partir d'un germe, basé sur ce critère peut conduire à une imprécision des pixels

contours qui sont supposés délimiter les MCs. Dans l’objectif de réduire cette imprécision, qui apparaît sous la forme d’une sous ou sur-segmentation, de nombreux autres critères doivent être pris en considération par le processus appliqué.

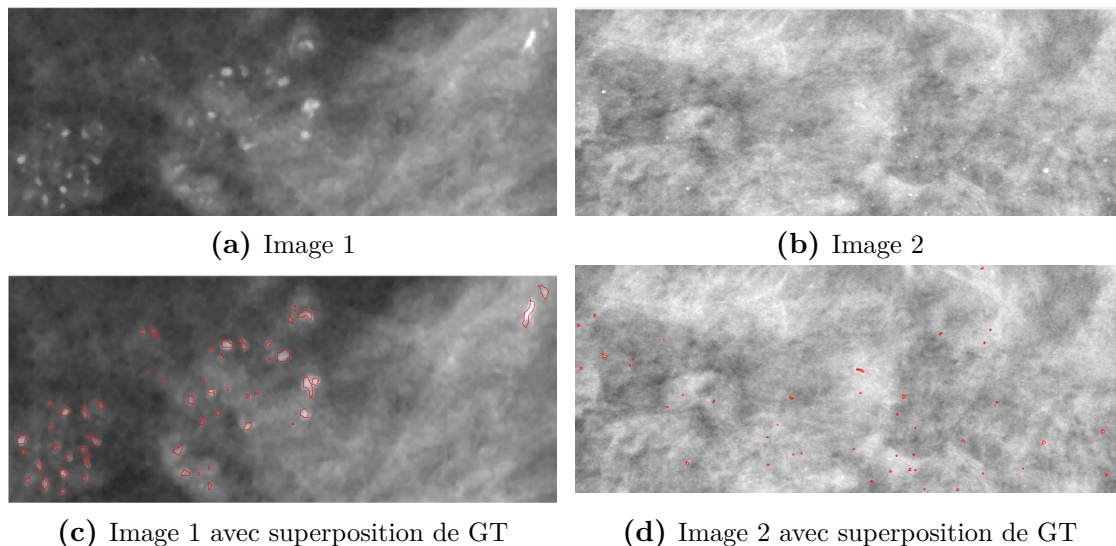


Figure 4.1 – Illustration de deux exemples de foyers de MCs avec des contrastes variables

En conclusion, l’idée principale de l’algorithme de croissance de région classique, qui consiste en une délinéation itérative des objets recherchés à partir d’un germe de départ, est pertinente pour la recherche des MCs. Son inconvénient majeur réside au fait que les informations spatiales et les connaissances *a priori* ne sont pas prises en considération pour l’estimation du critère d’arrêt. L’intégration de ces informations est nécessaire pour améliorer la qualité des régions segmentées et réduire la sensibilité au bruit et aux variations d’intensité [137]. En revanche, les caractéristiques particulières des images mammographiques et des MCs à segmenter imposent un ajustement des critères d’arrêt utilisés afin d’améliorer la précision des contours obtenus.

4.3.2 Démarche de la méthode proposée

Lorsque l’on demande à un radiologue de décrire l’apparition des MCs sur les images mammographiques, il répond que c’est un ensemble de zones connexes, sans chevauchement, plus brillantes en termes de niveaux de gris par rapport à leurs voisinages immédiats. Ainsi, la localisation de ces dernières peut commencer par la recherche des maxima régionaux à partir de l’image mammographique. Un maximum régional est un ou plusieurs pixels connexes ayant un niveau de gris supérieur à celui des pixels voisins. En revanche, un maximum régional ne décrit généralement qu’une partie d’une MC, mais pas sa totalité.

L’idée principale de notre contribution est de fournir une segmentation précise des contours en partant des germes de départ caractérisant des MCs. Pour ce faire, nous

proposons de procéder par une croissance de région afin de faire évoluer les maxima régionaux sélectionnés à partir de l'image mammographique. En revanche, un maximum régional ne correspond pas nécessairement à une MC. Il peut correspondre à un ensemble de pixels connexes décrivant un tissu mammaire sain (Figure 4.2). D'autre part, et en correspondance avec les propriétés des MCs décrites auparavant, une simple croissance de région peut ne pas être capable de correctement segmenter les contours exactes des MCs à cause de la variation du contraste qu'elles présentent. Il faut donc procéder par une croissance de région conditionnelle intégrant d'autres critères d'arrêts pour assurer une convergence sur les contours précis des MCs.

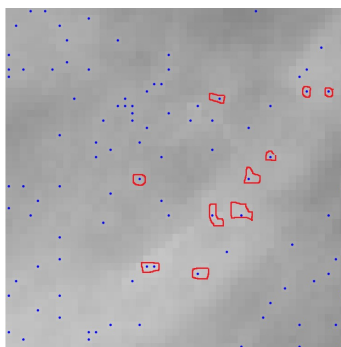


Figure 4.2 – Exemple d'une ROI avec superposition des maxima régionaux (points marqués en bleu) avec les contours vérité de terrain des MCs (contours rouge fermés)

Le schéma de la méthode proposée se divise en deux principales étapes (Figure 4.3) :

- Une étape de sélection automatique des germes de départ ;
- Une étape de croissance conditionnelle qui intègre des critères multiples estimés sur la base de connaissances *a priori* sur les MCs.

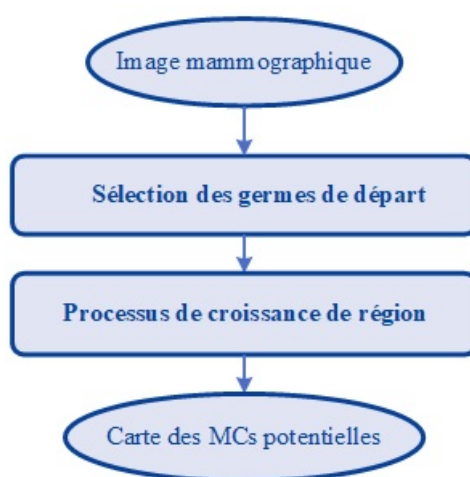


Figure 4.3 – Schéma global de la deuxième contribution proposée

Pour l'étape de sélection des germes de départ, nous commençons par considérer les maxima régionaux sur l'image comme les premiers germes candidats. Par la suite, nous

filtrons ces candidates afin de ne retenir que ceux qui peuvent faire partie d'une MC avec une forte probabilité.

Pour l'étape de croissance, le concept de base est similaire au processus de croissance classique qui fonctionne par agrégation des pixels voisins afin d'évoluer à partir du germe de départ initial. En revanche, elle utilise plus qu'un simple critère d'homogénéité et intègre d'autres critères issus des connaissances *a priori* sur les MCs. Cette intégration permettra au processus de croissance proposé d'être plus robuste à la variation des densités mammaires ainsi qu'à la variation du contraste des MCs.

4.4 Sélection des germes de départ

4.4.1 Maximas régionaux comme germes initiaux

Afin de correctement représenter une MC, un germe de départ doit satisfaire deux propriétés principales :

- il doit avoir l'une des valeurs de niveaux de gris les plus élevées de la MC en question ;
- il ne doit pas se trouver sur ou proche de pixels contours de la MC en question.

La représentation de la distribution statistique des valeurs des niveaux de gris d'une MC à l'intérieur de sa fenêtre circonscrite (Figure 4.4), à partir d'une image mammographique, montre une forme gaussienne. Son centre fait référence à celui de la MC étudiée. Il constitue l'ensemble de pixels connexes ayant la valeur de niveau de gris la plus élevée dans cette fenêtre. En s'éloignant du centre, les valeurs de niveaux de gris diminuent jusqu'à ce qu'elles atteignent les pixels décrivant un autre type de tissu mammaire. La pente de la courbe des niveaux de gris obtenue varie en fonction du contraste avec les pixels voisins. Cette description fait parfaitement référence à un maximum régional.

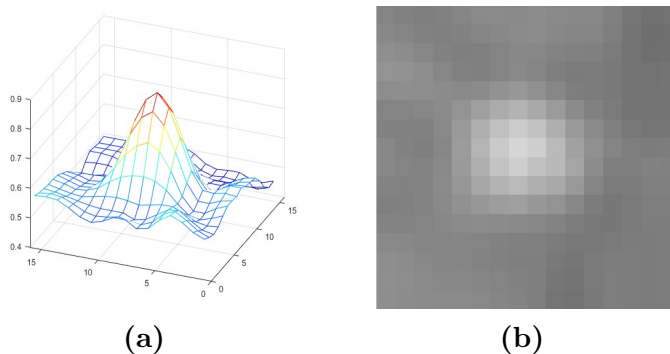


Figure 4.4 – Exemple d'une distribution statistique (a) des intensités des pixels d'une MC dans une fenêtre de taille 15×15 (b)

En partant de cette constatation, nous proposons de considérer les maxima régionaux

comme germes de départ pour démarrer le processus de croissance de région. Avec cette initialisation, en plus d'une localisation stratégique des MCs, nous sommes en mesure de contrôler le processus de croissance en analysant l'évolution des valeurs des niveaux de gris dans le voisinage du germe de départ. La démarche suivie, dans ce travail, pour la détermination des maxima régionaux est décrite par l'algorithme 2.

Algorithme 2 : Algorithme de détection des maxima régionaux

Entrées : I : Image en entrée

Sorties : **Out** : Image étiquetée (une étiquette pour chaque maximum régional)

Initialisation :

$Label(P) \leftarrow 1$ /* P : pixel de bord de l'image I */

$Label(Q \neq P) \leftarrow 2$ /* Q : pixels de I avec $Q \neq P$ */

début

```

1  pour chaque pixel  $P$  faire
    | /* Vérifier si  $P$  est maximum par rapport à son voisinage */
    |  $out\_img(P) \leftarrow neighbourCheck(P)$ 
    fin
2   $lbl\_img \leftarrow ConnectedComponent(out\_img)$  /* Étiquetage des régions
    connexes groupant des maxima régionaux */
3  pour chaque pixel  $P$  où  $(lab == lbl\_img)$  faire
4  | si  $lbl\_img(P) \neq 0$  /* Fait partie d'une région maximale ? */
    | alors
        | /* Vérifier si  $P$  fait partie de la même région maximale de ses
        | voisins (pixel plateau) */
        |  $Est\_Plateau \leftarrow neighbourCleanup(P)$ 
        | si  $neighbourCleanup(P) == 1$  alors /* fait partie d'une région
        | maximale */
        | |  $bad\_labels \leftarrow [bad\_labels, P]$ 
        | fin
    | fin
    fin
5  fin
7  Ordonner et supprimer les labels dupliquées dans le vecteur  $bad\_labels$ 
8  pour chaque pixel  $P \in I$  faire
9  | si  $label(P) \notin bad\_labels$  alors
    | |  $Out[P] \leftarrow 0$  /*  $P$  est considéré comme un maximum */
    | fin
    fin
fin

```

L'algorithme proposé commence par attribuer pour chaque pixel P de l'image I une label 0 ou 1 en utilisant la fonction `neighbourCheck`. Cette fonction consiste à vérifier si un pixel donné est maximum par rapport à son voisinage. Une valeur 0 signifie que P n'est pas un maximum et qu'il existe un pixel Q dans son voisinage immédiat $V(P)$ ayant une valeur de niveau de gris plus élevée. Une valeur 1 signifie que P est un pixel maximum.

Ensuite, la fonction **ConnectedComponent** qui consiste à étiqueter les régions connexes groupant des maximas régionaux. Elle attribue le même label pour chaque ensemble de pixels maximas connectés de la *out_img* précédemment obtenue. Le résultat, de cette deuxième étape, est une image étiquetée *lbl_img*. La troisième étape consiste à vérifier si les pixels dont le label affecté est différent de zéro décrivent le même maxima régional. La fonction **neighbourCleanup** qui consiste à identifier et supprimer les pixels P maximas qui font partie de la même région maximale de leurs voisins. Elle cherche ainsi un pixel au voisinage de P qui a une valeur 0 dans *out_img* et qui présente le même niveau de gris que P . Elle renvoie une valeur 1 si ce test est vrai et 0 dans le cas contraire. Si la valeur retournée est 1, alors le label de ce pixel est considéré comme '*mauvais label*' qui doit être supprimé. En dernier lieu, les pixels dans I dont la label ne figure pas dans le vecteur *bad_label* sont conservés comme des maximas.

Avec cette initialisation, nous assurons que chaque MC, présente dans l'image mammographique, est représentée par un maxima. Cependant, du fait que les maximas régionaux peuvent ne pas correspondre à des MCs mais à d'autres types de tissus mammaires sains (Figure 4.2), la croissance à partir de tous les maximas entraîne la génération d'un nombre élevé de faux positifs. Elle peut également créer des régions connexes avec chevauchement. En effet, le caractère multi-tissulaire des images mammographiques est à la base de la 'grande' variation des niveaux de gris sur une petite région de l'image qui, à son tour entraîne des maximas géométriquement très proches. Il est donc essentiel de passer par une autre étape supplémentaire de sélection pour diminuer le nombre de germes obtenus. Cette sélection consiste à filtrer l'ensemble des maximas pour ne garder, comme germes de départ pour le processus de croissance à appliquer, que ceux qui peuvent appartenir à des MCs.

4.4.2 Sélection des maximas germes de départ

L'étape de sélection des maximas à considérer comme germes de départ pour le processus de croissance consiste à choisir un certain nombre de maximas régionaux parmi l'ensemble de tous les maximas obtenus à partir de l'image. Dans cette étape, nous définissons un ensemble de conditions pour deux types de superpixels :

- *Superpixels SLIC* générés par l'algorithme SLIC [131] et respectent un critère d'homogénéité déterminé (Figure 4.5a) ;
- *Superpixels LPE* générés par l'algorithme LPE (*Ligne de Partage des Eaux*) [138] appliqué à l'image mammographique complémentaire.

$$I_C(P) = \max((I) - I(P)) \quad (4.1)$$

L'algorithme LPE décompose une image donnée en plusieurs bassins versants tout en partant de ses minimas. Un maximum dans l'image mammographique originale se transforme en minimum local dans l'image complémentaire I_C (Équation (4.1)). Les limites entre deux bassins versants voisins définissent les limites d'un superpixel LPE. Ainsi, un superpixel LPE (Figure 4.5b), généré à partir de l'image mammographique, est une région homogène connexe qui peut contenir au maximum une MC puisqu'elle présente un seul maxima régional. Par contre, un superpixel SLIC (Figure 4.5a) peut présenter un ou plusieurs maxima et peut donc contenir une voir plusieurs MCs.

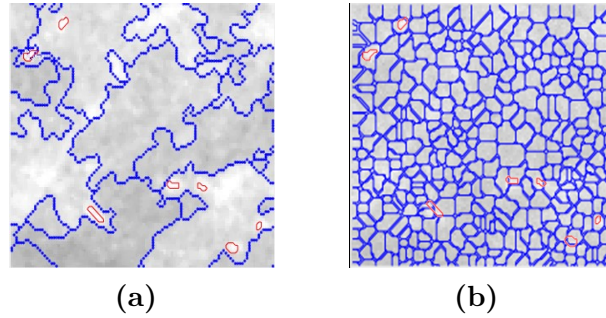


Figure 4.5 – Exemple de superpixels SLIC (a) et LPE (b) générés à partir d'une image mammographique (contours bleu) avec superposition des contours GT des MCs (contours rouge)

Les conditions utilisées pour décider de retenir ou de rejeter un maximum sont liées aux relations entre les MCs avec les deux types de superpixels que nous venons de présenter. La description et la justification de ces conditions sont discutées dans les paragraphes suivants.

4.4.2.1 Sélection des germes de départ à partir des superpixels LPE

Les superpixels LPE générés à partir de l'image mammographique décrivent les bassins versants partant des maximas et limités par des crêtes. La position des crêtes pour un superpixel donné, dépend de la pente des niveaux de gris au voisinage du maximum qu'il contient. En revanche, la distribution gaussienne qui décrit le voisinage d'une MC apparaît symétrique par rapport à son centre. Ceci implique que la pente de cette distribution est symétrique et que le centre de la MC (le maximum) se situe au milieu de la ligne de partage des eaux qui le limite et donc au centre de gravité de son superpixel LPE.

Par conséquent, un maximum régional n'est retenu comme germe de départ (pixels jaunes dans la Figure 4.6) pour notre processus de croissance appliqué que s'il est situé au niveau du centre de gravité de son superpixel LPE.

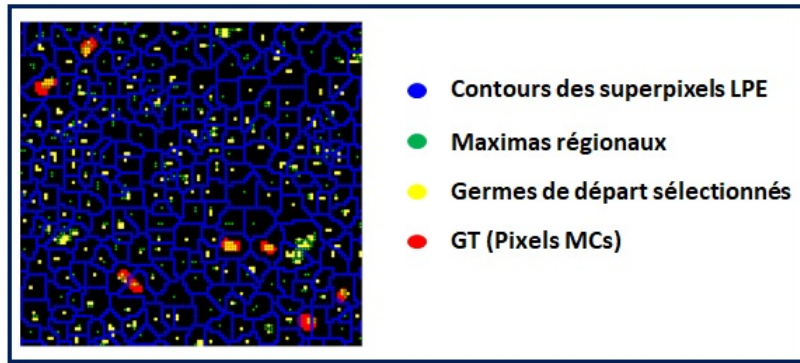


Figure 4.6 – Exemple de maxima régionaux sélectionnés (jaune) à partir de tous les maxima initialement générés (vert) avec la superposition des contours des superpixels LPE (bleu) et GT des MCs (rouge)

4.4.2.2 Sélection des germes de départ à partir des superpixels SLIC

Sur le même superpixel SLIC, on peut avoir une dizaine voir parfois centaine de maxima. Cela dépend de la taille ainsi des caractéristiques du superpixel en question. Ces maxima peuvent être géométriquement proches de sorte à ce que les régions segmentées, suite à la croissance appliquée à partir de deux maxima voisins, peuvent se coller l'une à l'autre et générer de 'grandes' régions présentant une ou plusieurs MCs.

La sélection des maxima sur les superpixels LPE, s'est basée sur l'analyse de la distribution des niveaux de gris dans le voisinage de chaque maximum. Elle permet de réduire le nombre de maxima régionaux qui font référence à des FPs. Cependant, ce nombre reste important étant donnée la petite taille de ces superpixels et la variété des structures anatomiques dans les images mammographiques. La sélection des maxima par superpixel SLIC, se base sur l'analyse de la nature du superpixel en question décrite par la distribution des niveaux de gris de ses maxima. L'idée sous-jacente de cette analyse est d'étiqueter les superpixels SLIC en superpixels homogènes et hétérogènes. Les superpixels considérés comme homogènes présentent des maxima régionaux avec des valeurs de niveaux de gris très proches. Ceci induit une mauvaise discrimination entre les maxima qui correspondent à des MCs des autres. Il sera donc avantageux de retenir la majorité d'entre eux comme germes de départ, ce qui permet d'éviter le risque de génération des FNs. Au contraire, les maxima régionaux qui pourraient correspondre à des MCs, dans un superpixel considéré comme hétérogène, présentent nécessairement des valeurs de niveau de gris écartés par rapport à la majorité des maxima qu'il contient. La discrimination est généralement plus facile.

Pour décider du type d'un superpixel SLIC donné, nous déciderons deux paramètres qui sont la moyenne et l'écart-type des niveaux de gris des maxima régionaux qu'il contient. A partir de ces paramètres, nous estimons deux valeurs de seuils $S1_k$ et $S2_k$.

La procédure d'estimation de ces seuils, décrite par l'algorithme 3, se déroule en deux principales étapes.

Algorithme 3 : Estimation des seuils $S1_k$ et $S2_k$

Entrées : SP : Superpixels SLIC
Sorties : Seuils $S1_k$, $S2_k$
début

```

1  pour  $k = 1$  à  $NbSP$  faire
    /*  $NbSP$ : nombre de superpixels SLIC */
2     $S1_k \leftarrow \text{mean}_{i=\{1..NbM_k\}} (NG_i)$ 
    /* L'ensemble des  $M_i$  dont le niveau de gris est supérieur à  $S1_k$  */
3     $G1_k \leftarrow \bigcup_{i=1..NbM_k} \{M_i / NG_i \geq S1_k\}$ 
4     $StdG1_k \leftarrow \text{std}_{i=\{1..|G1_k|\}} (NG_i) / M_i \in G1_k$ 
    fin
    /* Moyenne de l'écart-type de l'ensemble  $G1_k$  */
5     $moyG1 \leftarrow \text{mean}_{k:1..NbSP} (StdG1_k)$ 
6    pour  $k = 1$  à  $NbSP$  faire
7      si  $StdG1_k \leq moyG1$  alors
8         $S2_k \leftarrow Q_1$  /*  $SP_k$  est considéré comme homogène */
      sinon
9         $S2_k \leftarrow Q_3$  /*  $SP_k$  est considéré comme hétérogène */
      fin
10    $G2_k \leftarrow \{M_i / M_i \in G1_k \text{ et } NG_i \geq S2_k\}$ 
    fin
fin

```

Dans un premier temps, les maximas M_i dont la valeur de niveau de gris NG_i est inférieure à un premier seuil noté $S1_k$ sont écartés. $S1_k$ est choisi comme la valeur moyenne des niveaux de gris de tous les maximas dans SP_k . La supériorité, en termes de niveaux de gris, des pixels MCs par rapport aux pixels du tissu de superposition est à l'origine de ce choix. L'ensemble des maximas qui respectent cette condition sont retenus et enregistrés dans $G1_k$.

Une fois cette première sélection de maximas est effectuée, la deuxième étape décide de la pertinence d'un maximum en fonction du type de son superpixel d'appartenance. Une valeur de seuil $S2_k$ est définie, pour chaque SP_k , afin de distinguer les superpixels pouvant correspondre à des MCs des autres appartenant à des tissus mammaires sain. Elle est égale au premier Q_1 (respectivement troisième Q_3) quartile, calculé à partir de la distribution des niveaux de gris des maximas dans $G1_k$, selon le type du superpixel (homogène respectivement hétérogène). A la fin de cette étape, l'ensemble des maximas germes pour chaque superpixel SP_k sont regroupés dans $G2_k$.

En conclusion, un maximum régional M_i lié au $k^{\text{ème}}$ superpixel (SP_k) est considéré comme un germe de départ si et seulement si son niveau de gris NG_i respecte la condition suivante exprimée à l'aide de deux seuils $S1_k$ et $S2_k$ estimés pour SP_k .

$$NG_i \geq S1_k \text{ et } NG_i \geq S2_k \quad (4.2)$$

Cette étape de sélection risque de faire perdre quelques maximas faisant partie des 'vrais' MCs. Elle peut ainsi engendrer des faux négatifs dans la carte finale des MCs potentielles. En contrepartie, elle permet de gagner en terme de temps d'exécution ainsi que de la qualité des résultats. En effet, les régions dans la carte finale produite seront bien séparées (elles ne présentent pas des chevauchements), ce qui facilite leur interprétation.

4.4.3 Description du processus de croissance

Une fois les germes de départ déterminés, la deuxième étape consiste à faire évoluer, sous le contrôle de plusieurs critères, le contour autour de chaque germe afin de segmenter les régions qui seront considérées comme des MCs potentielles dans la carte de segmentation finale produite. Les modifications que nous avons apportées par rapport au processus de croissance de région classique concernent la définition des conditions d'arrêts et la sélection des voisins candidats. En effet, le processus de croissance classique ajoute tous les pixels voisins à la région en cours d'évolution tant qu'ils respectent un critère d'homogénéité prédéfini. Cependant, comme nous l'avons déjà mentionné auparavant, une telle condition est insuffisante pour la segmentation précise des contours des MCs. Les critères que nous avons utilisés sont issus des connaissances *a priori* sur les MCs et sont modélisés par deux types de contraintes : des contraintes spatiales et des contraintes de délinéation.

La démarche suivie par cette étape de croissance est comme suit (Figure 4.7). En partant d'un germe de départ, on identifie l'ensemble de tous les pixels candidats respectant certaines contraintes spatiales. Les premiers candidats sélectionnés pour une éventuelle croissance se situent dans la fenêtre du voisinage local du germe de taille 3×3 . Les pixels sélectionnés au cours des itérations suivantes seront considérés comme des candidats pour l'agrégation de la région R_c en cours d'évolution. Au cours de chaque itération, plusieurs tests relatifs à certaines contraintes de délinéation, détaillées dans la section suivante, sont appliqués et la liste des pixels candidats est mise à jour. Ce processus est répété tant que cette liste n'est pas vide.

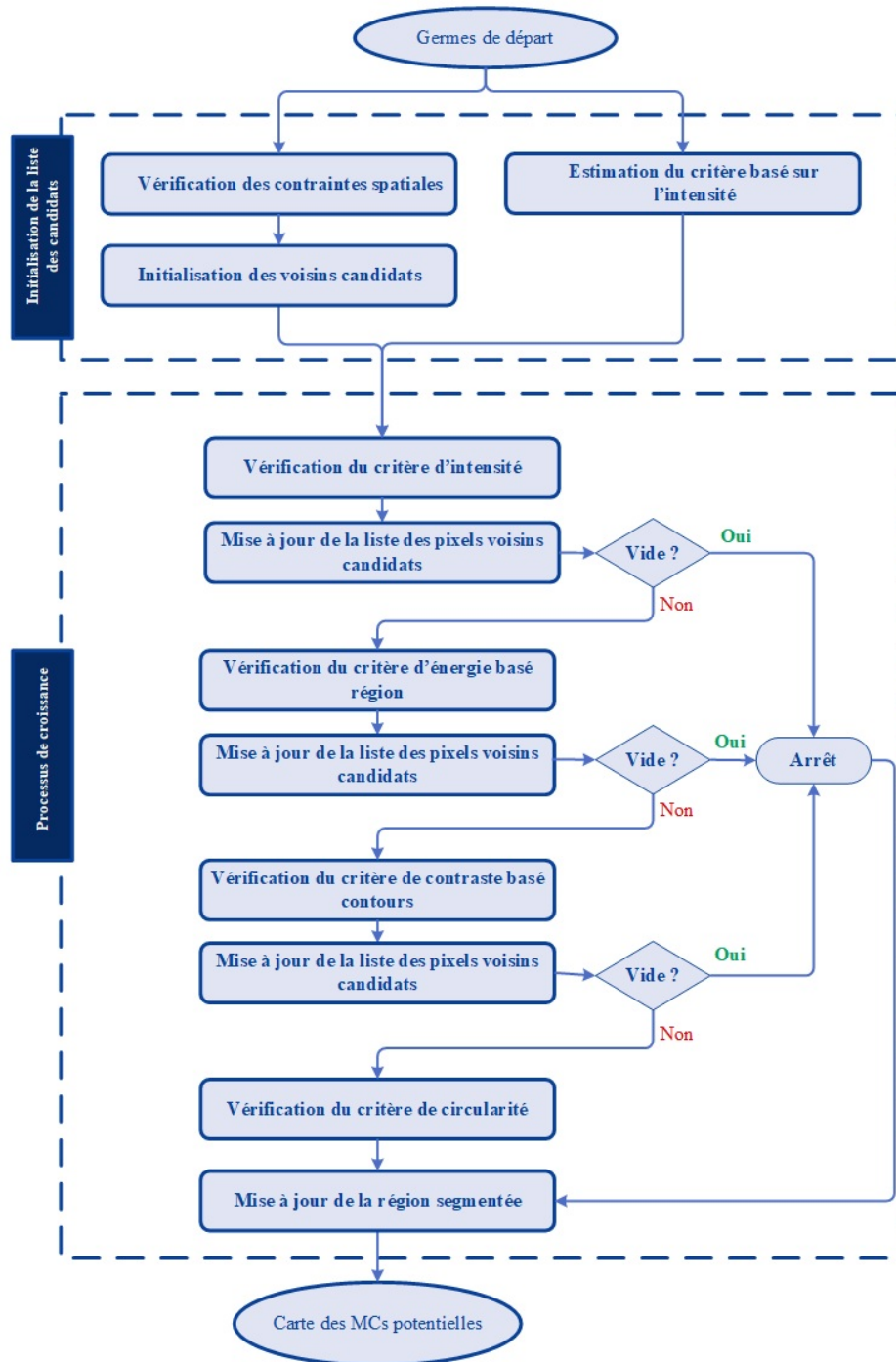


Figure 4.7 – Organigramme du processus de la croissance de région conditionnelle proposée

4.5 Contraintes définies pour la croissance de région conditionnelle

Dans ce chapitre, nous proposons d'intégrer des critères basés sur les connaissances *a priori* afin de contrôler le processus de croissance. Ces critères sont dérivés de l'expertise des radiologues et sont modélisés par deux types de contraintes : des contraintes spatiales et des contraintes de délimitation des MCs. Les contraintes spatiales concernent la taille

de la fenêtre de voisinage pour l'évolution des germes. Deux informations principales sont déduites de l'analyse de cette fenêtre et seront exploitées pour guider le processus de croissance. La première concerne une limite géométrique de l'espace local pour l'évolution des germes où elle permet de sélectionner les voisins candidats. La deuxième consiste à estimer automatiquement une valeur de seuil pour le critère basé sur l'intensité. Les contraintes de délimitation analysent les informations sur le gradient au voisinage des régions évoluées et suit le changement de leurs formes pendant le processus de croissance. Elles décrivent un ensemble de critères à tester à chaque itération qui sont : un critère basé sur l'énergie, un critère de contraste des contours et un critère de forme en plus du critère basé sur l'intensité. Nous décrivons en détails ces critères dans les paragraphes suivants.

4.5.1 Contraintes spatiales

Selon les radiologues, les MCs sur une image mammographique apparaissent avec une petite taille qui ne dépasse pas 1 mm de diamètre [139]. Cette information nous conduit à conclure que la zone de recherche d'une MC à partir d'un germe de départ doit être géométriquement limitée. Les limites à définir doivent présenter un compromis entre la résolution des images et les tailles des MCs.

Dans ce travail et dans l'objectif de segmenter des régions avec une taille proche de celle des MCs, nous avons proposé d'utiliser deux fenêtres d'analyse carrées, R_1 et R_2 , autour de chaque germe de départ. Ces fenêtres sont centrées au centre du germe en question et présentent de tailles 15×15 pour R_1 et 7×7 pour R_2 . Elles sont utilisées dans le processus de croissance comme des contraintes spatiales :

- La fenêtre R_2 est la fenêtre à ne pas dépasser par la région à faire croître R_c . Sa taille est déduite en fonction de la résolution des images mammographiques et la taille maximale des MCs ;
- La fenêtre R_1 est utilisée pour la détermination du critère d'intensité pour le germe en question. C'est à partir de l'analyse de l'histogramme des intensités dans cette dernière que l'on estime la valeur de seuil à utiliser.

L'évolution du contour doit toujours respecter les contours du superpixel SLIC dans lequel se situe le germe en question. Ainsi, les limites d'évolution de la région autour d'un germe de départ sont les pixels communs des fenêtres d'analyse R_1 et R_2 avec ceux du superpixel SLIC auquel il appartient (zone délimitée par un contour jaune dans la figure 4.8). Cette intersection nous permet d'éviter une croissance inutile dans une région voisine similaire ou une région de fond homogène. Elle permet ainsi de réduire la sensibilité de la croissance de région à la grande similarité, en termes d'intensité, des pixels localement voisins.

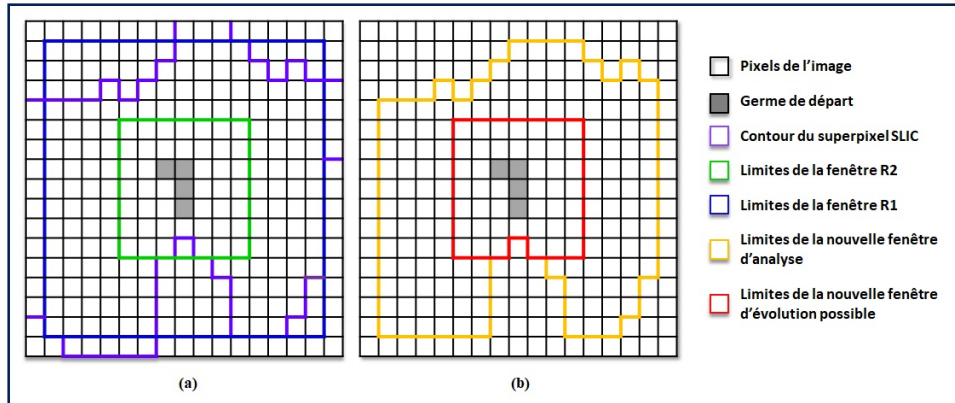


Figure 4.8 – Illustration d'un exemple des fenêtres d'analyses définies autour d'un germe de départ

L'utilisation de ces fenêtres nous permet d'estimer un critère d'intensité pour chaque germe de départ et de réduire la sensibilité à l'homogénéité locale. En effet, en partant d'un germe qui fait partie d'une MC, la fenêtre R_1 entoure la MC en question et une partie de son tissu de superposition. L'analyse de cette fenêtre nous offre donc une idée sur le contraste et la variation du niveau de gris autour de cette MC. Cela implique une valeur de seuil déduite à partir des constatations locales observées.

4.5.2 Contraintes de délinéation

Les limites principales des techniques de détection des MCs existantes sont liées à la nature floue des contours sur les images mammographiques. En effet, un foyer de MCs peut être dispersé sur la totalité du sein qui présente une grande variété tissulaire. La visibilité des MCs dépend du degré de luminosité du tissu de superposition du fait que ces dernières acquièrent souvent ses mêmes caractéristiques.

En suivant la description d'apparition des MCs sur les images mammographiques par un radiologue, nous constatons que les conditions d'adjacence et d'homogénéité de niveau de gris sont nécessaires mais insuffisantes pour une bonne délinéation. En effet, parmi les justifications utilisées par le radiologue pour argumenter son choix d'un pixel contour d'une MC par rapport à un autre, nous pouvons mentionner :

- Une MC est un groupement de pixels avec des valeurs de niveaux de gris supérieures à celles dans leurs voisinages ;
- Les pixels du contour externe d'une MC ont un contraste plus élevé que celui des pixels du contour interne ;
- Une MC présente une forme souvent circulaire ou elliptique.

A partir de ces descriptions, nous avons proposé d'utiliser les contraintes suivantes pour le contrôle de l'évolution d'une région Rc à partir d'un germe de départ. Elles sont

en nombre de quatre et sont appliquées dans le même ordre de citation. Leur description détaillée est donnée par les sous sections suivantes :

- Un critère d'intensité qui définit l'écart maximal toléré entre l'intensité du germe de départ et un voisin candidat à ajouter ;
- Un critère d'énergie basée-région qui cherche à faire évoluer une région R_c vers une solution précise en minimisant une expression d'énergie ;
- Un critère de contraste basé-contours qui cherche à ajouter par chaque nouvelle itération le meilleur pixel candidat ayant la valeur de contraste la plus élevée (pixel contour) ;
- Un critère de forme qui assure l'évolution d'une région R_c avec le pixel candidat qui présente le degré de circularité le plus élevé.

4.5.2.1 Critère d'intensité

Le critère basé sur l'intensité que nous avons défini est assez similaire au critère d'homogénéité de l'algorithme de croissance de région classique. En revanche, il définit la limite supérieure à atteindre par l'écart, en termes de niveau de gris, entre la région définie par le germe de départ (S_i) et ses pixels voisins qui pourraient être considérés comme des pixels candidats (P_c). Cette valeur est estimée sur la base de l'analyse de la distribution des intensités dans la fenêtre R_1 . Elle est exprimée comme la différence entre la valeur de l'intensité $I(S_i)$ du germe de départ (S_i) et le premier quartile ($Q_1(R_{1_{S_i}})$) dans la fenêtre R_1 définie autour de S_i comme suit :

$$I_{th} = I(S_i) - Q_1(R_{1_{S_i}}) \quad (4.3)$$

L'intersection de la région obtenue avec ce critère et de celle délimitée par la fenêtre R_2 définit la liste initiale des pixels candidats qui doivent remplir certaines autres contraintes pour être agrégée à une région en cours d'évolution R_c .

4.5.2.2 Critère d'énergie basée région

Les images mammographiques apparaissent avec une variation d'intensité significative entre les différents types de tissus mammaires. Les pixels MCs présentent, dans ces images, des valeurs d'intensité élevées avec des faibles contrastes par rapport aux pixels voisins qui peuvent appartenir à différents types de tissus. Ainsi, dans une même zone, nous pouvons avoir des pixels dont les valeurs d'intensité sont plus proches aux pixels des MCs qu'aux autres types de tissus. Avec ce cas de figure, le processus de croissance classique est incapable de distinguer les pixels de ces deux régions similaires. C'est pour cette raison que nous proposons de définir un critère d'énergie basée-région. Ce critère exploite le formalisme d'une croissance de région variationnelle pour distinguer les pixels

des MCs des autres pixels du tissu de superposition voisin. Ce formalisme exprime la croissance de région comme un processus d'optimisation qui vise à minimiser une fonction d'énergie [140, 141]. Ce concept variationnel a été largement utilisé par les algorithmes de contours actifs et de level-sets. Cependant, il a été exploité pour la première fois pour la croissance de région en 2010 par Rose *et al.* [142]. Il peut être formulé comme indiqué dans l'équation (4.4) où Φ^* est la solution optimale du processus de minimisation de l'énergie E .

$$\Phi^* = \arg \min(\Phi(E)) \quad (4.4)$$

Dans ce travail, nous avons utilisé une expression d'énergie basée région afin de réduire la sensibilité à la diversité du contraste des MCs ainsi qu'à la variation des intensités des pixels voisins. Elle est exprimée comme la moyenne des valeurs d'intensité dans une région R donnée. Pour une itération t , dans le processus de croissance proposé, la région R correspond à la région en cours d'évolution Rc . Ainsi, afin de quantifier la variation de l'énergie dans cette dernière, liée à l'ajout d'un candidat P_c , nous calculons deux valeurs d'énergie : une énergie interne (E_{In}) (Équation (4.5)) et une énergie externe (E_{Out}) (Équation (4.6)) pour les pixels du contour interne (Rc_{In}) et externe (Rc_{Out}) de la région Rc (Figure 4.9).

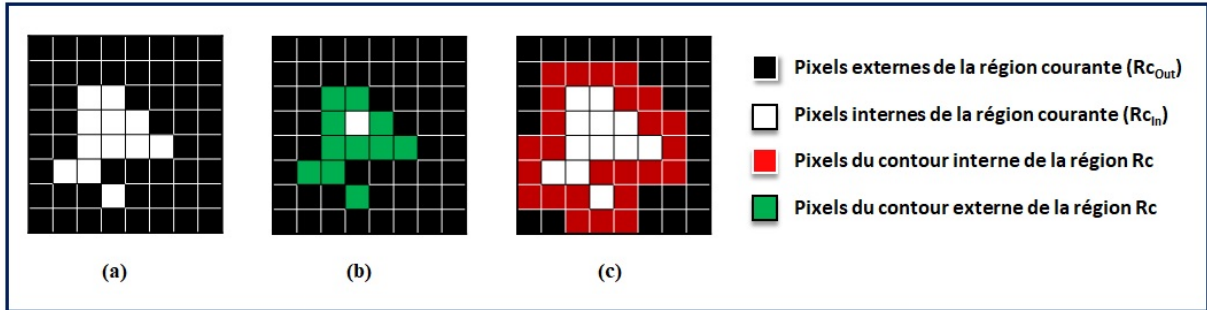


Figure 4.9 – Exemple des contours internes (b) et externes (c) d'une région Rc

$$E_{In}^t(P_c) = moy(Rc_{In}(P_c)) \quad (4.5)$$

$$E_{Out}^t(P_c) = moy(Rc_{Out}(P_c)) \quad (4.6)$$

L'énergie variationnelle $\Delta E(P_c)$ calculée par rapport à l'agrégation d'un pixel candidat P_c à la région Rc est exprimée par les variations des énergies interne (Équation (4.7)) et externe (Équation (4.8)). Ces dernières sont respectivement calculées comme la différence entre les énergies interne et externe pour deux itérations successives (t) et $(t-1)$.

$$\Delta E_{In}^t(P_c) = E_{In}^t(P_c) - E_{In}^{t-1} \quad (4.7)$$

$$\Delta E_{In}^t(P_c) = E_{In}^t(P_c) - E_{In}^{t-1} \quad (4.8)$$

Dans ce travail, nous avons adopté l'expression variationnelle de Chan et Vese [143] pour le calcul de l'énergie variationnelle (Équation (4.9)).

$$\Delta E_{Rc}^t(P_c) = (1 - 2 * Rc^t(P_c)) * (K1 - K2) \quad (4.9)$$

avec

$$K1 = |I(P_c) - E_{In}^t(P_c) - \Delta E_{In}^{t-1}|^2 \quad (4.10)$$

$$K2 = |I(P_c) - E_{Out}^t(P_c) - \Delta E_{Out}^{t-1}|^2 \quad (4.11)$$

Un pixel voisin candidat P_c est retenu comme pixel à ajouter à la région Rc uniquement si la valeur de son énergie variationnelle $\Delta E_{In}^t(P_c)$ est positive. Avec ce critère, nous traitons le problème de l'ajout des pixels appartenant à une région homogène (avec des valeurs d'intensités stables) ou à une nouvelle région voisine dont les valeurs d'intensité sont plus proches que la région décrite par le germe en question (MC ou autre). Dans ces cas de figure, l'énergie variationnelle est inférieure ou égale à zéro. Notre idée derrière l'utilisation de l'énergie variationnelle, est d'agréger, à chaque itération, les candidats les plus similaires à la région segmentée. Ce sont ceux qui présentent la plus faible variation d'énergie positive ΔE^t .

4.5.2.3 Critère de contraste basé contours

Le critère d'énergie précédent assure l'évolution d'une région Rc par le processus de croissance proposé avec des pixels de même type de tissu mammaire que les pixels qu'ils la constituent. En revanche, cette région peut correspondre à un tissu dense. Dans le cas où elle correspond ou présente une MC, le contour de cette dernière apparait très flou et non entièrement visible. Avec cette configuration, l'énergie modélisée peut ne pas délimiter correctement la MC recherchée puisque les valeurs des intensités des pixels internes de la région, qui la caractérise, sont très similaires à celles des pixels externes. Ainsi, la variation de l'énergie calculée peut présenter une faible valeur positive voir nulle. En revanche, malgré la diversité des contours et les contrastes faibles, les pixels contours d'une MC devraient présenter une similarité plus élevée aux pixels internes de la MC qu'à celle aux pixels externes. Ceci explique la nécessité d'intégrer un critère de contraste basé contours qui assure l'ajout, à une région Rc en cours d'évolution, du pixel candidat maximisant le contraste au niveau du contour.

Le troisième critère que nous intégrons modélise les contrastes des contours internes et externes relativement à une région donnée Rc . Compte tenu de la liste des voisins

candidats, retenus à partir de l'application des critères précédents, les valeurs de contraste des contours internes et externes liées à l'ajout d'un candidat P_c la région Rc sont calculées (Équations (4.12) et (4.13)).

$$InternB_{Rc}(P_c) = \frac{1}{N_1} \left(\sum_{p_i} Contr(P_i, Rc) \right) \quad (4.12)$$

$$ExternB_{Rc}(P_c) = \frac{1}{N_2} \left(\sum_{p_j} Contr(P_j, Rc) \right) \quad (4.13)$$

où :

- N_1 et N_2 : font respectivement référence au nombre de pixels internes et externes de Rc ;
- P_i et P_j : font respectivement référence aux pixels du contour interne et externe de Rc ,
- $Contr(P, Rc)$: fait référence à la valeur de contraste du pixel P par rapport à Rc calculée comme la différence des niveaux de gris.

Les pixels des contours internes (respectivement externes) sont définis comme l'ensemble des pixels les plus extérieurs de la région courante Rc (respectivement le fond). Enfin, la valeur du contraste du contour ($IEBound_{Rc}(P_c)$) de la région Rc suite à l'ajout du candidat P_c est calculée. Elle est exprimée comme la différence entre les contrastes du contour interne et externe précédemment calculés (Équation (4.14)). Les pixels candidats retenus pour l'évolution de la région Rc sont ceux qui présentent la valeur de contraste de contour la plus élevée.

$$IEBound_{Rc}^t(P_c) = InternB_{Rc}(P_c) - ExternB_{Rc}(P_c) \quad (4.14)$$

4.5.2.4 Critère basé sur la forme

Le quatrième et dernier critère que nous proposons traite de la forme des MCs qui est souvent circulaire ou elliptique. Il permet d'ajouter à la région segmentée le pixel voisin candidat qui respecte tous les critères précédemment décrits et présente le degré de circularité le plus élevé par rapport aux autres candidats. En fait, plus le degré de circularité tend vers 1, plus la forme de la région segmentée est circulaire. Le degré de circularité relatif à l'ajout d'un candidat (P_c) à une région (Rc) est exprimé comme suit :

$$Circ_{P_v}^t(Rc) = \frac{D_1}{D_2} \quad (4.15)$$

avec D_1 (respectivement D_2) correspond au plus petit (respectivement plus grand) diamètre de Rc .

4.6 Résultats et discussion

4.6.1 Méthodologie de validation et méthode de comparaison

Afin d'évaluer la fiabilité de la méthode proposée, les tests ont été réalisés sur l'ensemble des 50 images retenues à partir de la base INBreast. La même démarche de validation que nous avons présentée dans le chapitre précédent est suivie pour la validation de cette deuxième contribution.

En guise de comparaison, nous avons choisi la méthode de Melendez *et al.* [112] (notée ' M ' dans le reste de ce chapitre). Tel qu'il était décrit dans le deuxième chapitre, ce travail, utilisé pour la comparaison, a introduit le filtre Butterworth BBP [113] capable d'améliorer la qualité des pixels contours des MCs. Les résultats de l'application de ce filtre ont été comparés avec ceux obtenus avec les filtres DoG et LoG et ont montré une supériorité aussi bien dans des problématiques d'identification que de segmentation [112]. Notre choix est justifié par le fait que les deux méthodes (proposée et de comparaison) sont fondées sur un traitement local au niveau des pixels contours des MCs pour leur détection.

Dans cette étude expérimentale, nous allons examiner la fiabilité de la méthode proposée en termes d'identification et de segmentation. Pour ce faire, l'ensemble des résultats qualitatifs et quantitatifs qui seront exposés dans le reste de ce chapitre concernent uniquement les pixels de types TP ou FN. Les pixels de type FP ne seront pas considérés. Ce choix est fait dans l'idée de pouvoir comparer les taux de sur-segmentation de la méthode proposée par rapport à la méthode de comparaison utilisée.

4.6.2 Fiabilité de détection par rapport à la méthode de comparaison

Dans l'objectif d'évaluer la robustesse de la méthode proposée, il est important de montrer sa capacité à améliorer les performances obtenues par rapport à d'autres travaux. La figure 4.10 présente les cartes des MCs potentielles obtenues pour trois mammographies différentes avec différents types de MCs. Chaque colonne, dans cette figure, correspond à la ROI où apparaît les MCs sur une mammographie donnée. La première ligne montre les ROIs en niveaux de gris. La deuxième (respectivement troisième) ligne montre les résultats obtenus par la méthode de comparaison (' M ') (respectivement proposée).

Les résultats montrés sur cette figure présentent l'ensemble des pixels de types FN et TP issus de la superposition du masque de GT avec la carte des MCs potentielle retenue. Les pixels marqués en blanc correspondent aux pixels TP. Les pixels marqués en violet correspondent aux pixels FN. En revanche, les pixels noir regroupent les pixels du fond

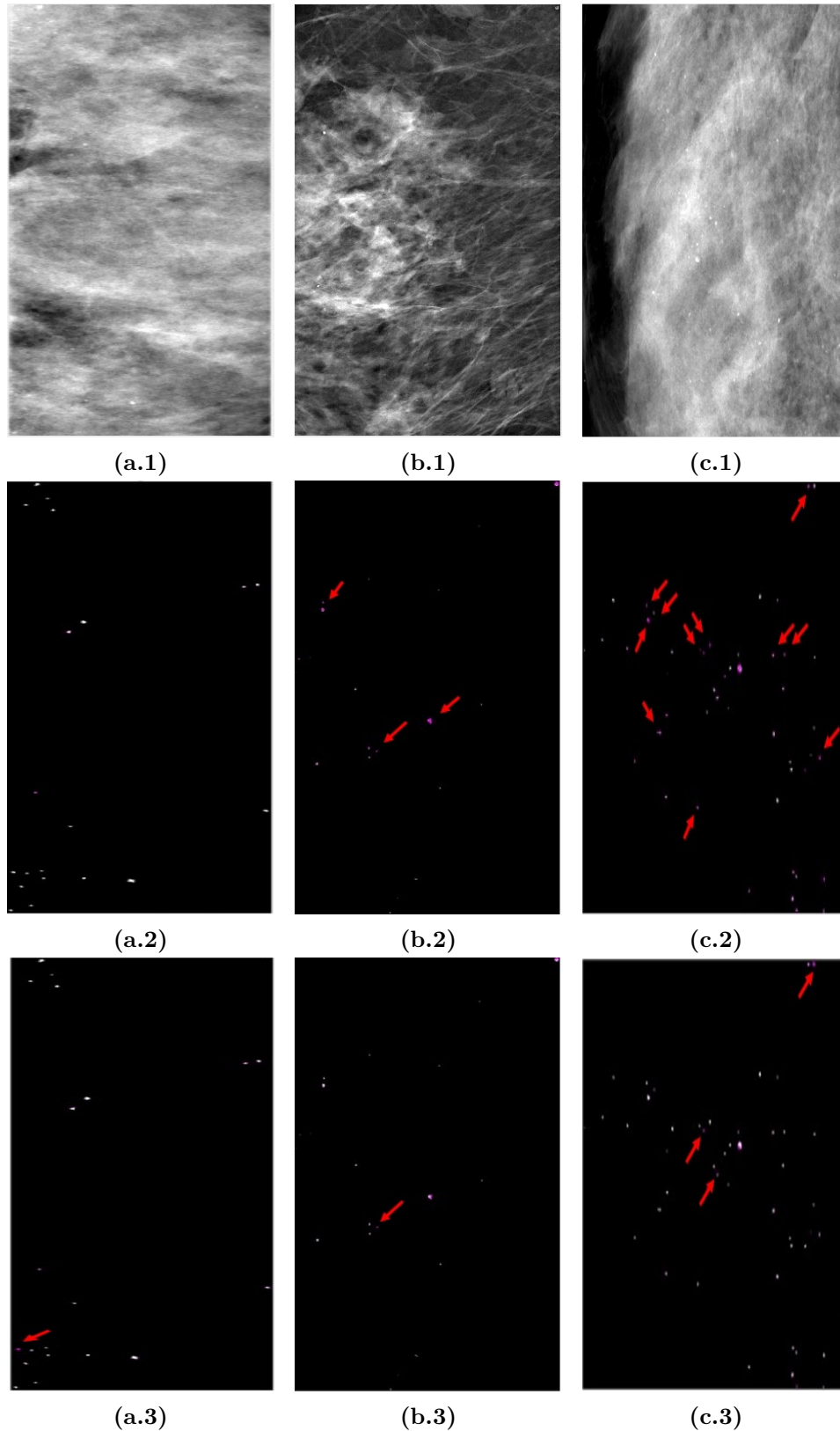


Figure 4.10 – Illustration des exemples de MCs détectées à partir de trois différentes mammographies (une mammographie par colonne) par la méthode ' M ' de comparaison (deuxième ligne) et la méthode ' CRC ' proposée (troisième ligne). Les flèches en rouge pointent sur les MCs qui n'ont pas été identifiées par une méthode donnée

de l'image ainsi que les pixels considérés, à tort, comme des pixels MCs par la méthode utilisée (les pixels FP).

L'analyse visuelle des résultats affichés montre que la méthode proposée est capable de réduire les détections de type FN même avec la présence des MCs de contrastes variables (Figure 4.10(b.3) et 4.10(c.3)).

Ces observations qualitatives doivent toujours être associées à des mesures quantitatives. Le tableau suivant présente une synthèse des résultats obtenus relativement à la mesure de sensibilité (SE) qui donne une information sur le taux d'identification des 'vraies' MCs par la méthode utilisée.

Tableau 4.1 – Description statistique de la répartition des valeurs de la sensibilité (SE) par la méthode de comparaison (' M ') et la méthode proposée (' CRC ') sur l'ensemble des images issues de la base INBreast

	MIN	MAX	Moyenne \pm écart-type
' M '	0.31428	0.97916	0.77251 \pm 0.17991
' CRC '	0.42857	1	0.75763 \pm 0.14797

La valeur moyenne et l'écart-type obtenus sont respectivement 0.75 et 0.14 pour la méthode proposée par rapport à 0.77 et 0.17 pour la méthode ' M '. Ces valeurs montrent que la méthode CRC proposée est capable d'identifier 75% des MCs présentes sur l'ensemble des mammographies de test comparant à 77% pour la méthode ' M '. Cette légère dégradation est liée à l'exclusion de certains maxima, correspondant réellement à des MCs, au niveau de l'étape de sélection. En revanche, ces mêmes valeurs supposent l'existence d'une concurrence en termes d'identification des MCs par les deux méthodes comparées avec un peu plus de stabilité pour la méthode proposée (une valeur d'écart-type légèrement inférieure à celle de la méthode ' M '). Nous allons infirmer, par les expérimentations décrites dans les sections suivantes, cette première constatation.

4.6.3 Fiabilité de segmentation

4.6.3.1 Évaluation qualitative

La figure 4.11 suivante illustre quelques exemples de MCs segmentées à partir de deux images mammographiques. Seules des ROIs présentant les MCs sur ces images sont montrées sur cette figure. A partir de chaque ROI, trois petites zones rectangulaires (contours respectivement jaune, vert et bleu sur l'image) sont sélectionnées. Ces sélections présentent des MCs avec des caractéristiques variables qui sont de type ACR3 pour la première ROI

et ACR4 pour la deuxième.

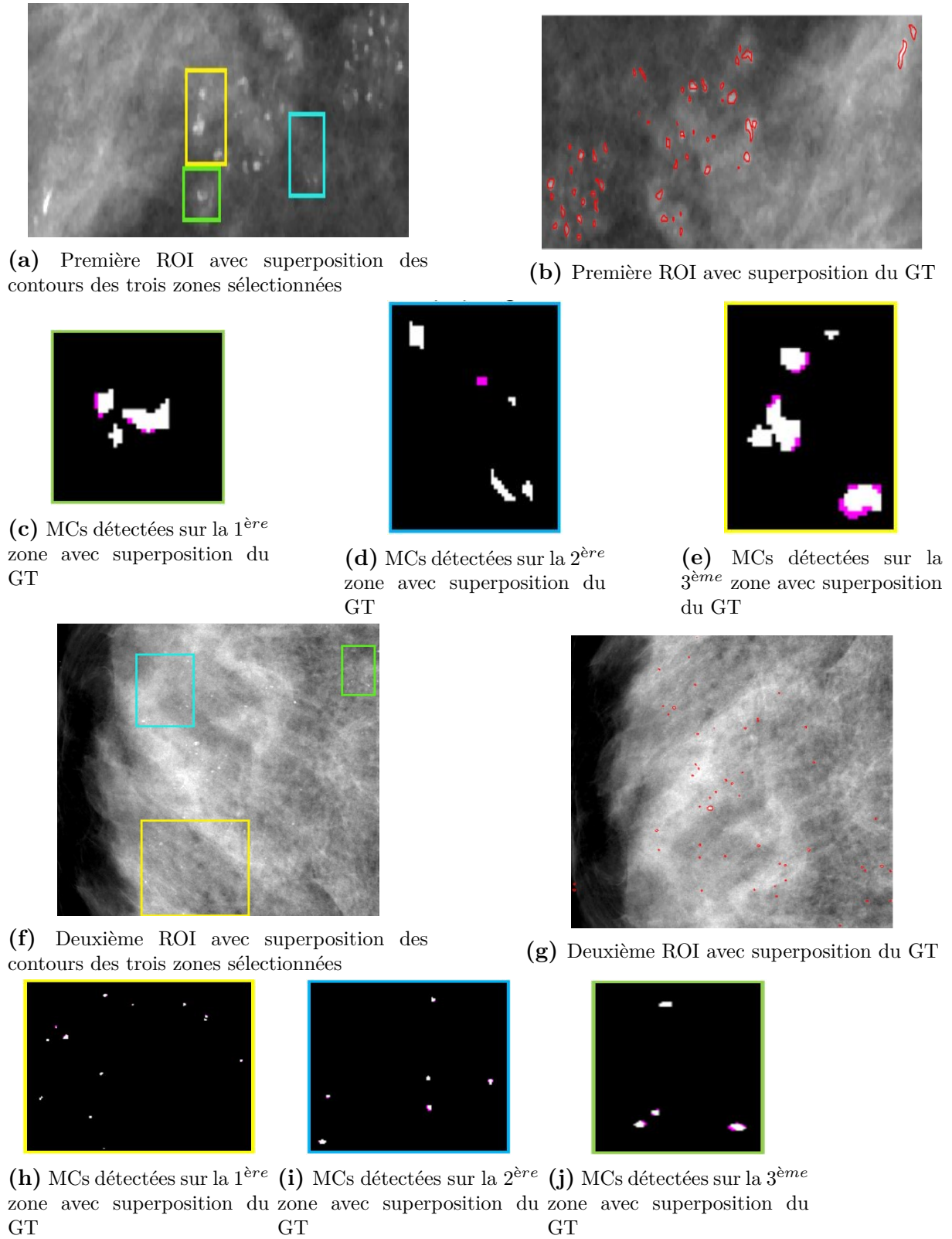


Figure 4.11 – Zoom sur les MCs détectées sur une sélection de trois zones différentes de deux ROIs

Les trois zones de la première ROI partagent la variation de taille et de formes des MCs présentes. En revanche, les MCs dans la deuxième zone (contour bleu) se distinguent par un contraste variable du fait qu’elles soient superposées sur deux différents types de tissus mammaires.

La deuxième ROI présente des MCs réparties sur une grande partie du tissu conjonctif homogène avec des petites tailles et de faibles contrastes. La taille variable très petite des MCs est le point commun des trois zones sélectionnées à partir de cette ROI. Par contre, la présence de contraste variable d’une MC à une autre est à noter des MCs dans la première (jaune) et la troisième (verte) zone.

L’analyse de ces résultats montre la fiabilité de la méthode que nous avons proposée pour correctement segmenter les MCs même si elles apparaissent avec des tailles et des contrastes différents.

4.6.3.2 Évaluation quantitative

L’évaluation quantitative que nous menons est basée sur les mêmes métriques précédemment définies et utilisées dans le chapitre précédent, à savoir l’indice de Dice (SI), la sous- (OF) et la sur- (EF) segmentation. Ces mesures permettent d’évaluer les similarités, en termes des taux d’intersections, entre le masque des MCs obtenu et le masque GT.

Afin d’étudier la fiabilité de la méthode proposée, nous avons calculé les valeurs moyennes, min, max et l’écart-type des indices SI, OF et EF. Ces derniers sont affichés dans le tableau 4.2. Une méthode fiable doit présenter des valeurs élevées (proches de 1) pour SI et OF et des valeurs faibles (proches de 0) pour EF. En d’autres termes, elle doit maximiser les taux d’identification (SI) et elle doit minimiser les taux de sous-segmentation (1-OF) et de sur-segmentation (EF).

Tableau 4.2 – Les valeurs Min, Max, Moyenne et Ecart-type des mesures d’évaluation. Les meilleurs résultats sont en gras

	<i>SI</i>		<i>OF</i>		<i>EF</i>	
	'M'	'CRC'	'M'	'CRC'	'M'	'CRC'
MOYENNE	0.75462	0.75728	0.84266	0.84883	0.06307	0.05104
ECART-TYPE	0.16598	0.09881	0.13894	0.05301	0.05171	0.0467
MIN	0.32489	0.54402	0.53038	0.72339	0.00355	0.00522
MAX	0.97185	0.94836	1	0.90804	0.21311	0.14413

Les valeurs affichées, dans ce tableau, sont calculées pour la méthode de comparaison 'M' ainsi que la méthode 'CRC' proposée à l’échelle de toute la base de données. Les meilleurs valeurs obtenues sont mises en gras. Les valeurs moyennes obtenues sont

d'environ 0.75 pour l'indice de similarité, 0.15 (1-0.85) pour le taux de sous-segmentation et 0.24 pour le taux de sur-segmentation. Les valeurs maximales montrent la capacité de la méthode proposée à correctement segmenter les MCs avec un minimum de sous-segmentation (1-OF = 0.1 = 10%) et un minimum de sur-segmentation (EF = 0.14 = 14%).

A première impression, ces pourcentages semblent peu satisfaisants. En revanche, n'oublions pas qu'une MC ne dépasse pas 1 mm de diamètre ce qui correspond à un environ 50 pixels. Un pourcentage de 10% (respectivement 14%) correspond à 5 (respectivement 7) pixels.

D'autre part, une première analyse suppose que les performances obtenues par la méthode proposée 'CRC' et la méthode 'M' de comparaison sont comparables. En revanche, nous remarquons que notre méthode présente une variance minimale sur l'ensemble des trois mesures évaluées. Cela signifie qu'elle offre une variabilité plus faible par rapport à la méthode de comparaison. Ceci est malgré la diversité des images de tests (en termes de types d'images et de MCs).

4.6.4 Discussion

Dans cette section, nous allons approfondir notre évaluation par l'étude des améliorations apportées par la méthode 'CRC' proposée par rapport à la méthode 'M' retenue pour la comparaison. Pour ce faire, nous avons distingué les images mammographiques utilisées pour les tests en quatre différentes familles en termes du contraste et des tailles des MCs qu'elles contiennent. *F1* (respectivement *F2*) regroupe l'ensemble des images avec des MCs caractérisées par un contraste faible (respectivement variable). *F3* (respectivement *F4*) regroupe l'ensemble des images avec des MCs caractérisées par des petites (respectivement diverses) tailles.

Tableau 4.3 – Pourcentages d'amélioration des métriques d'évaluation pour les différentes familles d'images mammographiques distinguées. Les couleurs rouge (respectivement bleu et vert) correspondent à OF (respectivement EF avec les familles F2 et F4)

	<i>Melandez et al. ('M')</i>				
	F1	F2	F3	F4	Totalité des images
SE	20%	50%	66.67%	42.85%	45%
SI	20%	50%	66.67%	42.85%	45%
OF	28.57%	16.67%	33.34%	71.43%	50%
EF	100%	100%	66.67%	85.71%	92.30

Le tableau 4.3 regroupe une analyse globale et une analyse détaillée, par famille ($F1$, $F2$, $F3$, $F4$), qui calcule les pourcentages où la méthode proposée présente des valeurs des métriques d'évaluations plus élevées par rapport à la méthode de comparaison 'M'. Dans l'idée de fournir une comparaison objective, les taux de sur- et de sous-segmentation sont comparées sur l'ensemble des MCs identifiées par les deux méthodes.

A partir de la colonne la plus à droite de ce tableau, la première constatation dégagée affirme la concurrence entre la méthode proposée et la méthode 'M'. En effet, elle montre que la méthode proposée a amélioré les taux d'identification globale des MCs et de sous-segmentation pour 45% respectivement 50% de la totalité des images de test utilisées. Par contre, la dernière case dans cette même colonne, affirme l'apport de notre méthode en termes de réduction des sur-segmentations. En effet, les valeurs de EF obtenues sont améliorées pour plus que 90% de la totalité des images de test. Ce pourcentage justifie une autre fois la capacité des critères intégrés d'assurer la convergence vers les meilleurs contours des MCs recherchés. D'autre part, l'analyse des pourcentages d'améliorations obtenues par famille confirment davantage nos premières constatations. En effet, les taux de sur-segmentation sont améliorés pour 67% des images qui présentent des MCs avec différents contrastes ($F3$) et pour 85% des images avec des MCs de très petites tailles ($F4$). Les meilleures améliorations en termes d'identification sont obtenues pour les familles $F3$, par 67%, et un peu moins pour $F2$, pour 50% des cas. En outres, les améliorations en termes de sous-segmentation sont obtenues pour plus que 70% des images de la famille $F4$.

A ce niveau, nous pouvons conclure que l'approche proposée pourrait traiter de manière adéquate toutes les caractéristiques des MCs et les délimiter correctement, même si elles apparaissent avec des tailles et des contrastes différents. En effet, nous sommes arrivés à réduire la sur-segmentation quelque soit le type du contraste des MCs. Par ailleurs, en plus de la réduction de la sur-segmentation pour la majorité des images avec des tailles variables, les taux d'identification (SI) (respectivement de sensibilité SE) sont améliorés pour les images avec des MCs de tailles petites (respectivement variables). Ces améliorations justifient la fiabilité des critères utilisés pour correctement distinguer les contours des MCs de leurs tissus mammaires de superposition. Elles prouvent également l'importance d'adopter une décision individuelle pour chaque MC en ce qui concerne ses caractéristiques spatiales.

4.7 Conclusion

Dans ce chapitre, nous avons conçu et évalué une méthode permettant la détection automatique des MCs fondée sur une croissance de région conditionnelle. Dans cette méthode, des maxima régionaux sont sélectionnés comme germes de départ pour le processus de croissance de région qui est contrôlée par des critères issus de connaissances *a priori* sur les MCs. L'intégration de ces connaissances est la principale contribution par rapport à l'algorithme classique de croissance de région. L'idée principale est d'appliquer une analyse spécifique pour chaque MC afin d'estimer les critères adéquats pour une délimitation précise, sans utiliser les mêmes paramètres et critères sur toute l'image. Les critères utilisés sont divisés en deux catégories. La première catégorie s'intéresse à la taille des MCs tandis que la deuxième traite de l'aspect du gradient et de l'évolution du contour. La précision des contours obtenus confirme la fiabilité des critères utilisés.

En revanche, il est intéressant de mentionner que malgré son efficacité en terme de précision de segmentation, cette méthode souffre de la production possible des faux négatifs. Ce problème soulève la question de la fiabilité de la sélection des germes de départ. En effet, malgré que nous nous sommes basés sur la description des radiologues pour effectuer cette sélection, elle nous a fait perdre des vrais MCs. La grande variété des tissus de superposition est à l'origine de cet échec.

A ce niveau, nous arrivons à répondre à la question initialement soulevée : *Est ce que l'on peut trouver une méthode capable de gérer, à elle seule, la diversité des images mammographiques et correctement interpréter les MCs y présentes ?*

Les deux premières contributions que nous avons présentées dans ce chapitre et dans le chapitre précédent étaient des propositions pour trouver et justifier la réponse à cette question. Avec la première contribution, nous avons proposé de traiter le problème de faible contraste pour la détection des MCs. L'idée de base était d'apporter des améliorations sur l'apparition de ces lésions dans l'objectif de faciliter la tâche de détection. En revanche, nous avons procédé par une analyse locale basée sur les connaissances *a priori* avec la deuxième contribution. L'idée était d'exploiter ces connaissances au niveau local afin de correctement converger vers les meilleurs contours de segmentation.

Dans les deux cas, nous avons pu apporter des améliorations par rapport aux travaux existants et justifier l'efficacité des raisonnements appliqués. Par contre, nous avons constaté des limites qui sont liées à la diversité des images mammographiques et des MCs.

Dans le chapitre suivant, nous allons présenter notre contribution principale qui consiste en une combinaison collaborative des résultats de plusieurs méthodes de détection. L'objectif principal consiste à faire collaborer ces résultats initiaux afin de produire une décision commune qui réduit le nombre de FNs par rapport à chaque méthode.

COMBINAISON COLLABORATIVE POUR LA DÉTECTION DES MICROCALCIFICATIONS

Sommaire

5.1	Introduction	126
5.2	Approche proposée	126
5.2.1	Motivations	126
5.2.2	Approche collaborative	127
5.2.3	Description générale	128
5.2.4	Espaces de représentation pixelique	129
5.3	Détection des zones suspectes	130
5.3.1	Principe et motivation	130
5.3.2	Représentation pixelique des résultats	131
5.4	Identification des objets candidats	133
5.4.1	Problèmes d'identification	133
5.4.2	Schéma d'identification proposé	136
5.5	Processus de classification collaborative	141
5.5.1	Description globale	141
5.5.2	Estimation des degrés de confiance	145
5.5.3	Ré-étiquetage des objets candidats	148
5.5.4	Convergence et représentation des résultats	151
5.6	Résultats expérimentaux et discussion	152
5.6.1	Détecteurs utilisés	153
5.6.2	Critères d'évaluation	155
5.6.3	Évaluation des performances	156
5.7	Conclusion	159

5.1 Introduction

Comme nous l'avons constaté d'après les chapitres précédents, il existe une grande variété de méthodes pour la détection de MCs dans la littérature. Cette multitude n'est pas restreinte à la problématique de détection des MCs, elle concerne tous les domaines de recherche. Pour une prise de décision, le choix de la 'meilleure' méthode est la tâche réservée à l'expert du domaine, qui trouve souvent lui-même une difficulté pour faire le bon choix, malgré son expertise.

Dans l'objectif d'améliorer la fiabilité en termes de résultats, les travaux de recherches se sont intéressés, ces dernières années, à l'utilisation conjointe de plusieurs méthodes pour traiter les mêmes données. C'est dans ce même ordre d'idée que nous proposons une approche de combinaison collaborative qui vise à détecter les MCs, à partir des images mammographiques, tout en minimisant le nombre de FNs. L'idée sous-jacente est d'opérer par un processus collaboratif qui permet de mettre en cause les décisions initiales de plusieurs méthodes de détection (appelées détecteurs) en fonction des informations partagées entre elles. Ce partage nous permet de tirer profit des avantages de chacune des méthodes afin de combler les faiblesses des autres.

Dans un premier temps, nous mettons en évidence les avantages de l'utilisation d'un processus collaboratif pour la détection des MCs. Ensuite, nous détaillons la démarche de l'approche proposée. Nous concluons ce chapitre par une évaluation de cette approche avec d'autres travaux dans la littérature.

5.2 Approche proposée

5.2.1 Motivations

Notre objectif principal, dans ce travail de recherche, consiste à développer une approche produisant un faible nombre de fausses détections de type FN. En effet, la morphologie et le nombre des MCs sont les facteurs les plus importants dans la décision prise par les radiologues [103, 144] à propos des degrés de malignité associés. Ainsi, une détection avec un faible nombre de faux négatifs est plus importante que celle avec des faux positifs (FPs). Un FP peut, à la limite, être filtré par un post-traitement. Par contre, une MC perdue ne pourra plus être récupérée, ce qui peut affecter la fiabilité des résultats obtenus.

Dans ce cadre, nous proposons un nouveau processus de classification collaborative dont l'idée est de tirer partie de différentes techniques de détection non supervisées, ap-

pelées *détecteurs*, afin de produire une décision commune plus pertinente que celle fournie par une technique individuelle. En effet, comme nous venons de justifier dans les chapitres précédents, une seule technique, basée sur un raisonnement particulier, est incapable de traiter la diversité des MCs en termes de caractéristiques, ce qui implique la génération de fausses détections. En revanche, du fait que les raisonnements diffèrent d'une technique à une autre, les fausses détections générées seront souvent différentes (Figure 5.1). D'où l'intérêt de procéder par une approche collaborative.

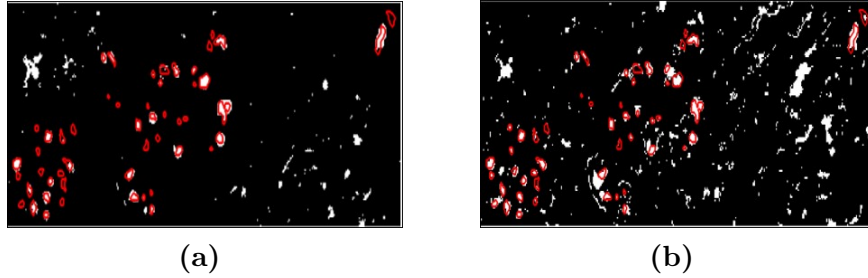


Figure 5.1 – Exemple de résultats de deux méthodes de détection de MCs avec superposition des contours GT (contours rouge). Les pixels blancs sont supposés appartenir à des MCs par une technique donnée

5.2.2 Approche collaborative

La collaboration est un processus dans lequel au moins deux acteurs collaborent ensemble et partagent leurs connaissances pour affiner les résultats initiaux et atteindre un objectif commun. Ce concept est commun pour toute approche menée dans le cadre d'un travail collectif de plusieurs acteurs. Outre la collaboration, on trouve deux autres types de processus : la coordination et la coopération [145]. La coordination consiste à construire une vision globale rassemblée des visions locales fournies par les différents acteurs participants. Prenons l'exemple des pièces d'un puzzle 2D, rassemblées ensemble, elles produisent un tableau, une image ou un dessin complet. La coopération, quand à elle, permet aux différents acteurs de participer à un travail commun en se basant sur certaines règles bien définies. Ces dernières permettent d'incorporer le travail de chaque acteur afin de valoriser la qualité de la production commune de l'ensemble.

Le choix du terme collaboratif pour l'approche proposée se justifie par la volonté de créer une décision commune issue de la communication et le partage de connaissances entre les différents acteurs participants, qui sont les différents détecteurs utilisés. On ne peut pas être dans un cadre de coordination du fait que chaque détecteur fournit une vision, propre mais différente, du problème global. Par ailleurs, on ne se situe pas non plus dans le cadre d'une coopération du fait qu'on ne peut pas prévoir ni décider de la fiabilité des détecteurs.

Dans la littérature, le *clustering* des données est le domaine d'application le plus fréquent des approches dites collaboratives [146]. Le *clustering* est un traitement dont l'objectif est de regrouper un ensemble de données, décrites par un vecteur d'attributs, en groupes appelés *clusters* ou *classes*. La problématique de détection des MCs à partir de l'image mammographique, dans ce travail de recherche, correspond parfaitement à un problème de clustering où on cherche à diviser les pixels de l'image en deux classes : la classe des pixels MCs et la classe de pixels de tissus mammaires sains. En revanche, les approches de clustering collaboratives, ont pour objectif de créer, via la collaboration, un consensus pour le partitionnement d'un ensemble de données. Le caractère clé qui diffère notre approche proposée des approches standards est la variation de l'ensemble des données issues de chaque collaborateur. En effet, les fausses détections générées par les détecteurs sont différentes.

5.2.3 Description générale

L'idée de base de l'approche proposée est d'opérer par un processus collaboratif itératif qui permet à différentes techniques de détection non supervisées (détecteurs) de partager des informations et de remettre en cause leurs décisions initiales en fonction de celles des autres détecteurs. Un échange d'informations et de connaissances est ainsi établi, à chaque itération, afin de pouvoir converger vers une décision commune et plus pertinente par rapport à celle fournie par un détecteur unique. Nous distinguons, sur le schéma global de cette approche, trois principales phases : la détection des zones suspectes, l'identification des objets candidats et la classification collaborative (Figure. 5.2).

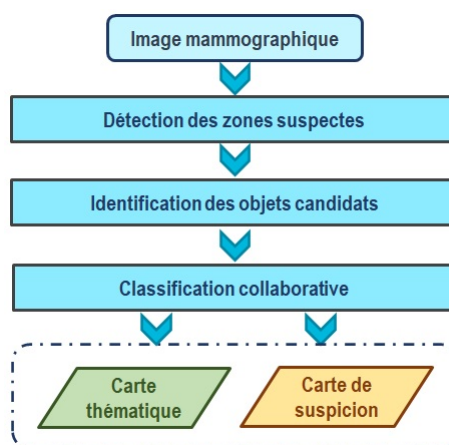


Figure 5.2 – Organigramme global de l'approche proposée

La phase de détection consiste en l'application séparée d'un ensemble de détecteurs sur l'image mammographique divisée en régions locales homogènes. Chaque détecteur fournit deux cartes pixeliques traduisant le résultat de son interprétation de l'image.

La phase d'identification constitue une transformation de l'espace d'interprétation pixelique, issu de la première phase, vers un espace d'interprétation objet qui sera exploité par la troisième phase de classification.

La phase de classification consiste à faire collaborer les résultats initiaux des détecteurs, interprétés dans la phase précédente et représentés sous la forme de listes d'objets candidats. Le partage des informations relatives aux objets ainsi que leurs étiquettes associées permet de raffiner les décisions initiales des détecteurs en fonction de celles des autres collaborateurs. Ce raffinement consiste à mettre à jour les étiquettes, initialement associées, aux objets candidats issus de chaque détecteur sur la base de deux analyses : locale et contextuelle.

La décision finale de l'approche proposée est simplifiée à une fusion conjonctive par l'application d'un opérateur d'intersection sur les nouvelles décisions fournies par le processus collaboratif appliqué.

5.2.4 Espaces de représentation pixelique

Le résultat final fourni pour les experts est modélisé sous la forme de deux cartes pixeliques : une carte thématique et une carte de suspicion.

5.2.4.1 Carte thématique

Une *carte thématique* (notée TM) est une carte qui associe une valeur binaire (0 ou 1) à chaque pixel P de l'image (Figure 5.3).

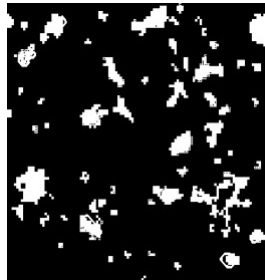


Figure 5.3 – Exemple d'une carte thématique. Les pixels blancs sont supposés appartenir à des MCs. Les pixels noir sont considérés comme des pixels de fond de l'image (pixels de tissus mammaire sain)

Les pixels ayant une valeur 1 sont supposés appartenir à une MC. Un ensemble de pixels connectés avec des valeurs égales à 1 est considéré comme une région d'intérêt que nous appelons *zone suspecte* et qui peut correspondre à une MC potentielle. En revanche, les pixels ayant la valeur 0 sont supposés être des pixels d'un tissu mammaire sain et

sont considérés comme des pixels de fond. Une telle représentation offre une segmentation binaire de l'image mammographique qui représente les MCs potentielles.

5.2.4.2 Carte de suspicion

Une *carte de suspicion* (notée SM) est une carte qui associe une valeur continue (comprise entre 0 et 1) à chaque pixel P (Figure 5.4). Les pixels présentant des valeurs de degré de suspicion faibles (proches de 0) correspondent à des pixels de tissu mammaire sain. Les pixels ayant des valeurs de degré de suspicion élevées (proche de 1) correspondent à un tissu suspect et doivent être considérés comme dignes d'intérêt par les radiologues. Une telle représentation est une sorte de passage du niveau de gris standard à un espace de représentation sémantique. Elle offre la possibilité de réduire l'impact des informations incertaines en niveau de gris sur le processus de prise de décision. Par ailleurs, elle peut être utilisée avec la représentation binaire pour classer les zones suspectes, à considérer comme des MCs, en bénignes ou malignes sur la base de leurs morphologies ainsi que leurs distributions.

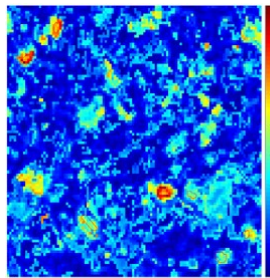


Figure 5.4 – Exemple d'une carte de suspicion. Plus les valeurs des pixels sont proches de '1' (couleur rouge), plus le pixel correspondant est suspicieux et est digne d'intérêt

5.3 Détection des zones suspectes

5.3.1 Principe et motivation

La phase de détection vise à identifier, à partir d'une image mammographique entière, les pixels connectés à considérer comme des *zones suspectes* pouvant correspondre à des MCs. On définit, dans ce travail, une zone suspecte tout ensemble de pixels adjacents qui répondent à la caractérisation des MCs, adoptée par une technique de détection donnée. Elle peut désigner une MC, une partie d'une MC ou une région plus grande qu'une MC. Le principe de base est d'interpréter l'image mammographique entière, divisée en régions homogènes appelées superpixels SP1, avec plusieurs techniques de détection (dites des détecteurs). L'idée sous-jacente est de travailler avec de petites régions présentant, chacune, un même type de tissu mammaire afin de traiter la diversité des MCs ainsi que

des tissus mammaires. Dans cette étude, les superpixels SP1 utilisés sont les superpixels SLIC (précédemment décrits dans le chapitre 4) générés à partir de l'image mammographique de sorte à respecter un critère spécifique d'homogénéité du niveau de gris et peuvent contenir plus d'une seule MC. L'ensemble des MCs qui appartiennent au même superpixel est appelé 'groupe de MCs'.

La figure 5.5 montre les contours de superpixels SP1 (contours bleus) générés pour une petite région d'une image mammographique présentant des MCs. La zone agrandie (délinée par le rectangle de contour rouge en pointillé) montre la relation, décrite ci-dessus, entre les MCs et les superpixels SP1.

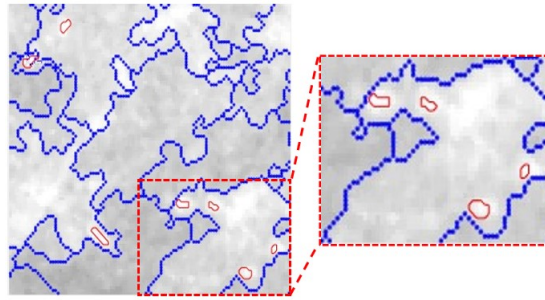


Figure 5.5 – Exemple de superpixels SP1 générés (contours bleu) avec superposition des contours GT (contours rouges)

5.3.2 Représentation pixelique des résultats

Dans cette première phase d'identification (Figure 5.6), les M détecteurs utilisés traitent les pixels des mammographies numériques et génèrent, chacun, deux cartes pixeliques : la carte thématique et la carte de suspicion.

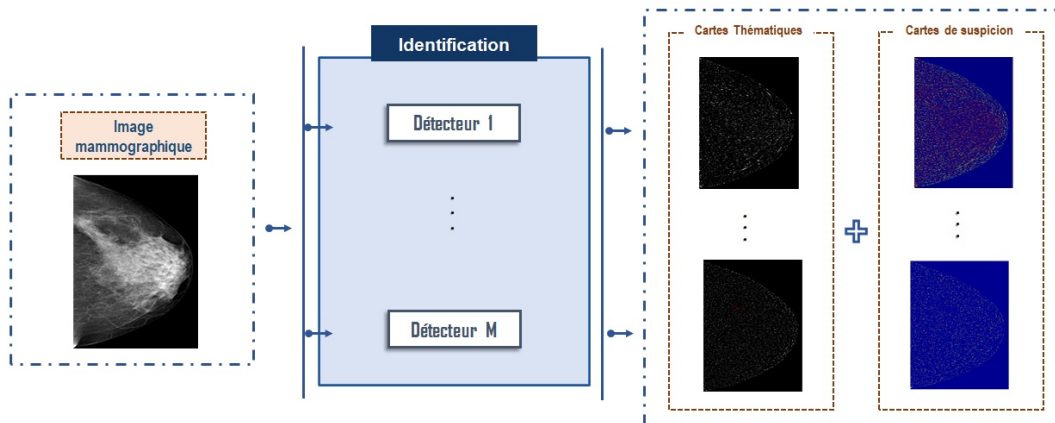


Figure 5.6 – Schéma global de la phase d'identification des zones suspectes

Il existe une relation implicite entre les cartes thématiques et de suspicion générées par chaque détecteur, où chacune d'entre elles peut être dérivée de l'autre. La conversion,

que nous proposons, d'une carte à l'autre est basée sur des connaissances *a priori* (Figure 5.7). Il s'agit d'analyser la distribution des niveaux de gris des pixels des zones suspectes dans la carte thématique et les distributions des degrés de suspicion dans les superpixels SP1 de la carte de suspicion.

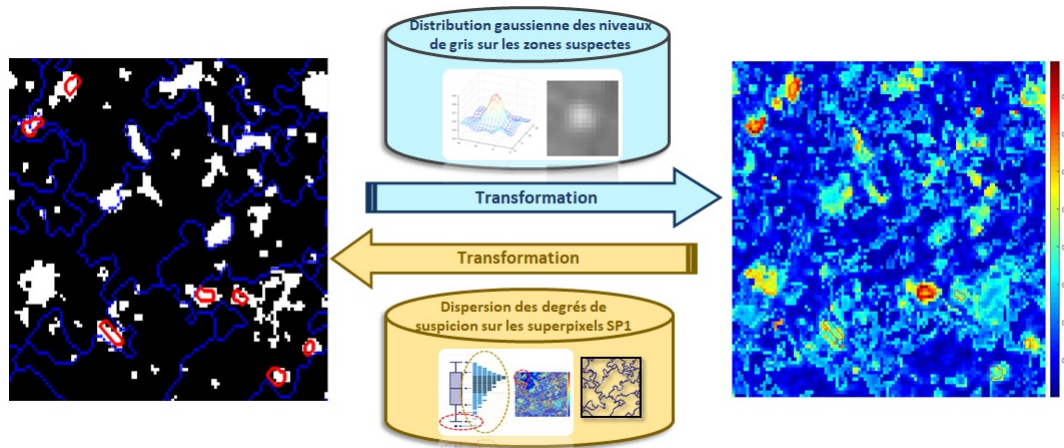


Figure 5.7 – Conversion des cartes thématique et de suspicion

À partir de la carte thématique, les zones suspectes obtenues sont sélectionnées et la distribution des niveaux de gris des pixels dans chacune d'elles est examinée. Nous avons déjà démontré (section. 4.4.1 du chapitre 4) que ces distributions correspondent bien à des distributions gaussiennes symétriques par rapport aux maxima régionaux. Nous proposons ainsi de transformer ces distributions en valeurs de probabilités à considérer comme degrés de suspicion. Pour le faire, nous avons utilisé la table de transformation z-score.

Pour transformer une carte de suspicion en une carte thématique, nous commençons par projeter les limites des superpixels SP1 sur cette dernière. Avec une telle projection, la carte de suspicion est divisée en régions locales ayant des valeurs de niveaux de gris homogènes. Les pixels adjacents, qui sont considérés comme des pixels MC potentielle et donc zone suspecte, sont les pixels outliers dans chaque superpixel SP1 considéré. Par conséquent, nous proposons d'estimer une valeur de seuil adaptative pour chaque superpixel SP1 en se basant sur la modélisation des pixels outliers de celui-ci. Dans ce travail, la valeur seuil, pour laquelle un pixel P est considéré comme un pixel outlier dans un superpixel donné, est fondée sur le critère de John Wilder Tukey [132] décrit dans le chapitre 3.

5.3.2.1 Schéma de prétraitement adopté

Nous avons discuté dans le deuxième chapitre l'utilité d'une étape de prétraitement dans l'objectif d'améliorer la qualité des images mammographiques. Nous avons également effectué une synthèse sur l'ensemble de techniques utilisées dans la littérature et associées

à la détection des MCs.

Dans ce travail, nous allons proposer une approche capable de correctement détecter les MCs tout en explorant les superpixels SP1 générés à partir de l'image. De ce fait, nous avons intérêt à générer des superpixels de meilleure qualité et qui soient capables de correctement caractériser ses objets recherchés (les MCs). De ce fait, le choix de la technique de prétraitement la plus adéquate exige l'amélioration de :

- la qualité d'apparition des MCs sur l'image (visibilité) ;
- la qualité des superpixels SP1 générés.

Nous avons, également, présenté dans le deuxième chapitre l'ensemble des prétraitements existants appliqués aux images mammographiques. Dans l'objectif d'adapter une technique à utiliser dans notre travail, nous avons évalué la pertinence de certaines techniques, les plus adaptées pour l'interprétation des MCs, en fonction de nos besoins. Les résultats de ces techniques sont illustrés dans l'annexe (A). Une discussion de leurs pertinences est aussi abordée dans le même annexe. Les conclusions tirées suite à ces études nous amènent à choisir le schéma de prétraitement donné dans la Figure 5.8.

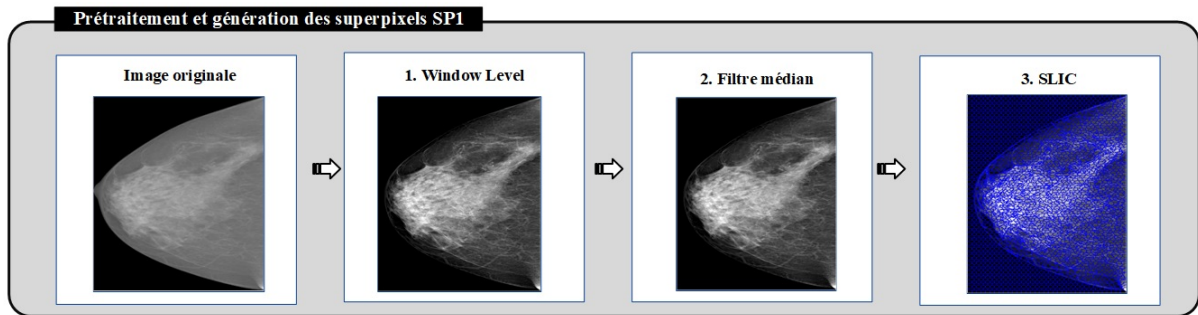


Figure 5.8 – Schéma de prétraitement adopté par l'approche proposée

Dans un premier temps, la technique Window/Level est appliquée pour transformer l'image mammographique en une image de 8 bits de profondeur et d'améliorer la distinction de ses différentes structures mammaires. Un filtre médian est ensuite appliqué afin de 'légèrement' débruiter l'image obtenue par la technique Window/Level, ce qui permet de préserver les structures mammaires fines sur l'image. Et finalement, l'algorithme SLIC est appliqué afin de générer l'ensemble des superpixels à partir de cette image prétraitée.

5.4 Identification des objets candidats

5.4.1 Problèmes d'identification

Dans la phase précédente, des traitements pixeliques sans contraintes sur les caractéristiques géométriques des régions obtenues (zones suspectes) ont été appliqués. Ainsi,

nous ne disposons d’aucune information sur les formes et les tailles des zones suspectes générées par rapport à celles des MCs. En revanche, ces dernières constituent dans notre travail le point de départ pour décider si leurs pixels correspondent ou non à des pixels de MCs. Il est donc dans notre intérêt d’identifier à partir d’elles le ou les ensembles de pixels adjacents pouvant faire référence à une MC.

L’objectif de base de cette phase d’identification consiste, ainsi, à transformer les zones suspectes, résultats de la phase précédente, en un ensemble d’*objets candidats* avec des caractéristiques associées. Un *objet candidat* est défini comme l’ensemble de pixels adjacents appartenant à une zone suspecte et pouvant faire référence à une MC potentielle. Ceci implique qu’il doit impérativement respecter les caractéristiques géométriques des MCs en termes de taille et de forme. Par ailleurs, un objet candidat peut correspondre à une zone suspecte entière ou à une partie d’une zone suspecte.

Cette transformation doit tenir compte des deux contraintes suivantes :

- Il faut s’assurer que la taille des objets identifiés soit conforme à celles des MCs. Cependant, on ne dispose d’aucun moyen pour juger la pertinence des contours des zones suspectes obtenues à partir de chaque détecteur ;
- Il faut prendre en considération la variabilité, en termes de taille et de forme, des zones suspectes issues des différents détecteurs.

5.4.1.1 Zones suspectes de diverse tailles et formes

Par la figure 5.9, nous allons étudier les liens possibles, en termes de taille et de forme, entre les MCs et les zones suspectes identifiées à partir d’une carte thématique issue d’un détecteur donné. Cette figure présente une même ROI d’une image mammographique sur laquelle sont superposées les zones suspectes identifiées à partir d’un même détecteur (régions blanches avec de contour bleu). Les contours rouges décrivent les contours vérité de terrain des MCs présentes sur la ROI étudiée.

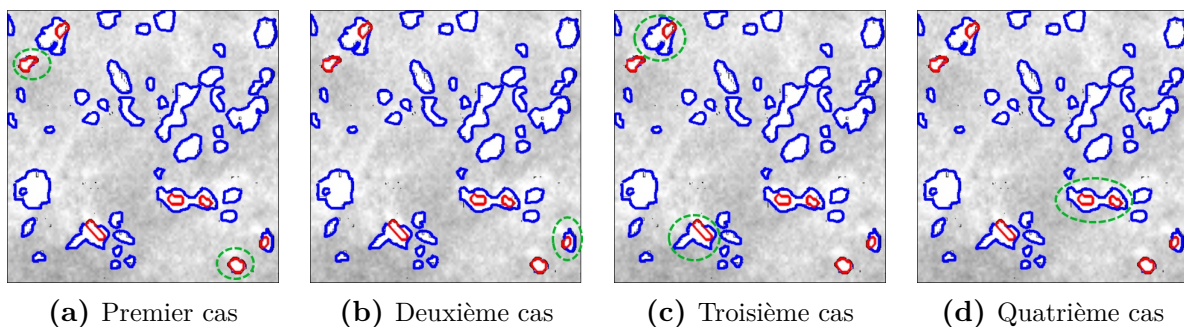


Figure 5.9 – Illustration d’un exemple de zones suspectes de différentes tailles et formes identifiées à partir d’une carte thématique d’un même détecteur

Nous distinguons sur cette figure quatre différents types de zones :

- Une zone suspecte qui correspond exactement à une MC : La région décrite par cette zone est conforme en taille et en forme avec la MC en question (Figure 5.9a) ;
- Une zone suspecte qui s'intersecte avec une seule MC (Figure 5.9b et 5.9c). La région définie par cette zone regroupe les pixels de la MC avec d'autres pixels du tissu mammaire sain. La taille de la région qui ne correspond pas à une MC peut varier de quelques (5.9b) à plusieurs (5.9c) pixels ;
- Une zone suspecte qui s'intersecte avec plusieurs MCs (Figure 5.9d). La région définie par cette zone regroupe plusieurs MCs sans qu'elle dispose forcément d'une grande taille.

5.4.1.2 Différentes zones suspectes d'un détecteur à un autre

Rappelons que les détecteurs utilisés, dans ce travail, exploitent différentes stratégies et paramètres pour identifier les pixels de MCs potentielles. Ainsi, deux régions R_i et R_j qui font référence à deux zones suspectes dans les cartes thématiques du $i^{\text{ème}}$ et $j^{\text{ème}}$ détecteur peuvent être :

- *Chevauchantes* : La région R_i obtenue par le $i^{\text{ème}}$ détecteur présente des pixels communs avec la région R_j issue du $j^{\text{ème}}$ détecteur (Figure 5.10a et 5.10b) ;
- *Différentes* : Les pixels qui représentent la région R_i dans la $i^{\text{ème}}$ carte thématique sont associés à la classe de fond dans la $j^{\text{ème}}$ carte thématique (Figure 5.10c) et vice-versa.

La figure 5.10 fournit une illustration des différents scénarios possibles lors de l'identification des objets à partir des cartes thématiques. Les pixels noirs de la carte thématique correspondent à des pixels de tissu mammaire sain (arrière-plan) tandis que les pixels blancs correspondent à des pixels de MCs potentielles. Chaque groupe de pixels adjacents blancs correspond à une zone suspecte. Sur ces cartes thématiques, nous représentons les zones suspectes qui résultent de deux détecteurs D_i et D_j délimitées par des contours respectivement rouge et vert.

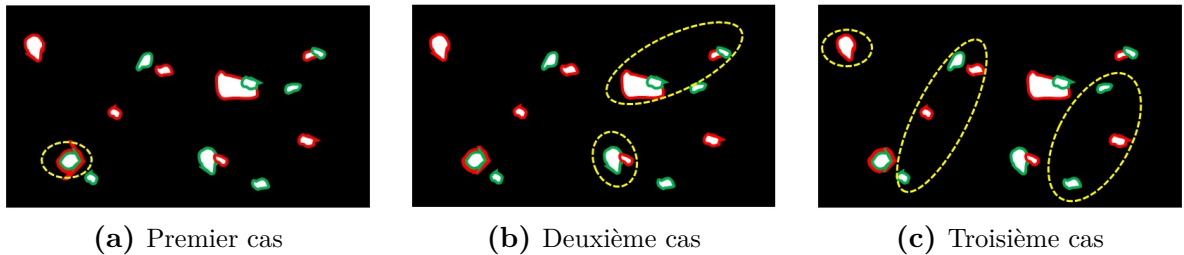


Figure 5.10 – Les différents scénarios possibles lors de la comparaison des zones suspectes issues de deux détecteurs différents

À ce niveau, les questions auxquelles nous devons apporter une réponse sont les suivantes :

- Les régions obtenues (zones suspectes) respectent-elles à la granularité des MCs ?
- Quels sont les contours de l'objet ou des objets à considérer si deux détecteurs présentent des régions avec chevauchement sachant que nous ne sommes pas en mesure de juger la pertinence d'un détecteur donné par rapport aux autres ?
- Quelles sont les régions à considérer comme objets candidats dans le cas où les détecteurs génèrent des régions différentes (l'ensemble de toutes les régions ou une sélection et sur quelle base ces décisions seront prises) ?

5.4.2 Schéma d'identification proposé

Afin de traiter les problèmes évoqués ci-dessus, nous proposons un schéma d'identification (Figure 5.11) qui utilise trois types d'information différentes : les superpixels SP2, les cartes thématiques et les cartes de suspicion. Les superpixels SP2 sont les superpixels LPE déjà décrits dans le chapitre 4 précédent.

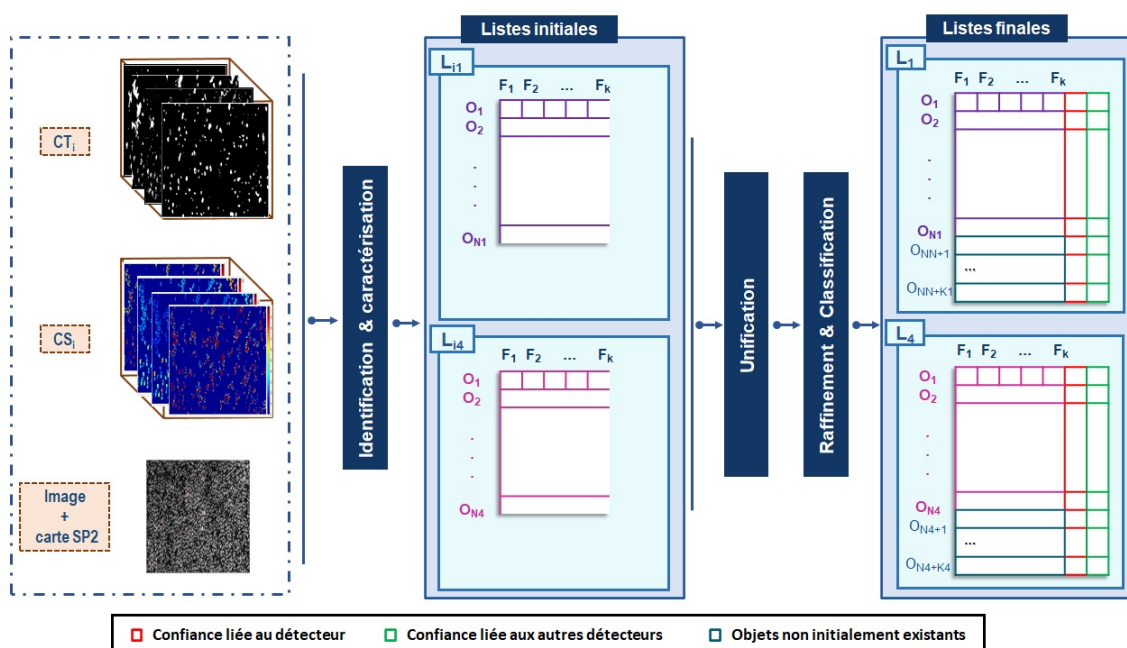


Figure 5.11 – Schéma du déroulement de la phase d'identification des objets candidats

Les zones suspectes issues de chaque détecteur, et identifiées à partir de leurs cartes thématiques correspondantes, sont dans un premier temps converties, chacune, en une liste d'objets candidats. La liste globale de tous les objets est ensuite générée à partir de l'union de ces listes partielles précédemment générées. En se basant sur cette nouvelle liste, les listes initiales des objets candidats, relativement à chaque détecteur sont raffinées et les objets candidats qu'elles présentes sont étiquetés en fonction de la fréquence de leur

occurrence dans tous les détecteurs.

Nous distinguons ainsi trois principales étapes, décrites dans les paragraphes suivants :

- Génération des listes des objets candidats initiaux ;
- Unification des listes des objets candidats ;
- Raffinement et classification initiale des objets.

5.4.2.1 Étape 1 : Génération des listes des objets candidats initiaux

La transformation des zones suspectes en listes d'objets candidats est basée sur les superpixels SP2 générés suite à l'application de l'algorithme LPE sur l'image mammographique. En effet, et comme nous l'avons montré dans le chapitre précédent, ces superpixels présentent la particularité de partage de propriétés géométriques (taille et forme) avec les MCs (Figure 4.5b du chapitre 4). En outre, les points de départ pour la génération de ces superpixels sont les maximas régionaux sur l'image (Figure 4.2 du chapitre 4). Ces maximas désignent souvent les centres de gravité des superpixels générés mais aussi les centres de gravité des MCs, s'ils l'appartiennent, (Figure 4.4 du chapitre 4).

La transformation proposée consiste à projeter la carte de superpixels SP2 sur chacune des cartes thématiques générées par les différents détecteurs. La carte de superpixels SP2 est utilisée pour définir les contours des objets candidats. L'utilisation de cette carte nous permet de surpasser les problèmes liés aux chevauchements et à la taille variable des zones suspectes générées. Afin de correctement déterminer les contours des objets identifiés à partir des zones suspectes, nous proposons d'analyser les régions occupées par ces zones par rapport à celles des superpixels SP2 auxquels elles appartiennent.

La figure 5.12 présente trois exemples de zones suspectes (régions blanches avec contour rouge) sur une carte thématique (image à gauche dans chaque ligne). Elle illustre les deux situations possibles lors de la comparaison d'une zone suspecte avec les superpixels SP2 (contours bleu en pointillés) :

- la zone suspecte s'intersecte avec un seul superpixel SP2 (première ligne) ;
- la zone suspecte s'intersecte avec deux (deuxième ligne) ou plusieurs (troisième ligne) superpixels SP2.

Pour la première situation, nous considérons la zone suspecte comme un seul objet candidat de contours bien définis. Il s'agit des contours de la zone suspecte en question du fait qu'elle est bien circonscrite dans son superpixel SP2.

Pour la deuxième situation, nous divisons la zone suspecte en différents objets. En effet, nous distinguons deux (respectivement neuf) objets sur l'exemple illustré par la deuxième

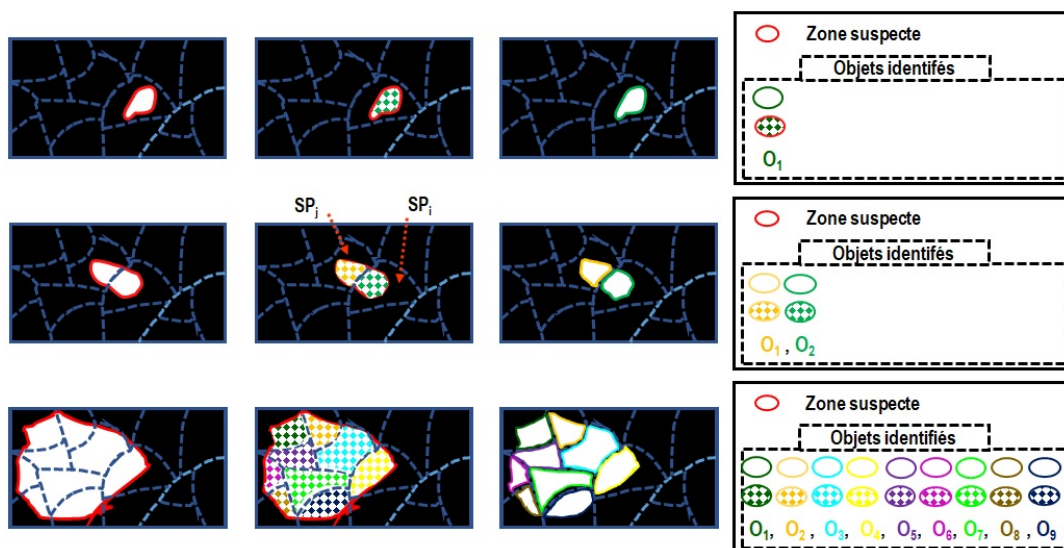


Figure 5.12 – Illustration des exemples d’objets candidats identifiés à partir de trois différentes zones suspectes

(respectivement troisième) ligne. Les intersections des contours délimitant les superpixels SP_2 avec la zone suspecte définissent les contours associés aux objets identifiés. A titre d’exemple, la zone suspecte (région blanche avec contours rouge), sur la deuxième ligne de la figure 5.12, intersecte deux superpixels différents SP_i et SP_j . Le contour divisant ces deux superpixels (contour en pointillés bleu) est celui qui définit les contours des deux objets identifiés (régions avec contours verts et jaunes dans l’image à droite).

Cette configuration nous permet de traiter tous les problèmes liés à la sélection des meilleurs contours d’un objet à partir d’une zone suspecte. Elle permet également d’uniformiser le processus d’identification des objets pour tous les détecteurs. Elle permet, par ailleurs, de s’assurer que tous les objets considérés soient conformes ou du moins présentent des formes et des tailles très proches de celles des MCs. Ainsi, le plus grand objet que nous acceptons correspond à un superpixel SP_2 (O_3 , O_5 et O_7 de la troisième ligne de la figure 5.12 par exemple).

Chaque objet O_i sera identifié par son centre de gravité et une étiquette héritée de son superpixel d’appartenance SP_2 (l’étiquette du superpixel sur la carte SP_2). Une fois les objets identifiés, certaines caractéristiques sont calculées à partir des cartes de suspicion et de niveaux de gris (image mammographique) et seront utilisées dans le processus de collaboration.

5.4.2.2 Étape 2 : Unification

La liste globale des objets candidats est l’union des listes précédemment identifiées à partir des différents détecteurs. Pour chaque objet dans cette liste, nous calculons le

même ensemble de caractéristiques que pour les objets de chaque détecteur. De point de vue théorique, les objets ayant la même étiquette SP2 (c'est à dire situés dans le même superpixel SP2) devront être considérés comme un objet unique dans cette liste.

En revanche, nous présentons par la figure 5.13 les différentes situations que nous pouvons rencontrer lors de l'analyse de deux objets O_1 et O_2 , ayant la même étiquette SP2 et identifiés à partir de deux zones suspectes relatives à deux détecteurs différents D_1 et D_2 . Les régions blanches de contours jaune et vert, sur cette figure, représentent ces objets. Les contours en pointillés bleu représentent les contours des superpixels SP2.

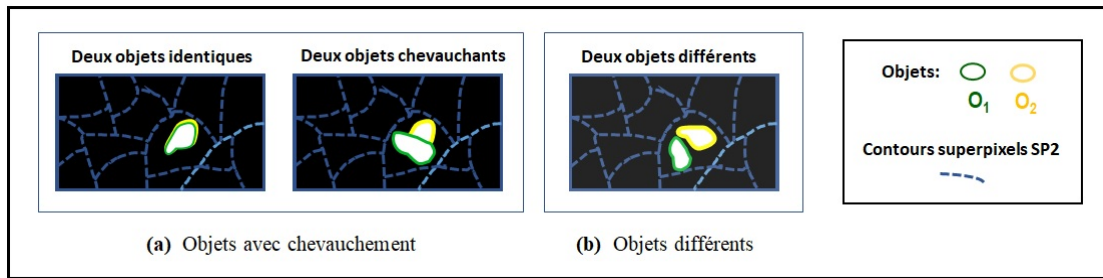


Figure 5.13 – Les différents scénarios possibles lors de l'unification des listes des objets

Nous pouvons ainsi distinguer :

- Deux objets *chevauchants* : Les régions caractérisant les objets O_1 et O_2 présentent des pixels en commun (Figure 5.13 a). Elles peuvent être identiques ou chevauchantes avec quelques pixels en commun ;
- Deux objets *différents* : Les régions caractérisant les objets ne présentent aucun pixel en commun (Figure 5.13b).

La configuration que nous proposons est la suivante. On enregistre, pour toutes ces situations, un seul objet O dans la liste globale des objets candidats et il correspond à :

- la région R issue de l'union de O_1 et O_2 pour le cas de chevauchement (Figure 5.14) ;

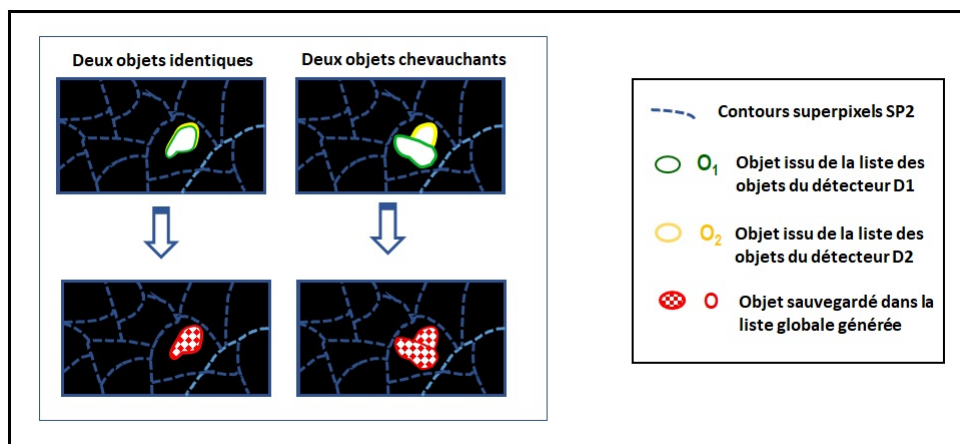


Figure 5.14 – Illustration d'un exemple de détermination de l'objet O à considérer dans la liste globale des objets à partir de deux objets présentant un chevauchement

- la région R qui caractérise un de ces objets (Figure 5.15 a ou 5.15 b) à condition qu'il possède la distance géométrique ($D(C_G(O_i), C_G(SP2))$, $i = 1$ ou 2) entre son centre de gravité ($C_G(O_2)$) et centre du superpixel SP2 ($C_G(SP2)$) (Équation (5.1)). Dans le cas contraire, l'objet à considérer est le superpixel SP2 (Figure 5.15 c).

$$O = \begin{cases} O_1 & \text{Si } D(C_G(O_1), C_G(SP2)) < D(C_G(O_2), C_G(SP2)) \\ O_2 & \text{Si } D(C_G(O_2), C_G(SP2)) < D(C_G(O_1), C_G(SP2)) \\ SP2 & \text{Sinon} \end{cases} \quad (5.1)$$

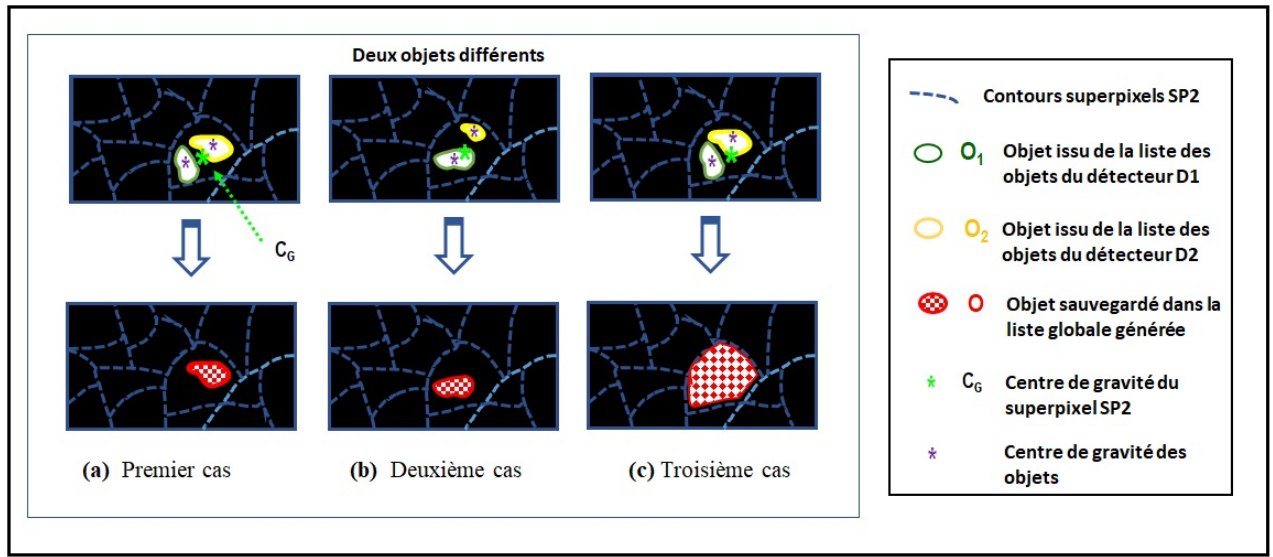


Figure 5.15 – Illustration d'un exemple de détermination de l'objet O à considérer dans la liste globale des objets à partir de deux objets différents

Avec cette configuration, le même objet peut être caractérisé différemment s'il n'est pas détecté uniformément par les différents détecteurs. Prenons l'exemple décrit dans la figure 5.14 précédente qui montre deux objets O_1 et O_2 ayant des pixels en communs mais qui ne sont pas identiques. Bien que l'objet considéré dans la liste globale porte la même étiquette de O_1 et O_2 , il présente des caractéristiques différentes à celles de O_1 et de O_2 , telles que, par exemple, les caractéristiques géométriques ou les caractéristiques issues des cartes de suspicion ou de l'image mammographique en elle même.

De même, un objet qui n'a pas été détecté par un détecteur donné héritera ses propriétés de celles associées dans la liste globale. En résumé, cette liste représente l'ensemble de tous les objets issus des différents détecteurs.

5.4.2.3 Étape 3 : Raffinement et classification initiale

Une fois la liste globale des objets candidats générée, un retour en arrière est effectué afin de :

- rajouter aux listes initiales de chaque détecteur les objets qui ne sont pas initialement identifiés. La caractérisation de ces derniers est celle donnée par la liste globale ;
- associer chaque objet dans les différentes listes à une classe bien définie.

Dans ce travail, nous définissons trois différentes classes (Figure 5.16) en fonction de l'occurrence des objets dans les résultats des détecteurs :

- L'étiquette de la classe *Certitude Totale* (C_{CT}) est attribuée aux objets candidats détectés par au moins $M - m + 1$ détecteurs. M représente le nombre de détecteurs utilisés et m (nombre entier avec $m < M$) le paramètre de qualité sur la base duquel la décision concernant un objet candidat peut être considérée comme certaine. Dans ce travail, M est égal à 4 et m est égal à 1 ;
- L'étiquette de la classe de *Certitude Partielle* (C_{CP}) qui est attribuée aux objets candidats détectés par au moins $m + 1$ et maximum $M - m$ détecteurs.
- L'étiquette de la classe d'*Incertitude* (C_I) qui est attribuée aux objets candidats détectés par au maximum m détecteurs.

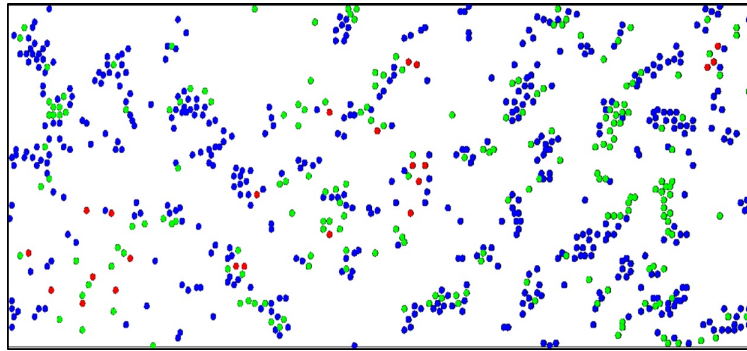


Figure 5.16 – Illustration d'un exemple de classification initiale des objets candidats représentés par leurs positions géométriques sur l'image mammographique (centre de gravité) : Les objets associés aux classes C_{CT} , C_{CP} and C_I sont respectivement en rouge, vert et bleu

Cette classification initiale suit le raisonnement ordinaire qui considère une décision comme *fiable* si elle apparait en accord avec la majorité des détecteurs ($m = 1$).

Le résultat finale de cette phase est, ainsi, un ensemble de M listes d'objets candidats présentant le même ensemble d'objets mais associées à des classes différentes relativement à chaque détecteur.

5.5 Processus de classification collaborative

5.5.1 Description globale

Le processus de classification collaborative que nous proposons est présenté dans la Figure 5.17. Il est décomposé en trois étapes principales qui seront détaillées dans les

sous-sections suivantes.

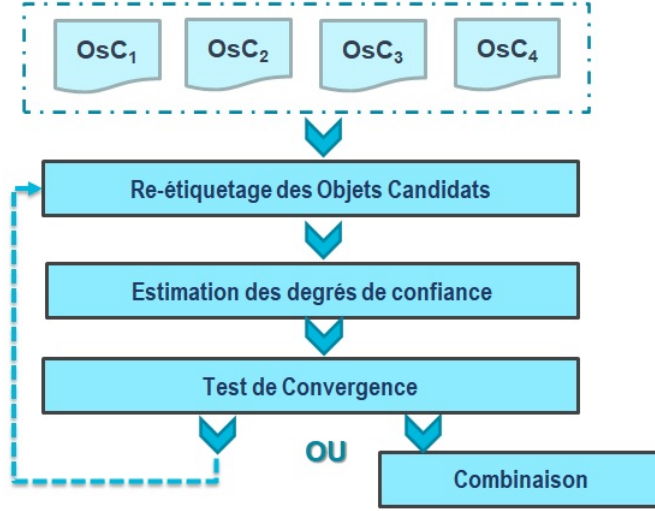


Figure 5.17 – Organigramme global du processus de classification collaborative proposé

Par ce processus collaboratif, nous cherchons à converger vers une décision collaborative unique en fonction des décisions initiales proposées par les différents détecteurs utilisés. Pour ce faire, les différentes listes d'objets candidats retenus à partir de la phase précédente sont comparées entre elles. Ces comparaisons permettent d'évaluer les similarités ainsi que les différences, par paire de détecteurs, sur chaque superpixel SP1. Ces études de similarité constituent la première étape pour analyser et ré-étiqueter les objets candidats. Ce ré-étiquetage consiste à mettre à jour les étiquettes des classes associées aux objets en désaccord par les détecteurs. L'idée étant de tirer partie des objets en accord par les détecteurs du fait qu'ils décrivent le premier noyau des MCs potentielles sur un superpixel donné. Cela implique que les objets candidats en désaccord doivent présenter une certaine similarité à ces derniers pour qu'ils changent de classes.

L'analyse des similarités, en termes de caractéristiques, des objets en accord nous donne une idée sur le noyau de départ censé caractériser l'ensemble des MCs potentielles présentes sur un superpixel SP1 donné et appelé *groupe de MCs*. Dans ce travail, nous proposons de représenter ce noyau sous la forme d'un graphe $G = (N, E)$ avec des nœuds N et des arêtes E (Figure 5.18). G est initialisé comme un graphe non pondéré. N (Équation (5.2)) est défini comme l'ensemble des nœuds qui correspondent aux objets candidats O_c de la liste des objets (L_{Obj}) appartenant aux classes CP ou CT . La position de chaque nœud correspond à celle du centre de gravité de l'objet qu'il désigne. E est défini comme l'ensemble des arêtes (E_{ij}) entre les paires de nœuds n_i et n_j dans N (Équation (5.3)). Les poids de ces arêtes seront évalués, pendant le ré-étiquetage, en tant que les similarités entre deux nœuds connectés dans le graphe.

$$N = \{O_c \in L_{Obj} \text{ et } O_c \in C_{CT} \text{ ou } O_c \in C_{CP}\} \quad (5.2)$$

$$E = \{E_{ij}, \text{ où } n_i \text{ et } n_j \in N\} \quad (5.3)$$

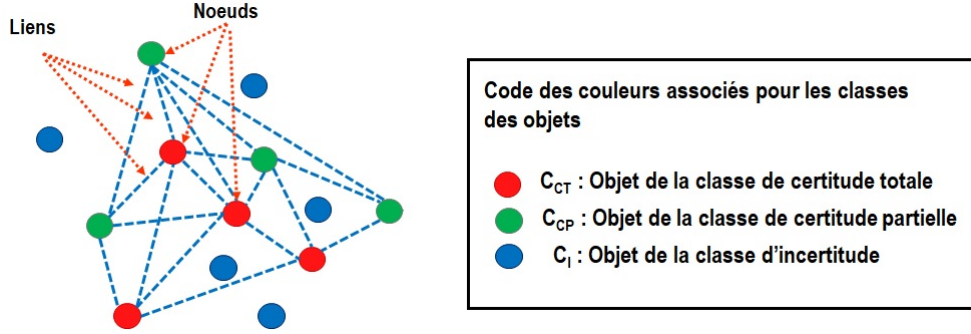


Figure 5.18 – Illustration d'un exemple de graphe généré à partir d'un ensemble d'objets candidats. A ce stade, les objets de la classe d'incertitude (cercles bleu) ne sont pas considérés comme étant des nœuds du graphe

Cette représentation, bien qu'elle puisse augmenter la complexité du traitement, est pertinente dans notre contexte. En effet, la représentation géométrique des nœuds dans le graphe permet de conserver la connaissance issue du niveau pixelique et, également, de tirer profit de cette connaissance dans le processus à suivre.

La mise à jour des étiquettes des classes associées aux différents objets (Figure. 5.19) est liée à la mise à jour du graphe initial. Pour mettre à jour le graphe initial, généré à partir de chaque détecteur, nous procédons à deux types de raffinements. Ils consistent à rajouter au graphe courant les objets candidats initialement associés à la classe C_I à condition qu'ils respectent certains critères de similarité avec les nœuds déjà connectés. Une fois connectés, leurs étiquettes précédemment associées sont mises à jours. Les critères de similarité que nous définissons sont inspirés des connaissances *a priori* sur les groupes de MCs. Les deux raffinements proposés dans ce travail sont :

- un raffinement local qui s'applique sur les objets candidats détectés dans un même superpixel SP1. Il décide en fonction des similarités géométriques calculées d'ajouter de nouveaux objets de la classe C_I au graphe courant. A ce niveau, notre raisonnement est fondé sur le fait que les MCs d'un groupe donné sont souvent réparties à des distances régulières les unes des autres [8]. Par ailleurs, il prend en considération la sensibilité des détecteurs liée aux petites variations de niveau de gris, qui peuvent affecter les similarités des MCs présentes, sur un même superpixel (comme le contraste) ;
- un raffinement contextuel qui s'applique sur les objets candidats dans le voisinage immédiat d'un superpixel SP1. Il opte vers un compromis de similarité géométrique

et numérique (de niveau de gris) pour décider l'ajout d'un objet au sein du graphe. Cette mesure de similarité définie permet d'étendre le raisonnement effectué au niveau local, à savoir la régularité de la distance géométrique entre les objets, au niveau contextuel des superpixels. En revanche, elle réduit l'importance de cette similarité géométrique si elle n'est pas accompagnée par une similarité des niveaux de gris entre les objets dans les superpixels voisins. Cette similarité, si elle existe, traduit une homogénéité entre les superpixels eux-mêmes.

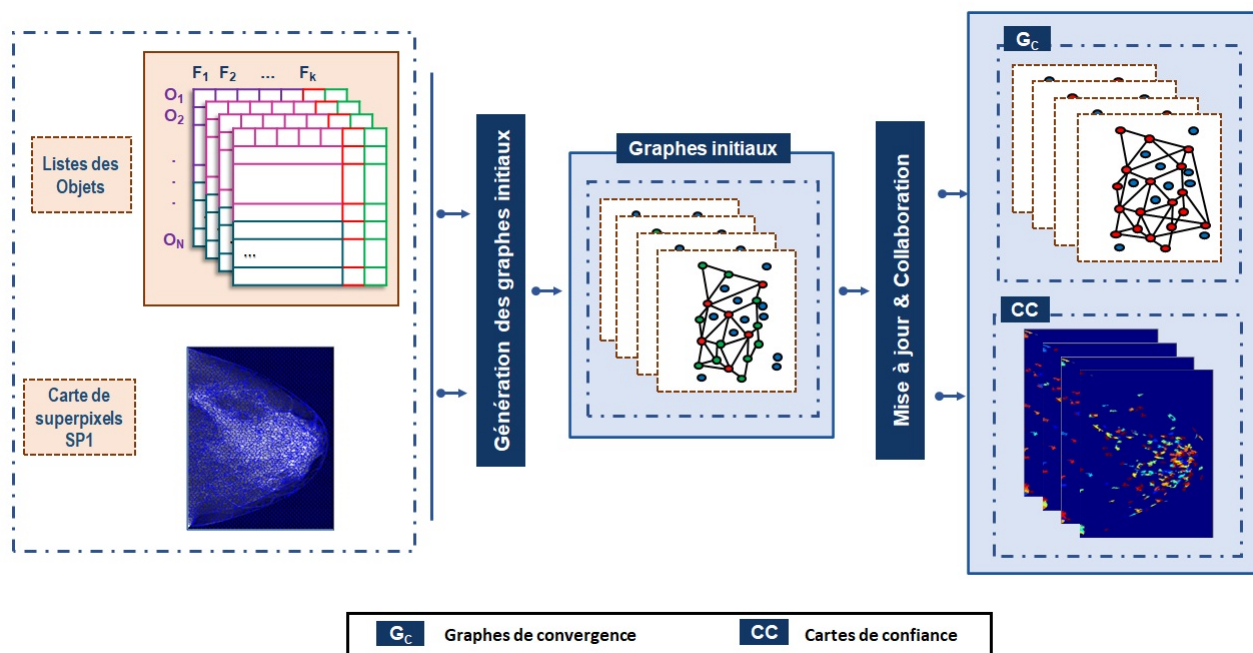


Figure 5.19 – Schéma de déroulement de la phase de classification collaborative

Contrairement au raffinement local, le raffinement contextuel prend en considération le fait qu'un groupe de MCs peut être dispersé sur une région qui dépasse les limites d'un seul superpixel $SP1$. Il caractérise, ainsi, un groupe de MCs sur la base des connaissances spatiales issues du voisinage des superpixels $SP1$. En effet, il impose que les MCs d'un même groupe aient des caractéristiques à la fois numériques et géométriques similaires et qu'elles apparaissent dans des superpixels homogènes. Une telle caractérisation offre l'opportunité de réduire la probabilité de sélection des objets équidistants indépendamment de leurs caractéristiques numériques et donc de réduire le nombre de détections de type FP.

Les raffinements proposés permettent une révision des décisions de chaque détecteur en se basant sur les informations partagées par les autres tout en changeant les étiquettes de leurs objets. En fonction de ce changement, des degrés de confiances ($W_i^k(t)$) sont associés à chaque superpixel $SP1$ (k) et détecteur (i). Ces raffinements sont appliqués d'une manière itérative jusqu'à une convergence. La convergence d'un graphe donné est obtenue si ce dernier ne subie aucun changement ou si la variation des degrés de confiances

$(W_i^k(t))$ (Équation (5.4)) associés aux superpixels SP1 entre deux itérations (t) et $(t-1)$ est inférieure à un seuil donné (ϵ) .

$$|W_i^k(t) - W_i^k(t-1)| = \epsilon \quad (5.4)$$

$$W_i^k(t) = T_i^k(t) * T_i^{V(k)}(t) * CC_i^k \quad (5.5)$$

$$T_i^k(t) = \frac{|E(G(t-1)^k)|}{|E(G(t)^k)|} \quad (5.6)$$

Dans les équations ci-dessus :

- L'indice k fait référence au $k^{\text{ème}}$ superpixel SP1 ;
- L'indice i fait référence au $i^{\text{ème}}$ détecteur ;
- $|E|$ désigne la cardinalité de l'ensemble des arêtes E d'un graphe donné
- $T_i^k(t)$ désigne le poids des nouvelles arêtes ajoutées au graphe précédent. Il correspond au rapport de la cardinalité du graphe à l'itération $(t-1)$ par rapport à celle de l'itération (t) ;
- CC_i^k désigne le degré de conflit maximal associé au superpixel SP_k et au détecteur D_i . Le calcul de ce coefficient sera détaillé dans la section suivante.

La description algorithmique de la démarche appliquée par le processus proposé est donnée par l'algorithme 4. Les termes utilisés dans cet algorithme seront définis au fur et à mesure dans le reste de ce chapitre.

5.5.2 Estimation des degrés de confiance

Cette étape consiste à observer les étiquettes des objets, dans chaque superpixel SP1, relativement à chaque détecteur et à comparer leurs caractéristiques associées. L'objectif est d'estimer un degré de confiance par détecteur et par superpixel SP1 afin de mettre à jour les étiquettes associées à ses objets (classes) et réduire le désaccord qu'ils présentent avec les autres détecteurs.

L'application de ce traitement pour chaque superpixel renforce l'intérêt des informations sémantiques produites pour une meilleure prise de décision. En effet, ces régions présentent une certaine homogénéité de niveaux de gris qui traduit une homogénéité sémantique sur l'image (les pixels du même type de tissu mammaire apparaissent avec des valeurs de niveau de gris similaires sur l'image mammographique). Et par ce fait, les MCs qui apparaissent dans un même superpixel présentent 'normalement' des caractéristiques similaires, du moins celles qui ont été détectées par le même détecteur (Figure 5.20).

Algorithme 4 : Algorithme du processus de classification proposé

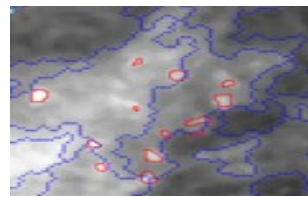
Entrées : I : Image mammographique
Sorties : G : Graphe de convergence

1 **Initialisation :**
 $t \leftarrow 0$
 $converg \leftarrow 'faux'$
début

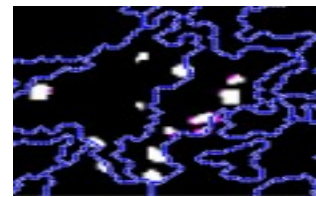
2 Génération des cartes de superpixels (cartes SP1 et SP2)
3 Application des détecteurs D_i sur les superpixels SP1
pour $i = 1$ **à** 4 **faire**
4 Génération de la liste des objets L_{obj}^i
5 Initialisation des classes des objets
6 Identification de la liste des superpixels à traiter (SP_{Cand}^i)
pour chaque $SP_k \in SP_{Cand}^i$ **faire**
7 Identification des superpixels voisins $V(SP_k)$
8 Initialisation de degré de confiance (W_i^k) et du graphe initial $G_{Init}^{i,k}$
9 Initialisation du graphe : $G_i \leftarrow G_{Init}^{i,k}$
10 **tant que** $converg \leftarrow 'faux'$ **faire**
11 $t \leftarrow t + 1$
12 $G_{LU} \leftarrow$ raffinement local de G^i
13 $G_{CU} \leftarrow$ raffinement contextuel de G_{LU}
14 Mise à jour des classes des objets
15 Mise à jour des degrés : W_i^k, T_i^k, CC_i^k
16 $G^i \leftarrow G_C$
17 **fin**
fin
fin
return Graphe final
fin



(a) ROI avec des MCs



(b) ROI avec superposition des contours des MCs (GT) et superpixels SP1



(c) MCs détectées par un détecteur (pixels blancs)

Figure 5.20 – Exemple d'un groupe de MCs qui apparait sur le même superpixel SP1

Le point de départ des estimations du degré de confiance consiste à comparer les couples de listes d'objets candidats pour chaque superpixel SP1 et chaque couple de détecteurs (D_i et D_j).

5.5.2.1 Définition de la similarité

La matrice de similarité (S_k) relativement au $k^{\text{ème}}$ superpixel (Équation (5.7)) que nous définissons, permet d'évaluer la concordance entre chaque couple de détecteurs (D_i et D_j). Elle est exprimée comme la proportion de leurs objets candidats (Os) détectés en commun. Il est important de noter que l'ordre de comparaison est différent et que la similarité $Sim(D_i, D_j)$ est différente de $Sim(D_j, D_i)$.

$$S^k = \begin{pmatrix} S_{1,1}^k & \cdot & \cdot & S_{1,M}^k \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ S_{M,1}^k & \cdot & \cdot & S_{M,M}^k \end{pmatrix} \quad (5.7)$$

avec

$$S_{ij}^k = Sim(D_i^k, D_j^k) = \frac{|Os_i^k \cap Os_j^k|}{|Os_i^k|} \quad (5.8)$$

5.5.2.2 Définition du degré de conflit

Une fois la similarité calculée, il est possible de définir un degré de conflit (C_{ij}^k), qui évalue le désaccord entre les détecteurs D_i et D_j (Équation (5.9)), ainsi que le conflit maximal CC_i^k (Équation (5.10)) relativement au $k^{\text{ème}}$ superpixel SP1 et donc d'initialiser son degré de confiance W_i^k (Équation (5.11)) relativement à chaque détecteur D_i donné. Ce dernier est exprimé en fonction du degré de confiance moyen entre le détecteur D_i et les autres détecteurs D_j avec ($j \neq i, i=1..M$).

$$C^k = \begin{pmatrix} C_{1,1}^k & \cdot & \cdot & C_{1,4}^k \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ C_{4,l}^k & \cdot & \cdot & C_{4,4}^k \end{pmatrix} \quad \text{où } C_{ij}^k = 1 - S_{ij}^k \quad (5.9)$$

5.5.2.3 Définition du degré de conflit maximal

Le degré de conflit maximal CC_i^k , pour un superpixel k donné, met en correspondance un détecteur D_i et son correspondant le moins similaire D_j en termes d'objets communs détectés et sera utilisé dans le processus de ré-étiquetage.

$$CC_i^k(D_i^k, D_j^k) = \arg \max_{l \neq i} (C_{il}^k) \quad (5.10)$$

$$W_i^k(t=0) = 1 - \text{moy}_j(C_{ij}^k) \quad (5.11)$$

Le degré de confiance estimé joue un rôle important dans la sélection de l'ordre de traitement des superpixels, et plus encore, dans la combinaison finale des résultats.

5.5.3 Ré-étiquetage des objets candidats

L'étape de ré-étiquetage itératif s'applique pour chaque superpixel SP1 et pour chaque détecteur D_i . Elle a pour objectif d'améliorer la qualité des résultats de chaque détecteur afin de réduire le conflit qu'il présente avec les autres et, ainsi, d'avoir des objets avec des étiquettes identiques. Pour ce faire, elle commence par sélectionner les superpixels SP1 (SP_{Cand}^i) contenant des objets appartenant à la classe C_{CT} ou C_{CP} (Equation 5.12).

$$SP_{Cand}^i = \{SP_k, k \leq Nk / \exists ! O \in Os_c^k \text{ qui appartient } C_{CP} \text{ et/ou } C_{CT}\} \quad (5.12)$$

Elle est basée sur le fait que les superpixels avec des objets présentant un accord total ou partiel par les détecteurs sont capables de fournir les informations pertinentes pour une meilleure prise de décision. Elle consiste à mettre à jour les graphes initiaux obtenus $G_{Init}^{i,k}$ générés sur chaque superpixel SP_k de SP_{Cand}^i en appliquant des analyses intra et inter-superpixels. Ces analyses quantifient à chaque fois les similarités entre les objets candidats (Os_c), à relier, et les nœuds du graphe à raffiner pour étudier la possibilité de leur ajout à ce dernier. Ces similarités constituent les poids des nouveaux liens ajoutés entre les candidats sélectionnés et les anciens nœuds du graphe.

5.5.3.1 Analyse intra-superpixel

La figure 5.21 illustre un exemple de raffinement intra-superpixel.

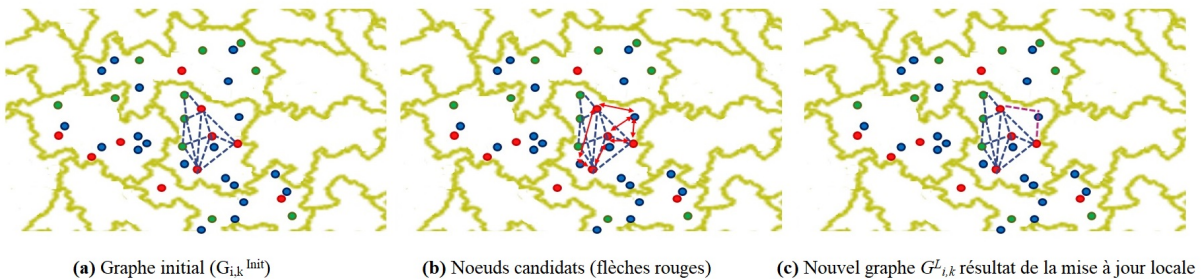


Figure 5.21 – Illustration d'un exemple de raffinement intra-superpixel

Un objet candidat O_c dans un superpixel SP_k , associé à la classe C_I est ajouté au graphe initial ($G_{Init}^{i,k}$) si et seulement si la condition suivante est satisfaite :

$$SimL^{i,k}(O_n, O_c) \leq \underset{\substack{O_c \in C_I \\ O_n \in G_{Init}^{i,k}}}{moy} (SimG^k(O_{n1}, O_{n2})) \quad (5.13)$$

$$SimG^k(O_{n1}, O_{n2}) = d(O_{n1}, O_{n2}) \quad (5.14)$$

où :

- $SimL^{i,k}$ désigne la similarité locale entre les objets O_n et O_c ;
- $SimG^k$ désigne la distance euclidienne géométrique normalisée entre les objets O_{n1} et O_{n2} de la liste des nœuds du graphe $G_{i,k}^{Init}$ (Équation. (5.14)).

L'algorithme 5 présente les étapes du raffinement local appliqué pour une itération t donnée. Pour alléger le texte, des notations simples sont utilisées. Les indices faisant référence au numéro du détecteur (i) et du superpixel (k) n'apparaissent pas dans les notations utilisées. A titre exemple, la notation $G(t)$ fait référence au graphe à utiliser à l'itération t au lieu de $G_{(N,E)}^{i,k}(t)$. Les notations N et E désignent respectivement les nœuds et les arêtes d'un même graphe.

Algorithme 5 : Algorithme de raffinement local pour une itération t , un détecteur i et un superpixel k

Entrées : $G(t-1)$

Sorties : $G_{LU}(t)$

1 Initialisation :

$G \leftarrow G(t-1)$ (si $t==1$ alors $G \leftarrow G_{Init}$)

$N_{Cand} \leftarrow \{O_c \in SP_k \text{ où } O_c \in C_I \text{ et } O_c \notin G\}$

début

- 2** Calcul, pour chaque paire d'objets dans N , les valeurs de similarités locales $SimL^k$
- 3** Calcul pour chaque objet O_c dans N_{Cand} les valeurs de similarité locales avec les nœuds O_n dans N
- 4** // Raffinement
Sélection des objets O_c , qui respectent la condition de similarité locale, à rajouter à G :
 - $N_{Add} \leftarrow \{O_c \in N_{Cand} \text{ et } O_c \text{ vérifie la condition dans (5.13)}\}$
 - $E_{Add} \leftarrow \{E_{cn}, O_c \in N_{Add}, O_n \in N\}$
- 5** Mise à jour du graphe :
 - $N_{LU} \leftarrow N \cup N_{Add}$
 - $E_{LU} \leftarrow E \cup E_{Add}$
 - $G_{LU}(t) \leftarrow G(N_{LU}, E_{LU})$

return G_{LU}

fin

5.5.3.2 Analyse inter-superpixel

Le raffinement inter-superpixel (Figure 5.22) s'applique pour le graphe $G_L^{i,k}$ résultat du raffinement local précédent.

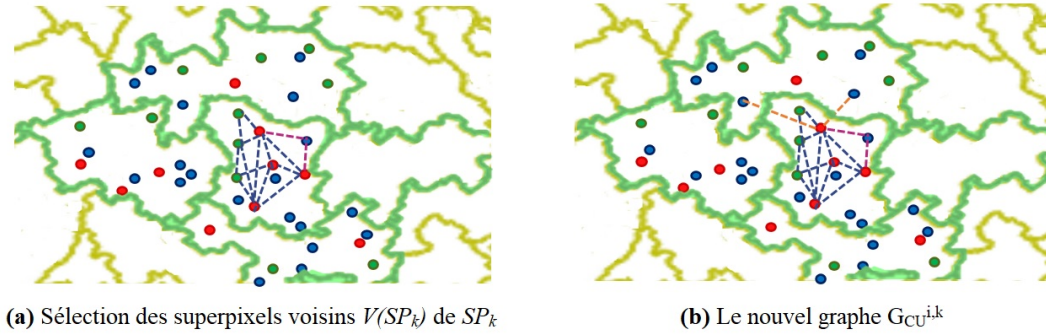


Figure 5.22 – Exemple d'un raffinement inter-superpixel

Un nouveau nœud (objet candidat) qui appartient aux voisins immédiats ($V(SP_k)$) du superpixel SP_k et associé à la classe C_I est accepté pour être ajouté au graphe $G_{LU}^{i,k}$, précédemment raffiné, si et seulement s'il satisfait le critère de similarité. Ce critère est défini comme la moyenne des similarités géométriques normalisées ($SimG^k$) et des similarités de caractéristiques ($SimF^k$) (Équation (5.15)) :

$$SimC^{i,k}(O_n, O_c) \leq \text{moy}_{\substack{O_c \in V(SP_k) \\ O_c \in C_I}} (SimG^k(O_{n1}, O_{n2}) + SimF^k(O_{n1}, O_{n2})) \quad (5.15)$$

$O_{n1}, O_{n2} \in G_{LU}^{i,k}$
 $n1 \neq n2$

$$SimF^k(O_{n1}, O_{n2}) = |Feat(O_{n1}) - Feat(O_{n2})| \quad (5.16)$$

où $Feat(O_c)$ correspond à la mesure de caractéristique de l'objet candidat O_c donné. A ce niveau, nous avons adopté l'écart-type du niveau de gris des pixels constituant un objet comme caractéristique pour cette analyse inter-superpixel.

L'algorithme 6 décrit les étapes de ce raffinement contextuel que nous proposons sur la base de l'analyse inter-superpixels.

Algorithme 6 : Algorithme de raffinement contextuel pour une itération t , un détecteur i et un superpixel k

Entrées : $G_{LU}(t)$
Sorties : $G_{CU}(t)$

1 Initialisation :
 $G \leftarrow G_L(t)(N, E)$
 $N_{Cand} \leftarrow \{O_c \in V(SP_k) \text{ où } O_c \in C_I \text{ et } O_c \notin G_L\}$
début

2 Calcul, pour chaque pair d'objets dans N , les valeurs de similarités contextuelles $SimC^k$

3 Calcul pour chaque objet O_c dans N_{Cand} les valeurs de similarités contextuelles aux nœuds O_n dans N
 // Raffinement contextuel

4 Sélection des objets O_c , qui vérifient la condition de similarité contextuelle, à rajouter à G :
 — $N_{Add} \leftarrow \{O_c \in N_{Cand} \text{ et } O_c \text{ vérifie la condition dans (5.15)}\}$
 — $E_{Add} \leftarrow \{E_{cn}, O_c \in N_{Add}, O_n \in N\}$

5 Mise à jour du graphe :
 — $N_{CU} \leftarrow N \cup N_{Add}$
 — $E_{CU} \leftarrow E \cup E_{Add}$
 — $G_{CU}(t) \leftarrow G(N_{CU}, E_{CU})$

return G_{CU}
fin

5.5.4 Convergence et représentation des résultats

Avec la convergence, l'ensemble des nœuds dans chaque graphe présente la nouvelle liste des objets candidats retenus comme des *objets d'intérêts* par le détecteur correspondant. Un objet d'intérêt est un objet candidat associé à la classe C_{CT} par un détecteur donné. Théoriquement, ces ensembles sont plus similaires que les ensembles initiaux. Ainsi, la combinaison finale de ces nouveaux résultats pourrait être simplifiée à l'aide d'un opérateur conjonctif alors que la concordance entre eux augmente. Les autres objets appartenant à la classe C_{CP} ou C_I sont ignorés et considérés comme des objets de non-intérêts. Cette liste finale d'objets est projetée sur un espace pixelique afin de construire la carte thématique finale à fournir pour l'expert.

En plus de la liste des objets d'intérêts finale, le processus collaboratif que nous avons proposé attribue des degrés de confiances (W_i^k) relativement à chaque superpixel SP1 (SP_k) et détecteur D_i . De même, la projection de ces degrés de confiance sur la carte des superpixels SP1 permet de générer une carte de confiance pixelique (CC_i) par détecteur D_i . Elle présente des valeurs continues appartenant à l'intervalle $[0..1]$. Les pixels d'un superpixel SP_k donné présentent la même valeur qui correspond à son degré de confiance

associé par un détecteur donné.

En résumé, avec la convergence, nous obtenons avec trois différentes cartes pixeliques :

- les cartes de suspicion initiales ;
- les cartes de degrés de confiance ;
- la carte thématique finale.

En partant de toutes les cartes pixeliques, que nous disposons, on génère une carte de suspicion finale à fournir pour l'expert. Nous distinguons trois types de pixels sur la carte générée :

- *Pixel Objet d'Intérêt* (noté P_{OI}) : C'est un pixel appartenant à un objet d'intérêt retenu par le processus collaboratif appliqué. Il présente également une valeur 1 dans la carte thématique finale ;
- *Pixel Doute* (noté P_D) : C'est un pixel appartenant à un objet associé à la classe C_{CP} par le processus appliqué ;
- *Pixel Fond* (noté P_F) : C'est un pixel du fond de l'image mammographique ou appartenant à un objet associé à la classe C_I par le processus appliqué.

Nous proposons l'estimation des degrés de suspicion $CS(P)$ pour chaque pixel de l'image sur la base des expressions décrites ci-dessous de la manière suivante :

$$CS(P) = \begin{cases} \arg \max_i (CS_i(P)) & \text{Si } P \text{ est } P_{OI} \\ CS_j(P) \quad \text{avec } j = \arg \max_i (CC_i(P)) & \text{Si } P \text{ est } P_D \\ \arg \min_i (CS_i(P)) & \text{Si } P \text{ est } P_F \end{cases} \quad (5.17)$$

où $\arg \min$, $\arg \max$ et moy correspondent respectivement aux valeurs minimale, maximale et moyenne des degrés de suspicion sur les cartes CS_i initialement générées par les différents détecteurs.

5.6 Résultats expérimentaux et discussion

Dans cette section, nous proposons d'évaluer les performances de l'approche décrite dans ce chapitre. De manière similaire aux chapitres précédents, la démarche d'évaluation a été également menée sur l'ensemble des 50 images sélectionnées à partir de la base IN-Breat [38].

Elle consiste à étudier la pertinence des résultats obtenus et elle se divise en deux parties principales. La première concerne une évaluation globale de la pertinence des résultats

obtenus par rapport à chaque détecteur. La seconde concerne la comparaison par rapport aux résultats des opérateurs de combinaison standard tels que l'union et l'intersection.

Avant de procéder à l'évaluation, nous présentons dans la section suivante les détecteurs que nous avons utilisés dans cette étude.

5.6.1 Détecteurs utilisés

Pour la mise en œuvre de l'approche proposée, nous avons utilisé quatre détecteurs différents. Chaque détecteur suit un raisonnement spécifique pour modéliser et chercher des pixels de MCs potentielles à partir des superpixels SP1 :

- un détecteur basé sur la morphologie qui utilise des éléments structurants et des opérations morphologiques pour la recherche des pixels connectés pouvant correspondre à des MCs ;
- un détecteur de Croissance de Région Conditionnelle (CRC) qui suit un processus itératif afin de délimiter les régions, qui peuvent correspondre à des MCs, en partant des germes de départ sélectionnés. Il intègre des critères basés sur des connaissances *a priori* au processus de croissance à la place d'un simple critère d'homogénéité ;
- un détecteur basé sur l'algorithme PFCM [107] qui modélise les MCs en tant que des pixels adjacents 'aberrants' par rapport aux pixels du tissu mammaire de superposition ;
- un détecteur basé sur un filtre passe-bande appliqué dans le domaine de Fourier afin d'améliorer les contours des MCs, vers une meilleure segmentation.

Les deux premiers détecteurs sont les deux contributions déjà présentées par les deux derniers chapitres 3 et 4 (Elles étaient le sujet de deux publications scientifiques respectivement dans [124] et [40]). Le troisième et quatrième détecteur ([109] et [112]) sont sélectionnés depuis la littérature et adaptés pour être appliqués à l'image mammographique entière en respectant les superpixels SP1.

A titre d'exemple, la méthode de référence pour le troisième détecteur a été proposée par Quintanilla *et al.* [109] s'applique à une ROI sélectionnée par les radiologues. De même, elle procède par un seuillage fixe sur la carte des degrés de typicalité de la classe objet obtenue à partir de la ROI. En revanche, l'utilisation d'un seuil fixe peut engendrer la génération des fausses détections dues à la diversité des caractéristiques des superpixels où apparaissent les MCs, ce qui se traduit par une diversité de caractéristiques d'apparition. Dans l'idée d'éviter cette situation, nous avons proposé une démarche pour l'estimation automatique d'un seuil par superpixel SP1. L'utilisation du seuillage a pour objectif de trouver les pixels MCs à partir de la classe *objet* segmentée sur un superpixel donné. Toutefois, les niveaux de gris de ces pixels étant similaires, ils ont une faible oc-

currence dans la classe objet segmentée.

Le tableau suivant montre une synthèse des avantages et des inconvénients des détecteurs que nous avons utilisé dans ce travail de recherche.

Tableau 5.1 – Comparaison des détecteurs utilisés

Détecteur	Principe, Avantages et Inconvénients
D_1 [124]	Détection basée sur une analyse spatiale et des connaissances sur la forme <ul style="list-style-type: none"> • Réduire la sensibilité à la variation de contraste ; • Fiabilité relative au choix des éléments structurants.
D_2 [40]	Segmentation basée sur un processus itératif et une analyse locale <ul style="list-style-type: none"> • Analyse très fine et contrôlée par des conditions basées à base de connaissances <i>a priori</i> ; • Fiabilité liée au choix des germes de départ et aux conditions d'arrêt.
D_3 [109]	Modélisation des MCs comme des pixels 'outliers' qur la classe 'objet' segmentée <ul style="list-style-type: none"> • Modélisation mathématique des pixels 'outliers' ; • Sensibilité au faible contraste.
D_4 [112]	Détection basée sur l'amélioration des contours <ul style="list-style-type: none"> • Réduction de la sensibilité au faible contraste ; • Sensibilité à la diversité de taille et à la présence du bruit.

La figure 5.23 montre des exemples de cartes thématiques et de suspicion obtenues par les quatre détecteurs à partir d'une image mammographique. L'analyse des différentes cartes présentées montre la complémentarité en termes de résultats des détecteurs utilisés. De même, elle justifie l'intérêt de l'utilisation conjointe de plusieurs détecteurs.

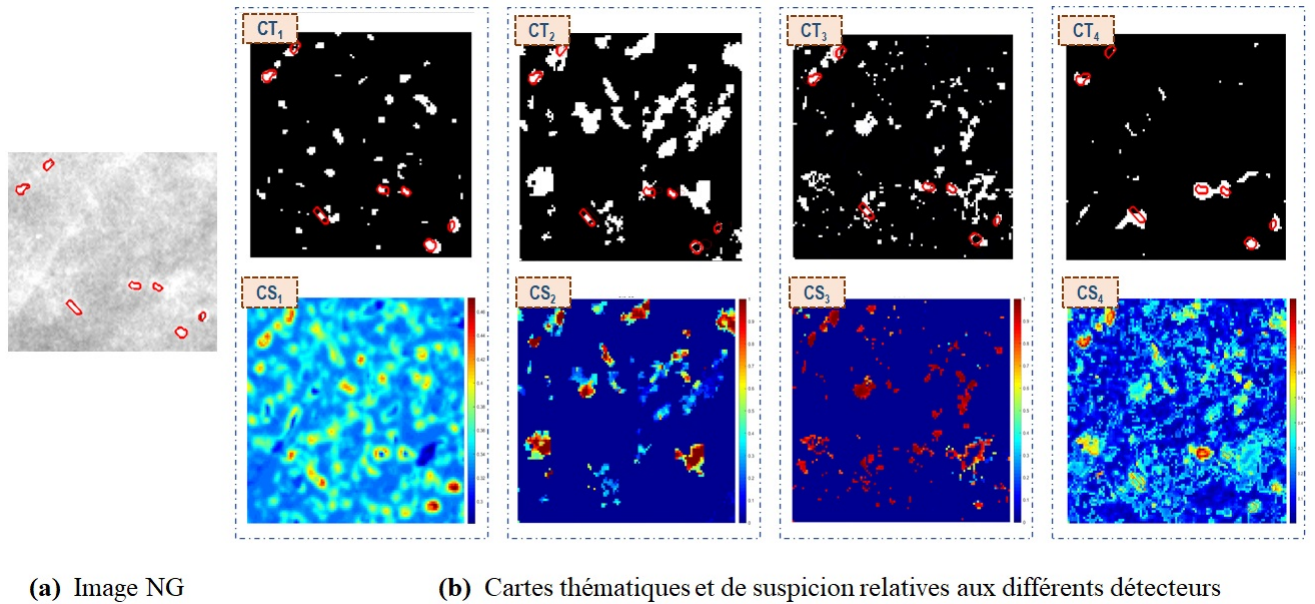


Figure 5.23 – Illustration des cartes pixeliques (b) générées par les différents détecteurs à partir une ROI d’une image mammographique en NG (a)

5.6.2 Critères d’évaluation

L’évaluation se fait sur la base de trois différentes critères illustrées par la figure 5.24 et qui portent les mêmes noms que ceux déjà utilisés dans les chapitres précédents : Vrai Positif (TP), Faux Positif (FP) et Vrai Négatif (TN) :

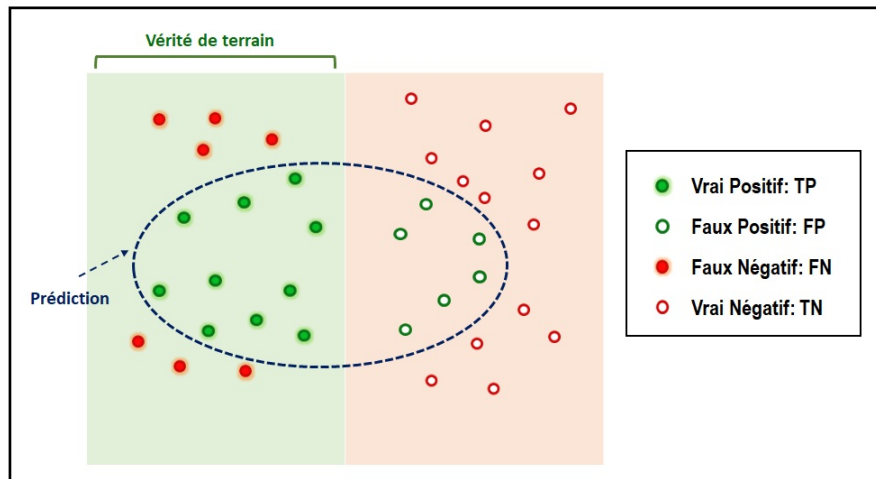


Figure 5.24 – Illustration des différents types possibles des objets identifiés

- Un vrai positif décrit une MC correctement identifiée comme objet candidat par l’approche proposée ;
- Un faux positif décrit une région connexe caractérisant un tissu mammaire normal mais considérée, à tort, comme un objet candidat par l’approche proposée ;

- Un faux négatif décrit une vraie MC considérée, à tort, comme un tissu mammaire normal par l'approche proposée.

En revanche, il est à noter à ce niveau, que nous avons migré, avec cette approche, de l'espace pixelique vers l'espace objet. Dans cet espace, un objet fait référence à une MC potentielle qui n'est plus caractérisée par un ensemble de pixels adjacents mais par sa position géométrique et son vecteur caractéristique. Cette transition justifie la différence en termes de définition entre les termes TP, FP et FN utilisés dans les chapitres précédents pour désigner les types de pixels distingués suite à la superposition des cartes de pixeliques de segmentation GT et candidate.

5.6.3 Évaluation des performances

5.6.3.1 Évaluation globale

L'objectif principal de l'approche proposée, dans ce chapitre, est de diminuer le nombre de faux négatifs tout en augmentant le nombre de vrais positifs. Dans la figure 5.25, les taux de vrais positifs de l'application du processus collaboratif proposé sont données. Les mesures obtenues prouvent la fiabilité de cette contribution pour l'identification des MCs. En effet, elle est capable d'identifier plus de 70% (respectivement 90%) des MCs pour 58% (respectivement 80%) des images de test utilisées. Il est important de rappeler ici que le nombre moyen de MCs présent par image est d'environ 125 MCs. Par ailleurs, le taux de détection des MCs est inférieur à 50% pour moins de 4% des images.

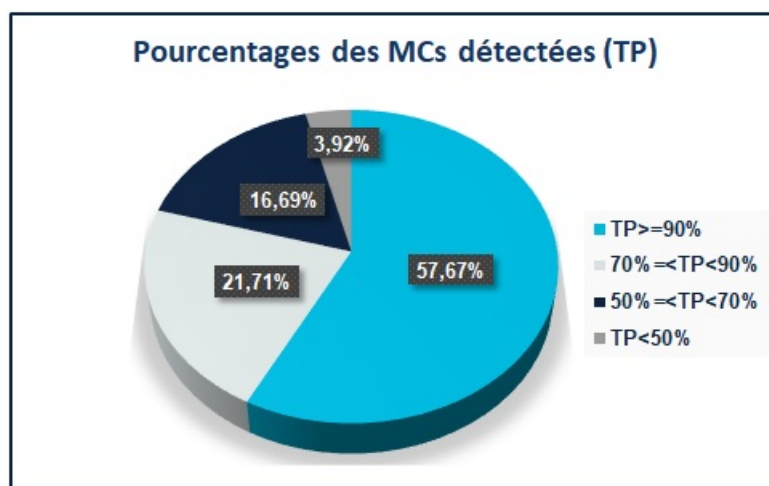


Figure 5.25 – Pourcentages des taux de TP obtenus suite à l'application de l'approche proposée

5.6.3.2 Évaluation des performances par rapport aux détecteurs

Afin d'évaluer la robustesse de l'approche proposée, il est important de prouver sa fiabilité par rapport aux détecteurs utilisés. Le tableau 5.2 montre les taux de détection

moyens atteint par chaque détecteur (D_i , $i=1..4$). Il montre que la valeur la plus élevée (85%) est atteinte par l'approche proposée. Ceci justifie l'amélioration apportée par le processus de classification collaboratif appliqué sur le résultat de chaque détecteur.

Tableau 5.2 – Comparaison des taux moyens de TPs obtenus par l'approche proposée par rapport à chaque détecteur

	D_1 [124]	D_2 [40]	D_3 [109]	D_4 [112]	Notre contribution
Taux moyens de TPs	75.29%	68.42%	83%	79.22%	85%

Dans l'idée de quantifier cette amélioration, le tableau 5.3 montre les pourcentages des taux de TPs améliorés par l'approche proposée par rapport à chacun des détecteurs utilisés. Les résultats dans ce tableau montrent que le processus collaboratif proposé a réussi d'améliorer le taux d'identification des MCs, dans plus de 50% des images de test, par rapport aux résultats des différents détecteurs utilisés. Il montre également que les taux de TPs obtenus sont améliorés dans 44% des cas par rapport au meilleur détecteur qui a fourni les résultats les plus pertinents pour une image de test donnée. Par ailleurs, ces résultats justifient l'intérêt d'adapter ce schéma de fusion collaborative, pour la sélection finale des objets considérés comme MCs potentielles, au lieu d'un simple détecteur. En effet, même si le troisième détecteur (D_3) présente une moyenne de taux de TP comparable à celle de l'approche proposée (Tableau 5.2), nous avons réussi à améliorer ses taux de TPs pour 52% des images mammographiques.

Tableau 5.3 – Pourcentages des taux de TPs améliorés en comparaison avec chaque détecteur utilisé

	D_1 [124]	D_2 [40]	D_3 [109]	D_4 [112]	Meilleur détecteur
Pourcentages d'amélioration des TPs	68%	72%	52%	56%	44%

5.6.3.3 Évaluation des performances par rapport aux opérateurs de fusion standard

Dans ce travail, les opérateurs de fusion standards disjonctif et conjonctif sont utilisés pour les comparaisons. Ce choix se justifie notamment par les caractéristiques des détecteurs ainsi que par la diversité des images mammographiques. En effet, aucune information préalable n'est présente sur la précision ou la fiabilité globale des décisions réalisées par les détecteurs. Elle dépend de nombreux facteurs tels que la densité du sein

et la diversité des caractéristiques des MCs. Par ailleurs, la fiabilité des détecteurs utilisés dépend de leur raisonnement appliqué par rapport aux types et caractéristiques des MCs. A titre d'exemple, le troisième détecteur D_3 décrit une MC comme un ensemble de pixels adjacents *outliers* dans un superpixel SP1. Bien qu'il présente un taux moyen de TP concurrent à celui de l'approche proposée, il montre néanmoins quelques faiblesses en ce qui concerne les MCs avec des contrastes variés [112].

La figure 5.26 compare les boîtes à moustaches des taux de TP obtenus par l'application de l'opérateur coonjonctif pour combiner les décisions initiales et raffinées (par le processus proposé) des détecteurs. Ces boîtes à moustaches montrent clairement la capacité de l'approche proposée à réduire les désaccords initiaux entre les détecteurs et à améliorer la fiabilité de l'opérateur d'intersection.

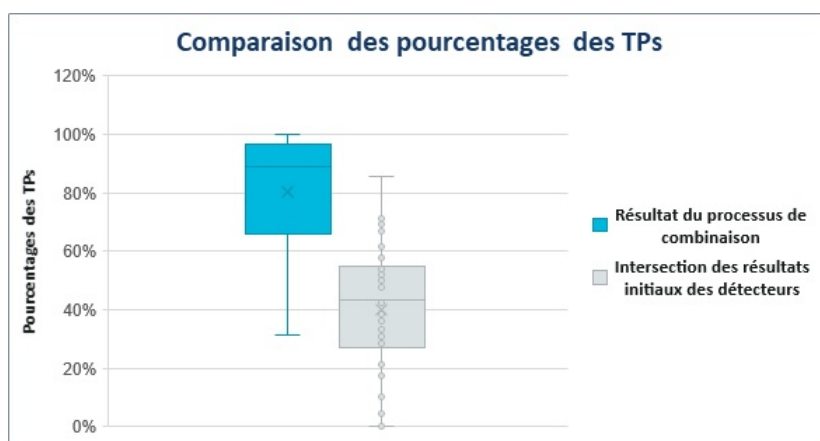


Figure 5.26 – Boîtes à moustaches des taux des TPs obtenus suite à l'application de l'opérateur de fusion conjonctif avant et après l'application du processus collaboratif proposé

La première impression, suite à l'analyse du tableau 5.4, suppose que l'approche proposée est moins fiable qu'une simple combinaison disjonctive des résultats initiaux des détecteurs. En effet, le taux moyen de TPs obtenus par l'opérateur de fusion disjonctif est de l'ordre de 92% par rapport à 85% par l'approche proposée.

Tableau 5.4 – Taux moyens de TP obtenus en comparaison avec les opérateurs d'union et d'intersection

	Union	Intersection	Approche proposée
Taux de TP moyen	92.05%	42.2%	85%

En revanche, l'analyse quantitative illustrée dans le tableau 5.5, montre que l'opérateur conjonctif que nous avons appliqué est un véritable concurrent à l'opérateur union. Il est même plus fiable que cet opérateur disjonctif pour 48% des cas.

Tableau 5.5 – Pourcentages d’amélioration des taux de TP en comparaison par rapport aux opérateurs union et intersection ainsi que le meilleur détecteur

	Union	Intersection	Meilleur détecteur
Pourcentages d’amélioration	48%	96%	44%

En plus d’être compétitive par rapport à l’opérateur d’union, l’approche proposée est capable de réduire d’en moyenne 32% le taux de faux positifs (FP) pour toutes les images de test. Cette réduction permet d’écarter les objets de type FP présentant des caractéristiques similaires aux MCs. Cela implique, ainsi, une amélioration de la fiabilité d’une éventuelle étape de classification dont l’objectif est de distinguer les objets correspondants à des vrais MCs des autres faux positifs. Ces pourcentages prouvent également que l’approche proposée a pu atteindre son objectif principal, à savoir la réduction des FNs avec un taux acceptable de TP, concurrent à celui d’un opérateur d’union.

5.7 Conclusion

Nous avons présenté dans ce chapitre une approche qui aborde le problème de la détection fiable des MCs, à partir d’images mammographiques, tout en minimisant les détections de type FN. Elle consiste en un processus collaboratif itératif qui fait collaborer différents détecteurs dans l’idée de mettre en cause leurs décisions initiales et converger vers une nouvelle décision commune qui minimise leurs désaccords initiaux et donc les détections de type FN. Il applique, à chaque itération, des raffinements, pour chaque superpixel SP1, basés sur une analyse locale et une autre contextuelle guidées par des connaissances *a priori* sur les liens entre les superpixels et les MCs. Ces analyses estiment également, de manière itérative, les termes de fiabilité locaux par superpixel. Le résultat final est simplifié à une combinaison conjonctive des nouvelles décisions du détecteur.

Les résultats obtenus, relatifs à la base de données INBreast, ont prouvé la capacité du processus collaboratif proposé à améliorer les taux de détection des MCs par rapport à chacun des détecteurs ainsi qu’à des opérateurs de fusion standard.

CONCLUSIONS ET PERSPECTIVES

Dans ce travail de recherche, nous avons abordé la problématique de l'interprétation des images mammographiques en vue de proposer un système capable de détecter les microcalcifications. Le choix de ce type de lésions cancéreuses se justifie par deux principaux facteurs. Le premier étant que ces lésions présentent le premier signe de production du cancer du sein et leur détection précoce contribue à améliorer l'espérance de vie pour les femmes affectées. Le deuxième concerne leur diversité anatomique qui, en s'ajoutant au critère multi-tissulaire du sein et la faible qualité des images mammographiques, rend difficile la détection des microcalcifications surtout si elles sont superposées à des tissus denses.

Contributions et bilan

Pour correctement concevoir un tel système, il est crucial d'étudier les connaissances relatives au cancer du sein et l'ensemble des lésions cancéreuses qu'il peut générer. Cette étude était l'objectif du premier chapitre dans ce manuscrit. Dans la première partie de ce chapitre, nous avons présenté une description anatomique du sein suivie par une autre du lexique médical utilisé pour décrire ce problème de santé souvent diagnostiqué à l'aide d'une mammographie. Ensuite, une étude détaillée des caractéristiques des microcalcifications est aussi abordée et a permis de mettre en lumière les difficultés ainsi que les intérêts de proposer un système de détection fiable. Dans sa deuxième partie, ce chapitre a étudié les caractéristiques des différentes bases d'images qui existent dans la littérature. Il y est également décrit la démarche que nous avons suivie pour concevoir la vérité-terrain utilisée pour la validation de notre système sur la base INBreast.

Le deuxième chapitre a présenté un état de l'art sur les méthodes mises en œuvre pour traiter la problématique de détection des microcalcifications à partir des images mammographiques. Il nous a permis d'évoquer les avantages et les inconvénients de chacune de ces méthodes et de mettre l'accent sur l'importance d'une étape de prétraitement pour améliorer la qualité des images mammographiques. Sur la base de ces études, nous en

avons conclu que les difficultés rencontrées, par ces méthodes, sont liées à la diversité des caractéristiques des microcalcifications ainsi qu'à l'absence des connaissances sémantiques dans le processus de détection appliqué. Ceci nous conduit à présenter l'idée générale de notre contribution principale dans ce travail de recherche. Elle consiste à proposer une combinaison collaborative des résultats de plusieurs méthodes de détection afin de converger vers une décision collaborative optimale.

Le troisième chapitre s'est intéressé à présenter la première contribution de ce travail de recherche qui consiste à proposer une nouvelle méthode de détection basée sur la morphologie mathématique et les indices de similarité structurelle SSIM. Composée de trois étapes, elle vise à traiter le problème de faible contraste des microcalcifications par rapport aux tissus de superposition. En plus de la combinaison des opérations morphologiques avec les indices de similarité, cette méthode génère une carte de suspicion comme nouvel espace d'interprétation en se basant sur l'indice SSIM. Des analyses locales sont effectuées sur cette dernière afin d'identifier les zones suspectes faisant référence à des microcalcifications potentielles. Les résultats évalués ont prouvé sa capacité à traiter la diversité des microcalcifications en termes de contrastes et de formes. En revanche, sa fiabilité a montré une sensibilité à la forte similarité des microcalcifications à leurs tissus de superposition.

La deuxième contribution dans ce travail de recherche, méthode appelée CRC, a été introduite dans le quatrième chapitre. Il s'agit d'une méthode qui intègre des critères issus des connaissances *a priori* afin de faire croître et converger un germe de départ vers les meilleurs contours. L'évaluation de cette contribution a été menée avec une méthodologie similaire à celle de la première contribution. La pertinence en termes de segmentation offerte se trouve accompagnée par des fausses détections liée à la diversité des types de tissus de superposition des microcalcifications.

Les discussions des résultats obtenus par les deux premières contributions nous ont amené à conclure que quelque soit le raisonnement appliqué, l'application d'une seule méthode de détection est incapable de gérer la diversité des microcalcifications. La contribution mise en œuvre dans le cinquième chapitre part de ce constat. Elle consiste en une nouvelle méthode qui opère par un processus collaboratif qui permet aux différentes méthodes utilisées de partager leurs connaissances et de mettre en cause leurs décisions initiales en fonction de celles des autres. Le partage s'effectue à deux niveaux : le niveau local et le niveau contextuel des superpixels afin de raffiner les décisions initiales. Il se base sur les connaissances *a priori* sur les liens entre les microcalcifications et les superpixels. L'évaluation et la comparaison de cette approche ont prouvé sa pertinence en termes de la qualité des résultats et des améliorations apportées par rapport à certaines méthodes

de détection dont les deux premières contributions font partie.

Perspectives

Le système décrit par la contribution principale de ce travail de recherche semble avoir atteint une maturité nécessaire pour pouvoir apporter de l'aide aux médecins radiologues. En revanche, de nombreuses perspectives d'amélioration et d'extension peuvent lui être apportées.

Ainsi, dans l'idée de prouver la pertinence de la démarche proposée, une première perspective consiste en l'augmentation de la taille de la base d'images utilisées pour l'évaluation. Ceci aura pour objectif de non seulement se comparer avec des approches supervisées de détection des microcalcifications, qui sont de plus en plus utilisées ces dernières années, mais également d'exploiter ces techniques comme de nouveaux détecteurs dans le processus collaboratif proposé.

Une première piste d'extension peut concerner l'intégration de connaissances *a priori* de plus haut niveau d'interprétation sémantique (l'image mammographique entière) afin d'affiner le résultat obtenu. Rappelons que la démarche proposée s'est uniquement concentrée sur l'exploration des connaissances pixéliques, locales et contextuelles relatifs aux régions homogènes que sont les superpixels SLIC. Une analyse des dispersions géométriques des différents groupes de microcalcifications potentielles obtenus peut, à titre d'exemple, être envisagée.

Par ailleurs, l'objectif principal de ce travail de recherche a été de réduire le nombre de détections de type faux négatifs (FN) par rapport à une méthode individuelle. En revanche, le temps d'exécution doit être considéré en parallèle avec l'avantage important de la réduction des FNs. Une piste d'amélioration possible consiste ainsi à réduire le temps d'exécution afin qu'il soit acceptable pour les radiologues. Une autre piste d'extension consiste, ainsi, à envisager une étape de classification qui permet de réduire le nombre de détections de type faux positifs (FP). Elle aurait comme objectif de distinguer les vraies microcalcifications des détections de type FP retenues comme microcalcifications potentielles par le système proposé.

Une autre perspective peut-être lointaine mais aussi intéressante serait d'appliquer le même concept collaboratif proposé sur d'autres domaines d'applications. Ceci implique une généralisation du concept de combinaison collaborative pour une prise de décision tout en partageant des connaissances apprises et exploitant d'autres *a priori*.

LISTE DES PUBLICATIONS SCIENTIFIQUES

Revues

- [Touil *et al.* 2020] (IF : **2.537**) : Touil, A., Kalti, K., Conze, P.-H., Solaiman, B., & Mahjoub, M.A. (2020). Automatic detection of microcalcification based on morphological operations and structural similarity indices. *Biocybernetics and Biomedical Engineering*, Vol.40, Num.3, pp. 1155-1173. (**Accepté**)
- [Touil *et al.* 2021] (IF : **2.116**) : Touil, A., Kalti, K., Conze, P.-H., Solaiman, B., & Mahjoub, M.A. (2021). A new conditional region growing approach for microcalcification delineation in mammograms. *Medical and Biological Engineering and Computing journal*, In Press. (**Accepté**)
- [Touil *et al.* 2021] (IF : **7.196**) : Touil, A., Kalti, K., Conze, P.-H., Solaiman, B., & Mahjoub, M.A. (2021). Collaborative microcalcification detection based on graphs and superpixel knowledge propagation. *Pattern Recognition*. (**En cours de révision - Soumission le 23/08/2021**)

Conférences

Classées

- [Touil *et al.* 2020] (Classe **C**) : Touil, A., Kalti, K., Conze, P.-H., Solaiman, B., & Mahjoub, M.A. (2020 October). A new conditional region growing approach for an accurate detection of microcalcifications from mammographic images. 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE), IEEE, 2020, pp. 777-784.

Internationales

- [Touil *et al.* 2020] : Touil, A., Kalti, K., Conze, P.-H., Solaiman, B., & Mahjoub, M.A. (2020 September). Morphological-based microcalcification detection using adaptive

thresholding and structural similarity indices. In 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), IEEE, 2020, pp. 1-6.

- [Touil *et al.* 2018] : Touil, A., Kalti, K., Conze, P.-H., Solaiman, B., & Mahjoub, M.A. (2018 March). Microcalcifications detection from mammographic images based on region growing and variational energy convergence. In 2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), IEEE, 2018, pp. 1-6.

BIBLIOGRAPHIE

- [1] L. Vivona, D. Cascio, F. Fauci, and G. Raso, “Fuzzy technique for microcalcifications clustering in digital mammograms,” *BMC medical imaging*, vol. 14, no. 1, p. 23, 2014.
- [2] E. Zhang, F. Wang, Y. Li, and X. Bai, “Automatic detection of microcalcifications using mathematical morphology and a support vector machine,” *Bio-medical materials and engineering*, vol. 24, no. 1, pp. 53–59, 2014.
- [3] M. A. Duarte, A. V. Alvarenga, C. M. Azevedo, M. J. G. Calas, A. F. Infantsi, and W. C. Pereira, “Evaluating geodesic active contours in microcalcifications segmentation on mammograms,” *Computer Methods and Programs in Biomedicine*, vol. 122, no. 3, pp. 304–315, 2015.
- [4] G. Veni, E. Regentova, and L. Zhang, “Detection of clustered microcalcifications with susan edge detector, adaptive contrast thresholding and spatial filters,” in *Image Analysis and Recognition*, pp. 837–843, Springer, 2008.
- [5] W. Eziddin, *Segmentation itérative d’images par propagation de connaissances dans le domaine possibiliste : application à la détection de tumeurs en imagerie mammographique*. PhD thesis, Mines Telecom, 2012.
- [6] J. Wang, X. Yang, H. Cai, W. Tan, C. Jin, and L. Li, “Discrimination of breast cancer with microcalcifications on mammography by deep learning,” *Scientific reports*, vol. 6, p. 27327, 2016.
- [7] M. I. Nounou, F. ElAmrawy, N. Ahmed, K. Abdelraouf, S. Goda, and H. Syed-Sha-Qhattal, “Breast cancer : conventional diagnosis and treatment modalities and recent patents and technologies,” *Breast cancer : basic and clinical research*, vol. 9, pp. BCBCR–S29420, 2015.
- [8] M. Alsheh Ali, M. Eriksson, K. Czene, P. Hall, and K. Humphreys, “Detection of potential microcalcification clusters using multivendor for-presentation digital mammograms for short-term breast cancer risk estimation,” *Medical physics*, vol. 46, no. 4, pp. 1938–1946, 2019.
- [9] S. Buseman, J. Mouchawar, N. Calonge, and T. Byers, “Mammography screening matters for young women with breast carcinoma : Evidence of downstaging among

- 42–49-year-old women with a history of previous mammography screening,” *Cancer*, vol. 97, no. 2, pp. 352–358, 2003.
- [10] B. I. Reporting, “Data system atlas (bi-rads atlas) reston,” *VA : American College of Radiology*, 2003.
- [11] M. Le Gal, G. Chavanne, and D. Pellier, “Valeur diagnostique des microcalcifications groupées découvertes par mammographies : à propos de 227 cas avec vérification histologique et sans tumeur du sein palpable,” *Bulletin du cancer*, vol. 71, no. 1, pp. 57–64, 1984.
- [12] C. Nickson, Y. Arzhaeva, Z. Aitken, T. Elgindy, M. Buckley, M. Li, D. R. English, and A. M. Kavanagh, “Autodensity : an automated method to measure mammographic breast density that predicts breast cancer risk and screening outcomes,” *Breast Cancer Research, BioMed Central*, 2013.
- [13] U. Sovio, J. Li, Z. Aitken, K. Humphreys, K. Czene, S. Moss, P. Hall, V. McCormack, and I. dos Santos-Silva, “Comparison of fully and semi-automated area-based methods for measuring mammographic density and predicting breast cancer risk,” *British journal of cancer*, 2014.
- [14] N. Boyd, “Mammographic density and risk of breast cancer,” *Am Soc Clin Oncol Educ Book*, pp. 57–62, 2013.
- [15] J. J. Heine, M. J. Carston, C. G. Scott, K. R. Brandt, F.-F. Wu, V. S. Pankratz, T. A. Sellers, and C. M. Vachon, “An automated approach for estimation of breast density,” *Cancer Epidemiology Biomarkers & Prevention*, vol. 17, no. 11, pp. 3090–3097, 2008.
- [16] A. Cheddad, K. Czene, J. A. Shepherd, J. Li, P. Hall, and K. Humphreys, “Enhancement of mammographic density measures in breast cancer risk prediction,” *Cancer Epidemiology Biomarkers & Prevention*, pp. cebp–1240, 2014.
- [17] N. Saidin, H. A. Mat Sakim, U. K. Ngah, and I. L. Shuaib, “Computer aided detection of breast density and mass, and visualization of other breast anatomical regions on mammograms using graph cuts,” *Computational and mathematical methods in medicine*, vol. 2013, 2013.
- [18] W. He, A. Juettenb, E. R. Dentonb, A. Oliverc, R. Martić, and R. Zwiggelaar, “A review on automatic mammographic density and parenchymal segmentation,” *Hindawi Publishing Corporation, International Journal of Breast Cancer*, 2015.
- [19] W. He, P. Hogg, A. Juetten, E. R. Denton, and R. Zwiggelaar, “Breast image pre-processing for mammographic tissue segmentation,” *Computers in biology and medicine*, vol. 67, pp. 61–73, 2015.
- [20] J. N. Wolfe, “Breast patterns as an index of risk for developing breast cancer,” *American Journal of Roentgenology*, vol. 126, no. 6, pp. 1130–1137, 1976.

- [21] E. A. Sickles, “Wolfe mammographic parenchymal patterns and breast cancer risk,” *American Journal of Roentgenology*, vol. 188, no. 2, pp. 301–303, 2007.
- [22] L. Tabár and P. B. Dean, “Mammographic parenchymal patterns : risk indicator for breast cancer?,” *Jama*, vol. 247, no. 2, pp. 185–189, 1982.
- [23] L. Tabár, T. Tot, and P. B. Dean, *Breast cancer : the art and science of early detection with mammography : perception, interpretation, histopathologic correlation*. Thieme, 2005.
- [24] N. Boyd, J. Byng, R. Jong, E. Fishell, L. Little, A. Miller, G. Lockwood, D. Tritchler, and M. J. Yaffe, “Quantitative classification of mammographic densities and breast cancer risk : results from the canadian national breast screening study,” *Journal of the National Cancer Institute*, vol. 87, no. 9, pp. 670–675, 1995.
- [25] N. F. Boyd, L. J. Martin, L. Sun, H. Guo, A. Chiarelli, G. Hislop, M. Yaffe, and S. Minkin, “Body size, mammographic density, and breast cancer risk,” *Cancer Epidemiology Biomarkers & Prevention*, vol. 15, no. 11, pp. 2086–2092, 2006.
- [26] A. C. of Radiology, B.-R. Committee, *et al.*, *ACR BI-RADS breast imaging and reporting data system : breast imaging atlas*. American College of Radiology, 2003.
- [27] E. S. Burnside, E. A. Sickles, L. W. Bassett, D. L. Rubin, C. H. Lee, D. M. Ikeda, E. B. Mendelson, P. A. Wilcox, P. F. Butler, and C. J. D’Orsi, “The acr bi-rads® experience : learning from history,” *Journal of the American College of Radiology*, vol. 6, no. 12, pp. 851–860, 2009.
- [28] D. O. Kang and R. et al., “Breast calcifications,” 2015.
- [29] D. R. Ericeira, A. C. Silva, A. C. De Paiva, and M. Gattass, “Detection of masses based on asymmetric regions of digital bilateral mammograms using spatial description with variogram and cross-variogram functions,” *Computers in biology and medicine*, vol. 43, no. 8, pp. 987–999, 2013.
- [30] G. Palma, I. Bloch, and S. Muller, “Detection of masses and architectural distortions in digital breast tomosynthesis images using fuzzy and a contrario approaches,” *Pattern Recognition*, vol. 47, no. 7, pp. 2467–2480, 2014.
- [31] A. Mencattini and M. Salmeri, “Breast masses detection using phase portrait analysis and fuzzy inference systems,” *International journal of computer assisted radiology and surgery*, vol. 7, no. 4, pp. 573–583, 2012.
- [32] R. F. Cox, A. Hernandez-Santana, S. Ramdass, G. McMahon, J. H. Harmey, and M. P. Morgan, “Microcalcifications in breast cancer : novel insights into the molecular mechanism and functional consequence of mammary mineralisation,” *British journal of cancer*, vol. 106, no. 3, pp. 525–537, 2012.
- [33] “Programme de dépistage du cancer du sein en communauté française : Classification des anomalies mammographiques,” 2005. 2005.

- [34] C. Imene and K. Djemal, *Description et classification des masses mammaires pour le diagnostic du cancer du sein*. PhD thesis, Université d'evry val d'essonne, 2012.
- [35] S. Banik, R. M. Rangayyan, and J. L. Desautels, "Computer-aided detection of architectural distortion in prior mammograms of interval cancer," *Synthesis Lectures on Biomedical Engineering*, vol. 8, no. 1, 2013.
- [36] P. SUCKLING J, "The mammographic image analysis society digital mammogram database," *Digital Mammo*, pp. 375–386, 1994.
- [37] K. Bowyer, D. Kopans, W. Kegelmeyer, R. Moore, M. Sallam, K. Chang, and K. Woods, "The digital database for screening mammography," in *Third international workshop on digital mammography*, vol. 58, p. 27, 1996.
- [38] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "Inbreast : toward a full-field digital mammographic database," *Academic radiology*, vol. 19, no. 2, pp. 236–248, 2012.
- [39] S. Bae, J. H. Yoon, H. J. Moon, M. J. Kim, and E.-K. Kim, "Breast microcalcifications : diagnostic outcomes according to image-guided biopsy method," *Korean journal of radiology*, vol. 16, no. 5, pp. 996–1005, 2015.
- [40] A. Touil, K. Kalti, P.-H. Conze, B. Solaiman, and M. A. Mahjoub, "A new conditional region growing approach for microcalcification delineation in mammograms," *Medical & Biological Engineering & Computing*, 2021.
- [41] S. Khouaja, H. Jlassi, and K. Hamrouni, "An automated method for breast mass segmentation," in *Soft Computing and Pattern Recognition (SoCPaR), 2014 6th International Conference of*, pp. 180–185, IEEE, 2014.
- [42] R. M. Rangayyan, S. Banik, and J. L. Desautels, "Detection of architectural distortion in prior mammograms via analysis of oriented patterns," *JoVE (Journal of Visualized Experiments)*, no. 78, pp. e50341–e50341, 2013.
- [43] A. Fanizzi, T. M. Basile, L. Losurdo, R. Bellotti, U. Bottigli, R. Dentamaro, V. Didonna, A. Fausto, R. Massafra, M. Moschetta, *et al.*, "A machine learning approach on multiscale texture analysis for breast microcalcification diagnosis," *BMC bioinformatics*, vol. 21, no. 2, pp. 1–11, 2020.
- [44] T. Basile, A. Fanizzi, L. Losurdo, R. Bellotti, U. Bottigli, R. Dentamaro, V. Didonna, A. Fausto, R. Massafra, M. Moschetta, *et al.*, "Microcalcification detection in full-field digital mammograms : A fully automated computer-aided system," *Physica Medica*, vol. 64, pp. 1–9, 2019.
- [45] S. Sapate, S. Talbar, A. Mahajan, N. Sable, S. Desai, and M. Thakur, "Breast cancer diagnosis using abnormalities on ipsilateral views of digital mammograms," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 290–305, 2020.

- [46] M. Posso, M. Carles, M. Rué, T. Puig, and X. Bonfill, “Cost-effectiveness of double reading versus single reading of mammograms in a breast cancer screening programme,” *PloS one*, vol. 11, no. 7, p. e0159806, 2016.
- [47] Z. Wang, J. Xin, Q. Zhang, S. Gao, C. Ma, J. Ren, H. Zhang, W. Qian, W. Zhu, X. Zhang, *et al.*, “Breast microcalcifications detection based on fusing features with dtcwt,” *Journal of X-Ray Science and Technology*, no. Preprint, pp. 1–22, 2020.
- [48] A. Bria, C. Marrocco, J.-J. Mordang, N. Karssemeijer, M. Molinara, and F. Tortorella, “Lut-qne : Look-up-table quantum noise equalization in digital mammograms,” in *International Workshop on Digital Mammography*, pp. 27–34, Springer, 2016.
- [49] A. Torrent, A. Oliver, X. Lladó, R. Martí, and J. Freixenet, “A supervised microcalcification detection approach in digitised mammograms,” in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pp. 4345–4348, IEEE, 2010.
- [50] A. Oliver, A. Torrent, X. Lladó, M. Tortajada, L. Tortajada, M. Sentís, J. Freixenet, and R. Zwigelaar, “Automatic microcalcification and cluster detection for digital and digitised mammograms,” *Knowledge-Based Systems*, vol. 28, pp. 68–75, 2012.
- [51] R. Bhanumathi and G. Suresh, “Detection of microcalcification in mammogram images using support vector machine based classifier,” *TSI Transactions on Electrical and Electronics Engineering (ITSI-TEEE)*, vol. 1, no. 2, pp. 27–32, 2013.
- [52] L. Wei, Y. Yang, R. M. Nishikawa, and Y. Jiang, “A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications,” *IEEE transactions on medical imaging*, vol. 24, no. 3, pp. 371–380, 2005.
- [53] X. Liu, M. Mei, J. Liu, and W. Hu, “Microcalcification detection in full-field digital mammograms with pfc clustering and weighted svm-based method,” *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, pp. 1–13, 2015.
- [54] A. Bria, C. Marrocco, A. Galdran, A. Campilho, A. Marchesi, J.-J. Mordang, N. Karssemeijer, M. Molinara, and F. Tortorella, “Spatial enhancement by dehaizing for detection of microcalcifications with convolutional nets,” in *International Conference on Image Analysis and Processing*, pp. 288–298, Springer, 2017.
- [55] G. Valvano, D. Della Latta, N. Martini, G. Santini, A. Gori, C. Iacconi, A. Ripoli, L. Landini, and D. Chiappino, “Evaluation of a deep convolutional neural network method for the segmentation of breast microcalcifications in mammography imaging,” in *EMBECE & NBC 2017*, pp. 438–441, Springer, 2017.
- [56] J.-J. Mordang, A. Gubern-Mérida, G. den Heeten, and N. Karssemeijer, “Reducing false positives of microcalcification detection systems by removal of breast arterial calcifications,” *Medical physics*, vol. 43, no. 4, pp. 1676–1687, 2016.

- [57] J.-J. Mordang, A. Gubern-Mérida, A. Bria, F. Tortorella, G. Heeten, and N. Karssemeijer, “Improving computer-aided detection assistance in breast cancer screening by removal of obviously false-positive findings,” *Medical Physics*, vol. 44, no. 4, pp. 1390–1401, 2017.
- [58] N. Wahab, A. Khan, and Y. S. Lee, “Two-phase deep convolutional neural network for reducing class skewness in histopathological images based breast cancer detection,” *Computers in biology and medicine*, vol. 85, pp. 86–97, 2017.
- [59] K. Hu, W. Yang, and X. Gao, “Microcalcification diagnosis in digital mammography using extreme learning machine based on hidden markov tree model of dual-tree complex wavelet transform,” *Expert Systems with Applications*, 2017.
- [60] J. Wang and Y. Yang, “A context-sensitive deep learning approach for microcalcification detection in mammograms,” *Pattern recognition*, vol. 78, pp. 12–22, 2018.
- [61] F. Zhang, L. Luo, X. Sun, Z. Zhou, X. Li, Y. Yu, and Y. Wang, “Cascaded generative and discriminative learning for microcalcification detection in breast mammograms,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 12578–12586, 2019.
- [62] L. Wei, Y. Yang, and R. M. Nishikawa, “Microcalcification classification assisted by content-based image retrieval for breast cancer diagnosis,” *Pattern recognition*, vol. 42, no. 6, pp. 1126–1132, 2009.
- [63] H. Cai, Q. Huang, W. Rong, Y. Song, J. Li, J. Wang, J. Chen, and L. Li, “Breast microcalcification diagnosis using deep convolutional neural network from digital mammograms,” *Computational and mathematical methods in medicine*, vol. 2019, 2019.
- [64] A. Cruz-Bernal, M. M. Flores-Barranco, D. L. Almanza-Ojeda, S. Ledesma, and M. A. Ibarra-Manzano, “Analysis of the cluster prominence feature for detecting calcifications in mammograms,” *Journal of healthcare engineering*, vol. 2018, 2018.
- [65] Lowis, Hendra, and Lavinia, “The use of dual-tree complex wavelet transform (dtcwt) based feature for mammogram classification,” *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 8, no. 3, pp. 87–96, 2015.
- [66] N. Arikidis, K. Vassiou, A. Kazantzi, S. Skiadopoulos, A. Karahaliou, and L. Costaridou, “A two-stage method for microcalcification cluster segmentation in mammography by deformable models,” *Medical physics*, vol. 42, no. 10, pp. 5848–5861, 2015.
- [67] A. S. Bharadwaj and M. Celenk, “Detection of microcalcification with top-hat transform and the gibbs random fields,” in *Engineering in Medicine and Biology Society*

- (EMBC), *2015 37th Annual International Conference of the IEEE*, pp. 6382–6385, IEEE, 2015.
- [68] M. A. Guerroudji and Z. Ameer, “A new approach for the detection of mammary calcifications by using the white top-hat transform and thresholding of otsu,” *Optik*, vol. 127, no. 3, pp. 1251–1259, 2016.
- [69] M. Ciecholewski, “Microcalcification segmentation from mammograms : A morphological approach,” *Journal of Digital Imaging*, pp. 1–13, 2016.
- [70] A. Mencattini, G. Rabottino, M. Salmeri, F. Caselli, R. Lojacono, and G. Frondizi, “Features extraction to classify microcalcification clusters in mammography,” in *16th IMEKO TC4 Symposium, Exploring New Frontiers of Instrumentation and Methods for Electrical and Electronic Measurements*, pp. 22–24, Citeseer, 2008.
- [71] M. Rizzi, M. D’Aloia, and B. Castagnolo, “Computer aided detection of microcalcifications in digital mammograms adopting a wavelet decomposition,” *Integrated Computer-Aided Engineering*, vol. 16, no. 2, pp. 91–103, 2009.
- [72] T. Balakumaran, I. Vennila, and C. G. Shankar, “Detection of microcalcification in mammograms using wavelet transform and fuzzy shell clustering,” *International Journal of Computer Science and Information Security (IJCSIS)*, vol. 7, no. 1, pp. 121–125, 2010.
- [73] R. N. Dave, “Fuzzy shell-clustering and applications to circle detection in digital images,” *International Journal Of General System*, vol. 16, no. 4, pp. 343–355, 1990.
- [74] E. Malar, A. Kandaswamy, D. Chakravarthy, and A. Giri Dharan, “A novel approach for detection and classification of mammographic microcalcifications using wavelet analysis and extreme learning machine,” *Computers in biology and medicine*, vol. 42, no. 9, pp. 898–905, 2012.
- [75] N. B. Hamad, K. Taouil, and M. S. Bouhlef, “Mammographic microcalcifications detection using discrete wavelet transform,” *International Journal of Computer Applications*, vol. 64, no. 21, 2013.
- [76] K. P. Shetty, D. V. Udupi, and M. B. K. Saptalakar, “Wavelet based microcalcification detection on mammographic images,” *IJCSNS International Journal of Computer Science and Network Security*, vol. 9, no. 7, pp. 213–217, 2009.
- [77] M. Sundaram, K. Ramar, N. Arumugam, and G. Prabin, “Efficient edge emphasized mammogram image enhancement for detection of microcalcification,” *Biomedical Engineering : Applications, Basis and Communications*, vol. 26, no. 05, p. 1450056, 2014.
- [78] R. Lakshmanan and V. Thomas, “Microcalcification detection by morphology, singularities of contourlet transform and neural network,” *Bonfring International Journal of Networking Technologies and Applications*, vol. 1, no. 1, pp. 14–19, 2012.

- [79] R. Lakshmanan and V. Thomas, "Mammographic feature enhancement using singularities of contourlet transform," *International Journal on Recent Trends in Engineering and Technology*, vol. 8, no. 2, pp. 87–90, 2013.
- [80] W. B. Vignesh and M. Sundaram, "Effect of contourlet transform in detect of microcalcification in noisy environment," *IEEE Sponsored 9th International Conference on Intelligent Systems and Control (ISCO)2015, At COIMBATORE*, 2015.
- [81] Y. Guo, M. Dong, Z. Yang, X. Gao, K. Wang, C. Luo, Y. Ma, and J. Zhang, "A new method of detecting micro-calcification clusters in mammograms using contourlet transform and non-linking simplified pcnn," *Computer methods and programs in biomedicine*, vol. 130, pp. 31–45, 2016.
- [82] V. Alarcon-Aquino, O. Starostenko, J. Ramirez-Cortes, R. Rosas-Romero, J. Rodriguez-Asomoza, O. J. Paz-Luna, and K. Vazquez-Muñoz, "Detection of microcalcifications in digital mammograms using the dual-tree complex wavelet transform," *Engineering Intelligent Systems*, vol. 17, no. 1, p. 49, 2009.
- [83] L. Najman and H. Talbot, *Mathematical morphology : from theory to applications*. John Wiley & Sons, 2013.
- [84] G. Valvano, G. Santini, N. Martini, A. Ripoli, C. Iacconi, D. Chiappino, and D. Della Latta, "Convolutional neural networks for the segmentation of microcalcification in mammography imaging," *Journal of healthcare engineering*, vol. 2019, 2019.
- [85] M. d. A. Duarte, A. V. Alvarenga, C. M. Azevedo, M. J. G. Calas, A. F. C. Infantosi, and W. C. d. A. Pereira, "Segmenting mammographic microcalcifications using a semi-automatic procedure based on otsu's method and morphological filters," *Revista Brasileira de Engenharia Biomédica*, vol. 29, no. 4, pp. 377–388, 2013.
- [86] P. Soille, *Morphological image analysis : principles and applications*. Springer Science & Business Media, 2013.
- [87] C. C. Diaz-Huerta, E. M. Felipe-Riveron, and L. M. Montaña-Zetina, "Quantitative analysis of morphological techniques for automatic classification of microcalcifications in digitized mammograms," *Expert Systems with Applications*, vol. 41, no. 16, pp. 7361–7369, 2014.
- [88] S. Beucher, "Use of watersheds in contour detection," in *Proceedings of the International Workshop on Image Processing*, CCETT, 1979.
- [89] D. Betal, N. Roberts, and G. Whitehouse, "Segmentation and numerical analysis of microcalcifications on mammograms using mathematical morphology.," *The British Journal of Radiology*, vol. 70, no. 837, pp. 903–917, 1997.
- [90] P. Bougioukos, D. Glotsos, S. Kostopoulos, A. Daskalakis, I. Kalatzis, N. Dimitropoulos, G. Nikiforidis, and D. Cavouras, "Fuzzy c-means-driven fhce contextual

- segmentation method for mammographic microcalcification detection,” *The Imaging Science Journal*, vol. 58, no. 3, pp. 146–154, 2010.
- [91] M. P. Guisado and D. M. Gómez-Allende, “Wood texture analysis by combining the connected elements histogram and artificial neural networks,” in *Bio-inspired applications of connectionism*, pp. 160–167, Springer, 2001.
- [92] J. Quintanilla-Domínguez, B. Ojeda-Magaña, A. Marciano-Cedeño, M. G. Cortina-Januchs, A. Vega-Corona, and D. Andina, “Improvement for detection of microcalcifications through clustering algorithms and artificial neural networks,” *EURASIP J. Adv. Sig. Proc.*, vol. 2011, p. 91, 2011.
- [93] J. Quintanilla-Dominguez, B. Ojeda-Magaña, M. Cortina-Januchs, R. Ruelas, A. Vega-Corona, and D. Andina, “Image segmentation by fuzzy and possibilistic clustering algorithms for the identification of microcalcifications,” *Scientia Iranica*, vol. 18, no. 3, pp. 580–589, 2011.
- [94] Z. Suhail, M. Sarwar, and K. Murtaza, “Automatic detection of abnormalities in mammograms,” *BMC medical imaging*, vol. 15, no. 1, p. 53, 2015.
- [95] A. A. Malek, W. E. Z. W. A. Rahman, A. Ibrahim, R. Mahmud, S. S. Yasiran, and A. K. Jumaat, “Region and boundary segmentation of microcalcifications using seed-based region growing and mathematical morphology,” *Procedia-Social and Behavioral Sciences*, vol. 8, pp. 634–639, 2010.
- [96] R. Adams and L. Bischof, “Seeded region growing,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 16, no. 6, pp. 641–647, 1994.
- [97] V. Caselles, R. Kimmel, and G. Sapiro, “Geodesic active contours,” *International journal of computer vision*, vol. 22, no. 1, pp. 61–79, 1997.
- [98] P. Perona and J. Malik, “Scale-space and edge detection using anisotropic diffusion,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 12, no. 7, pp. 629–639, 1990.
- [99] M. Kass, A. Witkin, and D. Terzopoulos, “Snakes : Active contour models,” *International journal of computer vision*, vol. 1, no. 4, pp. 321–331, 1988.
- [100] B. Sahiner, H.-P. Chan, L. M. Hadjiiski, M. A. Helvie, C. Paramagul, J. Ge, J. Wei, and C. Zhou, “Joint two-view information for computerized detection of microcalcifications on mammograms,” *Medical physics*, vol. 33, no. 7, pp. 2574–2585, 2006.
- [101] H. Moradmand, S. Setayeshi, A. R. Karimian, M. Sirous, and M. E. Akbari, “Comparing the performance of image enhancement methods to detect microcalcification clusters in digital mammography,” *Iranian Journal of Cancer Prevention*, vol. 5, no. 2, pp. 61–68, 2012.

- [102] O. Intharasombat and P. Piamsa-nga, "Adaptive window and adaptive threshold method for microcalcification detection," in *Computer Science and Engineering Conference (ICSEC), 2014 International*, pp. 446–451, IEEE, 2014.
- [103] P. L. A. Hernández, T. T. Estrada, A. L. Pizarro, and M. L. D. Cisternas, "Breast calcifications : description and classification according to bi-rads 5th edition," *Revista Chilena de Radiología*, vol. 22, no. 2, pp. 80–91, 2016.
- [104] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, pp. 281–297, Oakland, CA, USA, 1967.
- [105] J. Quintanilla-Domínguez, J. Ruiz-Pinales, J. M. Barrón-Adame, and R. Guzmán-Cabrera, "Microcalcifications detection using image processing," *Computación y Sistemas*, vol. 22, no. 1, pp. 291–300, 2018.
- [106] J. C. Bezdek, R. Ehrlich, and W. Full, "Fcm : The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2-3, pp. 191–203, 1984.
- [107] N. R. Pal, K. Pal, J. M. Keller, and J. C. Bezdek, "A possibilistic fuzzy c-means clustering algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 4, pp. 517–530, 2005.
- [108] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE transactions on fuzzy systems*, vol. 1, no. 2, pp. 98–110, 1993.
- [109] J. Quintanilla-Domínguez, B. Ojeda-Magaña, A. Marcano-Cedeño, J. Barrón-Adame, A. Vega-Corona, and D. Andina, "Automatic detection of microcalcifications in roi images based on pfcmm and ann," *International Journal of Intelligent Computing in Medical Sciences & Image Processing*, vol. 5, no. 2, pp. 161–174, 2013.
- [110] B. Ojeda-Magaña, J. Quintanilla-Domínguez, R. Ruelas, and D.-M. Andina, "Images sub-segmentation with the pfcmm clustering algorithm," in *Industrial Informatics, 2009. INDIN 2009. 7th IEEE International Conference on*, pp. 499–503, IEEE, 2009.
- [111] H.-D. Cheng, J. Wang, and X. Shi, "Microcalcification detection using fuzzy logic and scale space approaches," *Pattern Recognition*, vol. 37, no. 2, pp. 363–375, 2004.
- [112] E. L. Meléndez and G. Urcid, "Mammograms calcifications segmentation based on band-pass fourier filtering and adaptive statistical thresholding," *European International Journal of Science and Technology*, 2016.
- [113] S. Butterworth *et al.*, "On the theory of filter amplifiers," *Wireless Engineer*, vol. 7, no. 6, pp. 536–541, 1930.
- [114] X. Zhang, N. Homma, S. Goto, Y. Kawasumi, T. Ishibashi, M. Abe, N. Sugita, and M. Yoshizawa, "A hybrid image filtering method for computer-aided detection

- of microcalcification clusters in mammograms,” *Journal of Medical Engineering*, vol. 2013, 2013.
- [115] A. Touil and K. Kalti, “Iterative fuzzy segmentation for an accurate delimitation of the breast region,” *Comput. Methods Programs Biomed.*, vol. 132, pp. 137–147, 2016.
- [116] J. Wang, Y. Yang, and R. M. Nishikawa, “Reduction of false positive detection in clustered microcalcifications,” in *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pp. 1433–1437, IEEE, 2013.
- [117] B. Joseph, B. Ramachandran, and P. Muthukrishnan, “Intelligent detection and classification of microcalcification in compressed mammogram image,” *Image Analysis & Stereology*, vol. 34, no. 3, pp. 183–198, 2015.
- [118] Y.-T. Kim, “Contrast enhancement using brightness preserving bi-histogram equalization,” *IEEE transactions on Consumer Electronics*, vol. 43, no. 1, pp. 1–8, 1997.
- [119] S. Wu, S. P. Weinstein, E. F. Conant, and D. Kontos, “Automated fibroglandular tissue segmentation and volumetric density estimation in breast mri using an atlas-aided fuzzy c-means method,” *Medical physics*, vol. 40, no. 12, p. 122302, 2013.
- [120] A. A. Kayode, B. S. Afolabi, and B. O. Ibitoye, “An explorative survey of image enhancement techniques used in mammography,” *IJCSI International Journal of Computer Science Issues*, vol. 12, January 2015.
- [121] K. Akila, L. Jayashree, and A. Vasuki, “Mammographic image enhancement using indirect contrast enhancement techniques—a comparative study,” *Procedia Computer Science*, vol. 47, pp. 255–261, 2015.
- [122] Z. Wu, J. Yuan, B. Lv, and X. Zheng, “Digital mammography image enhancement using improved unsharp masking approach,” in *Image and Signal Processing (CISP), 2010 3rd International Congress on*, vol. 2, pp. 668–672, IEEE, 2010.
- [123] J. Mohanalin *et al.*, “A new wavelet algorithm to enhance and detect microcalcifications,” *Signal Processing*, vol. 105, pp. 438–448, 2014.
- [124] A. Touil, K. Kalti, P.-H. Conze, B. Solaiman, and M. A. Mahjoub, “Automatic detection of microcalcification based on morphological operations and structural similarity indices,” *Biocybernetics and Biomedical Engineering*, 2020.
- [125] M. N. Gurcan, Y. Y. Mc, and A. E. Cetin, “Microcalcification segmentation and mammogram image enhancement using nonlinear filtering,” 1999.
- [126] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, *et al.*, “Image quality assessment : from error visibility to structural similarity,” *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

- [127] G. P. Renieblas, A. T. Nogués, A. M. González, N. G. León, and E. G. Del Castillo, “Structural similarity index family for image quality assessment in radiological images,” *Journal of Medical Imaging*, vol. 4, no. 3, p. 035501, 2017.
- [128] I. Bloch, “Fuzzy methods in medical imaging,” in *Handbook of Biomedical Imaging*, pp. 25–44, Springer, 2015.
- [129] W. Liu, “Conflict analysis and merging operators selection in possibility theory,” in *ECSQARU*, pp. 816–827, Springer, 2007.
- [130] X. Ren and J. Malik, “Learning a classification model for segmentation,” in *null*, p. 10, IEEE, 2003.
- [131] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, “Slic super-pixels,” tech. rep., 2010.
- [132] S. Seo, *A review and comparison of methods for detecting outliers in univariate data sets*. PhD thesis, University of Pittsburgh, 2006.
- [133] C. H. Yu, “Exploratory data analysis,” *Methods*, vol. 2, pp. 131–160, 1977.
- [134] A. Touil, K. Kalti, P.-H. Conze, B. Solaiman, and M. A. Mahjoub, “A new conditional region growing approach for an accurate detection of microcalcifications from mammographic images,” in *20th IEEE International Conference on BioInformatics And BioEngineering, BIBE, October 26-28, 2020 USA*, 2020.
- [135] J. L. Muerle, “Experimental evaluation of techniques for automatic segmentation of objects in a complex scene,” *Pictorial pattern recognition*, pp. 3–13, 1968.
- [136] G. M. Hadi, “Image segmentation based on single seed region growing algorithm,” *ZANCO Journal of Pure and Applied Sciences*, vol. 28, no. 2, 2016.
- [137] M. S. Chaibou, K. Kalti, B. Solaiman, and M. A. Mahjoub, “Image segmentation by contextual region growing based on fuzzy classification,” in *2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, pp. 489–493, IEEE, 2016.
- [138] H. Digabel and C. Lantuéjoul, “Iterative algorithms,” in *Proc. 2nd European Symp. Quantitative Analysis of Microstructures in Material Science, Biology and Medicine*, vol. 19, p. 8, Stuttgart, West Germany : Riederer Verlag, 1978.
- [139] A. Albiol, A. Corbi, and F. Albiol, “Automatic intensity windowing of mammographic images based on a perceptual metric,” *Medical physics*, vol. 44, no. 4, pp. 1369–1378, 2017.
- [140] C. Revol-Muller, T. Grenier, J.-L. Rose, A. Pacureanu, F. Peyrin, and C. Odet, “Region growing : When simplicity meets theory–region growing revisited in feature space and variational framework,” in *Computer Vision, Imaging and Computer Graphics. Theory and Application*, pp. 426–444, Springer, 2013.

- [141] A. Touil, K. Kalti, B. Solaiman, and M. A. Mahjoub, “Microcalcifications detection from mammographie images based on region growing and variational energy convergence,” in *4th International Conference on Advanced Technologies for Signal and Image Processing, ATSIP 2018, Sousse, Tunisia, March 21-24, 2018*, pp. 1–6, 2018.
- [142] J.-L. Rose, T. Grenier, C. Revol-Muller, and C. Odet, “Unifying variational approach and region growing segmentation,” in *2010 18th European Signal Processing Conference*, pp. 1781–1785, IEEE, 2010.
- [143] P. Getreuer, “Chan-veese segmentation,” *Image Processing On Line*, vol. 2, pp. 214–224, 2012.
- [144] L. Wilkinson, V. Thomas, and N. Sharma, “Microcalcification on mammography : approaches to interpretation and biopsy,” *The British journal of radiology*, vol. 90, no. 1069, p. 20160594, 2017.
- [145] G. Forestier, *Connaissances et clustering collaboratif d’objets complexes multi-sources*. PhD thesis, Strasbourg, 2010.
- [146] A. Cornuéjols, C. Wemmert, P. Gañçarski, and Y. Bennani, “Collaborative clustering : Why, when, what and how,” *Information Fusion*, vol. 39, pp. 81–95, 2018.
- [147] W. R. Gonzalez RC, *Digital image processing, 3rd edition*. Prentice Hall, Upper Saddle River, 2008.
- [148] L. C. Romualdo, M. A. Vieira, H. Schiabel, N. D. Mascarenhas, and L. R. Borges, “Mammographic image denoising and enhancement using the anscombe transformation, adaptive wiener filtering, and the modulation transfer function,” *Journal of digital imaging*, vol. 26, no. 2, pp. 183–197, 2013.
- [149] R. S. Saunders Jr, J. A. Baker, D. M. Delong, J. P. Johnson, and E. Samei, “Does image quality matter ? impact of resolution and noise on mammographic task performance,” *Medical physics*, vol. 34, no. 10, pp. 3971–3981, 2007.

ANNEXE A : SCHÉMA DE PRÉTRAITEMENT ADOPTÉ

Le choix d'une technique d'amélioration des mammographies numériques est une étape clé pour le bon fonctionnement du processus de détection appliqué. Dans cette annexe, nous présentons la démarche d'évaluation que nous avons suivie pour la comparaison de la pertinence de plusieurs techniques de prétraitement sélectionnées à partir de l'état de l'art.

A.1 Sélection des techniques de prétraitement

Dans ce travail de recherche, les techniques de prétraitement que nous avons étudié la fiabilité dans le contexte d'amélioration des MCs sur les images mammographiques sont les suivantes :

- La transformation Gamma ainsi que l'égalisation d'histogramme comme techniques de changement de contraste ;
- Le Window Level comme technique de changement de contraste et de luminosité ;
- Les filtres de Nagao et le Unsharp Masking comme des techniques de filtrages.

La liste détaillée des techniques appliquées ainsi que leurs notations sont comme suit :

- Transformation de Gamma avec trois différentes valeurs qui sont respectivement 0.5, 2 et 3. Elles sont respectivement notées $\mathbf{T1}_{G05}$, $\mathbf{T1}_{G2}$ et $\mathbf{T1}_{G3}$;
- Transformation d'égalisation d'histogramme qui sont :
 - * $\mathbf{T2}_{HE}$: Égalisation d'histogramme standard ;
 - * $\mathbf{T2}_{CLAHE}$: Égalisation d'histogramme adaptative ;
 - * $\mathbf{T2}_{BBHE}$: 'Brightness Preserving Bi-histogram Equalization'.

- Technique Window Level $\mathbf{T3}_{WL}$ proposée par Albiol *et al.* en 2017 [139] ;
- Les filtres de Nagao $\mathbf{T4}_N$ et Unsharp Masking $\mathbf{T4}_U$.

A.2 Démarche d'évaluation

L'application d'une étape de prétraitement, dans ce travail de recherche, avait comme objectif d'améliorer la qualité d'apparition des microcalcifications sur les images mammographiques. C'est une étape clé pour démarrer le processus de détection basé dans un premier temps sur la génération des superpixels SLIC (SP1).

De ce fait, la fiabilité d'une technique est liée à l'évaluation de la qualité des superpixels SLIC générés. On peut, à ce niveau, distinguer deux aspects d'évaluation : Le premier aspect concerne le lien entre le prétraitement appliqué et la qualité d'apparition des microcalcifications sur l'image. Le deuxième aspect concerne l'impact du prétraitement sur la qualité des superpixels SP1 générés à partir de l'image.

A.2.1 Critères d'évaluation

A.2.2 Images de test

Pour l'évaluation des techniques de prétraitement, nous avons choisis quatre catégories d'images. Ces dernières correspondent aux différentes classes de densité et présentent différents types de microcalcifications. Une image de chacune de ces catégories est illustrée dans la figure suivante (Figure A.1).

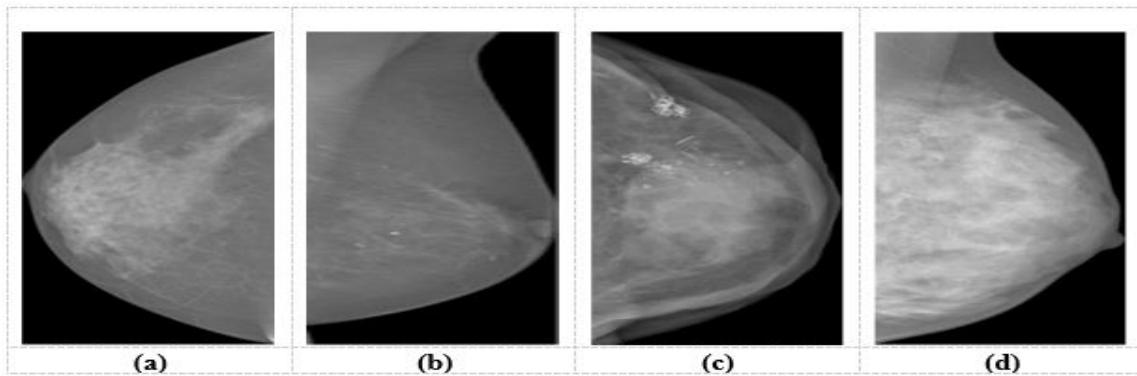


Figure A.1 – Echantillons d'images mammographiques pour l'évaluation des techniques de prétraitement appartenant à différentes catégories

La première catégorie concerne des seins gras hétérogènes. L'image donnée dans la figure A.1.a appartient à cette catégorie et présente deux foyers de microcalcifications superposés sur deux zones mammaires avec différentes densités.

La deuxième catégorie concerne des seins grassex homogènes. L'image donnée dans la figure A.1.b appartient à cette catégorie et présente quelques microcalcifications avec des contrastes élevés et de tailles moyennes.

La troisième catégorie présente des seins grassex homogènes. L'image donnée dans la figure A.1.c appartient à cette catégorie et présente des macrocalcifications de grandes tailles éparpillées sur la moitié supérieure du sein. Ces dernières apparaissent avec des zones brillantes et d'autres sombres.

La quatrième catégorie concerne des seins denses homogènes. L'image donnée dans la figure A.1.d appartient à cette catégorie et présente des microcalcifications de petites tailles avec des contours flous et superposées sur des tissus denses. Elles apparaissent ainsi avec des contours et des contrastes flous.

La figure suivante présente les régions d'intérêts sur ces images où apparaissent les microcalcifications.

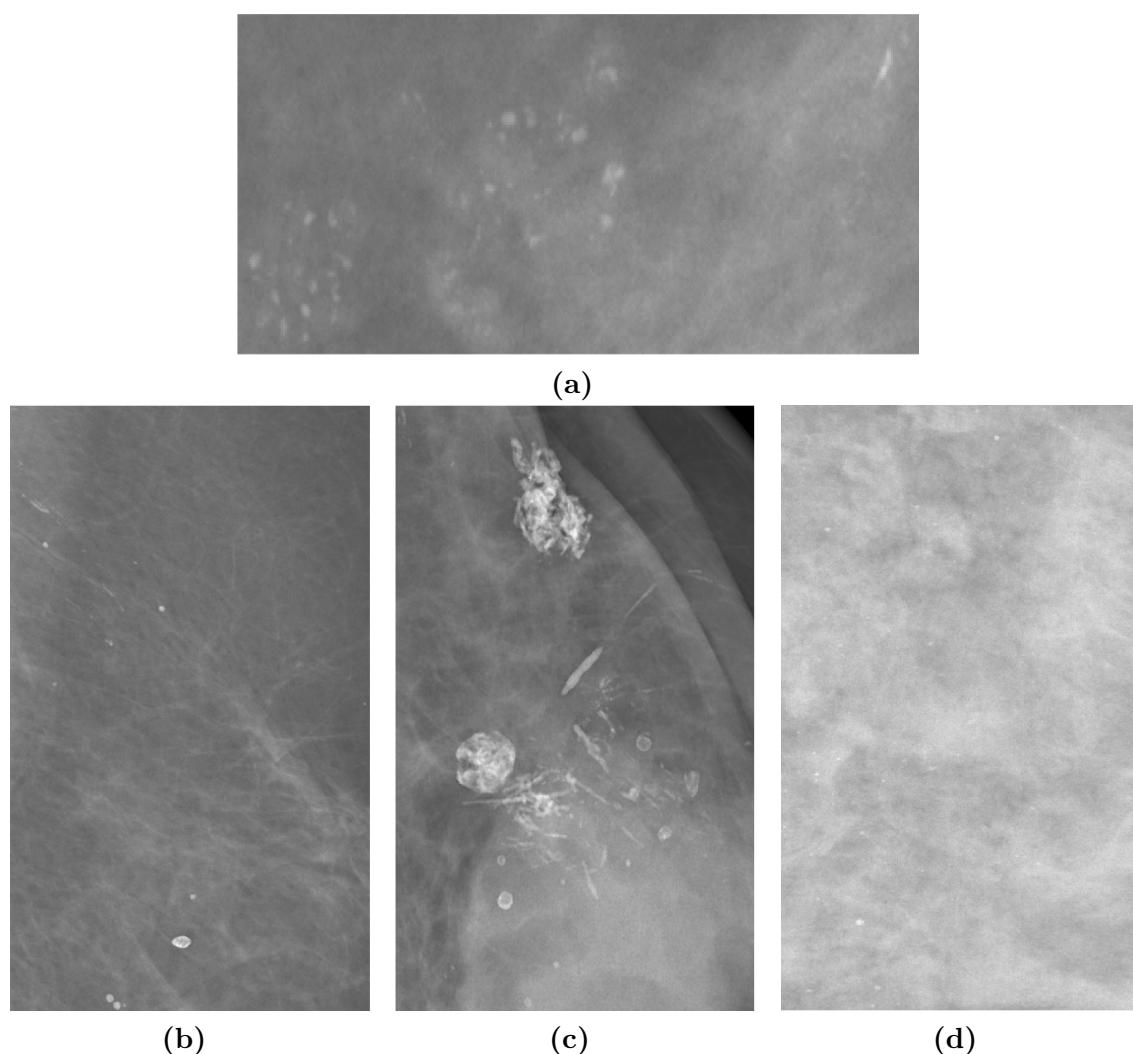


Figure A.2 – Les ROIs sur les images mammographiques sélectionnées

A.3 Évaluation relative à la qualité d'apparition des microcalcifications

A.3.1 Critères d'évaluation

Les critères d'évaluation à ce niveau concernent l'amélioration de la qualité d'apparition des microcalcifications sur l'image. On distingue les critères suivants :

- C_1 : Préservation des microcalcifications sans la génération de fausses alarmes de type Faux Négatifs (FNs) ;
- C_2 : Amélioration de la qualité des contours des microcalcifications ;
- C_3 : Amélioration de la qualité du contraste des microcalcifications ;
- C_4 : Pas de génération des fausses alarmes de type Faux Positifs (FPs) ;
- C_5 : Préservation des structures mammaires de la mammographie.

Le schéma de la figure A.3 présente l'ordre d'importance associé à ces critères.

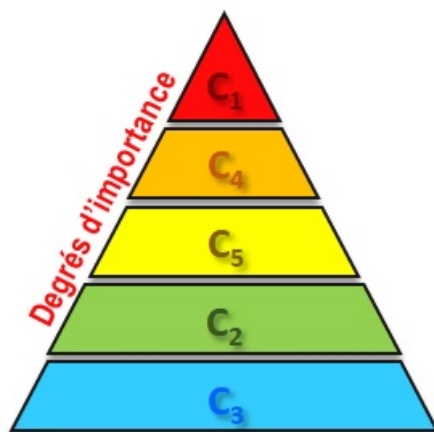


Figure A.3 – Degrés d'importance des critères de validation des techniques de prétraitement

A.3.2 Protocoles d'évaluation

L'évaluation s'effectue selon deux repères bien définis :

- La qualité des superpixels SLIC générés à partir de l'image ;
- La vérité de terrain des microcalcifications validée par le médecin radiologue.

Avec l'absence de mesures d'évaluations quantitatives, nous avons défini un repère d'interprétation basé sur des symboles illustrés dans la figure A.4 suivante afin d'étudier

le degré de validité d'une technique de prétraitement donnée : Les symboles \odot , \bullet et \bullet désignent les degrés de validité du critère en question. Le symbole \odot (respectivement \bullet) désigne que le critère est très peu (respectivement très bien) validé. Par ailleurs, les symboles \ominus , \ominus et \bullet désignent les degrés d'invalidité du critère en question.

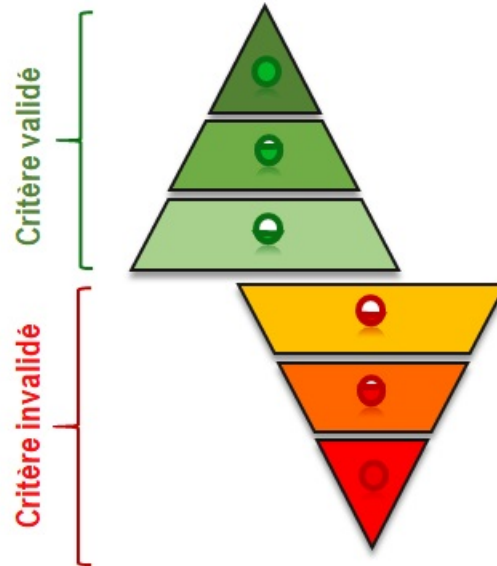


Figure A.4 – Symboles utilisés pour évaluer la validation des critères

A.3.3 Résultats d'évaluation

Dans les figures A.5 et A.6, nous avons porté les résultats de l'application des différentes techniques de prétraitement sur les échantillons d'images que nous avons sélectionnées. La figure A.5 présente ces résultats sur les images mammographiques de test entières. La figure A.6 présente un agrandissement des résultats obtenus sur les régions d'intérêts de chaque image.

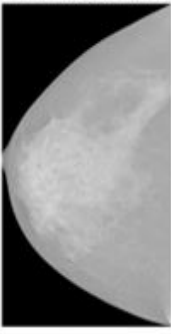
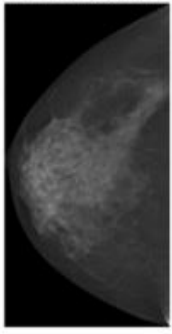
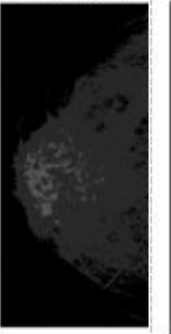
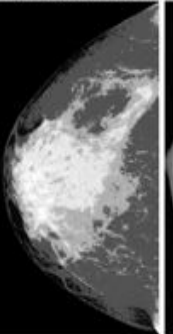
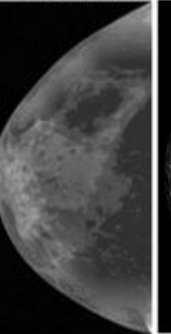
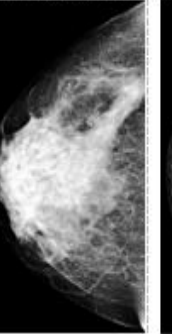
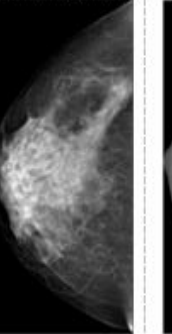
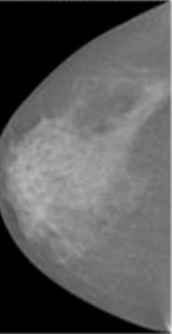
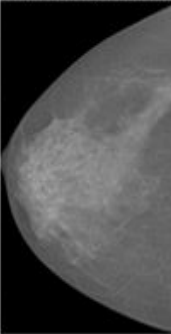
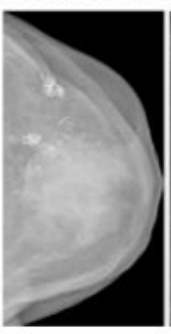
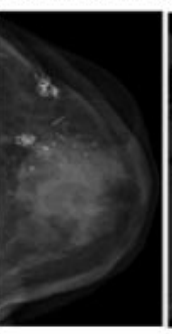
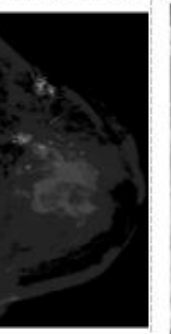
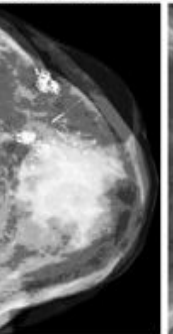
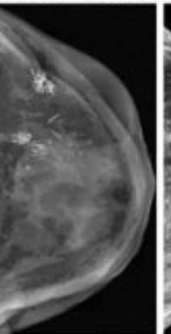
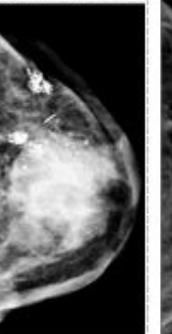
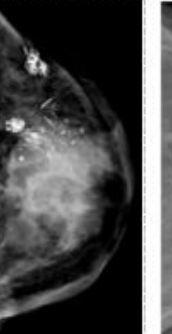
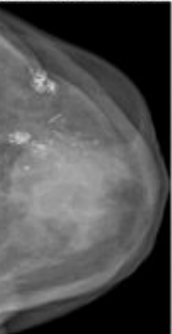
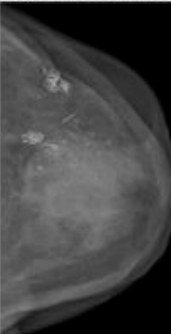
		Techniques								
		Amélioration de contraste			Egalisation d'histogramme			W/L	Filtrage	
		$T1_{G05}$	$T1_{G2}$	$T1_{G3}$	$T2_{HE}$	$T2_{CLAHE}$	$T2_{BBHE}$	$T3_{WL}$	$T4_N$	$T4_U$
Echantillons d' images	Image 1									
		(a.1)	(b.1)	(c.1)	(d.1)	(e.1)	(f.1)	(g.1)	(h.1)	(i.1)
Echantillons d' images	Image 2									
		(a.2)	(b.2)	(c.2)	(d.2)	(e.2)	(f.2)	(g.2)	(h.2)	(i.2)

Figure A.5 – Résultats des différents prétraitements appliqués sur les échantillons des images considérées

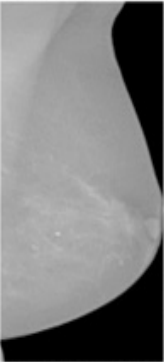
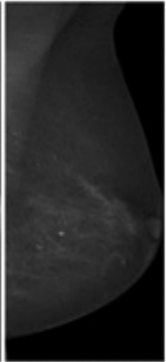
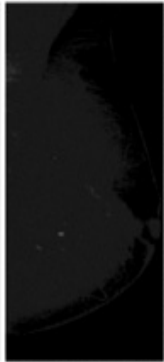
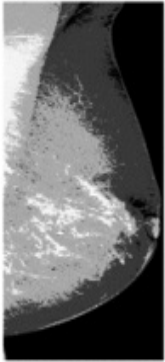
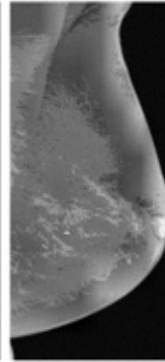
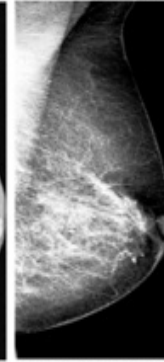

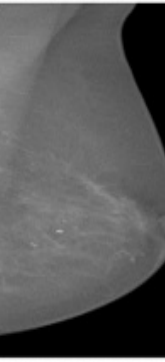
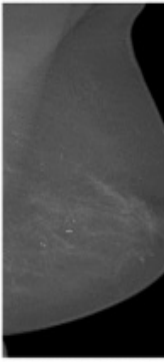
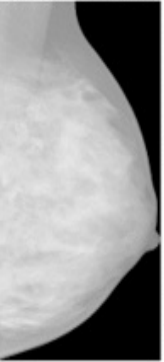
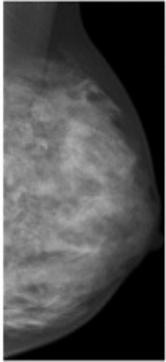
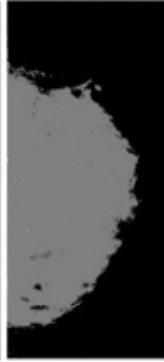
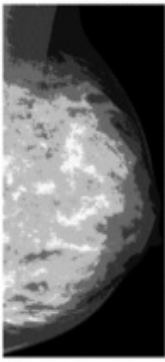
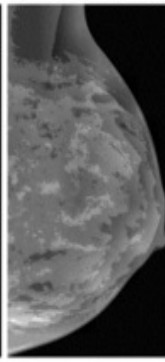
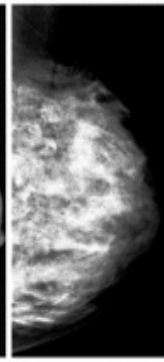
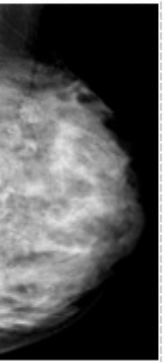
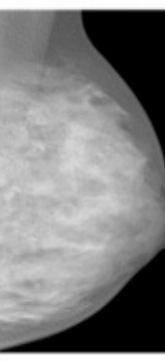
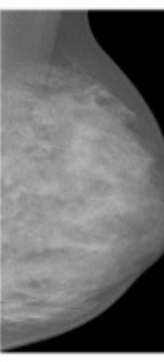
Echantillons d' images										
	Image 3									
		(a.3)	(b.3)	(c.3)	(d.3)	(e.3)	(f.3)	(g.3)	(h.3)	(i.3)
	Image 4									
		(a.4)	(b.4)	(c.4)	(d.4)	(e.4)	(f.4)	(g.4)	(h.4)	(i.4)

Figure A.5 – Résultats des différents prétraitements appliqués sur les échantillons des images considérées (cont.)





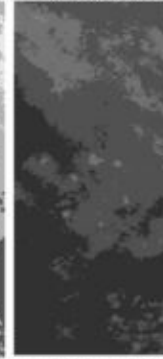
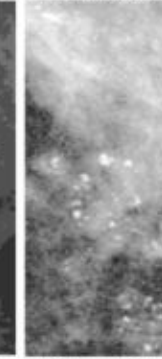
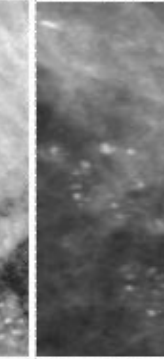
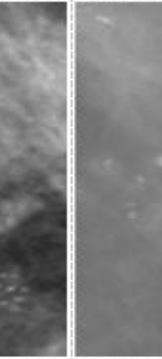
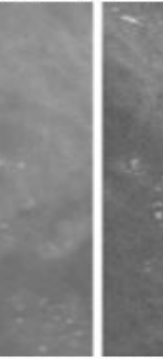





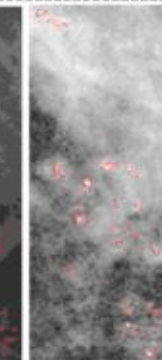
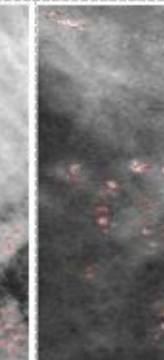
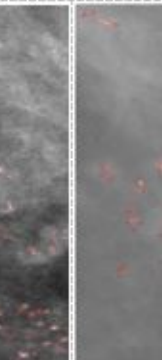
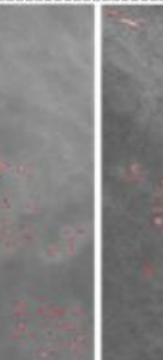
		<i>Amélioration de contraste</i>			<i>Egalisation d'histogramme</i>			<i>W/L</i>	<i>Filtrage</i>	
		<i>T1_{G03}</i>	<i>T1_{G1}</i>	<i>T1_{G3}</i>	<i>T2_{HE}</i>	<i>T2_{CLAHE}</i>	<i>T2_{BBHE}</i>	<i>T3_{WL}</i>	<i>T4_N</i>	<i>T4_U</i>
<i>ROI Image 1</i>	<i>ROI prétraitée</i>									
	<i>ROI prétraitée avec superposition GT</i>									
		(a.1.1)	(b.1.1)	(c.1.1)	(d.1.1)	(e.1.1)	(f.1.1)	(g.1.1)	(h.1.1)	(i.1.1)
		(a.1.2)	(b.1.2)	(c.1.2)	(d.1.2)	(e.1.2)	(f.1.2)	(g.1.2)	(h.1.2)	(i.1.2)

Figure A.6 – Résultats des différents prétraitements appliqués sur les ROIs des échantillons des images considérées

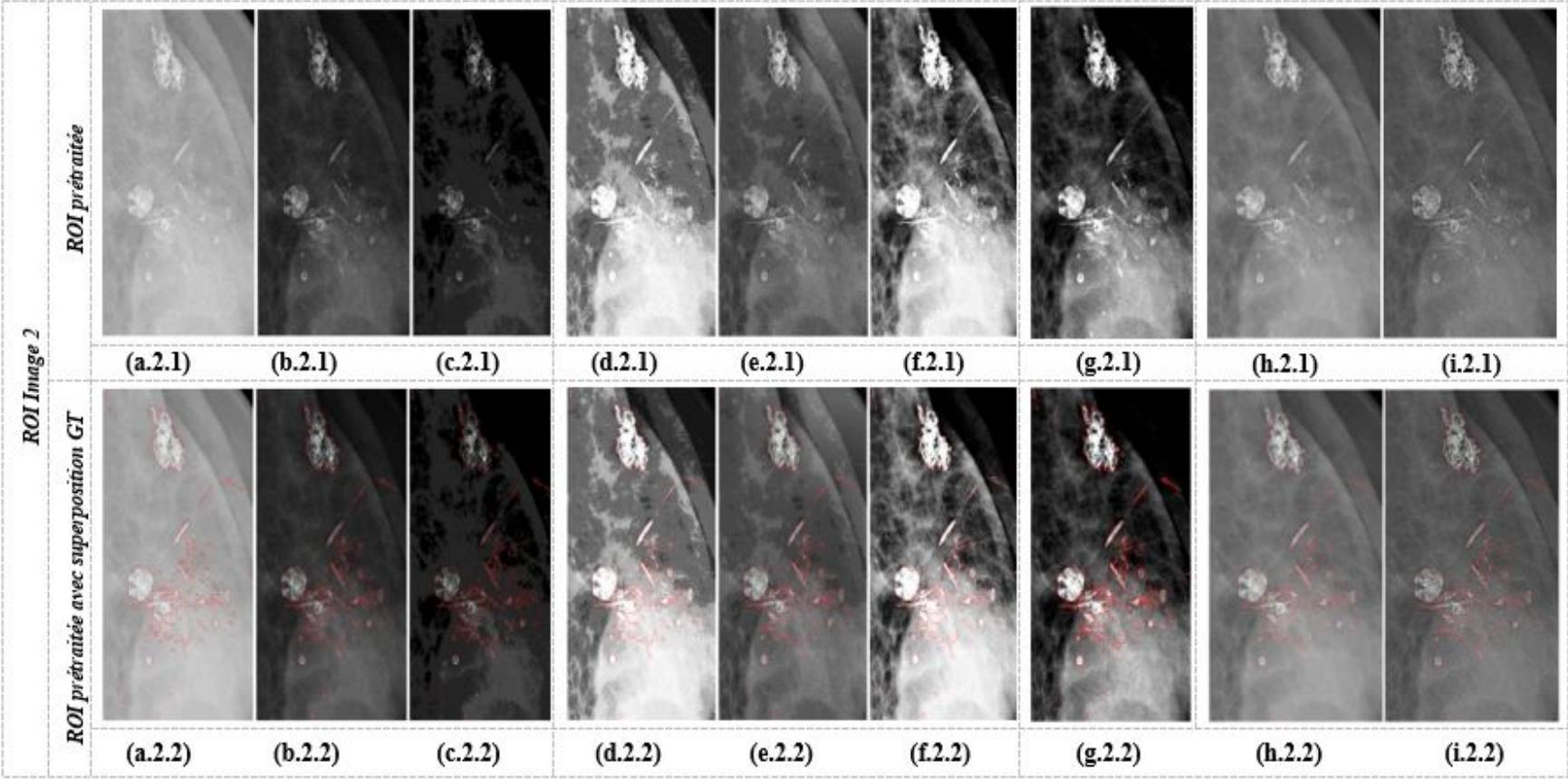


Figure A.6 – Résultats des différents prétraitements appliqués sur les ROIs des échantillons des images considérées (cont.)

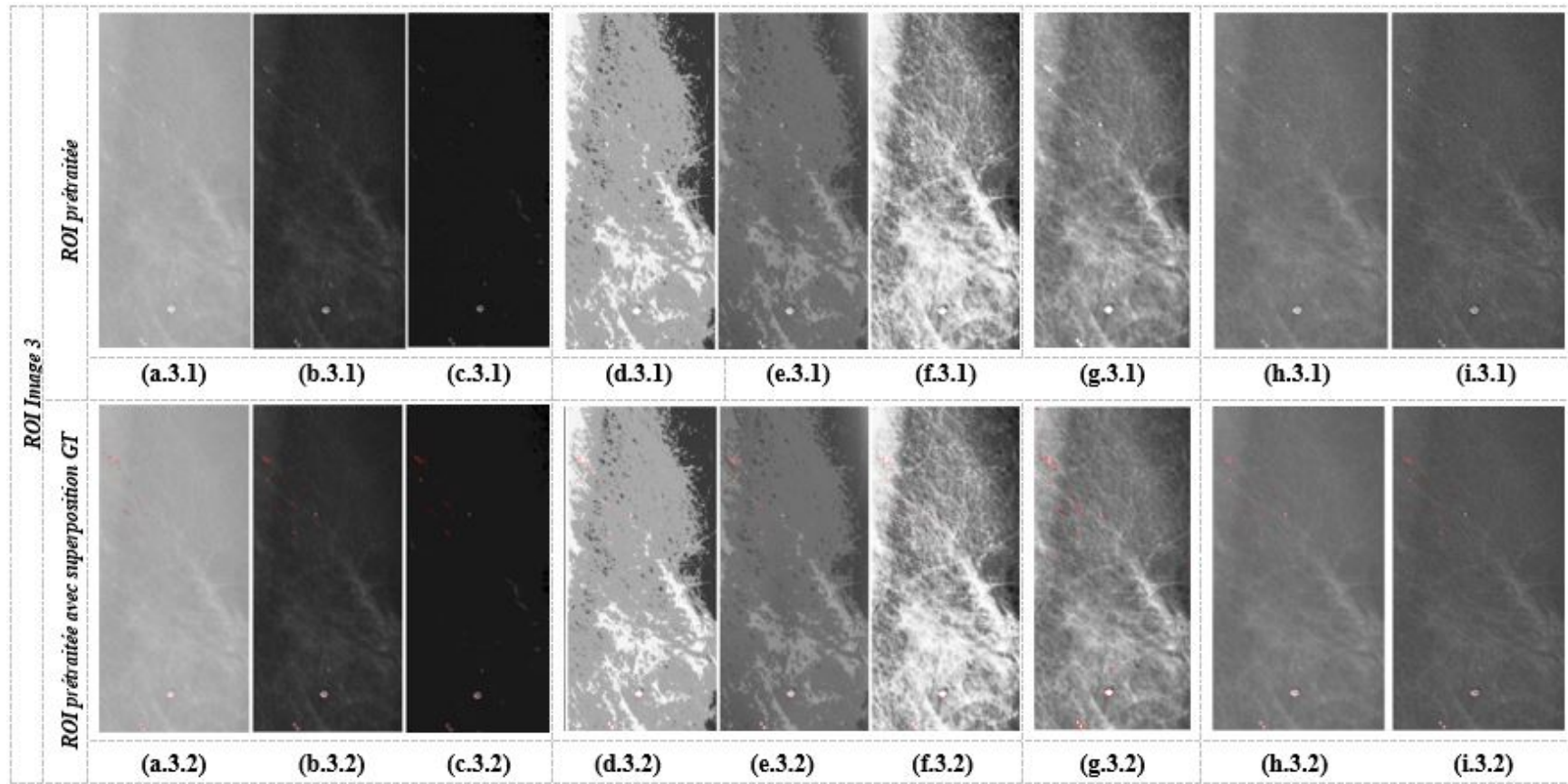


Figure A.6 – Résultats des différents prétraitements appliqués sur les ROIs des échantillons des images considérées (cont.)

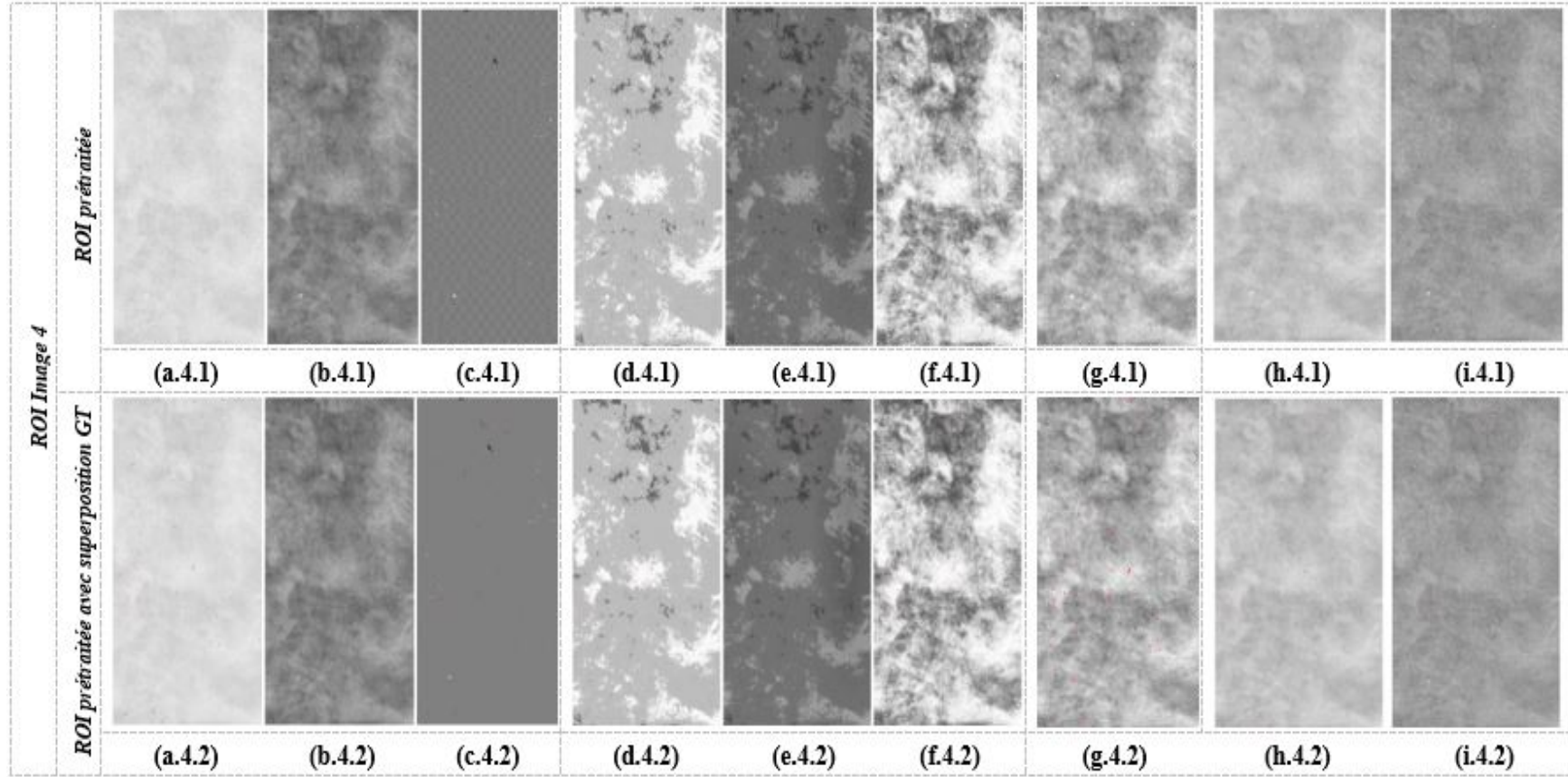


Figure A.6 – Résultats des différents prétraitements appliqués sur les ROIs des échantillons des images considérées (cont.)

L'analyse visuelle de ces résultats montre une supériorité des techniques d'amélioration de contraste $\mathbf{T1}_{G2}$ et Window/Level $\mathbf{T3}_{WL}$ par rapport aux autres techniques. Ce sont pratiquement les seules techniques qui ont validé les différents critères d'évaluations, sachant qu'à ce stade, ces critères font partie du premier protocole d'évaluation qui concerne la qualité d'apparition des microcalcifications sur l'image.

Les résultats obtenus par les autres techniques d'amélioration de contraste ($\mathbf{T1}_{G05}$ et $\mathbf{T1}_{G3}$) présentent affectent la structure anatomique projetée par l'image et peuvent créer de fausses alarmes des deux types FPs et FNs. La création des FPs (figure A.6.c.4) peut être expliqué par le fait que les microcalcifications, sur ces images, apparaissent superposées sur des tissus non denses (des structures linéaires ou un sein non dense). Les niveaux de gris des pixels de ces microcalcifications sont 'moyennes' sur l'intervalle des niveaux de gris dans l'image entière. La transformation de ces valeurs par une 'grande' valeur de $1/\text{Gamma}$ (ex. 3 avec $\mathbf{T1}_{G3}$) résulte de nouvelles valeurs d'intensités comparables et parfois supérieures à celles des tissus mammaires sur l'image (Figure A.5.c.4). Cette 'concurrence' de valeurs de niveaux de gris des pixels microcalcifications avec des pixels des tissus mammaires sera une source de confusion et de génération de fausses alarmes dans les prochains traitements à appliquer.

Les techniques d'égalisation d'histogramme sont toutes capables de préserver les microcalcifications tout en améliorant la qualité de leurs contours sur l'image. Cependant, elles échouent dans l'amélioration du contraste des microcalcifications présentes sur l'image (figure A.5.d.4) et elles augmentent souvent le taux de FPs (figure A.6.f.4). En effet, dans l'égalisation de l'histogramme classique, les pixels se dispersent uniformément sur l'intervalle des niveaux de gris. Toutefois les images produites à travers cette technique auront un contraste global un peu amélioré, la structure des tissus mammaires n'est plus visible. En outre, les images améliorées à travers la technique d'égalisation adaptative ($\mathbf{T2}_{CLAHE}$) sont dégradées en qualité visuelle (figure A.5.e.1). On peut clairement constater que le bruit est agrandi. Les tissus mammaires de basse fréquence (tissus adipeux) sont ignorés et les tissus de hautes fréquences sont mélangés avec les autres (tissu fibroglandulaire et tissu dense) (figure A.6.e.3 et figure A.6.e.4). Ce problème se réduit par la technique d'égalisation BBHE ($\mathbf{T2}_{BBHE}$) qui effectue une égalisation par partie. Ceci, peut être justifié par le fait que l'image mammographique est une image bipartite. Les microcalcifications apparaissent dans la deuxième partie droite (le sous intervalle avec des valeurs d'intensités élevées) avec un contraste faible. Ainsi, l'égalisation par partie de l'image permet de renforcer le contraste des microcalcifications par rapport aux autres types d'égalisations. En contrepartie, pour la catégorie d'images mammographiques dense et homogène, le problème persiste toujours (figure A.5.d.4 et figure A.6.d.4) de sorte que nous ne serons pas capables de distinguer les microcalcifications des autres tissus denses.

Les résultats obtenus par les techniques de filtrage montrent que ces dernières ne sont pas toujours capables d'améliorer la qualité d'apparition des microcalcifications. En effet, la technique à base de filtre de Nagao ($\mathbf{T4}_N$) affecte un caractère flou à l'image tout en conservant les contours. Ce caractère flou réduit le contraste entre les différents tissus mammaires de l'image. En revanche, l'effet négatif de cet adoucissement apparaît clairement dans le cas où les microcalcifications se superposent sur un tissu dense (figure A.5.h.4 et figure A.6.h.4).

En contrepartie, le filtre de masque flou améliore d'avantage la qualité des contours des tissus mammaires sur l'image, en particulier ceux des microcalcifications (figure A.5.h). Cependant, il renforce la dégradation des intensités des pixels voisins qui correspondent normalement à des pixels du même type de tissu. Cette dégradation atténue la qualité des images mammographiques et augmente le nombre de fausses alarmes de type FPs (figure A.5.h.1).

A.4 Synthèse et discussions

Le tableau A.1 synthétise les résultats de l'application des techniques de prétraitements en fonction des critères précédemment définis. Plusieurs constatations peuvent être déduite suite à l'analyse de ce tableau.

Pour les techniques d'amélioration du contraste, seule la technique de correction Gamma ($\mathbf{T1}_{G2}$) valide les différents critères d'évaluations qui justifient l'amélioration de la qualité d'apparition des microcalcifications. Une correction Gamma avec une valeur 3 ($\mathbf{T1}_{G3}$) offre une validation 'moyenne' des trois premiers critères, à savoir l'amélioration de la qualité des contours (\mathbf{C}_2) et du contraste (\mathbf{C}_3) et la préservation des microcalcifications (\mathbf{C}_1). En revanche, elle laisse créer des fausses alarmes de type FPs. Cet échec est dû au fait que les microcalcifications peuvent apparaître superposées sur des tissus non denses (des structures linéaires ou un sein non dense). Les intensités des pixels de ces microcalcifications se voient 'moyennes' sur l'intervalle des niveaux de gris de l'image. La transformation Gamma appliquée engendre des valeurs d'intensités comparables et parfois supérieures à celles de ces tissus par rapport à celles des pixels des microcalcifications (Figure A.6.c.4). Cette 'concurrence' de valeurs de niveaux de gris des pixels microcalcifications avec des pixels des tissus mammaires sera une source de confusion et de génération de fausses alarmes dans les prochains traitements à appliquer.

Les techniques d'égalisation d'histogramme sont toutes capables de préserver les mi-

Tableau A.1 – Évaluation des résultats des prétraitements appliqués sur les images mammographiques

Critères Techniques		C_1	C_2	C_3	C_4	C_5
Amélioration contraste	$T1_{G05}$	●	●	●	●	●
	$T1_{G2}$	●	●	●	●	●
	$T1_{G3}$	●	●	●	●	●
Egalisation histogramme	$T2_{HE}$	●	●	●	●	●
	$T2_{AHE}$	●	●	●	●	●
	$T2_{BBHE}$	●	●	●	●	●
W/L	$T3_{WL}$	●	●	●	●	●
Filtres spatiaux	$T4_N$	●	●	●	●	●
	$T4_U$	●	●	●	●	●
	$T4_{FA}$	●	●	●	●	●

crocalcifications (C_1) tout en améliorant la qualité de leurs contours sur l'image (C_2) puisqu'elles renforcent le contraste. Cependant, elles échouent dans la préservation des microcalcifications présentes sur l'image et qui apparaissent sur une partie dense. Ce qui implique souvent l'augmentation du taux de faux positifs (C_4).

La technique de fenêtrage 'Window/Level' est capable d'améliorer la qualité d'apparition des microcalcifications (C_2 , C_3) ainsi que la visualisation des différents tissus mammaires sur l'image (C_5).

Les techniques à base de filtrage associent un caractère flou aux images mammographiques. Ce caractère adoucit le contraste entre les tissus mammaires et nuit à la détection des détails fins, à savoir les microcalcifications, et des petites structures. Ce qui explique la réduction de taux des FPs (C_1) et l'augmentation du taux des FNs (C_4). Cette constatation revient aux principes de ces techniques qui cherchent principalement à supprimer le bruit à partir des images [147]. Par ailleurs, les techniques de réduction du bruit sont

généralement basées sur l'hypothèse que le bruit est additif et indépendant du signal [148]. Elles supposent une absence de corrélation entre les valeurs des pixels objets sur l'image et celles des pixels bruits. Cependant, les images mammographiques sont le résultat de l'exposition du sein mammaire à des radiations. Cette exposition produit des images avec un bruit de type quantum [57]. Ce dernier est un bruit non additif et dépend du signal. En effet, les valeurs des intensités des pixels bruits est sont corrélées avec l'intensité du rayonnement. Ainsi, traiter ce type de bruit est une tâche beaucoup plus complexe car sa nature doit être prise en compte dans ce processus de débruitage. En effet, une étude en 2007 [149] a montré que le bruit de type 'quantum' est le facteur de qualité le plus dominant des images mammographiques et qu'il exerce un effet plus important que la résolution spatiale pour les radiologues en interprétant les microcalcifications, ce qui nuit aux performances diagnostiques.

En guise de conclusion, les performances de la technique de correction Gamma ($\mathbf{T1}_{G2}$) et la technique Window/Level ($\mathbf{T3}_{WL}$) sont comparables en termes d'amélioration de la qualité d'apparition des microcalcifications (\mathbf{C}_2) ainsi que la conservation de la structure anatomique de l'image (\mathbf{C}_5). En revanche, d'après l'analyse des images résultats de ces deux techniques, nous constatons que $\mathbf{T3}_{WL}$ est capable d'améliorer, en plus de la qualité d'apparition des microcalcifications, la qualité d'apparition des structures mammaires sur l'image. En effet, elle montre un contraste élevé entre les différentes structures mammaires de l'image, qui est un avantage demandé par les radiologues du fait qu'il facilite la distinction des éléments denses intra-mammaires et donc améliore la qualité de diagnostic.

Pour ces raisons, nous allons adopter la technique Window/Level $\mathbf{T3}_{WL}$ comme la technique de prétraitement à appliquer sur les images mammographiques.

A.5 Evaluation relative au lien entre les superpixels et les microcalcifications

A.5.1 Schémas de génération des superpixels

Les superpixels SLIC (nommés SP1) dans ce travail de recherche consistent à des régions 'locales' générées à partir de l'image mammographique. Elles sont conçues dans l'objectif de détecter les groupements de pixels adjacents, dans ces dernières, qui peuvent correspondre à des microcalcifications.

Le deuxième repère d'évaluation des résultats de prétraitements concerne l'impact de ces derniers sur la génération des superpixels. On distingue trois schémas possibles, qui

seront évalués :

- Schéma 1 : Génération des superpixels SP1 à partir de l'image originale ;
- Schéma 2 : Génération des superpixels SP1 à partir de l'image prétraitée ;
- Schéma 3 : Application d'un postraitement sur les superpixels SP1 générées à partir de l'image originale.

A.5.2 Critères d'évaluation

Rappelons que, dans ce travail, l'idée de base derrière la génération des superpixels est de concevoir des régions 'locales' et 'homogènes' qui regroupent une voir plusieurs microcalcifications. A partir de ces régions locales, nous chercherons, dans une deuxième étape, les groupements de pixels adjacents qui peuvent correspondre à des microcalcifications.

L'évaluation des différents schémas que nous venons de citer dans la section précédente s'effectue ainsi selon deux types de critères. Certains critères sont liés aux superpixels eux-mêmes et d'autres sont liés aux liens entre les superpixels et les microcalcifications. Ils concernent :

- C'_1 : Diminution du nombre de superpixels ;
- C'_2 : Augmentation de l'homogénéité des superpixels ;
- C'_3 : Diminution du nombre de superpixels qui délimitent une seule microcalcification ;
- C'_4 : Diminution du nombre de superpixels qui délimitent une 'partie' d'une seule microcalcification ;
- C'_5 : Augmentation du nombre de superpixels qui regroupent plusieurs microcalcifications ;
- C'_6 : Diminution du nombre des superpixels FPs.

Rappelons qu'on appelle un superpixel faux positif tout superpixel de 'petite taille' mais qui ne délimite pas une 'vraie' microcalcification. Il correspond à une partie du tissu mammaire avec un contraste 'fort' par rapport à son voisinage.

A.5.3 Synthèse et discussions

Notre évaluation est une évaluation qualitative qui peut être renforcée par une évaluation quantitative. Les différents critères que nous venons de citer seront vérifiés par rapport aux résultats obtenus suite à l'application du premier schéma (schéma 1) et discutés dans les sections précédentes.

Pour des raisons de visibilité, on ne va pas montré des figures qui illustrent les résultats des superpixels générés par les différents schémas. En effet, le nombre de superpixels générés à partir d’une image est grand (de l’ordre de 3000 et même plus). Ils montrent des formes très variées et des tailles qui peuvent être très petites.

Le tableau suivant synthétise les performances des différents résultats obtenus par les deux schémas précédemment cités. Les mêmes symboles illustrés par la figure A.3 sont utilisés. En revanche, avec cette évaluation, ils permettent d’affirmer (\odot , \bullet ou \bullet) ou nier (\ominus , \bullet ou \bullet) l’amélioration des critères étudiés par rapport au premier schéma.

Tableau A.2 – Synthèse des résultats de générations des superpixels SP1 à partir des différents schémas

Schémas \ Critères	C'1	C'2	C'3	C'4	C'5	C'6
	schéma 1					
Schéma 2	\ominus	\bullet	\odot	\ominus	\bullet	\bullet
Schéma 3	\bullet	\bullet	\odot	\bullet	\ominus	\odot

L’analyse des résultats obtenus à la suite de l’application d’un certain nombre de pré et des post-traitements pour la génération des superpixels SP1 montre que la validation d’un ou de plusieurs critères est relative au principe du traitement appliqué. Le caractère multi-tissulaire et le type de la densité mammaire sont les principales causes de l’instabilité de la qualité des résultats de ces pré et pos-traitements.

Prenons l’exemple du critère d’homogénéité numérique (C'1). Ce dernier constitue le critère de base pour la génération des superpixels. La validation de ce critère reste toujours dépendante de la diversité tissulaire du sein (densité du sein et donc le contraste entre les différents types de tissus mammaires). Pour les images avec des contrastes variables (seins non denses), la validation de ce critère par un prétraitement peut engendrer une augmentation du nombre de superpixels FP.

Le postraitement sur lequel se base le troisième schéma, sert à réduire le nombre de superpixels de type FP. Ces techniques opèrent nécessairement par fusion des superpixels de petites tailles avec les superpixels voisins. Cette fusion est capable de réduire le nombre de superpixels FP (C'6), augmenter le nombre des superpixels qui regroupent plusieurs microcalcifications (C'5) ainsi que diminuer le nombre total de superpixels générés (C'1). En revanche, elle diminue leurs homogénéité (C'2) ainsi que le nombre de ceux qui délimitent une seule microcalcification (C'3) ou une 'partie' d’une même microcalcification

(C'₄).

Nous pouvons conclure, d'après cette analyse que nous avons fait, que la qualité des superpixels générés par les deux schémas dépend de la répartition des structures anatomiques sur l'image. En effet, un prétraitement s'il réussit à réduire le nombre de primitives visuelles de type faux positifs, il affecte nécessairement la structure anatomique projetée sur l'image. D'autre part, un postraitement, même s'il réduit le nombre de superpixels FPs, leur nombre reste important par rapport à ceux obtenues à partir de l'image originale.

A.6 Génération des superpixels : schéma adopté

À la suite de ces constatations, nous avons décidé d'utiliser le schéma suivant pour la génération des superpixels à partir d'une image mammographique (Figure A.7).

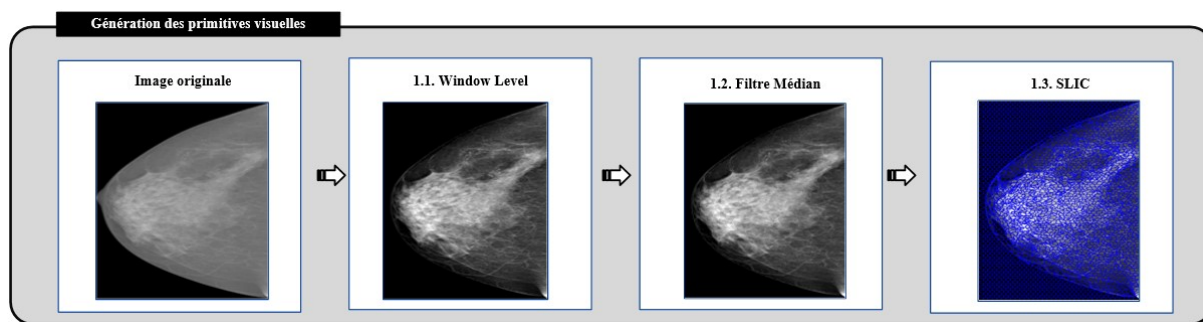


Figure A.7 – Schéma adopté pour la génération des superpixels

Le choix du filtre médian est dans l'objectif de 'légèrement' débruiter l'image obtenue par la technique Window/level' sans affecter les détails fins sur cette dernière (les structures mammaires). Le choix de la technique $T3_{WL}$ appliquée, dans un premier temps, sur l'image mammographique permet de :

- Transformer l'image en une image de 8 bits de profondeur ;
- Améliorer la distinction des différentes structures mammaires ainsi que l'apparition des microcalcifications.

Il est aussi intéressant de mentionner que nous avons, à ce niveau, préféré de garder le nombre élevé de superpixels tant qu'on n'affecte pas la structure anatomique présentée sur l'image.

Titre : Fusion collaborative de classifieurs : Application à la détection de pathologies en mammographie

Mots clés : Mammographie, Détection des microcalcifications (MCs), Morphologie mathématique, Croissance de régions conditionnelle, Combinaison collaborative, Propagation des connaissances

Résumé: Ce travail s'intéresse à l'interprétation d'images mammographiques pour la détection des MCs. Partant d'un prétraitement, nous proposons une première méthode basée sur la morphologie mathématique et l'indice de similarité structurelle SSIM. Elle se trouve capable de gérer la diversité des MCs en termes de formes et de contrastes. Dans l'idée d'étudier l'effet d'une analyse locale à base de connaissances *a priori*, nous avons proposé une deuxième méthode. Exploitant le concept de croissance de régions conditionnelle, elle part des maxima régionaux et intègre des connaissances *a priori* sous la forme de contraintes. Dans l'idée d'améliorer davantage la qualité des résultats obtenus, nous avons conçu une nouvelle approche de combinaison collaborative qui vise à minimiser le nombre de

détection faux négatif. Elle opère sur l'image entière divisée en superpixels utilisés pour identifier des zones suspectes et des objets candidats à partir des résultats des méthodes de détection (*détecteurs*). Les deux méthodes proposées ainsi que deux autres issues de la littérature sont utilisées comme détecteurs. Un processus de classification collaborative s'applique et permet de faire collaborer les détecteurs et de mettre en cause leurs décisions initiales. Ce raffinement met à jour les étiquettes des objets candidats à l'aide d'une analyse locale et contextuelle. D'une itération à une autre, ce processus réduit le désaccord entre les détecteurs et estime des termes de fiabilité locaux. Le résultat final est obtenu par combinaison conjonctive des nouvelles décisions obtenues.

Title : Collaborative combination of classifiers: Application to microcalcification detection in mammograms

Keywords : Mammography, Microcalcification (MC) detection, Mathematical morphology, Conditional region growing, collaborative combination, Knowledge propagation

Abstract : This thesis aims at detecting MCs. Starting from a pre-processing scheme, we propose a first method, based on mathematical morphology and the SSIM structural similarity index, that deals with MC diversities in terms of shapes and contrasts. In order to examine the efficiency of a local knowledge-based analysis, we have proposed a second detection method. Based on the conditional region growing concept, it starts from selected regional maxima and integrates *prior* knowledge as constraints. With the idea of further improving the quality of the obtained results, we proposed the third contribution that deals with a collaborative combination approach and aims at minimizing the false negative detections. It operates on the entire image divided into superpixels used to identify both suspicious

areas and candidate objects from the results of the used detection methods (called *detectors*). The two proposed methods as well as two other existing methods are used as detectors. A collaborative classification process is, then, applied and makes the different MC detectors collaborate to refine their initial results on the basis of the shared information and knowledge. The applied refinement consists of iteratively updating the candidate object labels of each detector following a local and contextual analyses based on *prior* knowledge. This process iteratively reduces the disagreement between different detectors and estimates local reliability terms for each superpixel. The final result is obtained by a conjunctive combination of the new detector decisions reached by the collaborative process.