



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

UNIVERSITE IBN KHALDOUN - TIARET

MEMOIRE

Présenté à :

FACULTÉ DES MATHÉMATIQUES ET DE L'INFORMATIQUE
DÉPARTEMENT D'INFORMATIQUE

Pour l'obtention du diplôme de :

MASTER

Spécialité : Génie Logiciel

Par :

LABRI Ahmed Younes
MEROUANE Imad Eddine

Sur le thème

Détection et Classification du Cancer du Sein dans les Images Mammographiques en utilisant des Caractéristiques Locales

Soutenu publiquement le 26/06/2022 à Tiaret devant le jury composé de :

Mr DAOUD Bachir

MAA Université Ibn Khaldoun Tiaret Président

Mr GAFOUR Yacine

MCB Université Ibn Khaldoun Tiaret Encadrant

Mr CHENINE Abdelkader

MAA Université Ibn Khaldoun Tiaret Examineur

2021-2022

Abstract

Recent studies have shown that breast cancer is still the number one cause of death among women. If detected early, 9 out of 10 cases can be cured, and there is currently no effective method of long-term reduction other than prevention. For this purpose, screening campaigns for asymptomatic women are organized on a regular basis, where hundreds of mammograms are analyzed. However, analysis of these images by the radiologist's visual perception tends to miss true positive readings, which requires the introduction of automated tools to support these large quantities of data. Detection aids are a very important tool that radiologists can use as a second reader for mammography to detect any signs of cancer early by looking for very specific signs of pathology, such as breast lumps.

This thesis tackles this problem by creating an automated system to assist in mass detection in mammography. In this work, we show how to improve performance to detect masses in mammograms by choosing a solution inspired by Shannon entropy, while exploiting the LBP descriptor that will be applied to the representation of an image. The result is a new descriptor called Entropy-Local Binary Pattern (H-LPB). We evaluated the effectiveness of the proposed approach on MIAS datasets. The experimental results show that our approach provides encouraging results in addressing this problem.

Keywords:

Mammography, Breast masses, Detection assistance system, LPB, MIAS, SVM.

Résumé

Des études récentes ont prouvé que le cancer du sein demeure la principale cause de décès chez les femmes. S'il est détecté tôt, 9 cas sur 10 peuvent être guéris, et il n'existe actuellement aucune méthode efficace de réduction à long terme autre que la prévention. Pour cela, des campagnes de dépistage des femmes asymptomatiques sont régulièrement organisées, où des centaines de mammographies sont analysées. Cependant, l'analyse de ces images par la perception visuelle du radiologue a tendance à manquer de véritables lectures positives, ce qui nécessite l'introduction d'outils automatisés pour prendre en charge ces grandes quantités de données. Les aides à la détection sont un outil très important que les radiologues peuvent utiliser comme deuxième lecteur pour la mammographie afin de détecter précocement toute trace de cancer en recherchant des signes très spécifiques de pathologie, tels que des nodules mammaires.

Ce mémoire aborde ce problème en construisant un système automatisé pour aider à détecter les masses dans les mammographies. Dans ce travail, nous montrons comment améliorer les performances pour détecter les masses dans les mammographies en choisissant une solution inspirée de l'entropie de Shannon, tout en exploitant le descripteur LBP qui sera appliqué à la représentation d'une image. Cette exploitation permet de donner naissance à un nouveau descripteur appelé Entropy-Local Binary Pattern (H-LPB). Nous avons évalué l'efficacité de l'approche proposée sur les ensembles de données MIAS. Les résultats expérimentaux montrent que notre approche donne des résultats encourageants pour résoudre cette problématique.

Mots-clés

Mammographie, Seins, Masses mammaires, Système d'aide à la détection, LPB, MIAS, SVM.

ملخص

أظهرت الدراسات الحديثة أن سرطان الثدي لا يزال السبب الأول للوفاة بين النساء. إذا تم اكتشافه مبكرًا، يمكن علاج 9 من كل 10 حالات، ولا توجد حاليًا طريقة فعالة للتخفيض طويل الأجل بخلاف الوقاية. ولهذا الغرض، تنظم بانتظام حملات لفحص النساء اللاتي لا تظهر عليهن أعراض، حيث يجري تحليل مئات التصوير الشعاعي للثدي. ومع ذلك، فإن تحليل هذه الصور من خلال الإدراك البصري لأخصائي الأشعة يميل إلى تفويت القراءات الإيجابية الحقيقية، مما يتطلب إدخال أدوات آلية لدعم هذه الكميات الكبيرة من البيانات. تعد أدوات الكشف أداة مهمة جدًا يمكن لأخصائي الأشعة استخدامها كقارئ ثانٍ للتصوير الشعاعي للثدي للكشف عن أي علامات للسرطان مبكرًا من خلال البحث عن علامات محددة جدًا لعلم الأمراض، مثل كتل الثدي.

تعالج هذه الأطروحة هذه المشكلة من خلال إنشاء نظام آلي للمساعدة في الكشف عن الكتلة في تصوير الثدي بالأشعة. في هذا العمل، نوضح كيفية تحسين الأداء لاكتشاف الكتل في تصوير الثدي بالأشعة السينية من خلال اختيار حل مستوحى من إنتروبي شانون، مع استغلال واصف LBP الذي سيتم تطبيقه على تمثيل الصورة. والنتيجة هي وصف جديد يسمى النمط الثنائي المحلي الانتروبي (H-LPB). قمنا بتقييم فعالية النهج المقترح على مجموعات بيانات MIAS. تظهر النتائج التجريبية أن نهجنا يوفر نتائج مشجعة في معالجة هذه المشكلة.

الكلمات المفتاحية

التصوير الشعاعي للثدي ، تكتلات الثدي ، نظام مساعدة الكشف ، LPB ، MIAS ، SVM.

Sommaire

Introduction généraleI

CHAPITRE I : Généralités sur les images

1. Introduction.....	1
2. Définition d'une image	1
3. Image numérique	1
4. Représentation des images	2
4.1 Les images vectorielles	2
4.2 Les images matricielles (ou bitmap)	2
4.3 Images en niveaux de gris.....	3
4.4 Images en couleur.....	3
5. Caractéristiques d'une image numérique.....	4
5.1 Pixel	4
5.2 Dimension	5
5.3 Résolution	5
5.4 Bruit	6
5.5 Histogramme	6
5.6 Modes colorimétriques	7
5.6.1 Couleurs RVB	8
5.6.2 Couleurs CMJN	9
5.6.3 Couleurs Lab	9
6. Conclusion	10

CHAPITRE II : Traitement des images dans l'imagerie mammaire

1. Introduction.....	11
2. Anatomie de sein	11

3. Le cancer du sein.....	13
4. Images mammographiques	14
5. Les masses.....	15
6. Diagnostic de cancer du sein	16
7. Détection et description des caractéristiques	16
7.1 <i>Caractéristiques locales</i>	16
7.2 <i>Caractéristiques globales</i>	16
7.3 <i>Descripteurs</i>	17
7.3.1 SIFT	17
7.3.2 SURF	17
7.3.3 HOG (Histogram of Oriented Gradients)	17
7.3.4 LBP (Local Binary Patterns).....	18
8. Classification	19
8.1 <i>Classification non supervisée</i>	19
8.1.1 K-means.....	19
8.1.2 Méthodes hiérarchiques	20
8.2 <i>Classification supervisée</i>	20
8.2.1 SVM	20
8.2.2 Random Forest	22
8.2.3 K-Nearest Neighbor	22
8.3 <i>Classification par renforcement</i>	23
8.3.1 Sarsa	23
8.3.2 Méthode de Monte Carlo (MC)	24
9. Conclusion	24

CHAPITRE III : Approche proposée pour classifier des images mammographiques

1. Introduction.....	25
2. Le modèle proposé pour classifier des images mammographies	25
a. <i>Base de données (MIAS)</i>	26
b. <i>Prétraitement des images mammographies</i>	28
c. <i>Shannon entropie pour l'extraction des caractéristiques</i>	29
3. H_LBP (Shannon entropie LBP).....	29

3.1	<i>Explication</i>	30
3.2	<i>Exemples</i>	32
4.	Expériences et Résultats obtenus	33
a.	<i>Confusion matrice</i>	34
b.	<i>Terminologie et dérivations d'une matrice de confusion</i>	34
5.	Discussion de résultats	36
6.	Conclusion	36
	Conclusion générale	

List de figures

Figure 1 Représentation d'image numérique.....	2
Figure 2 Différence entre une image matricielle et vectorielle	2
Figure 3 Image Monochrome	3
Figure 4 Image polychrome.....	4
Figure 5 Groupe de pixel formant la lettre A.....	5
Figure 6 Résolution d'une image	5
Figure 7 Image avec bruit.....	6
Figure 8 Image sans bruit.....	6
Figure 9 Une image de niveau de gris et son histogramme.....	7
Figure 10 Histogramme d'une image sombre (Gauche) et d'une image claire (Droit)	7
Figure 11 Espace de couleur RVB.....	8
Figure 12 Espace de couleur CMJN	9
Figure 13 La structure du sein	12
Figure 14 Les ganglions lymphatiques de sein.....	13
Figure 15 Le Mammographe	15
Figure 16 Les types de masses.....	15
Figure 17 Exemple de calcul du LBP d'un pixel donne : $1 + 16 + 32 + 64 + 128 = 241$	18
Figure 18 Exemple de calcul de LBP par les nombres binaire	19
Figure 19 Détermination d'un hyperplan via les vecteurs de supports.....	21
Figure 20 Linear separation et non_linear separation de SVM.....	22
Figure 21 Système proposé pour la classification des images mammographiques.....	26
Figure 22 Echantillons des images mammographiques de la base MIAS(mdb011.pgm et mdb058.pgm)	28
Figure 23 Schéma expliquant le fonctionnement de l'algorithme	30
Figure 24 Matrice d'image en niveaux de gris.....	30
Figure 25 Matrice de probabilité.....	30
Figure 26 Méthode de calcul de l'entropie des pixels.....	31
Figure 27 Matrice d'entropie	31
Figure 28 Matrice d'image après application de LBP	31
Figure 29 Lenna image , gauche (original) , milieu (niveaux de gris) , droite (H_LBP).....	32
Figure 30 geeksforgeeks image , gauche (original) , milieu (niveaux de gris) , droite (H_LBP)	32
Figure 31 MIAS images(mdb011.pgm et mdb058.pgm) prétraitées avec LBP.....	32
Figure 32 MIAS images(mdb011.pgm et mdb058.pgm) prétraitées avec H_LBP	33
Figure 33 Matrice de confusion.....	35

List des tables

Tableau 1 Critères de distinction entre tumeurs bénignes/ malignes.....	14
Tableau 2 comparaison de SVM linéaire et SVM non linéaire.....	21
Tableau 3 Paramètres des modules LBP et H_LBP.....	34
Tableau 4 Résultats des modules LBP et H_LBP	36

List des équations

Équation (1) LBP	Error! Bookmark not defined.
Équation (2) SARSA	Error! Bookmark not defined.
Équation (3) Shannon entropie	Error! Bookmark not defined.
Équation (4) Accuracy.....	Error! Bookmark not defined.
Équation (5) Precesion.....	Error! Bookmark not defined.
Équation (6) Recall	35
Équation (7) score F1	35

Introduction générale

Le cancer du sein est l'un des problèmes majeurs de santé publique. En fait, c'est un cancer bien connu qui menace la vie de la plupart des femmes. Environ une femme sur dix sera touchée par la maladie au cours de sa vie. Cependant, seul le traitement des tumeurs aux premiers stades de leur apparition peut réduire les taux de mortalité et améliorer les chances de guérison de ce type de cancer. Pour assurer une détection précoce de cette tumeur, les radiologues ont dû augmenter la fréquence des mammographies, en particulier pour les tranches d'âge les plus touchées. Par exemple, à l'échelle mondiale, de nombreux pays mènent des campagnes de dépistage systématique tous les deux ans. Cette approche s'est avérée très efficace et réduit la mortalité.

Selon les campagnes de test des maladies, deux à quatre mammographies sont effectuées par patiente à raison d'une ou deux mammographies par sein. Ceci a engendré une augmentation exponentielle du nombre de mammographies réalisées. Ainsi, la tâche d'interprétation est devenue difficile à gérer par les radiologues. L'interprétation est une tâche difficile et dépendante de l'expertise du radiologue. Par ailleurs, le taux de détection du cancer du sein s'améliore de 15% environ en utilisant une deuxième lecture.

Face à l'augmentation du nombre de mammographies au cours ces dernières décennies, différents travaux de recherche sont proposés pour détecter automatiquement les lésions mammaires à travers des systèmes assistés par Ordinateur. La vision par ordinateur est devenue un outil incontournable pour l'aide au diagnostic des experts médicaux. Elle se base sur des fondements mathématiques, le traitement des images, la reconnaissance de formes et sur la compression des images.

Notre travail consiste à présenter une aide au dépistage du cancer du sein en développant une approche de classification des images mammographies. Pour l'élaboration de notre approche, nous avons utilisé les H_LBP (Entropy Local Binary Pattern) afin d'extraire les vecteurs de caractéristique afin de l'exploité par le SVM pour faire classification des images mammographies.

Chapitre I

1. Introduction

L'image est une représentation d'une personne ou d'un objet par la peinture, le dessin, la photographie, cinéma, etc.... C'est aussi un ensemble structuré d'informations qui est affiché sur un écran. Cette présentation a une signification pour l'œil humain. Le traitement d'image peut être défini comme l'ensemble des méthodes et techniques opérant sur l'image afin d'extraire des informations les plus pertinentes ou tout simplement pour fournir une image plus perceptible à l'œil humain.

Dans ce chapitre nous présentons quelques notions de base du domaine de traitement d'image numériques tels que : la définition d'image, les types d'image, les caractéristiques d'image.

2. Définition d'une image

Les images sont faites à travers la peinture, le dessin, Photographie, film, etc... C'est aussi un ensemble structuré d'informations qui, après affichage sur l'écran, ont une signification pour l'œil humain.

Il peut être décrit comme une fonction analogique continue (x, y) , définie comme un domaine borné tel que x et y sont les coordonnées spatiales d'un point de l'image et I est Fonctions d'intensité lumineuse et de couleur. À cet égard, l'image est inexploitable par la machine, ce qui nécessite sa numérisation. ^[1]

3. Image numérique

Une image numérique est une image dont la surface est divisée en éléments de taille fixe appelées cellules ou pixels, chacune ayant pour caractéristique un niveau de gris ou Couleur. ^[1]

Numériser une image, c'est la convertir d'un état analogique en une image. Numérique est représenté par une matrice bidimensionnelle (x, y) de valeurs numériques, Comme le montre la figure 1 :

- x, y : Coordonnées cartésiennes du point image.
- $f(x, y)$: niveau d'intensité.

La valeur de chaque point représente la mesure d'intensité lumineuse perçue par le capteur.

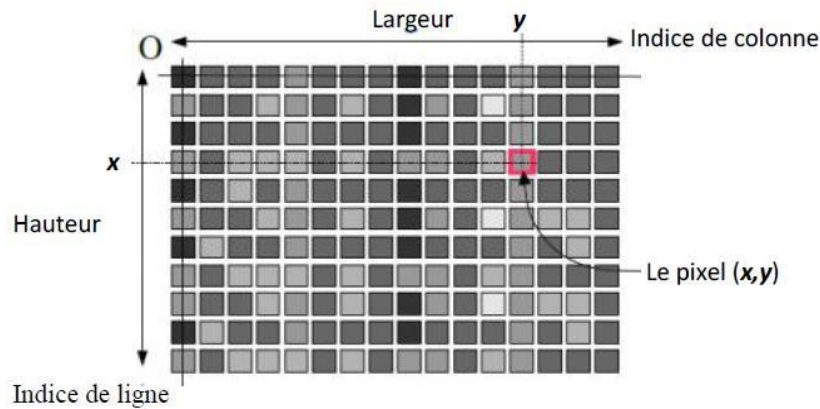


Figure 1 Représentation d'image numérique

4. Représentation des images

4.1 Les images vectorielles

Elle est créée à partir de formules mathématiques et elle est essentiellement constituée de formes géométriques. Elle peut être agrandie ou rétrécie à volonté sans perdre de sa qualité, car contrairement à l'image matricielle, elle ne contient pas de pixels. L'ordinateur réajustera simplement ses calculs pour produire l'image en question à la taille souhaitée.

Les professionnels (graphistes, illustrateurs ou concepteurs) réalisent la majorité de leurs visuels en vectoriel afin de pouvoir les modifier à volonté sans les altérer [19].

4.2 Les images matricielles (ou bitmap)

Elle est composée de petits points appelés « pixels » que l'on ne voit pas à l'œil nu. Ce type d'image est adapté à l'affichage sur écran mais peu adapté pour l'impression car bien souvent la résolution est faible. [19]

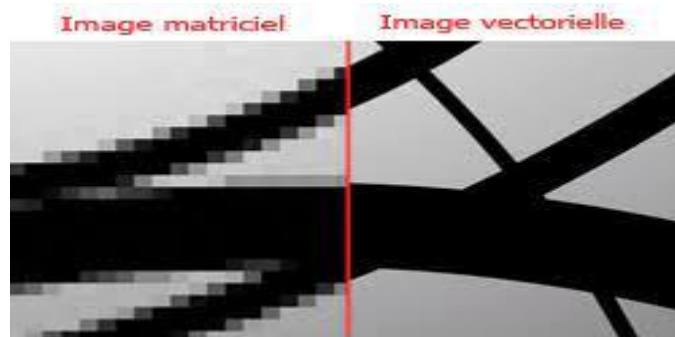


Figure 2 Différence entre une image matricielle et vectorielle

4.3 Images en niveaux de gris

Le niveau de gris est la valeur de l'intensité lumineuse en un point. La couleur du pixel peut prendre des valeurs allant du noir au blanc en passant par un nombre fini de niveaux intermédiaires. Donc pour représenter les images à niveaux de gris, on peut attribuer à chaque pixel de l'image une valeur correspondant à la quantité de lumière renvoyée. Cette valeur peut être comprise par exemple entre *0 et 255*. Chaque pixel n'est donc plus représenté par *1 bit*, mais par *1 octet*. Pour cela, il faut que le matériel soit utilisé pour afficher l'image, soit capable de produire les différents niveaux de gris correspondant.

Le nombre de niveaux de gris dépend du nombre de bits utilisés pour décrire la " couleur " de chaque pixel de l'image. Plus ce nombre est important, plus les niveaux possibles sont nombreux^[1].



Figure 3 Image Monochrome

4.4 Images en couleur

C'est une combinaison de trois couleurs dites primaires, le rouge, le vert et le bleu (RVB). Chaque couleur est codée comme une image en niveaux de gris, avec des valeurs Plage de 0 à 255. Pour $R=G=B=0$ nous aurons du noir pur et pour $R=G=B=255$ nous aurons du blanc pur. Ainsi, la représentation d'une image couleur se fait soit par une image dont la valeur en pixel est une combinaison linéaire des valeurs des trois composantes couleur, soit par trois images différentes représentant chacune une composante couleur. On distingue généralement deux types d'images^[2]:

- Images 24 bits
- Images dans la palette



Figure 4 Image polychrome

5. Caractéristiques d'une image numérique

Comme nous l'avons vu, l'image est un ensemble structuré d'informations parmi ses caractéristiques nous pouvons citer les paramètres suivants:

5.1 Pixel

L'image est définie comme une fonction bidimensionnelle $f(x, y)$, où x et y sont les coordonnées spatiales, les amplitudes de tous les points (x, y) correspondent à l'intensité ou en niveaux de gris. Lorsque le point (x, y) et la magnitude sont discrétisés, on parle d'une image numérique ou numérisée. Dans ce dernier cas, la fonction f est remplacée par une lettre, et le couple (x, y) est donné par paire (i, j) [3].

Une image numérique est constituée d'un ensemble de points appelés pixels. Les pixels sont approximés sous forme de rectangles, parfois carré. Leur taille peut être modifiée en ajustant l'écran ou la carte graphique. Pixels Représente ainsi le plus petit élément constitutif d'une image numérique. Tous ces Les pixels sont contenus dans un tableau à deux dimensions qui constitue l'image [3]. Tous ces les pixels sont contenus dans un tableau à deux dimensions qui constitue l'image [3]. Représente un groupe de pixel formant la lettre A.

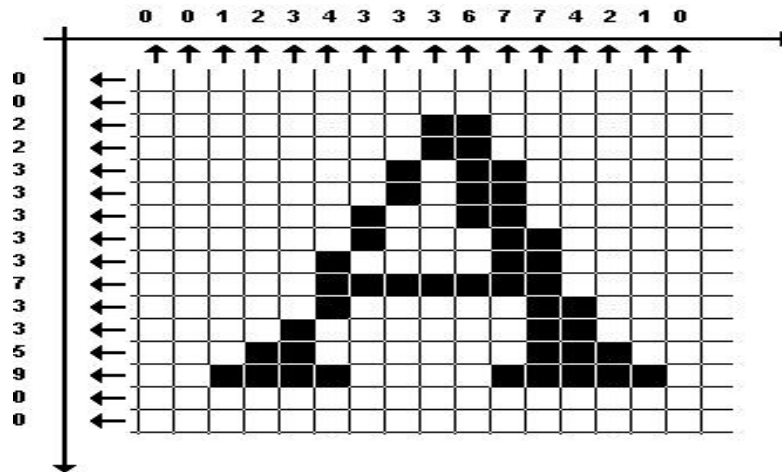


Figure 5 Groupe de pixel formant la lettre A

5.2 Dimension

C'est la taille de l'image. Cette dernière se présente sous forme de matrice dont les éléments sont des valeurs numériques représentatives des intensités lumineuses (pixels). Le nombre de lignes de cette matrice multiplié par le nombre de colonnes nous donne le nombre total de pixels dans une image^[4].

5.3 Résolution

La résolution d'une image correspond au niveau de détail qui va être représenté sur cette image. C'est le nombre de pixels par unité de longueur dans l'image à numériser. Elle est en dpi (dots per inch) ou en ppp (points par pouce).

Plus le nombre de pixels est élevé par unité de longueur de l'image à numériser, plus la quantité d'information qui décrit l'image est importante et plus la résolution est élevée^[18].

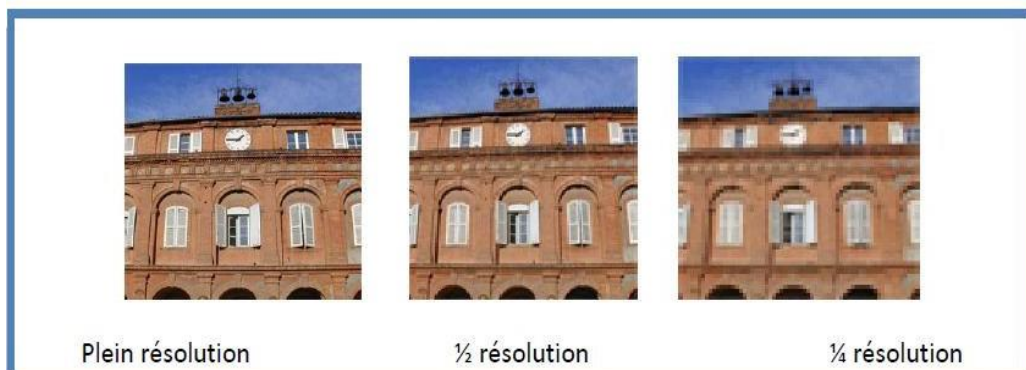


Figure 6 Résolution d'une image

5.4 Bruit

Le bruit en image ou vidéo numériques est un défaut parasite dégradant la qualité de l'image, Dans le domaine du traitement du signal et de l'image, le bruit correspond à un phénomène aléatoire qui se surajoute à l'image idéale. Probablement la meilleure approche pour comprendre le bruit est de réaliser que si l'on acquiert plusieurs fois l'image d'un même objet, immobile et inchangé, on n'observera pas exactement le même résultat. La différence est liée au bruit. Exemple, en lançant de la même manière plusieurs fois un dé, on n'obtient pas le même résultat ; c'est aléatoire. Comme tout phénomène aléatoire, le bruit est caractérisé par son écart type (racine carrée de la variance), qu'on note σ ^[4].



Figure 8 Image sans bruit



Figure 7 Image avec bruit

5.5 Histogramme

L'histogramme de niveaux de gris ou de couleurs d'une image est une fonction qui indique la fréquence à laquelle chaque niveau de niveaux de gris (couleur) apparaît dans l'image. Pour réduire les erreurs de quantification, comparez deux images obtenues sous un éclairage différent, ou mesurez certaines propriétés des images. Il fournit de nombreuses informations sur la distribution des niveaux de gris (couleurs) et voit entre quelles limites la distribution des niveaux de gris (couleurs) se situe dans le cas d'une image trop claire ou d'une image trop sombre ^[1].

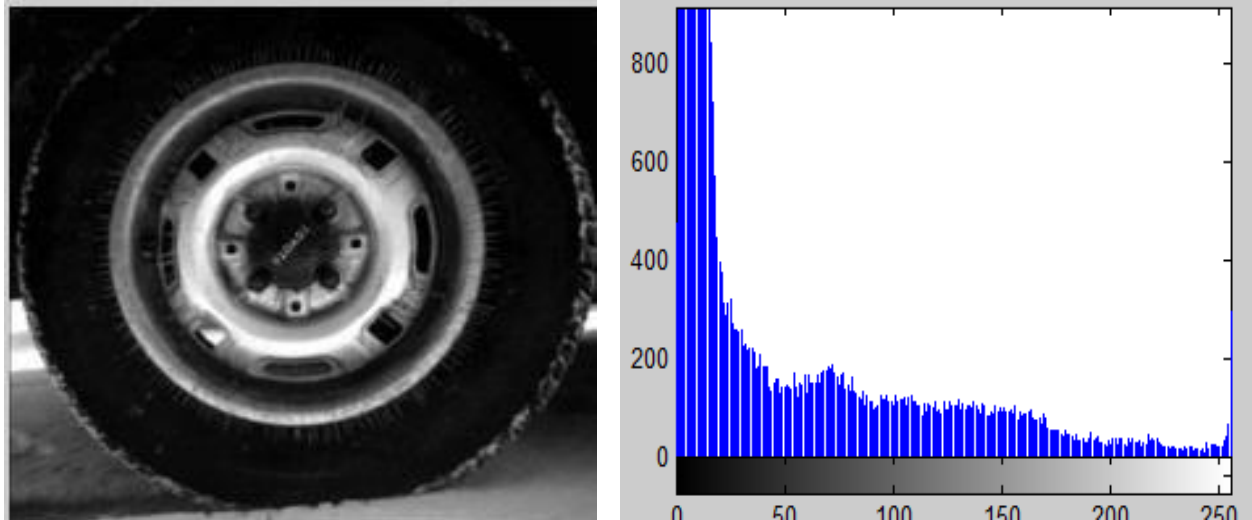


Figure 9 Une image de niveau de gris et son histogramme

La compilation d'une image se présente par son niveau de gris ou par son histogramme voir la figure 10

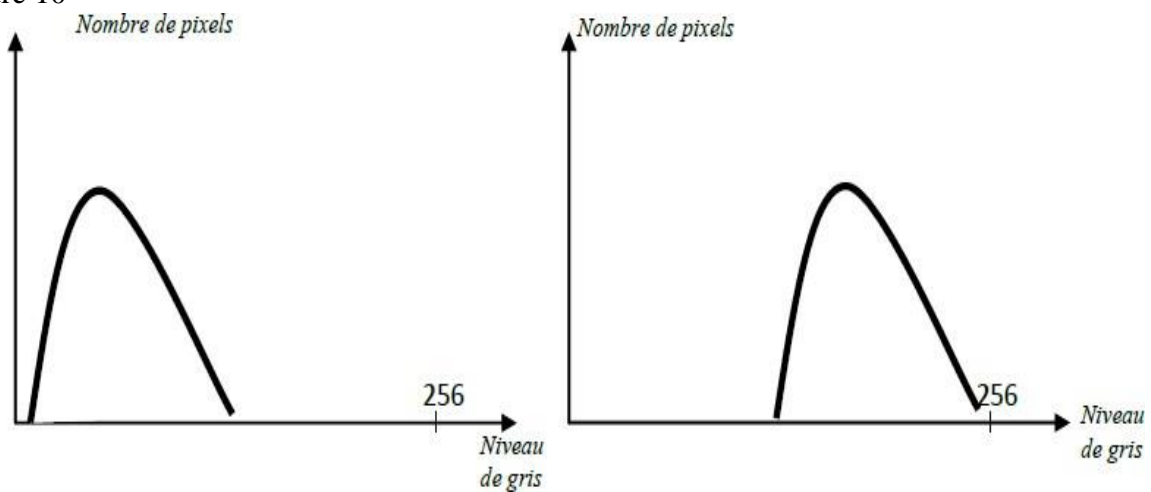


Figure 10 Histogramme d'une image sombre (Gauche) et d'une image claire (Droit)

5.6 Modes colorimétriques

Le mode colorimétrique ou le mode d'image déterminent comment combiner les couleurs en fonction du nombre de couches dans un modèle de couleur. Les différents modes colorimétriques correspondent aux différents niveaux de détail de la couleur et de taille du fichier. Par exemple, utilisez le mode couleurs CMJN pour les images dans une brochure en couleurs et utilisez le mode RVB pour le Web ou dans un courriel, afin de réduire la taille du fichier tout en conservant l'intégrité des couleurs.

5.6.1 Couleurs RVB

Ce mode dans Photoshop utilise le modèle RVB, attribuant une valeur d'intensité à chacun des pixels. Dans les images en 8 bits par couche, les valeurs d'intensité vont de 0 (noir) à 255 (blanc) pour chacune des composantes RVB (rouge, vert, bleu) d'une image en couleurs. Par exemple, un rouge vif aura une valeur R de 246, une valeur V de 20 et une valeur B de 50. Lorsque toutes les composantes ont une valeur égale, on obtient une nuance de gris neutre ; lorsqu'elles ont chacune une valeur de 255, on obtient un blanc pur. Une valeur égale à 0 pour chacune produit du noir. Les images RVB utilisent trois couleurs, ou couches, pour reproduire les couleurs à l'écran.

Dans les images en 8 bits par couche, les trois couches sont converties en 24 bits (8 bits x 3 couches) d'informations chromatiques par pixel. Pour les images 24 bits, les trois couches peuvent reproduire jusqu'à 16,7 millions de couleurs par pixel. Les images 48 bits (16 bits par couche) et 96 bits (32 bits par couche) peuvent reproduire encore davantage par pixel. Le modèle RVB n'est pas simplement le mode par défaut des nouvelles images Photoshop. Il est utilisé par les moniteurs d'ordinateur pour afficher les couleurs. Cela signifie que, dans les modes colorimétriques autres que RVB (par exemple, CMJN), Photoshop convertit l'image CMJN en image RVB pour l'afficher à l'écran. Bien que le modèle RVB soit un modèle de couleur standard, la gamme des couleurs exacte représentée peut varier selon l'application ou le périphérique d'affichage. Le mode Couleurs RVB de Photoshop varie suivant la configuration de l'espace de travail spécifiée dans la boîte de dialogue Couleurs^[20].

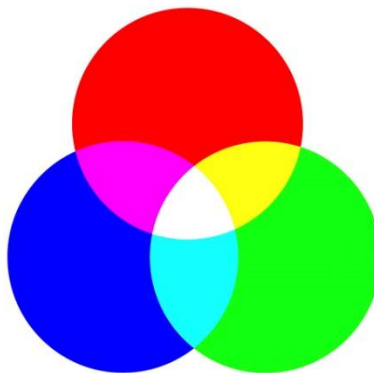


Figure 11 Espace de couleur RVB

5.6.2 Couleurs CMJN

Ce mode attribue à chaque pixel un pourcentage pour chacune des encres quadri. Aux couleurs les plus claires (tons clairs) sont attribués de faibles pourcentages de couleurs d'encre quadri, aux couleurs les plus foncées (tons foncés) de plus forts pourcentages. Par exemple, un rouge vif peut être obtenu avec 2 % de cyan, 93 % de magenta, 90 % de jaune et 0 % de noir.

Dans les images CMJN, le blanc pur est produit lorsque la valeur des quatre composantes est 0 %. Utilisez le mode CMJN lorsque vous préparez une image pour l'impression en quadrichromie. La conversion d'une image RVB en CMJN produit une séparation des couleurs. Si votre image de départ est en mode RVB, il vaut mieux la retoucher d'abord en RVB puis la convertir en CMJN, une fois que vous avez terminé. En mode RVB, vous pouvez utiliser les commandes Format d'épreuve pour simuler les effets d'une conversion CMJN, sans modifier les données image réelles. Vous pouvez également utiliser le mode CMJN pour travailler directement avec des images CMJN numérisées ou importées à partir de systèmes haut de gamme. Bien que le modèle CMJN soit un modèle de couleur standard, la gamme des couleurs exacte représentée peut varier selon la presse et les conditions d'impression. Le mode Couleurs CMJN de Photoshop varie suivant la configuration de l'espace de travail spécifiée dans la boîte de dialogue Couleurs^[20].



Figure 12 Espace de couleur CMJN

5.6.3 Couleurs Lab

Le modèle de couleur CIE L*a*b* (Lab) repose sur la perception de la couleur par l'œil humain. Les valeurs numériques du modèle Lab décrivent toutes les couleurs perceptibles à l'œil humain. Etant donné que le modèle Lab décrit l'aspect d'une couleur plutôt que la quantité de

colorant nécessaire pour la reproduire sur un périphérique donné (tel qu'un moniteur, une imprimante de bureau ou un appareil photo numérique), il est considéré comme indépendant du périphérique. Les systèmes de gestion des couleurs utilisent le modèle Lab comme référence pour transposer de manière fiable dans un autre espace colorimétrique une couleur d'un espace colorimétrique donné. La composante de luminance (L) du mode Couleurs Lab est comprise entre 0 et 100.

Dans le sélecteur de couleurs Adobe et dans le panneau Couleur, la composante a (axe vert-rouge) et la composante b (axe bleu-jaune) peuvent être comprises entre +127 et -128. Vous pouvez enregistrer les images Lab aux formats Photoshop EPS, Format de document volumineux (PSB), PDF, Photoshop Brut, TIFF, Photoshop DCS 1.0 et Photoshop DCS 2.0. Vous pouvez enregistrer les images Lab 48 bits (16 bits par couche) aux formats Photoshop, Format de document volumineux (PSB), Photoshop PDF, Photoshop Brut et TIFF^[20].

6. Conclusion

Dans ce chapitre, nous l'avons voulu à ce qu'il soit une brève introduction aux concepts liés au domaine du traitement d'images . Nous avons présenté l'image numérique d'une manière générale, et ses différentes représentations au point de vue de la machine, ainsi si que les caractéristiques que nous pouvons retirer de ces images.

Chapitre II

1. Introduction

L'approche médicale de l'imagerie mammaire est essentielle car elle permet d'acquérir les connaissances nécessaires à la conduite de la recherche. Il permet de diffuser l'expertise des radiologues de terrain sur des décennies de recherche. Afin de fournir des outils efficaces pour aider les radiologues à détecter le cancer du sein, les chercheurs scientifiques dans le domaine de l'imagerie médicale ont conçu des systèmes de détection et de classification automatiques des patientes et des non patientes. Nous examinerons ces systèmes dans ce chapitre, en nous concentrant sur la classification automatique des images mammographiques.

Dans ce chapitre, nous donnons d'abord le concept de cancer du sein. Nous discutons ensuite des outils d'imagerie médicale, notamment la mammographie, qui permet le dépistage et le diagnostic de ces cancers. Nous étudions ensuite l'anatomie du sein de manière non exhaustive. Et les méthodes utilisées dans la description et la classification d'images.

2. Anatomie de sein

La fonction biologique du sein est de produire du lait afin de nourrir un nouveau-né. Chaque sein contient une glande mammaire (elle-même composée de quinze à vingt compartiments séparés par du tissu graisseux) et du tissu de soutien qui contient des vaisseaux, des fibres et de la graisse. Chacun des compartiments de la glande mammaire est constitué de lobules et de canaux. Le rôle des lobules est de produire le lait en période d'allaitement. Les canaux transportent le lait vers le mamelon ^[21].

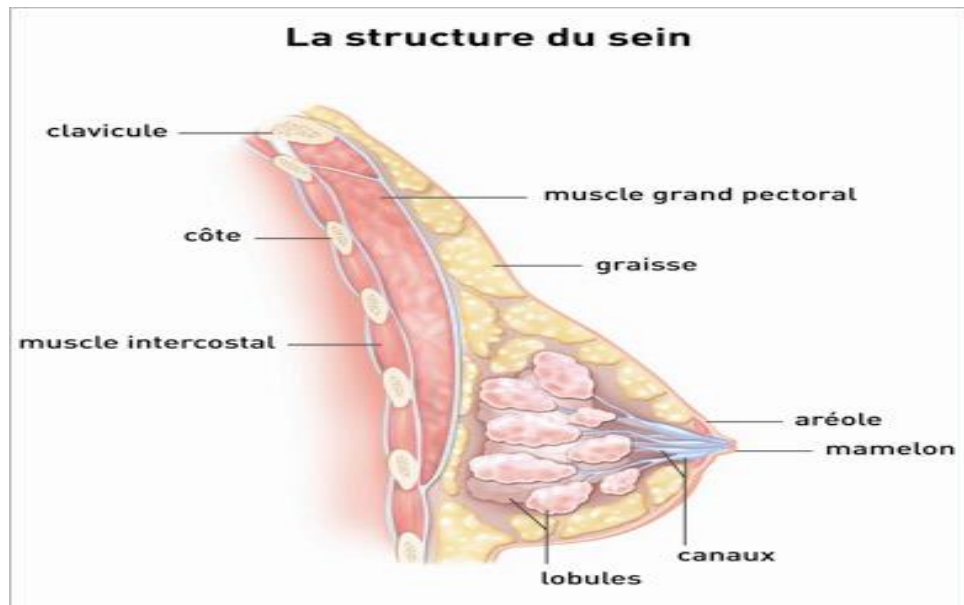


Figure 13 La structure du sein

La glande mammaire se développe et fonctionne sous l'influence des hormones sexuelles fabriquées par les ovaires. Ces hormones sont de deux types :

- Les œstrogènes, qui permettent notamment le développement des seins au moment de la puberté, stimulent les canaux en deuxième partie du cycle, joue un rôle important tout au long de la grossesse (assouplissement des tissus, augmentation du volume sanguin nécessaire à l'alimentation du bébé, pouvoir hydratant, etc.)
- La progestérone qui joue notamment un rôle dans la différenciation des cellules du sein et sur le cycle menstruel, en préparant par exemple l'utérus à une éventuelle grossesse (densification et développement de la vascularisation la muqueuse de l'utérus, etc.).

Le sein est parcouru de vaisseaux sanguins et de vaisseaux lymphatiques. Les ganglions et les vaisseaux lymphatiques composent le système lymphatique qui aide à combattre les infections.

Les ganglions lymphatiques du sein sont principalement situés :

- Au niveau de l'aisselle (creux axillaire), ce sont les ganglions axillaires
- Au-dessus de la clavicule : les ganglions sous-claviculaires
- Sous la clavicule : les ganglions sous-claviculaires (infra-claviculaires)

- À l'intérieur du thorax, autour du sternum : les ganglions mammaires internes [21].

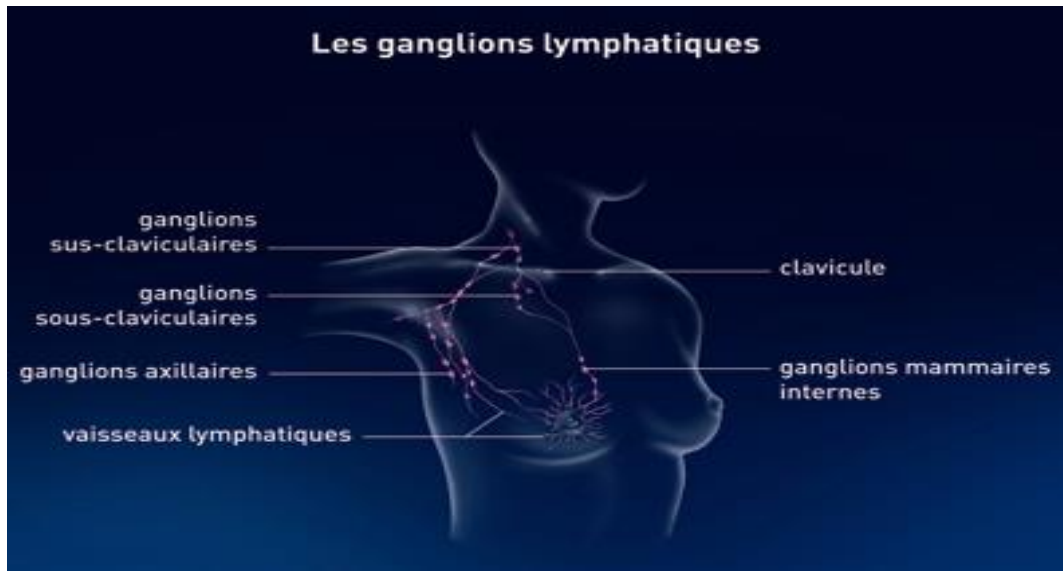


Figure 14 Les ganglions lymphatiques de sein

3. Le cancer du sein

Le cancer du sein est une tumeur maligne se développant à partir des cellules constituant la glande mammaire. Les cellules malignes se multiplient de manière anarchique et forment la tumeur. Lorsqu'un cancer du sein n'est pas traité, les cellules tumorales se propagent localement et envahissent les organes avoisinant (extension locale puis extension régionale). Elles peuvent également se propager par voie sanguine ou lymphatique (extension générale) pour atteindre les organes situés à distance (métastases) [22].

Les organes les plus fréquemment atteints par les métastases sont les os, les poumons, le foie et le cerveau.

On connaît deux catégories générales de tumeurs :

- Les tumeurs bénignes
- Les tumeurs malignes ou cancers

Tumeurs bénignes	Tumeurs malignes
Bien limitée	Mal limitée
Encapsulée	Non encapsulée
Histologiquement semblable au tissu d'origine	Plus ou moins semblable au tissu d'origine (Dédifférenciation, différenciation aberrante)
Cellules régulières	Cellules irrégulières (cellules cancéreuses)
Croissance lente	Croissance rapide
Refoulement sans destruction des tissus voisins	Envahissement des tissus voisins
Pas de récurrence locale après exérèse Complète	Récurrence possible après exérèse supposée Totale
Pas de métastase	Métastase(s)

Tableau 1 Critères de distinction entre tumeurs bénignes/ malignes

4. Images mammographiques

L'image mammographique est le résultat d'atténuation d'un faisceau de rayons X traversant les différents tissus mammaires. L'atténuation de ce faisceau dépend essentiellement de la composition des tissus traversés. En effet, la graisse est considérée comme une zone radio transparente vu qu'elle a une densité physique très légère. De ce fait, elle apparaît très sombre sur un cliché mammographique. En revanche, les zones radio opaques apparaissent claires et correspondent au tissu fibroglandulaire et au calcium qui est le composant essentiel des lésions mammaires. Pour les matières prédominantes dans le sein.

En rassemblant les informations concernant l'anatomie et la radio transparence, on peut confirmer que l'aspect général d'une mammographie est sombre alors que les zones contenant des microcalcifications ou des masses (composées de calcium) sont plus claires. L'appareil dédié à la réalisation d'une mammographie est le mammographe (figure 15). [5]

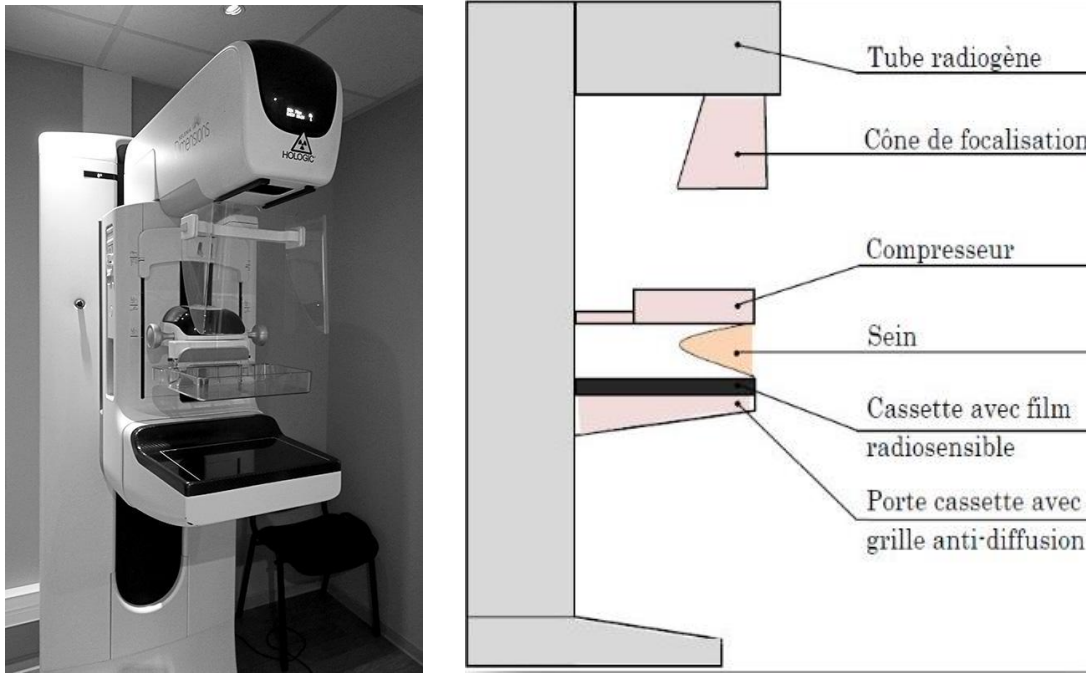


Figure 15 Le Mammographe

5. Les masses

Ce sont des lésions retrouvées en trois dimensions. Plus dense que le tissu Seins adjacents, qui apparaissent blancs sur une mammographie. Taille imprévisible Pas un trait malin à moins que les stéréotypes successifs montrent une augmentation régulière couper. On distingue traditionnellement cinq formes différentes de touffes rondes, ovales, Lobulaire, irrégulier ou structurellement déformé. La dernière forme comprend la lumière converge vers un point où la mise au point se rétrécit. La figure 4 montre un ensemble de lésions [6].

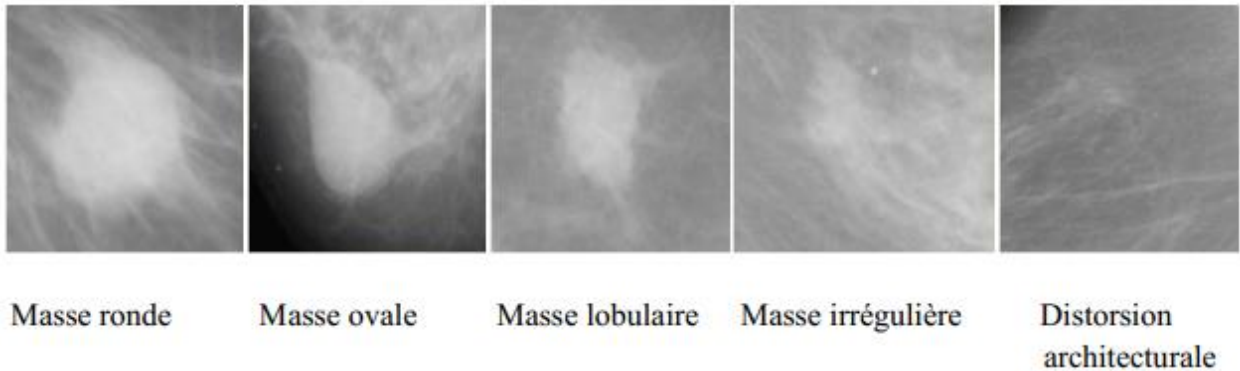


Figure 16 Les types de masses

6. Diagnostic de cancer du sein

Le diagnostic du cancer du sein repose sur le triplet diagnostique clinique mammographie-l'anatomopathologique.

- L'examen clinique : c'est un examen réalisé avant et après la mammographie pour mettre en évidence d'éventuelles anomalies des certaines régions et expliquer certains résultats, ainsi une corrélation entre clinique et imagerie est réalisée.
- Mammographie : est l'examen incontournable de l'exploration de la glande mammaire et peut être compléter par une échographie mammaire.
- L'anatomopathologie : Il donne des informations précises sur le type de cancer du sein. Bienfait à l'analyse au microscope des cellules et des tissus enlevés au niveau d'une anomalie du sein [7].

7. Détection et description des caractéristiques

L'extraction d'une caractéristique d'image peut être classée en deux catégories : les caractéristiques globales et les caractéristiques locales

7.1 Caractéristiques locales

Les caractéristiques locales visent à détecter les points d'intérêt (IP) dans une image et à les décrire par un ensemble de vecteurs, tels qu'un point, un bord ou un petit patch d'image. Récemment, de nombreux chercheurs se sont concentrés sur l'extraction des caractéristiques IP, de texture et de couleur d'une image. [17]

7.2 Caractéristiques globales

Les caractéristiques globales décrivent le contenu visuel de l'image entière par un seul vecteur. Ils représentent la texture, la couleur, les informations de forme qui sont les plus populaires pour la représentation d'image. [17]

7.3 Descripteurs

7.3.1 SIFT

Est l'algorithme proposé par Lowe aborde la rotation, la mise à l'échelle et la déformation affine des images en 2004, Changements d'angle de vue, bruit, changements d'éclairage, il y a aussi de forts robustesse. Pour l'algorithme SIFT, les étapes de détection peuvent être résumées comme suit :

- Sélection des pics dans l'espace d'échelles.
- Localisation des points d'intérêts (keypoints).
- Calcul de l'orientation.
- Description du point d'intérêt ^[9].

7.3.2 SURF

Les algorithmes SURF utilise des chemins légèrement différents détection de caractéristiques. SURF crée une "pile" qui échantillonne les couches supérieures de la pyramide, résultant en sont placés dans des images de même résolution. SURF utilise un filtre de boîte pour filtrer la pile en raison de l'utilisation d'une image intégrale d'une approximation en dérivée partielle du second ordre gaussien, c'est parce que l'image intégrale permet le calcul d'une boîte de filtre rectangulaire en presque le temps continu. Dans les points clés inégaux, le voisin le plus proche est défini comme vecteur descripteur de points clés avec distance euclidienne minimale invariant ^[10].

7.3.3 HOG (Histogram of Oriented Gradients)

L'histogramme des gradients orientés (HOG) est un descripteur de caractéristiques utilisé en vision par ordinateur et en traitement d'images à des fins de détection d'objets. La technique compte les occurrences d'orientation de gradient dans des parties localisées d'une image. Cette méthode est similaire à celle des histogrammes d'orientation des bords, des descripteurs de transformation de caractéristiques invariants à l'échelle et des contextes de forme, mais diffère en ce qu'elle est calculée sur une grille dense de cellules uniformément espacées et utilise une normalisation du contraste local qui se chevauche pour une précision améliorée^[23].

7.3.4 LBP (Local Binary Patterns)

LBP est l'un des descripteurs utilisés pour extraire les caractéristiques Local. Il est devenu l'un des descripteurs d'analyse de texture les plus importants. LBP est une Description de texture locale simple, efficace et robuste pour la variation de luminance. Cette La description de la texture se fait par un modèle d'histogramme basé sur la représentation binaire Calculé sur toute l'image ou une zone de l'image.

Le principe du descripteur LBP est d'étudier la relation entre le pixel central et ces pixels voisins qui permettent de construire des voisins binaires décrits autour du pixel central. Méthode basée sur LBP, il peut être utilisé pour différentes applications de vision par ordinateur, parmi lesquelles Ils montrent de meilleures performances dans la classification des textures^[11].

Le LBP est défini par :

$$LBP(x_c, y_c) = \sum_{m=0}^{p-1} s(i_n - i_c) \times 2^n \quad (1)$$

$$s(i_n - i_c) = \begin{cases} 1 & \text{si } i_n - i_c \geq 0 \\ 0 & \text{si } i_n - i_c < 0 \end{cases}$$

Avec p est le nombre de pixels voisins. Dans notre travail, nous considérons un voisinage de 3×3 d'où $p=8$ voisins. Nous obtenons donc, comme pour une image en niveaux de gris, une matrice contenant des valeurs des LBP comprises entre 0 et 255 pour chaque pixel. Un histogramme est calculé en se basant sur ces valeurs pour former le descripteur LBP ^[12].

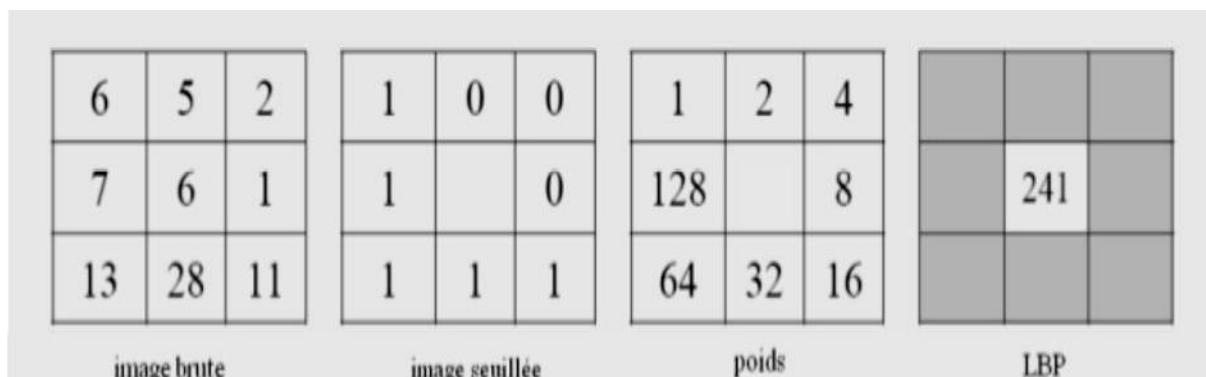


Figure 17 Exemple de calcul du LBP d'un pixel donne : $1 + 16 + 32 + 64 + 128 = 241$

Comme il peut se calculer par l'ensemble des nombres binaire obtenus dans l'image seuillée arrangée suivant le sens de l'aiguille horaire, ensuite les convertir en décimale.

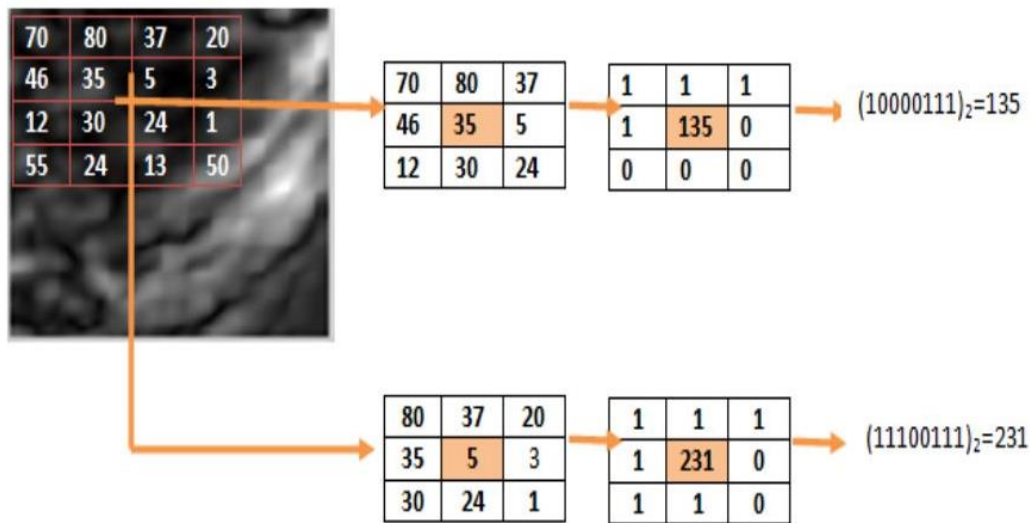


Figure 18 Exemple de calcul de LBP par les nombres binaire

8. Classification

La classification automatique consiste à regrouper divers objets (les individus) en sous-ensembles d'objets (les classes). Trois types de méthodes sont utilisés pour la classification, on trouve les méthodes supervisées, les méthodes non supervisées et les méthodes par renforcement.

8.1 Classification non supervisée

Lorsque le type d'objets dans une image est inconnu, en raison du manque d'informations sur la vérité terrain, il existe des algorithmes de classification qui créent un ensemble de pixels similaires. Dans le cas d'un classement non supervisé, il est souvent nécessaire d'indiquer le nombre de classes recherchées. Les textures similaires sont ensuite regroupées dans la même classe sans intervention préalable des connaissances, et sont estimées uniquement sur la base de la similarité entre les caractéristiques [13].

8.1.1 K-means

L'algorithme k-means mis au point par McQueen en 1967, un des plus simples algorithmes d'apprentissage non supervisé, il attribue chaque point dans un cluster dont

le centre est le plus proche. Le centre est la moyenne de tous les points du cluster, et ses coordonnées sont la moyenne arithmétique de chaque dimension séparée de tous les points du cluster, c'est-à-dire que chaque cluster est représenté par son centre de gravité [14].

8.1.2 Méthodes hiérarchiques

Le processus basique des méthodes hiérarchiques est un type de clustering, consiste à effectuer une suite de regroupements en Clusters de moins en moins fines en agrégeant à chaque étape les objets (simple élément) ou les groupes d'objets (un Cluster-partition-) les plus proches. Ce qui nous donne une arborescence de clusters. Cette approche utilise la mesure de similarité pour refléter l'homogénéité ou l'hétérogénéité des classes.[15]

8.2 Classification supervisée

La classification supervisée est une technique de classification automatique qui tente de générer automatiquement des règles à partir d'une base de données apprises contenant des échantillons pour regrouper les individus dans un certain nombre de classes.

L'utilisation de techniques supervisées nécessite une intervention de l'utilisateur et une première phase d'apprentissage où le système apprend les fonctionnalités associées à chaque classe de la base d'apprentissage.

Une fois cette étape terminée, ces méthodes supervisées sont capables de classer les individus dans la base de données en fonction des caractéristiques de texture qui leur sont associées.

Il existe plusieurs méthodes de classification supervisée telles que K plus proches voisins (K-Nearest Neighbor), SVM, Random Forest. [13]

8.2.1 SVM

Support Vector Machine (SVM) est un algorithme de classification. Les machines à vecteurs de support peuvent être utilisées pour lutter contre la discrimination. Autrement dit, il peut décider à quelle classe appartient l'échantillon. La technique SVM consiste à créer un hyperplan dans un espace de grande dimension afin de séparer les données des classes. Séparation Les données sont implémentées par l'hyperplan avec le point de données classifié le plus élevé classe la plus récente. Ces points de données sont appelés marges. Erreur La généralisation du classificateur SVM dépend

de la taille des bornes. La marge est la distance entre les frontières séparant les points les plus proches.^[13]

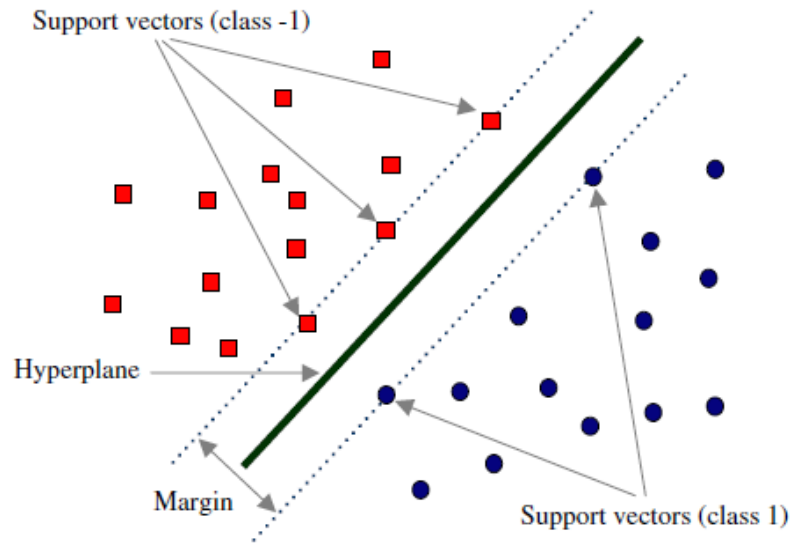


Figure 19 Détermination d'un hyperplan via les vecteurs de supports

- **Linear SVM vs Non-Linear SVM**

Lorsque nous pouvons facilement séparer les données avec un hyperplan en traçant une ligne droite, c'est Linear SVM. Lorsque nous ne pouvons pas séparer les données avec une ligne droite, nous utilisons le SVM non linéaire. En cela, nous avons les fonctions du noyau. Ils transforment des espaces non linéaires en espaces linéaires. Il transforme les données dans une autre dimension afin que les données puissent être classées ^[13].

SVM linéaire	SVM non linéaire
Il peut être facilement séparé avec une ligne linéaire.	Il ne peut pas être facilement séparé avec une ligne linéaire.
Les données sont classées à l'aide d'un hyperplan.	Nous utilisons des noyaux pour transformer des données non séparables en données séparables.
Les données peuvent être facilement classées en traçant une ligne droite.	Nous mappons les données dans un espace de grande dimension à classer.

Tableau 2 comparaison de SVM linéaire et SVM non linéaire

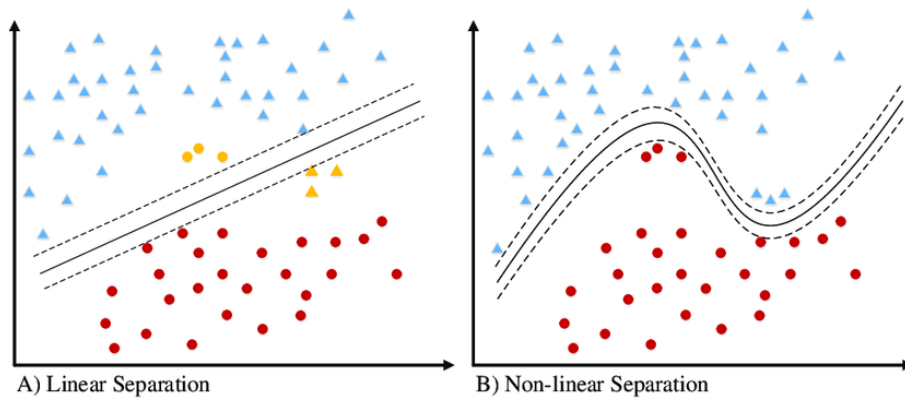


Figure 20 Linear separation et non_linear separation de SVM

8.2.2 Random Forest

Les forêts aléatoires ou bien *Random Forest* (RF) en anglais, est l'un des algorithmes utilisés pour la classification supervisée. Cet algorithme a été proposé au début des années 2000 et fait partie des algorithmes encore efficaces lorsqu'il est appliqué à de grandes quantités de données. Il combine plusieurs mécanismes insaisissables de manière complexe.

Les forêts aléatoires combinent des prédicteurs ou des estimateurs à partir d'arbres, ce qui donne ce que l'on appelle maintenant des arbres de décision. Plus généralement, ce sont des algorithmes établis dont le principe général est de construire des ensembles de prédicteurs puis d'agréger toutes leurs prédictions.

Dans la classification, l'agrégation inclut, par exemple, le vote à la majorité parmi les étiquettes de classe fournies par le prédicteur. Pour la régression, la cible de régression prédite pour l'échantillon d'entrée est calculée comme la cible de régression prédite moyenne pour les arbres de la forêt.^[11]

8.2.3 K-Nearest Neighbor

Le principe de l'algorithme K-Nearest Neighbor (K-NN) est de trouver un nombre prédéfini d'échantillons d'apprentissage avec la distance la plus proche d'une nouvelle entité (entrée), où chaque point de données est représenté par un ensemble de variables. Autrement dit, chaque point est tracé dans un espace de grande dimension, où chaque axe de l'espace correspond à une variable distincte. C'est-à-dire que k-plus proche voisin consiste à récupérer l'entité voisine la plus proche de la nouvelle entité et à l'affecter à une classe.

K-NN est sans doute le plus simple de tous les algorithmes de classification supervisée, K-NN continue de bien fonctionner à des échelles d'apprentissage plus grandes. Il lui suffit de choisir K, le nombre de voisins à prendre en compte lors de la classification. Le nombre K est généralement choisi comme la racine carrée de N, le nombre total de points dans l'ensemble de données d'apprentissage. (Donc si N est 400, alors K = 20).

Lorsque nous avons un nouveau point de données (test), nous voulons trouver les K voisins les plus proches qui sont les plus proches (c'est-à-dire les plus «similaires») en utilisant une distance. En général, la distance peut être n'importe quelle mesure métrique : la distance euclidienne standard est le choix le plus courant.^[11]

8.3 Classification par renforcement

Contrairement à l'apprentissage supervisé, seules les informations en entrée sont fournies au système. Celui-ci doit donc déterminer ses sorties en fonction des similarités détectées entre les différentes entrées, c'est-à-dire en fonction d'une règle auto - organisatrice. Le système est appelé donc à découvrir les régularités présentes dans ces configurations qui peuvent servir à les diviser en groupes de configurations semblables^[15].

8.3.1 Sarsa

SARSA est une méthode d'apprentissage par renforcement « on-policy » qui évalue et améliore la stratégie utilisée pour prendre les décisions d'action. La méthode « Sarsa » utilise la fonction valeur-état_action pour l'apprentissage. Ici, une expérience consiste non seulement au choix de l'action a depuis l'état s, mais aussi au choix de l'action a' qui va être choisie. Depuis l'état (en suivant la stratégie d'exploration adoptée). L'estimation des valeurs selon la stratégie courante s'écrit ^[24] :

$$Q_{t+1}(s, a) = Q_t(s, a) + \alpha \times [r_{t+1} + \gamma \times Q_t(s', a') - Q_t(s, a)]. \quad (2)$$

8.3.2 Méthode de Monte Carlo (MC)

Les méthodes de Monte-Carlo sont utilisées pour l'apprentissage des fonction $V^p(s)$ et $Q^p(s,a)$ et la découverte de la politique optimale. Les méthodes de Monte-Carlo utilisent seulement les résultats de l'interaction réelle ou simulée avec un environnement. Supposons que nous voulions évaluer $V^p(s)$, la valeur d'un état « s » selon la politique p. Pour estimer la fonction $V^p(s)$ il existe deux méthodes [24] :

- **« toute-visite »** : $V^\pi(s)$ est la moyenne des renforcements après toutes les visites à (s) dans un ensemble d'épisodes.
- **« Première visite »** : $V^\pi(s)$ est la moyenne des renforcements après les premières visites à (s) dans un ensemble d'épisodes.

9. Conclusion

Dans ce chapitre, nous avons présenté le contexte médical du cancer du sein et introduit quelques concepts généraux liés à l'anatomie du sein, aux facteurs de risque du cancer du sein et à l'apport de la mammographie.

Nous avons également parlé des caractéristiques locales et des caractéristiques globales, avec quelques descripteurs tels que SIFT, LBP ...etc.

De plus, nous avons introduit les types d'apprentissage automatique, supervisé, non supervisé et de renforcement. Avec une brève définition des méthodes et algorithmes les plus célèbres utilisés dans chaque type.

Chapitre III

1. Introduction

Les radiologues recherchent dans les mammographies, des microcalcifications et des masses, qui pour l'homme ordinaire ressemblent respectivement à des points et des taches blanches sur le cliché.

La tâche de détection des radiologues peut être facilitée par l'utilisation de détecteurs informatiques : les CAD's (Computer Aided Detection), qui sont des ensembles logiciel et matériel qui mettent en évidence des zones suspectes que l'expert aurait pu ne pas voir. Ces programmes sont fondés sur les algorithmes de l'intelligence artificielle.

L'objectif de ce travail est de proposer une nouvelle approche pour décrire les images mammographie. L'entropie de Shannon est utilisée afin que nos algorithmes d'apprentissage automatique puissent détecter plus facilement le cancer .

2. Le modèle proposé pour classifier des images mammographies

Notre modèle (figure 21) est composé essentiellement des étapes montrées dans ci-dessous :

1. Au début nous avons la base de données MIAS qui comporte les images à utiliser.
2. Prétraiter les images de mammographie avec la méthode Otsu pour supprimer le bruit.
3. Utilisation de l'opérateur LBP pour l'extraction des caractéristiques
4. Construction de la base d'apprentissage pour avoir le modèle de décision.
5. Application des SVM

Processus d'apprentissage automatique

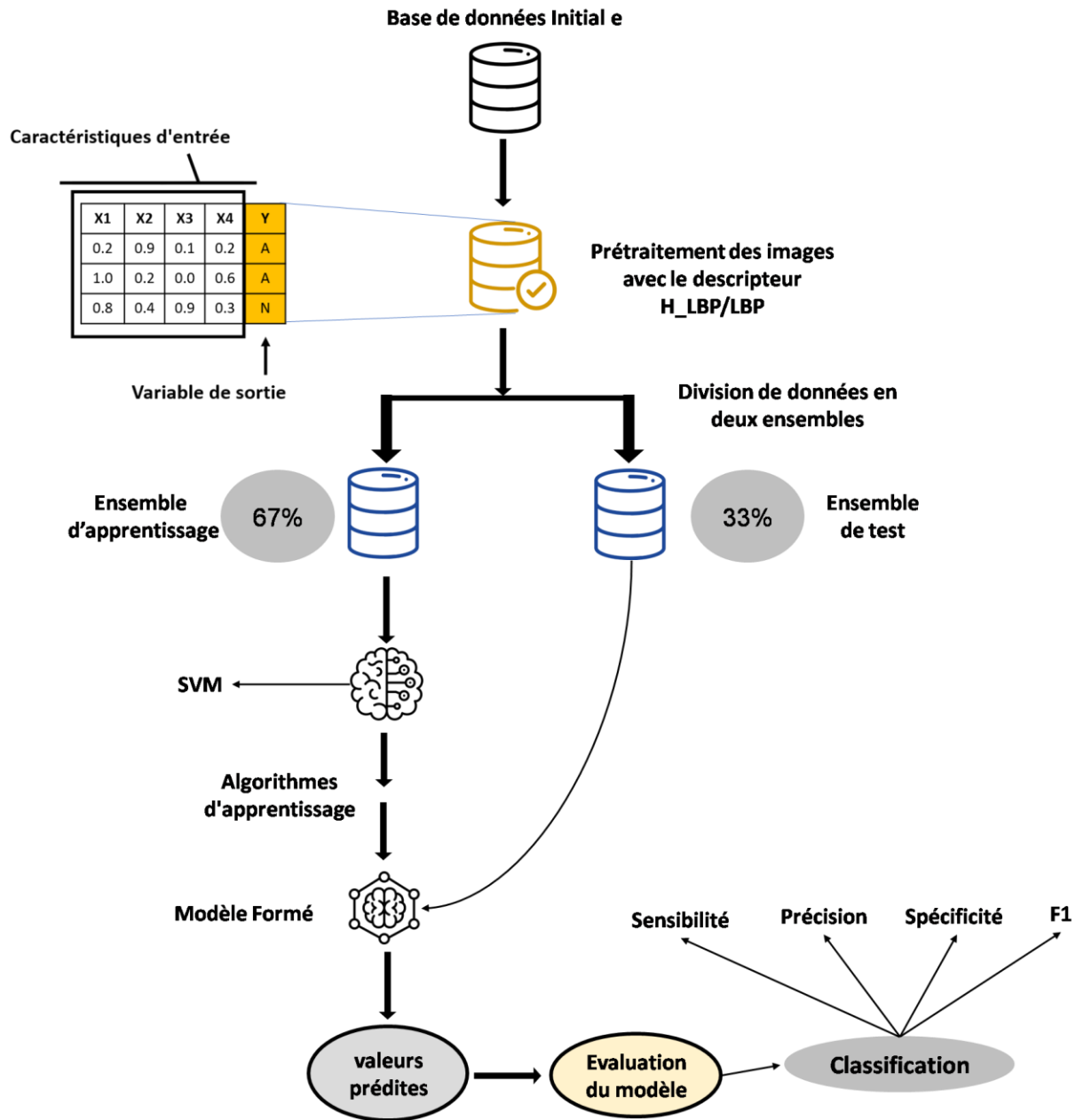


Figure 21 Système proposé pour la classification des images mammographiques

a. Base de données (MIAS)

La Mammographic Image Analysis Society (MIAS) [27] est une organisation de groupes de recherche britanniques intéressés par la compréhension des mammographies et a généré une base de données de mammographies numériques. Des films tirés du National Breast Screening Program

du Royaume-Uni ont été numérisés à 50 microns de bord de pixel avec un micro densitomètre à balayage Joyce-Loebl, un appareil linéaire dans la plage de densité optique de 0 à 3,2 et représentant chaque pixel avec un mot de 8 bits. La base de données contient 322 films numérisés et est disponible sur bande de 2,3 Go 8 mm (ExaByte). Il comprend également les marquages "vérité" du radiologue sur les emplacements de toutes les anomalies qui peuvent être présentes. La base de données a été réduite à un bord de pixel de 200 microns et rembourrée/coupée de sorte que toutes les images soient en 1024x1024. Les images mammographiques sont disponibles via le Pilot European Image Processing Archive (PEIPA) à l'Université d'Essex .

Chaque image est stockée au format brut: chaque numéro stocké dans le fichier correspond à la valeur du niveau de gris (de 0 à 255) du pixel correspondant dans l'image.

La base de données contient deux prises de vue du sein gauche et droit de 161 patients ; elle est constituée de 322 images classées en trois types :

- Type normal : il y a 208 cas.
- Type bénin : il y a 63 cas.
- Type malin : il y a 51 cas.

Un fichier joint à la base de données décrit en détails les images. Il est organisé en sept colonnes de la manière suivante :

1ère colonne : le numéro de référence de l'image dans la base de données MIAS

2ème colonne: caractérise le type des tissus du sein. Il y a trois types :

- F : Graisseux (Fatty)
- G : Glandes Graisseux (Fatty-glandular)
- D : Dense (Dense-glandular)

3ème colonne: Type de pathologie

- CALC : Calcification (Calcification)
- CIRC : Well-defined/circumscribed masses
- SPIC: Spiculated masses
- MISC: Other, ill-defined masses
- ARCH: Architectural distortion
- ASYM : Asymmetry

- NORM : Normal

4ème colonne: Sévérité de pathologie

- B Benin
- M Malin

5ème et 6ème colonne: les coordonnées (x,y) du centre de la pathologie

7ème colonne: Rayon approximatif (en Pixel) d'un cercle enfermant l'anomalie.

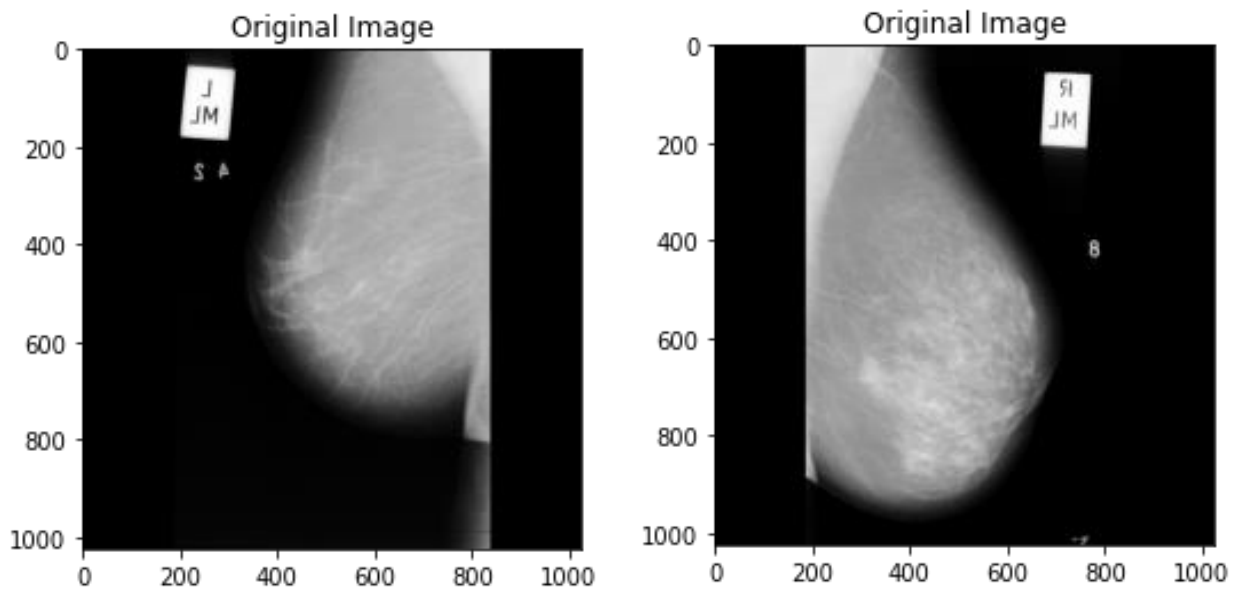


Figure 22 Echantillons des images mammographiques de la base MIAS(mdb011.pgm et mdb058.pgm)

b. Prétraitement des images mammographies

L'exploitation des clichés mammographiques nécessite souvent un prétraitement ayant pour objectif l'amélioration de la qualité de l'image considérée.

Dans la littérature parle souvent sur les différentes techniques de prétraitement d'images mammographiques, parmi eux nous trouvons la méthode OTSU.

La méthode OTSU (OTSU) est un algorithme de segmentation d'image à seuil de binaire global adaptatif, il est proposé par les chercheurs japonais OTSU en 1979. Cet algorithme prend la

variance interclasse maximale entre l'arrière-plan et l'image cible comme règle de sélection de seuil.

c. Shannon entropie pour l'extraction des caractéristiques

Shannon ^[16].a introduit le concept statistique d'entropie dans la théorie de la communication et de la transmission de l'information (Shannon 1948), mesurant la valeur moyenne de l'information manquante dans une source aléatoire. Étendu à l'entropie, ce qui donne des bornes sur les théorèmes d'encodage et de compression. Pour une variable aléatoire x à valeurs dans un ensemble fini X , l'entropie de Shannon est définie par :

$$H(X) = - \sum_{x \in X} P(x) \log_2 P(x) \quad (3)$$

3. H_LBP (Shannon entropie LBP)

Dans ce chapitre, nous montrons comment améliorer les performances de classification en choisissant une solution inspirée des Shannon entropie, tout en adaptant le descripteur LBP à l'entropie.

Le descripteur que nous avons proposé est appelée H_LBP, notre algorithme calcule l'entropie de chaque pixel de l'image (pixel central), en fonction des pixels entourant le pixel central (pixels voisins), et lui applique le LBP d'origine. Le principe de l'algorithme passe par quatre étapes :

i. Convertir l'image en matrice

Dans la première étape, nous prenons notre image et la convertissons en une matrice de la même taille que notre image.

ii. Calculer la probabilité de chaque pixel

Dans la deuxième étape, calculez la probabilité de chaque pixel et remplacez la valeur d'origine par sa probabilité.

iii. Calcul l'entropie de chaque pixel

Dans la troisième étape, nous calculons l'entropie (entropie de Shannon) de chaque pixel, en fonction de la matrice de probabilité que nous avons trouvée dans la deuxième étape. On peut avoir 8 pixels voisins dans le cas (3x3), et 24 pixels voisins dans le cas (5x5).

iv. Application de LBP à la matrice

Dans la dernière étape, nous prenons notre matrice d'entropie que nous avons trouvée dans la troisième étape, et nous y appliquons LBP.

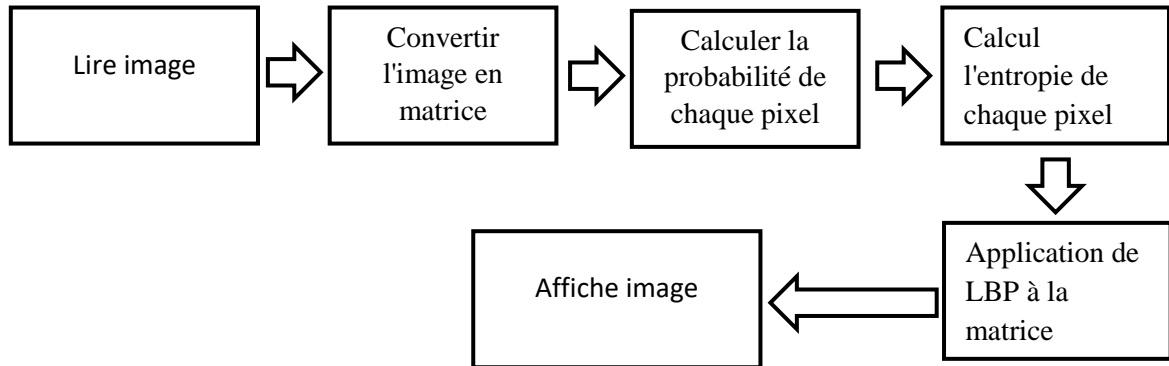


Figure 23 Schéma expliquant le fonctionnement de l'algorithme

3.1 Explication

- Nous avons par exemple, après l'avoir l'image et passer en niveaux de gris , nous avons obtenu cette matrice

255	255	255	14	255
68	14	100	255	99
99	255	60	14	68
255	60	200	255	160
60	68	14	200	60

Figure 24 Matrice d'image en niveaux de gris

- Après avoir calculé la probabilité de chaque pixel de l'image, nous avons cette nouvelle matrice ci-dessous

0.32	0.32	0.32	0.16	0.32
0.12	0.16	0.04	0.32	0.04
0.04	0.32	0.16	0.16	0.12
0.32	0.16	0.08	0.32	0.04
0.16	0.12	0.16	0.08	0.16

Figure 25 Matrice de probabilité

- À partir de la matrice de probabilité, nous créons une matrice d'entropie comme celle-ci

	Pv1	Pv2	Pv3	
	Pv8	Pc	Pv4	
	Pv7	Pv6	Pv5	

$$H(Pc) = -\sum Pv_i * \log (Pv_i)$$

$$H(Pc) = -[(Pv1 * \log(Pv1)) + (Pv2 * \log(Pv2)) + (Pv3 * \log(Pv3)) + (Pv4 * \log(Pv4)) + (Pv5 * \log(Pv5)) + (Pv6 * \log(Pv6)) + (Pv7 * \log(Pv7)) + (Pv8 * \log(Pv8))]$$

Figure 26 Méthode de calcul de l'entropie des pixels

3.241	3.241	3.165	3.087	1.849
3.265	3.265	3.796	3.059	2.265
3.003	2.825	3.324	2.690	1.846
2.900	3.165	3.402	2.827	2.030
1.924	2.086	1.899	1.849	1.003

Figure 27 Matrice d'entropie

- Dans la dernière phase , nous appliquons LBP sur la matrice d'entropie

56	240	224	192	225
8	152	0	195	193
8	152	0	195	193
10	12	0	129	129
15	7	135	135	131

Figure 28 Matrice d'image après application de LBP

3.2 Exemples



Figure 29 Lenna image , gauche (original) , milieu (niveaux de gris) , droite (H_LBP)



Figure 30 geeksforgeeks image , gauche (original) , milieu (niveaux de gris) , droite (H_LBP)

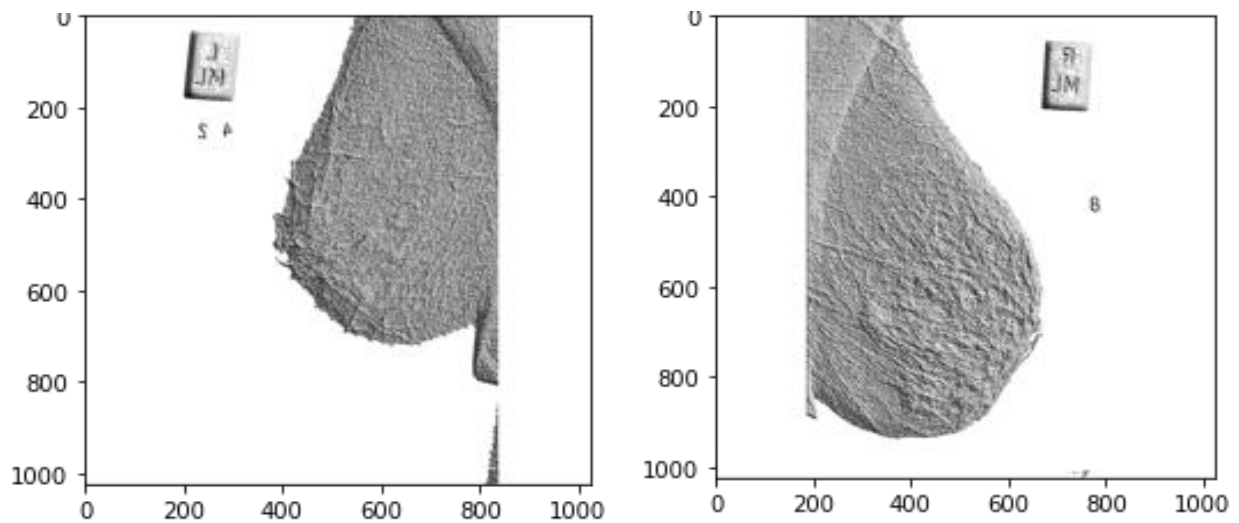


Figure 31 MIA5 images(mdb011.pgm et mdb058.pgm) prétraitées avec LBP

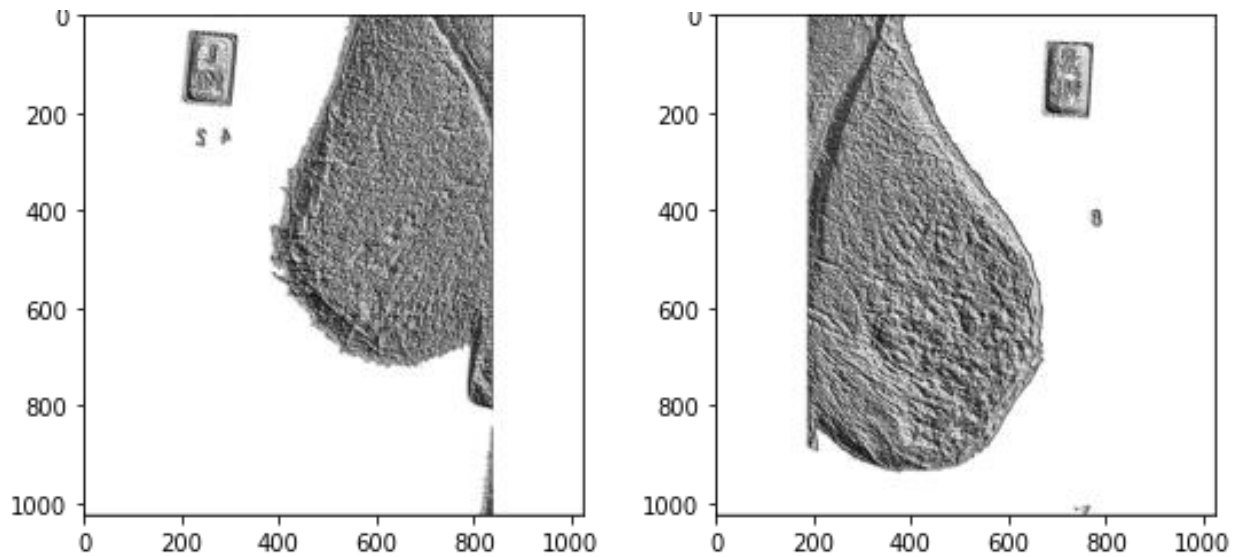


Figure 32 MIAS images(mdb011.pgm et mdb058.pgm) prétraitées avec H_LBP

4. Expériences et Résultats obtenus

Dans notre expérience, nous utilisons la base de données MIAS , nous choisissons 2 catégories (classes) des images , Normal et Anormal . Nous commençons le prétraitement sur ces images avec méthode OTSU. Ensuite, nous utilisons descripteur LBP Original, ou H_LBP.

L'application de descripteur LBP permet d'extraire des informations contenues dans les images sous forme une matrice que nous avons transformé en un vecteur que nous allons utiliser comme vecteur d'entrée des SVM.

Le vecteur caractéristique obtenu est le vecteur d'entrée des SVM.

Nous avons défini 2 classes soient : Normal et Anormal pour un ensemble de 322 images dont 115 anormaux et 207 normales.

L'apprentissage et l'évaluation se font selon la règle 33% , c'est-à-dire 33% pour l'évaluation et 67% pour l'apprentissage du modèle SVM (72 normales, 35 anormaux), avec paramètre de "random_state" égalera a 44 .

Nous avons fait 2 modèles, un pour les images prétraitées avec LBP, et le second pour les images prétraitées avec H_LBP.

Descripteur	LBP	H_LBP
Kernel	Poly	Poly
Itération maximale	500000	500000
Valeur C	0.001	11
Valeur coef0	690	10
Dégré	4	5
Valeur Gamma	Scale	Scale

Tableau 3 Paramètres des modules LBP et H_LBP

a. Confusion matrice

Une matrice de confusion est un résumé des résultats de prédiction sur un problème de classification. Le nombre de prédictions correctes et incorrectes est résumé avec des valeurs de comptage et ventilé par classe.

b. Terminologie et dérivations d'une matrice de confusion

- Condition positive (P) le nombre de vrais cas positifs dans les données
- Condition négative (N) le nombre de cas réels négatifs dans les données
- Vrai positif (TP) Un résultat de test qui indique correctement la présence d'une condition ou d'une caractéristique
- Vrai négatif (TN) Un résultat de test qui indique correctement l'absence d'une condition ou d'une caractéristique
- Faux positif (FP) Un résultat de test qui indique à tort qu'une condition ou un attribut particulier est présent
- Faux négatif (FN) Un résultat de test qui indique à tort qu'une condition ou un attribut particulier est absent

		Réponse de l'expert	
		p	n
Réponse du classifieur	Y	Vrai Positif	Faux Positif
	N	Faux Négatif	Vrai Négatif

Figure 33 Matrice de confusion

- Accuracy est un rapport entre les données correctement classées et le nombre total de classifications effectuées par le modèle.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

- Precision mesure la proportion de positifs, qui sont de vrais positifs.

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

- Recall

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

- Le score F1 est une mesure globale de la science des données de la précision d'un modèle qui mélange précision et rappel.

$$F1 = \frac{2 \times (precision \times recall)}{precision + recall} \quad (7)$$

Matrice de confusion pour le modèle LBP $\begin{matrix} 66 & 6 \\ 27 & 8 \end{matrix}$

Matrice de confusion pour le modèle H_LBP $\begin{matrix} 61 & 11 \\ 20 & 15 \end{matrix}$

Model	LBP	H_LBP
Accuracy	69.158 %	71.028 %
Precision	57.142 %	57.692 %
Recall	22.857 %	42.857 %
F1	32.653 %	49.180 %

Tableau 4 Résultats des modules LBP et H_LBP

5. Discussion de résultats

Le modèle LBP/ H_LBP est généré sur la base des images initiale mammographiques. Les résultats obtenus en utilisant H_LBP sont des résultats légèrement meilleurs en termes de Accuracy par rapport à LBP. En termes de Precision , les résultats obtenus sont similaires.

Recall nous aide à mesurer combien d'échantillons positifs ont été correctement classés par le modèle ML. Recall de H_LBP montre est 42.85% qui dépasse un peu de 20% de Recall de LBP. Le score F1 est une moyenne harmonique de précision et de score de rappel et est utilisé comme métrique dans les scénarios où le choix du score de précision ou de rappel peut entraîner un compromis en termes de modèle donnant respectivement des faux positifs et des faux négatifs élevés. Nous avons obtenu 49.18% avec H_LBP qui dépasse largement le score F1 de LBP.

6. Conclusion

Nous avons présentons aide au dépistage du cancer du sein en développant un système de classification des images de mammographies.

Pour l'élaboration de notre système nous avons utilisé les H_LBP afin d'extraire les vecteurs caractéristiques et faire la classification des image mammographiques en utilisant SVM. les résultats de classification obtenus pour les différentes catégories d'images mammographiques sont présentés dont le but de proposer une méthode de classification des images mammographiques pour extraire les lésions mammaires.

Conclusion générale

Le cancer du sein reste le type le plus meurtrier. Il est devenu un grave problème de santé ouvert dans de nombreux pays. Actuellement, la mammographie est utilisée pour le diagnostic du cancer du sein précoce. D'autre part, la croissance rapide de la mammographie a fait naître le besoin de développer des systèmes de diagnostic assistés par ordinateur qui aident les spécialistes des rayons X.

Notre travail a généralement porté sur l'étude du système de classification mammographique. Dans ce mémoire, nous avons essayé de participer à l'amélioration des performances des systèmes de détection des cancers. Nous avons utilisé le descripteur H_LBP pour décrire les informations pertinentes dans l'image. Ce descripteur a été créé sur la base du descripteur LBP pour générer un vecteur représentant l'image mammographique qui pourrait améliorer les performances de la classification finale de l'image. Des expérimentations menées sur la base de données MIAS avec l'algorithme SVM montrent l'efficacité de l'approche proposée. Bien que nos résultats soient modestes.

Pour cela, nous prévoyons d'améliorer notre modèle de classification en optimisant les images mammographiques en éliminant le bruit. Nous prévoyons également d'améliorer la complexité de notre algorithme de traitement d'image pour avoir des résultats rapides.

Bibliographie

[1] : Mohamed SANDELLI, traitement d'images par des approches bio-inspirées application à la segmentation d'images, diplôme de Magister en informatique, université Constantine 2. 2014 .

[2] : C.Houassine, segmentation d'images par une approche biomimétique hybride. université université m'Hamed bougera- Boumerdes. 2012.

[3] : Zerougui AissaMme , Sari nabil , Traitement d'images monochromes Détection de contours, Filtrage (spatial et fréquentiel) et Segmentation par Réseaux de Neurones, diplôme de Master ,Université Larbi Ben M'hidi, Oum El Bouaghi , Juin 2017.

[4] : MERSAOUI Imane , Etude sur l'imagerie médicale : prétraitement, segmentation et amélioration des exploitations , Mémoire de Master, Université SAAD DAHLAB de BLIDA ,2018.2019.

[5] : Imane-CHEIKHROUHOU Epse KACHOURI (2012). Description et classification des masses mammaires pour le diagnostic du cancer du sein ,Thèse doctorat. L'Université d'Evry-Val d'Essonne.

[6] : Fatima EDDAOUDI(2012). Aide à la détection des pathologies du sein par les techniques de segmentation et de codage de textures des images mammographiques. Thèse de doctorat. Université Mohammed V – AGDAL.

[7] : TERKI Hakima et BENYELLES Khadîdja, Analyse des images mammographiques en vue de la détection et la caractérisation des Microcalcifications mammaires, Diplôme de MASTER , Université Abou Bakr Belkaïd de Tlemcen, 23 juin 2013 devant le Jury.

[9] : Gueniaa fatma Zohra , Tourqui ikram , Ccomparaison de méthodes de classification appliquées à la détection d'objets, diplôme master 2, UNIVERSITÉ Echahid Hamma Lakhdar - EL-OUED ,2016.

[10] : Mr. DOUFENE Said , Mr. HADJAB Rachid, Présentation d'objets virtuels sur une scène réelle, diplôme Master 2, Université Mouloud MAMMERI de Tizi-Ouzou, Année 2012/2013.

[11] : Yacine GAFOUR, Apprentissage automatique pour une classification des images basée sur les descripteurs, Thèse de doctorat ,Université Djillali Liabes, 2020.

[12] : W. Ben Soltana, A. Porebski, N. Vandenbroucke, A. Ahmad et D. Hamad : Contribution des descripteurs de texture LBP à la classification d'images de dentelles ,article, Laboratoire d'Informatique Signal et Image de la Côte d'Opale Maison de la Recherche Blaise Pascal 50, rue Ferdinand Buisson BP 719, 62228 Calais Cedex France.

[13] : MOUDJEB Fatma IRSANE Feriel , Classification des images de mammographie, UNIVERSITE MOULOUD MAMMERI DE TIZI-OUZOU, 27 /09 / 2017.

[14] : Oumiloud Horiya ,Mokeddem Asma, Classification non supervisée : Application de k-means, diplôme de Licence en Informatique, Université Abou Bakr Belkaid– Tlemcen, 2013-2014

[15] : Mezaache Hatem « Les réseaux de Neurones formels et les systèmes Neuro-Flous pour l'apprentissage par renforcement » , Diplôme de Magister ,Université Hadj Lakhdar, Batna-Algérie. 2007.

[16] : ANNICK LESNE (2014). Shannon entropy: a rigorous notion at the crossroads between probability, information theory, dynamical systems and statistical physics . Mathematical Structures in Computer Science, 24, e240311 doi:10.1017/S0960129512000783,article .

[17] : Leila Kabbai1 · Mehrez Abdellaoui2 · Ali Douik1, Image classification by combining local and global features, article ,Springer-Verlag GmbH Germany, part of Springer Nature 2018.

Webographie

[18] : [http://www.memoireonline.com/12093040m_La-liaison-automatique-des-plusieurs-images-percues-sur-un-scanner4.html](http://www.memoireonline.com/12/09/30/40m_La-liaison-automatique-des-plusieurs-images-percues-sur-un-scanner4.html), visité en : Mai 2013.

[19] : <https://coffradoc.pistak.fr/data/niv3e/item42/Projet%20-%20Séance%202005/image%20matricielle%20ou%20vectorielle.pdf>

[20] : <https://onlyweb-formation.com/formation-creation-graphique/les-modes-colorimetrique.html>

[21] : <http://www.chirurgie-gynecologie.fr/page-chir-sein/anatomie-sein.html>

[22] : <https://www.docteur-benchimol.com/cancer-du-sein.html>

[23] : https://stringfixer.com/fr/Histogram_of_oriented_gradients

[24] : http://csidoc.insa-lyon.fr/these/2004/zennir/15_chapitre3.pdf

[25] : https://github.com/arsho/local_binary_patterns

[26] : <https://www.geeksforgeeks.org/create-local-binary-pattern-of-an-image-using-opencv-python/>

[27] : <http://peipa.essex.ac.uk/info/mias.html>.

[28] : <https://muthu.co/otsus-method-for-image-thresholding-explained-and-implemented/>