

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE MOULOUD MAMMERRI DE TIZI-OUZOU



FACULTE DU GENIE ELECTRIQUE ET D'INFORMATIQUE  
DEPARTEMENT D'ELECTRONIQUE

**Mémoire de Fin d'Etudes  
De MASTER ACADEMIQUE**

Domaine : Sciences et Technologies

Filière : **Electronique**

Spécialité : **Electronique biomédicale**

*Présenté par*  
**MOUDJEB Fatma**  
**IRSANE Ferial**

**Thème**

**Classification des images de mammographie.**

*Mémoire soutenu publiquement le 27 /09 / 2017 Devant le jury composé de :*

**Mr. MOHIA Yacine, Président**

**Mr. HAMEG Slimane, Examineur**

**Melle. KHERCHAOUI Sonia, Encadreur**

**2016/2017**

# *Remerciements*

*Tout d'abord, nous remercions le bon dieu qui nous avoir donné la santé, la volonté pour finir ce travail.*

*Nous tenons à adresser nos vifs remerciements à notre promotrice Melle KHERCHAoui pour nous avoir proposé le thème de ce mémoire et de nous avoir dirigés, orientés et conseillés tout le long de notre travail.*

*Nous remercions également les membres du jury qui nous ferons l'honneur de juger notre travail.*

*Sans oublier tous ceux de près ou de loin qui ont contribués à l'élaboration de ce mémoire.*

# *Dédicaces*

*À mes très chers parents*

*À mon Cher mari*

*À mes frères et sœurs*

*À ma binôme Fatma*

*À toute ma famille*

*À toutes mes amies*

*Irsane Ferial*

# ***Dédicaces***

*À mes très chers parents*

*À mes frères et sœurs*

*À ma binôme Feriel*

*À toute ma famille*

*À toutes mes amies*

*À mon cher ami karim*

*Moudjeb Fatma*

---

## TABLE DES FIGURES :

Figure1 : Anatomie du sein .....	4
Figure 2 : Le mammographe .....	6
Figure3: Les types de masses .....	6
Figure4: Les différents types de bords de masses .....	7
Figure 5: Exemple de micro-calcifications.....	8
Figure6 : Structure d'un neurone artificielle .....	10
Figure7 : Exemple d'illustration de k-plus proches voisins .....	11
Figure 8: Exemples de quartier utilisés pour définir une texture et calculer un modèle binaire local (LBP) .....	14
Figure 9 : Exemple de calcul du LBP d'un pixel donne .....	15
Figure 10: Exemple de calcul du LBP par les nombres binaire .....	15
Figure 11: Exemple de LBP appliqué à une image mammaire de la base MIAS .....	16
Figure12 : Détermination d'un hyperplan via les vecteurs de supports .....	19
Figure13: Illustration de la marge maximale et l'hyperplan optimal .....	20
Figure 14 : Cas linéairement séparable .....	21
Figure15: Hyperplan séparateur $\langle w, x \rangle + b = 0$ .....	22
Figure16 : Hyperplans canoniques et marge maximale .....	24
Figure17: Cas non linéairement séparable .....	25
Figure18: Système proposé pour la classification des images mammographiques.....	29
Figure 19 : Echantillon des images mammographiques de la base MIAS .....	30
Figure 20 : Exemple de l'égalisation d'histogramme de l'image mdb059.pgm .....	33

---

Figure 21 : Application de l'opérateur LBP d'une image normale mdb056.pgm.....	34
Figure 22 : Application de l'opérateur LBP d'une image maligne mdb111.pgm .....	35
Figure 23 : Application de l'opérateur LBP d'une image bénigne mdb150.pgm .....	35

---

## **LISTE DES TABLEAUX :**

Tableau 1 : Critères de distinction entre bénignes et malignes .....	5
Tableau 2 : Nombre de mammogrammes dans la base de données MIAS .....	31
Tableau 3 : Nombre de mammogrammes contenant des masses de la base de données MIAS .....	32
Tableau 4 : La matrice de confusion obtenue pour deux classes .....	36
Tableau 5 : La matrice de confusion obtenue pour trois classes .....	36

---

## SOMMAIRES

INTRODUCTION GENERALE .....	1
CHAPITRE I :	
I-1) Introduction .....	3
I-2) Anatomie de sein .....	3
I-3) Cancer de sein .....	4
I-4) La mammographie .....	5
I-5) Les masses .....	6
I-6) Les micro-calcifications .....	7
I-7) Les approches de classification .....	8
❖ Approche de classification supervisée .....	8
❖ Approches de classification non supervisée .....	9
❖ Quelques techniques de classification .....	9
1) Méthode de réseaux de neurones artificiels .....	9
2) Méthode K- plus proches voisins .....	10
CHAPITRE II :	
II-1) Introduction .....	13
II-2) Définition de la texture.....	13
II-3) Les LBP (Local Binary Pattern).....	13
II-4) Principe de LBP .....	13
II-5) Classification .....	17
II-6) Définition de SVM .....	17
II-7) Principe des SVM .....	19
II-8) Application des SVM .....	20
II-9) Classifieur linéaire .....	21
II-9-a) Marge maximale de l'hyperplan .....	23

---

II-10) SVM non linéaire .....	25
II-11) SVM pour le cas de multi classes .....	26
II-12) Conclusion .....	26
CHAPITRE III :	
III-1) Préambule .....	27
III-2) Environnement du logiciel .....	27
III-3) Langage adopté pour la programmation .....	27
III-4) Mammography Image Analysis Society (MIAS) .....	29
III-5) Le système de classification .....	32
III-5-1) Prétraitement.....	32
III-5-2) Extraction des caractéristiques .....	33
III-5-3) Application de SVM.....	35
III-6) interprétation .....	36
CONCLUSION GENERALE .....	39

# INTRODUCTION GENERALE

## Introduction générale

Le cancer du sein représente l'un des enjeux prépondérants dans le domaine de la santé publique. En effet, il s'agit d'un fameux cancer menaçant la vie de la plupart des femmes. Environ une femme sur dix est touchée par cette maladie durant son existence. Cependant, la réduction du taux de mortalité causée par ce type de cancer ainsi que la favorisation des chances de guérison ne sont possibles que si la tumeur a été prise en charge dès les premiers stades de son apparition. Dans l'objectif d'assurer le dépistage précoce d'une telle tumeur, les radiologues ont été amenés à augmenter la fréquence des mammographies notamment pour la tranche d'âge la plus concernée. Sur l'échelle mondiale, de nombreux pays ont mis en œuvre des campagnes de dépistage systématique tous les deux ans par exemple. Il a été démontré qu'une telle démarche est très efficace et permet de réduire le taux de décès. Selon les campagnes, deux ou 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 pendant ces dernières décennies, différents travaux de recherche travaillent sur ça, 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 un système de classification des images de mammographies. Pour l'élaboration de notre système nous avons utilisé les LBP (local binary pattern) afin d'extraire les vecteurs caractéristique et faire notre classification soit l'image est normale, bénigne ou contient des lésions maligne à travers les SVM (support vector machine).

## **Organisation du mémoire :**

Ce manuscrit est organisé en trois chapitres.

### ➤ **Chapitre I :**

- ✓ Ce chapitre contient la définition du cancer du sein et des images de mammographies et à l'anatomie du sein.
- ✓ Présentation des différentes approches de classification.

➤ **Chapitre II :** décrit la méthode d'analyse de texture LBP et donne sa définition, ainsi contient le principe de la classification par la méthode SVM.

➤ **Chapitre III :** présente la base de d'images utilisée et quelques détails sur sa constitution, l'algorithme qui réalise la classification avec des explications sur chaque étape, puis l'implantation de SVM et résultats obtenus.

# CHAPITRE I

L'Etat de l'art

## **I-1) INTRODUCTION**

L'approche médicale de l'imagerie mammaire est indispensable puisqu'elle nous permet d'acquérir un savoir nécessaire pour mener à bien notre étude. Elle permet de transmettre l'expertise des radiologues dans ce domaine durant des décennies de recherche. Afin de proposer des outils performants aidant les radiologues à dépister 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 personnes malades et non malades. Nous abordons dans ce chapitre l'étude de ces systèmes et nous nous focalisons sur la classification automatique des images de mammographie.

Dans ce chapitre, nous commençant par donner une notion du cancer du sein. Ensuite, nous abordons les outils d'imagerie médicale permettant le dépistage et le diagnostic de ce type de cancer notamment la mammographie. Nous étudions par la suite, d'une façon non exhaustive l'anatomie du sein. Et les méthodes utilisées dans la segmentation et la classification d'images.

## **I-2) ANATOMIE DU SEIN**

Le sein est essentiellement constitué d'un tissu adipeux gras plus ou moins important qui lui donne sa forme et son volume. La glande mammaire, noyée dans le tissu conjonctif, est composée d'une vingtaine de lobules qui deviennent actives en période de lactation. Le lait sécrété est déversé par des canaux galactophores séparés au niveau du mamelon. Le sein, parcouru par une multitude de vaisseaux sanguins, est maintenu par la peau qui le recouvre et par des fibres liées au muscle pectoral. La figure 1 détaille davantage ces différents composants mammaires.

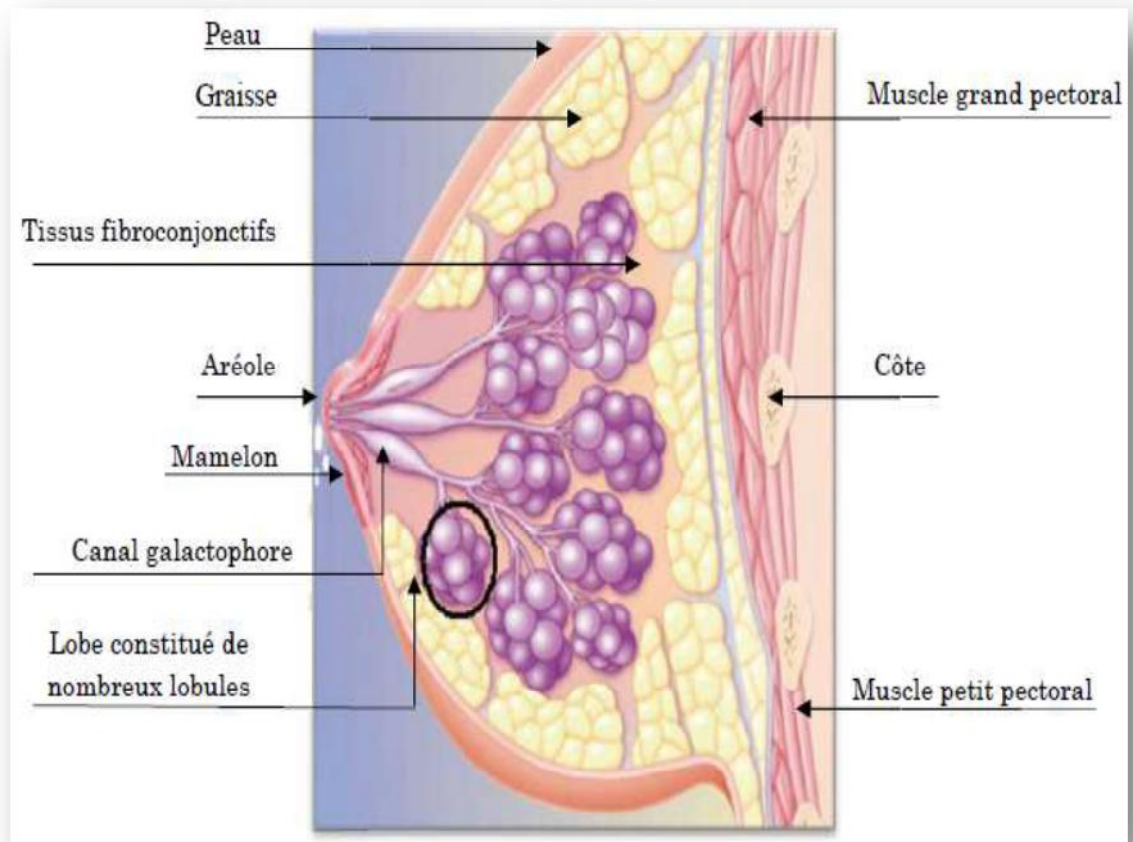


Figure1 : Anatomie du sein.

### I-3) 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). [1]

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

Il y a deux grandes catégories de tumeurs sont connues :

- Les tumeurs bénignes
- Les tumeurs malignes ou cancers

<b>Tumeurs bénignes</b>	<b>Tumeurs malignes</b>
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.

#### **I-4) LA MAMMOGRAPHIE**

La mammographie est une technique de radiographie, particulièrement adaptée aux seins de la femme. Elle a pour but de détecter au plus tôt des anomalies avant qu'elles ne provoquent des symptômes cliniques. La mammographie est non seulement pratiquée dans les campagnes de dépistage du cancer du sein, mais aussi pour le diagnostic et la localisation lors d'interventions chirurgicales (ponctions). Le point fort d'un tel examen est qu'il permet d'examiner la totalité du tissu mammaire avec une ou deux incidences seulement.

L'appareil dédié à la réalisation d'une mammographie est le mammographe (figure 2). Cet appareil se compose d'un tube radiogène générateur de rayons X de faible énergie (entre 20 et 50 keV) et d'un système de compression du sein. En premier temps, les deux seins sont comprimés à tour de rôle. Cette compression permet l'étalement des tissus mammaires ce qui facilite la visualisation des structures du sein et la réduction de la dose de rayons X délivrée. En deuxième temps, les deux seins sont exposés à une faible dose de rayons X. On obtient, alors, une projection du sein sur un détecteur plan. La radiographie est réalisée sur des films argentiques ou sur des systèmes de radiologie digitale de haute qualité. L'analyse de la glande mammaire est réalisée grâce aux différences de l'atténuation des différents types de tissu. [2]

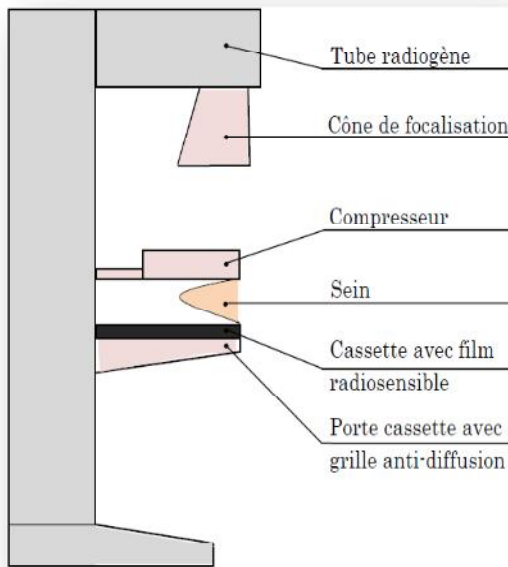
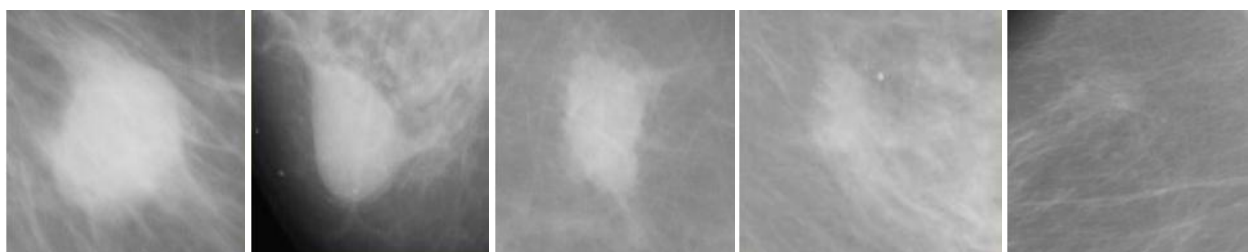


Figure 2 : Le mammographe.

**I-5) LES MASSES**

Ce sont des lésions retrouvées en trois dimensions. Etant plus denses que le tissu mammaire adjacent, elles apparaissent blanches sur les mammogrammes. La taille ne prédit pas le caractère malin, sauf si les clichés successifs montrent une augmentation régulière de la taille. On distingue traditionnellement cinq formes différentes de masses ronde, ovale, lobulaire, irrégulière ou une distorsion architecturale. Cette dernière forme inclut des rayons convergent vers un point avec une rétraction focale. La figure 3 présente un ensemble de lésions.[3]



Masse ronde

Masse ovale

Masse lobulaire

Masse irrégulière

Distorsion architecturale

Figure3: Les types de masses.

Les bords de la masse sont très importants à étudier lors de l'analyse sémiologique qui nécessite parfois l'utilisation de la loupe. On décrit 5 catégories de bords de type: circonscrits, effacés, micro-lobulés, mal définis et avec des spicules (Figure 4). Les bords effacés correspondent à une superposition de tissus adjacents. Les bords mal définis ou les spicules correspondent à une invasion dans le tissu sain.



Bords circonscrits

Bords effacés

Bords  
micro-lobulésBords  
mal définis

Bords spiculés

Figure4: Les différents types de bords de masses

### I-6) Les micro-calcifications

Les micro-calcifications sont des petits dépôts de calcium de taille inférieure à 1 mm, qui apparaissent comme des petits points lumineux situés dans la glande mammaire. Ces micro-calcifications sont caractérisées par leur type et les propriétés de leur distribution. [3]

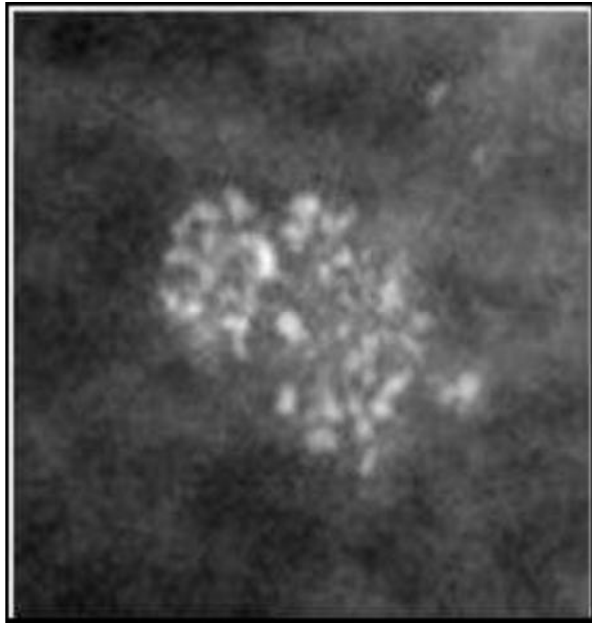


Figure 5 : Exemple de micro-calcifications

### **I-7) Les approches de classification**

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

#### **❖ Approche de classification supervisée :**

La classification supervisée est une technique de classification automatique où l'on cherche à produire automatiquement des règles pour regrouper des individus en un certain nombre de classe, à partir d'une base de données d'apprentissage contenant des échantillons.

L'utilisation des techniques supervisées nécessite l'intervention d'un utilisateur et aussi une première phase d'apprentissage durant laquelle le système apprend les caractéristiques associées à chaque classe d'une base d'apprentissage.

Une fois cette étape réalisée, ces méthodes supervisées sont capables de classer les individus d'une base en fonction des caractéristiques textures qui leurs sont associées.

Il existe plusieurs méthodes de classification supervisé telle que la méthode de K plus proche voisins, la méthode de maximum de vraisemblance.

### ❖ **Approche de classification non supervisé :**

Lorsque le type des objets dans l'image n'est pas connu, cela résulte d'un manque d'information sur la réalité du terrain, il existe des algorithmes de classification qui permettent de créer un regroupement de pixels similaires. Dans le cadre de la classification non-supervisé, il est généralement nécessaire d'indiquer le nombre de classe recherchées. Les textures similaires sont alors regroupées à l'intérieur d'une même classe sans l'intervention de connaissances a priori et simplement à partir d'estimations de similarité entre caractéristiques.

### ❖ **Quelques techniques de classifications**

#### **1. Méthode de réseaux de neurones artificiels**

Est un ensemble d'algorithmes dont la conception est à l'origine très schématiquement inspirée du fonctionnement des neurones biologiques, et qui par la suite s'est rapproché des méthodes statistiques.

Les réseaux de neurones sont généralement optimisés par des méthodes d'apprentissage de type probabiliste, en particulier bayésien. Ils sont placés d'une part dans la famille des applications statistiques, qu'ils enrichissent avec un ensemble de paradigmes permettant de créer des classifications rapides, et d'autre part dans la famille des méthodes de l'intelligence artificielle auxquelles ils fournissent un mécanisme perceptif indépendant des idées propres de l'implémenteur, et fournissant des informations d'entrée au raisonnement logique formel.

En modélisation des circuits biologiques, ils permettent de tester quelques hypothèses fonctionnelles issues de la neurophysiologie, ou encore les conséquences de ces hypothèses pour les comparer au réel. [4]

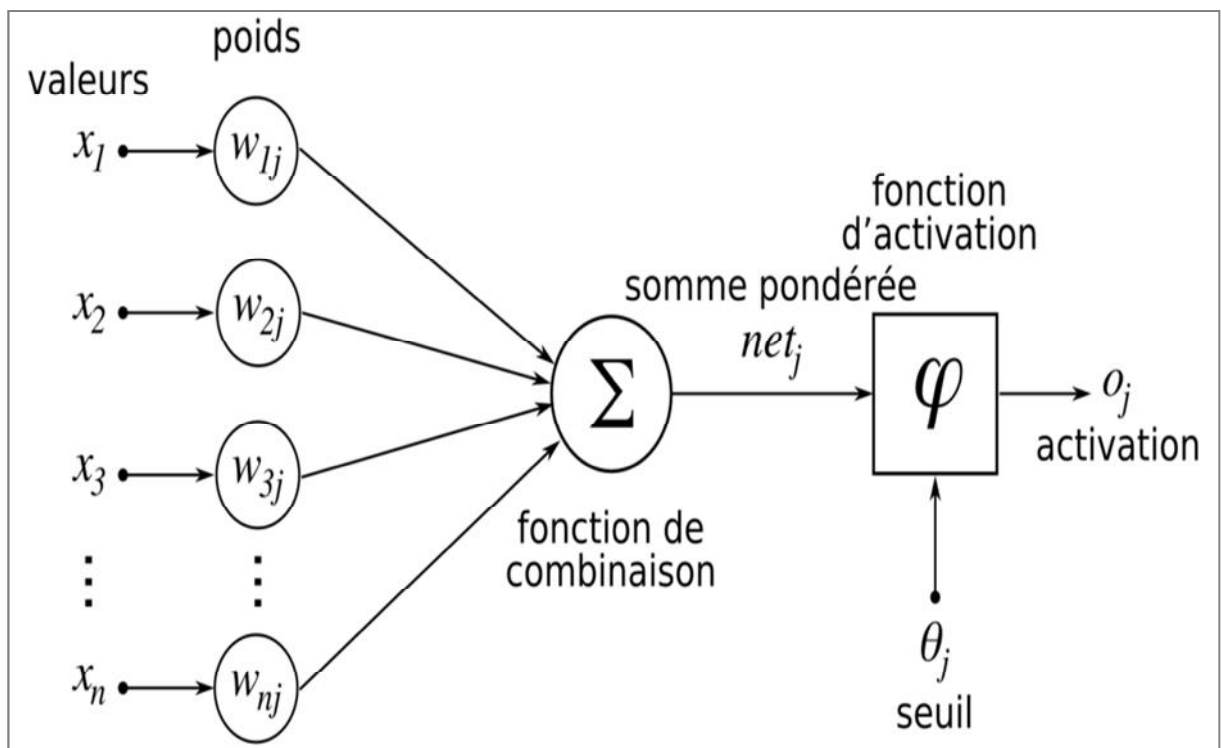


Figure 6: structure d'un neurone1 artificielle

## 2. méthode k plus proche voisins

La méthode des  $k$  plus proches voisins est une méthode d'apprentissage supervisé. En abrégé k-NN ou KNN, de l'anglais k-nearest neighbor. Dans ce cadre, on dispose d'une base de données d'apprentissage constituée de  $N$  couples « entrée-sortie ». Pour estimer la sortie associée à une nouvelle entrée  $x$ , la méthode des  $k$  plus proches voisins consiste à prendre en compte (de façon identique) les  $k$  échantillons d'apprentissage dont l'entrée est la plus proche de la nouvelle entrée  $x$ , selon une distance à définir. Par exemple, dans un problème de classification, on retiendra la classe la plus représentée parmi les  $k$  sorties associées aux  $k$  entrées les plus proches de la nouvelle entrée  $x$ . [5]

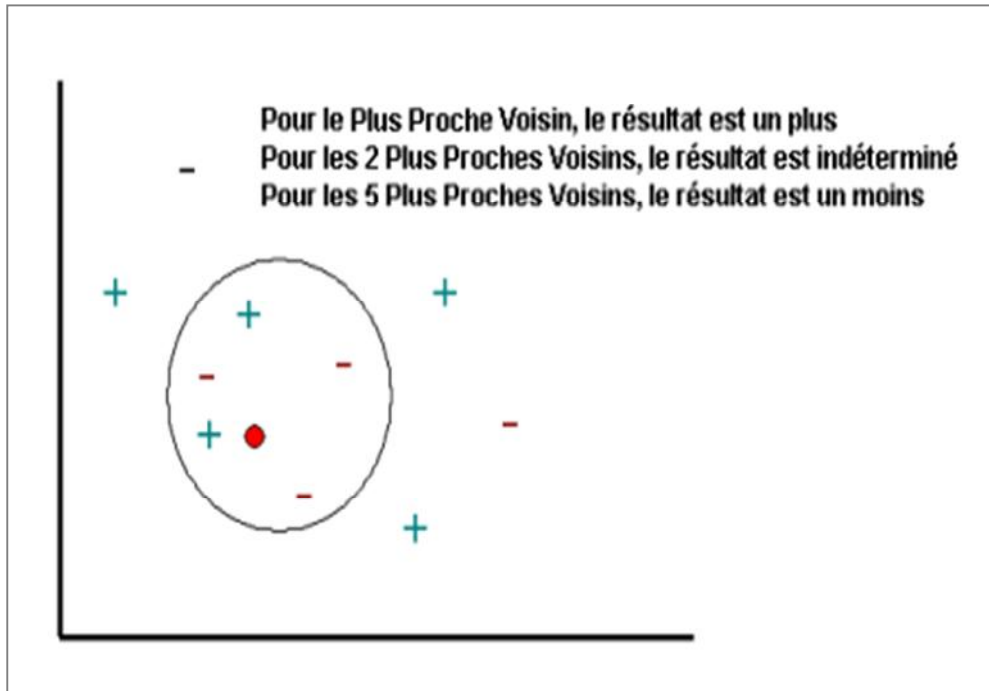


Figure 7: Exemple d'illustration de k-Plus proches voisins.

# CHAPITRE II

Définition de LBP et SVM

### II-1) Introduction

Dans ce chapitre nous allons présenter la technique de classification SVM et le descripteur de texture **LBP** ainsi que la base de données MIAS [6] utilisée dans notre système de classification des images mammographiques.

### II-2) Définition de la texture

La notion de texture est utilisée pour traduire un aspect homogène de la surface d'un objet sur une image. La texture est l'attribut qui représente la distribution spatiale des niveaux de gris des pixels d'une région, ou encore c'est l'ensemble des propriétés de voisinage locales de niveaux de gris d'une région de l'image.

### II-3) Les LBP (motifs binaires locaux)

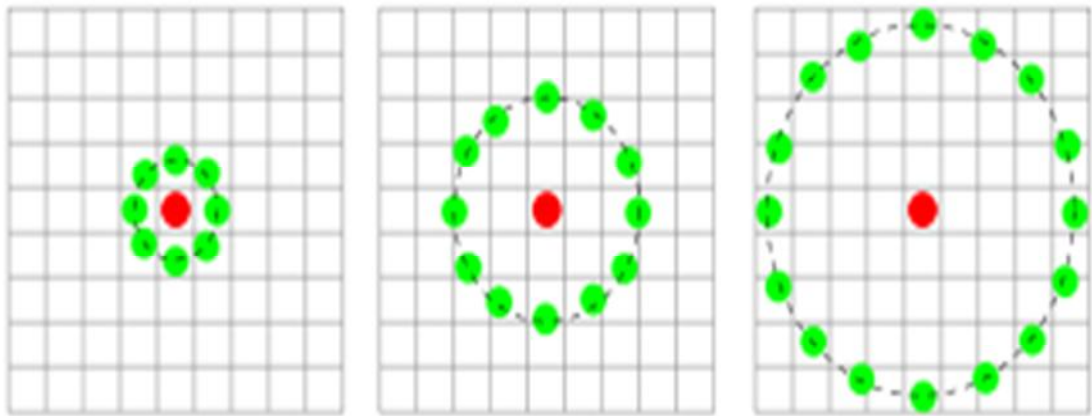
#### Définition

Est un opérateur de texture simple mais très efficace qui étiquette les pixels d'une image en seillant le voisinage de chaque pixel et le résultat est un nombre binaire. En raison de sa puissance discriminative et de sa simplicité de calcul, le descripteur LBP est devenu une approche populaire dans diverses applications. La propriété la plus importante de l'opérateur LBP dans les applications du monde réel est sa robustesse aux changements mono-toniques de l'échelle de gris provoqués, par exemple, par des variations d'illumination. Une autre propriété importante est sa simplicité de calcul, qui permet d'analyser des images dans des environnements en temps réel difficiles.

#### Principe des LBP

La méthode LBP, introduite par Ojala et Al (1996) [7] se définit comme une mesure de texture invariante, dérivée de la définition générale d'une texture dans un voisinage local. Un code binaire décrivant la texture locale d'une région est calculé par seuillage d'un voisinage avec le niveau de gris du pixel central. Tous les voisins prendront alors une valeur 1 si leur valeur est supérieure ou égale au pixel courant et 0 sinon. On va alors multiplier cette matrice composée de 0 et 1 par les poids LBP et sommer tous ses éléments pour obtenir la valeur LBP du pixel courant. On obtiendra donc des pixels dont l'intensité se situe entre 0 et 255 comme dans une image 8 bits ordinaire.

Pour calculer un LBP dans un voisinage de P pixels, dans un rayon R, on compte simplement les occurrences de niveaux de gris  $g_k$  plus grands que la valeur centrale.



P=8, R=1

p=12, R=2

p=16, R=4

Figure 8: Exemples de quartier utilisés pour définir une texture et calculer un modèle binaire local (LBP).

Nous utilisons la fonction suivante :

$$LBP_{P,R} = \sum_{k=0}^{p-1} 2^k s(g_k - g_c) \dots\dots\dots(1)$$

Où S est la fonction signe :

$$S(x) = \begin{cases} 1, & x > 0 \\ 0, & \text{sinon} \end{cases} \dots\dots\dots(2)$$

Avec :

$g_c$  : le niveau de gris de pixel central.

$g_k$  : le niveau de gris de pixel voisin.

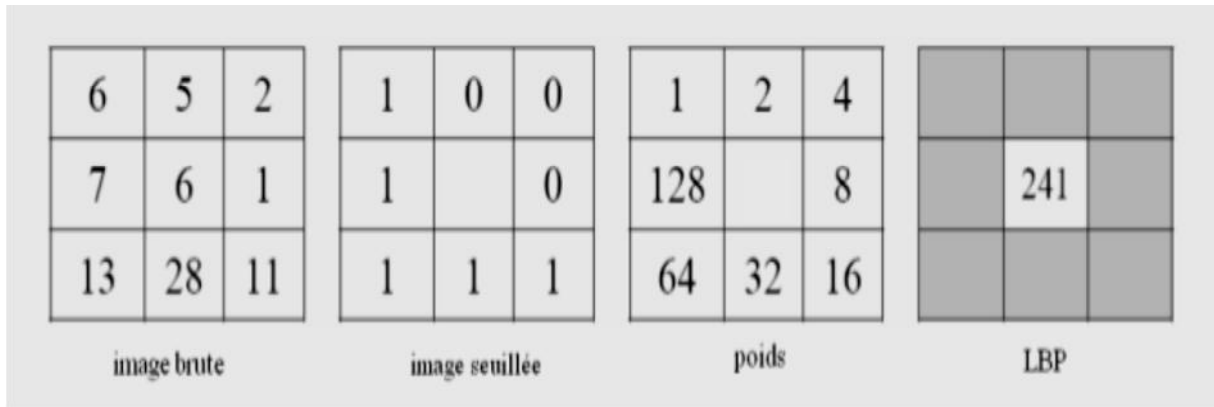


Figure 9: Exemple de calcul du LBP d'un pixel donne :  $1 + 16 + 32 + 64 + 128 = 241$

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.

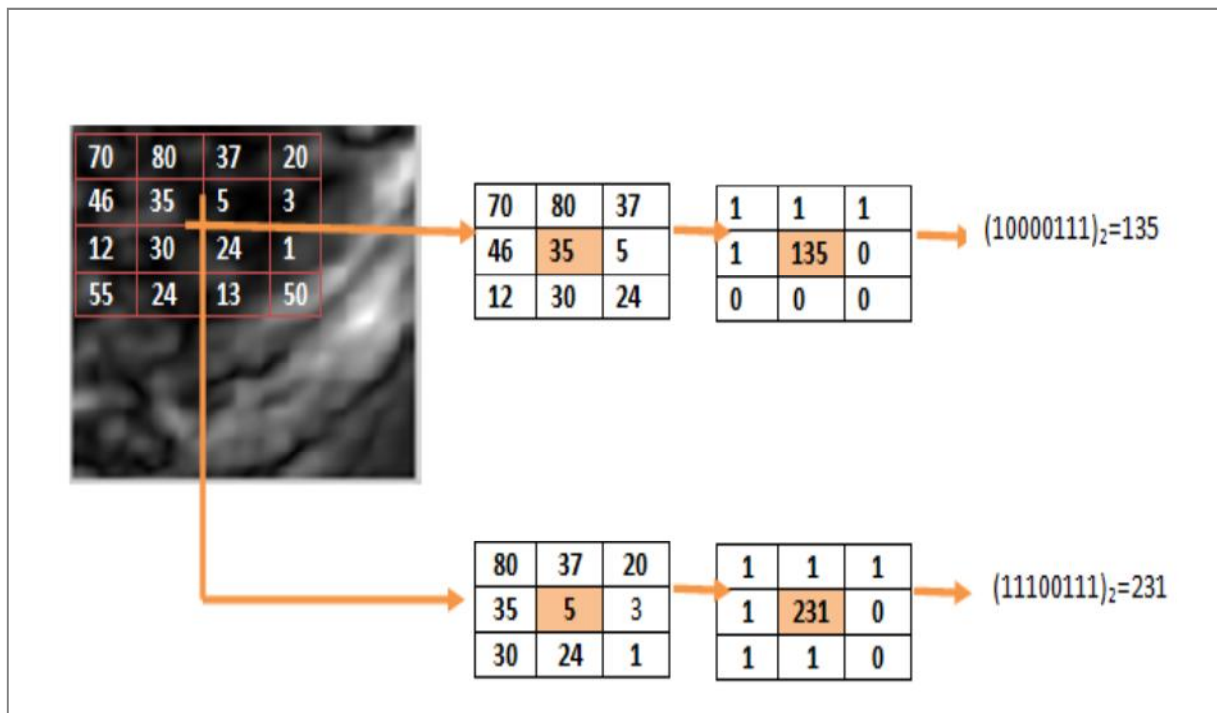


Figure10 : Exemple de calcul de LBP par les nombres binaire.

L'application de LBP sur une image mammographie donne le résultat suivant :

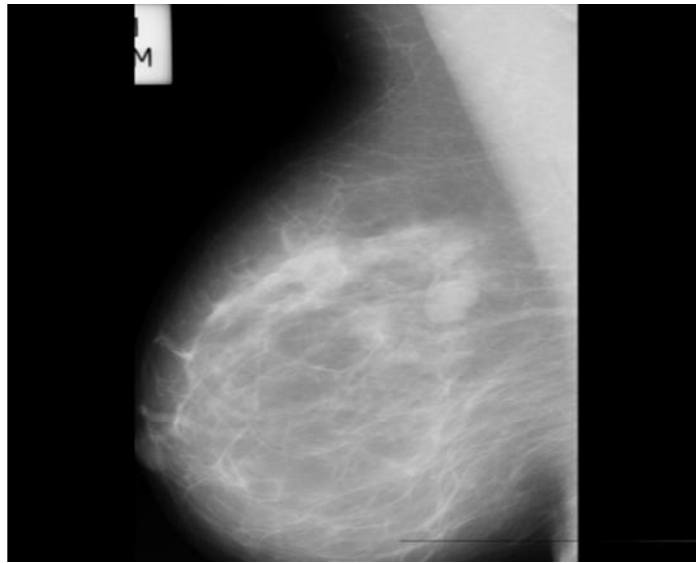


Image de la base MIAS: mdb019.pgm

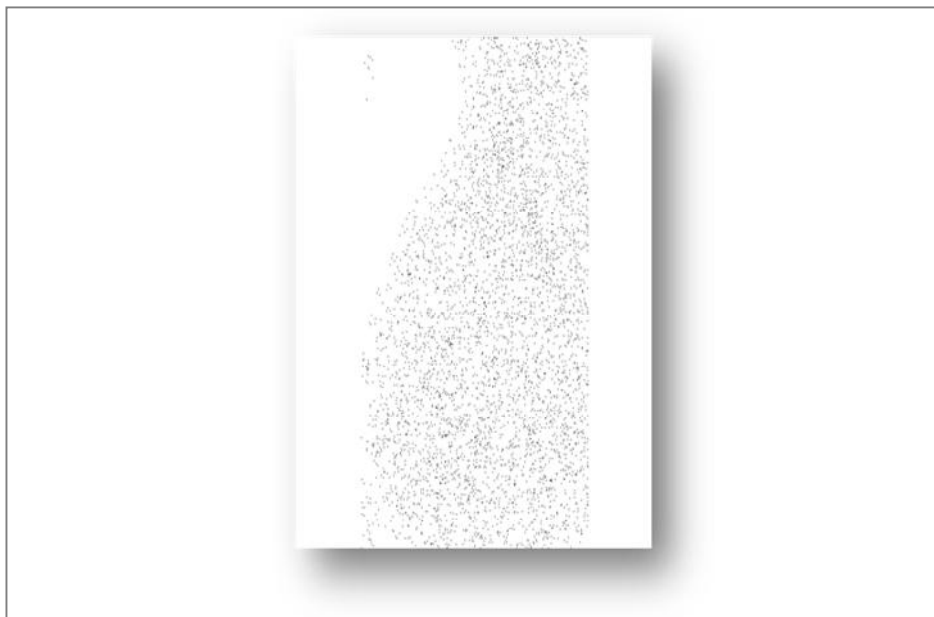
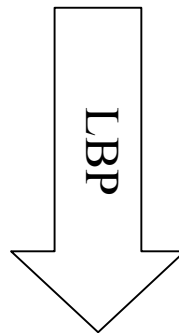


Figure 11: Application des LBP sur une image de la base de données MIAS.

### II-5) Classification

Nous nous intéressons à la classification supervisée, qui se base sur une règle de décision à partir d'une image déjà classifiée. Cette opération s'effectue en deux phases :

- La première est l'apprentissage et la génération du modèle de classification.
- La deuxième est l'évaluation.

### II-6) Définition de SVM

Machines à vecteurs supports (support vector machine, SVM) a été proposée en 1995 par Vapnik [8]. Depuis, elle est très activement utilisée dans beaucoup d'application. Son but est de chercher une surface capable de séparer au mieux les prototypes de deux classes issus de la base d'apprentissage.

Considérons le cas où l'ensemble d'apprentissage  $(X_a, Y_a) = (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  sont linéairement séparable en deux classes  $C_1$  et  $C_2$  telles que  $\mathbf{x}_i \in \mathbf{R}^d$  représente la  $i$ ème observation caractérisée par le vecteur de  $d$  attributs et  $y_i \in \{+1, -1\}$ ,

Avec :

$$\begin{cases} y_i = 1, & \mathbf{x}_i \in C_1 \\ y_i = -1, & \mathbf{x}_i \in C_2 \end{cases} \dots\dots\dots(3)$$

La méthode SVM consiste à chercher un hyperplan qui sépare les données des deux classes. Les points situés sur cet hyperplan satisfait l'équation [9] :  $\mathbf{w}^T \mathbf{x} + \mathbf{b} = 0$  où  $\mathbf{w} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \dots, \dots, \mathbf{w}_d]$  est le vecteur perpendiculaire à l'hyperplan et  $\mathbf{b}$  un scalaire appelé bias.

Dans le cas ou les observations sont séparables, la méthode SVM consiste à maximiser la marge représentant la distances entre un hyperplan  $\mathbf{H}_1$  définit par l'équation :  $\mathbf{w}^T \mathbf{x} + \mathbf{b} = -1$  et un L'hyperplan  $\mathbf{H}_2$  définit par l'équation  $\mathbf{w}^T \mathbf{x} + \mathbf{b} = 1$ . Ceci constitue un problème d'optimisation avec contrainte.

$$\begin{cases} \text{Min } f(x) = \frac{1}{2} w^T w \\ \text{sous contrainte } g_i(x) = y_i (w^T x_i + b) - 1 \geq 0 \quad i = 1, \dots, N \end{cases} \dots\dots\dots(4)$$

En introduisant les multiplicateurs  $\lambda_i \geq 0$  de Lagrange, ce problème devient :

$$\begin{cases} \text{Max } -\frac{1}{2} \sum_{i=1}^p |y_i y_j| x_i x_j + \sum_{i=1}^p \lambda_i \text{ avec } \lambda_i \geq 0, i = 1 \dots, N \\ \sum_{i=1}^N y_i \lambda_i = 0 \end{cases} \dots\dots\dots(5)$$

Dont la résolution fournit les valeurs  $\bar{\lambda}_i$  et par conséquent l'hyperplan optimal défini par la paire  $(\bar{w}, \bar{b})$  telles que :

$$\bar{w} = \sum_{i=1}^p \bar{\lambda}_i x_i y_i \dots\dots\dots(6)$$

Et 
$$\bar{b} = y_i - \sum_{i=1}^p \bar{\lambda}_i x_i y_i \dots\dots\dots(7)$$

La propriété la plus remarquable des SVM est que l'hyperplan trouvé doit passer au milieu des points des deux classes de façon à maximiser la distance qui sépare les points des deux classes. Cette distance est appelée une marge. L'hyperplan séparateur optimal est celui qui maximise la marge. Comme nous cherchons à maximiser cette marge, nous parlerons de séparateurs à vaste marge.

Pratiquement, le fait d'avoir une marge plus large procure plus de sécurité lorsque nous classons un nouvel exemple. De plus, si nous trouvons le classificateur qui se comporte le mieux vis-à-vis des données d'apprentissage, il est clair qu'il sera aussi celui qui permettra au mieux de classer les nouveaux exemples.

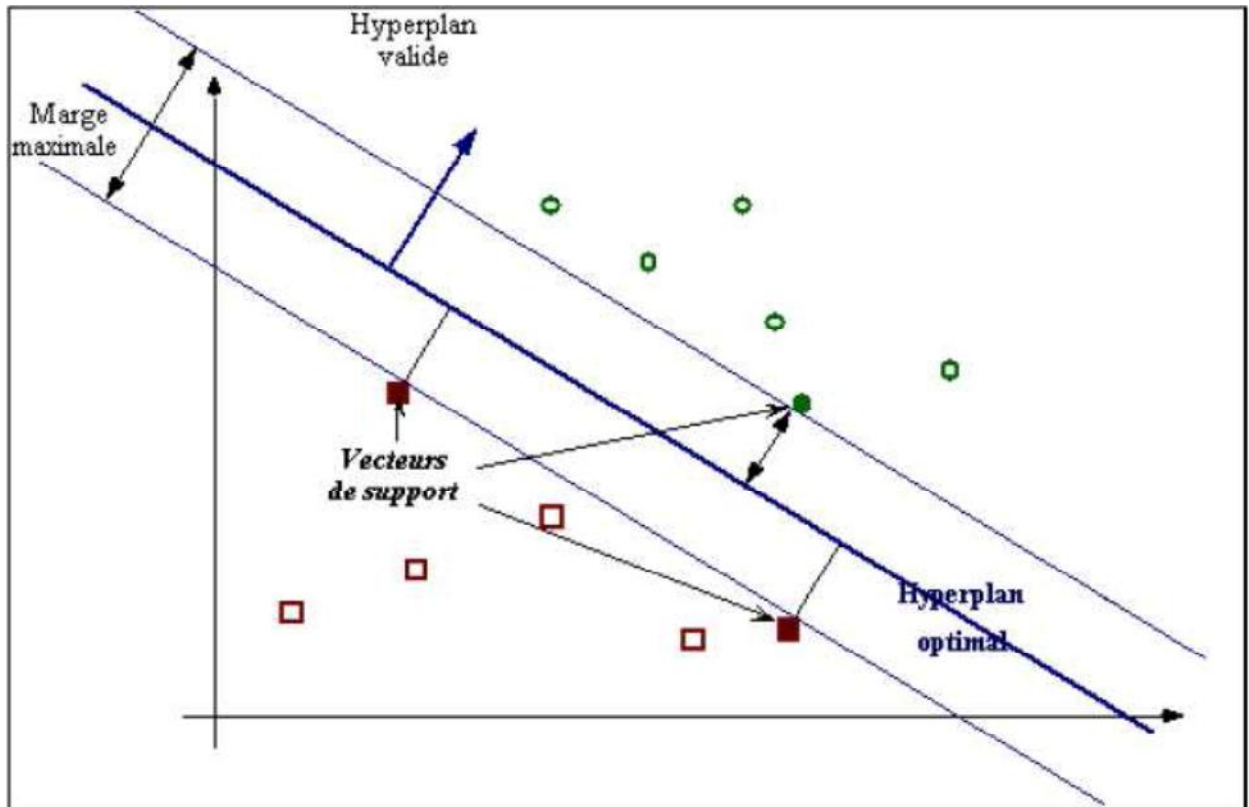


Figure12 : Détermination d'un hyperplan via les vecteurs de supports

### II-7) Principe des SVM

C'est une méthode de classification à deux classes qui tente de séparer les exemples positifs des exemples négatifs dans l'ensemble des exemples. La méthode cherche alors l'hyperplan qui sépare les exemples positifs des exemples négatifs, en garantissant que la marge entre le plus proche des positifs et des négatifs soit maximale. Cela garantit une généralisation du principe car de nouveaux exemples pourront ne pas être trop similaires à ceux utilisés pour trouver l'hyperplan mais être situés d'un côté ou l'autre de la frontière. L'intérêt de cette méthode est la sélection de vecteurs supports qui représentent les vecteurs discriminants grâce auxquels est déterminé l'hyperplan. Les exemples utilisés lors de la recherche de l'hyperplan ne sont alors plus utiles et seuls ces vecteurs supports sont utilisés pour classer un nouveau cas, ce qui peut être considéré comme un avantage pour cette méthode.

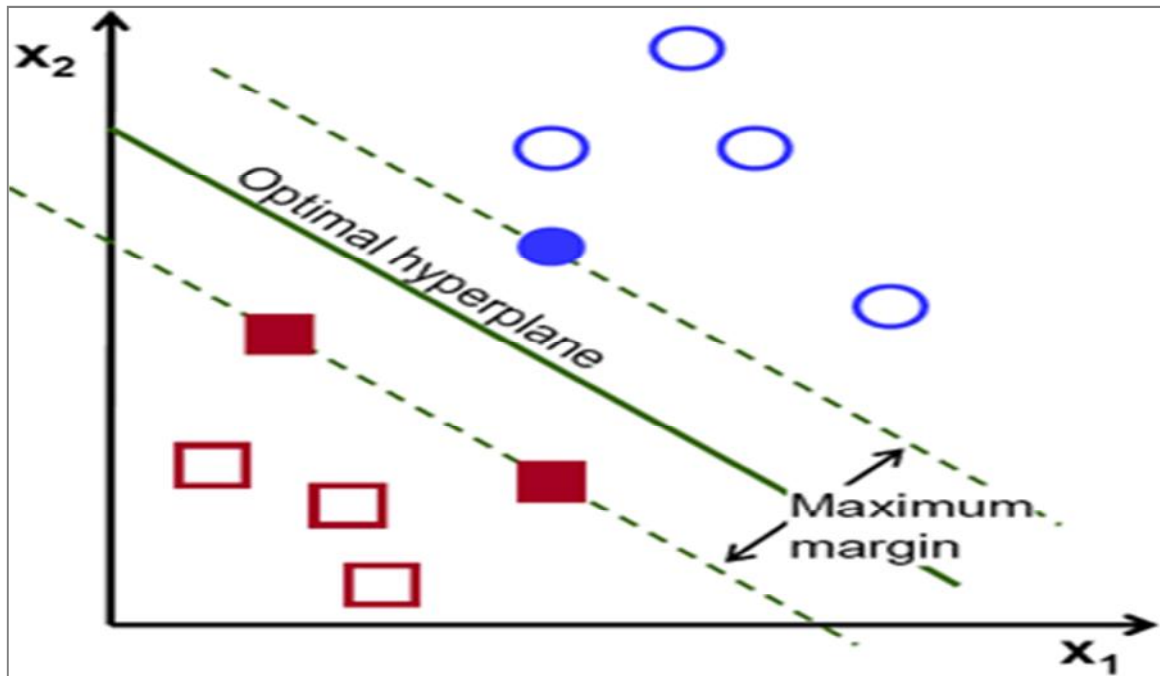


Figure 13 : Illustration de la marge maximale et l'hyperplan optimal

### II-8) Applications des SVM

Les SVM peuvent être utilisés pour résoudre divers problèmes du monde réel:

- Les SVM sont utiles dans la catégorisation de texte et d'hypertexte car leur application peut considérablement réduire le besoin d'instances de formation marquées dans les paramètres inductifs et transductifs standard.
- La classification des images peut également être effectuée à l'aide de SVM. Les résultats expérimentaux montrent que SVM atteignent une précision de recherche nettement supérieure à celle des schémas traditionnels de raffinement des requêtes après seulement trois à quatre tours de rétroaction. Cela vaut également pour les systèmes de segmentation d'image, y compris ceux utilisant une version modifiée SVM qui utilise l'approche privilégiée comme suggéré par Vapnik.
- Les caractères manuscrits peuvent être reconnus à l'aide de SVM.
- L'algorithme SVM a été largement appliqué dans les sciences biologiques et autres. Ils ont été utilisés pour classer les protéines avec jusqu'à 90% des composés classés correctement.

II-9) Classifieur linéaire

Un classifieur est dit linéaire lorsqu'il est possible d'exprimer sa fonction de décision par une fonction linéaire en  $x$ . On peut exprimer une telle fonction par:

$$h(x) = \langle w, x \rangle + b = \sum_{i=1}^n w_i x_i \dots\dots\dots(8)$$

où  $w \in \mathbb{R}^n$  est le vecteur de poids et  $b \in \mathbb{R}^0$  le biais, alors que  $x$  est la variable du problème.  $X$  est l'espace d'entrée et qui correspond à  $\mathbb{R}^n$ , où  $n$  est le nombre de composantes des vecteurs contenant les données. Notons que l'opérateur  $\langle \rangle$  désigne le produit scalaire usuel dans  $\mathbb{R}^n$ .  $w$  et  $b$  sont les paramètres à estimer de la fonction de décision  $h(x)$ .

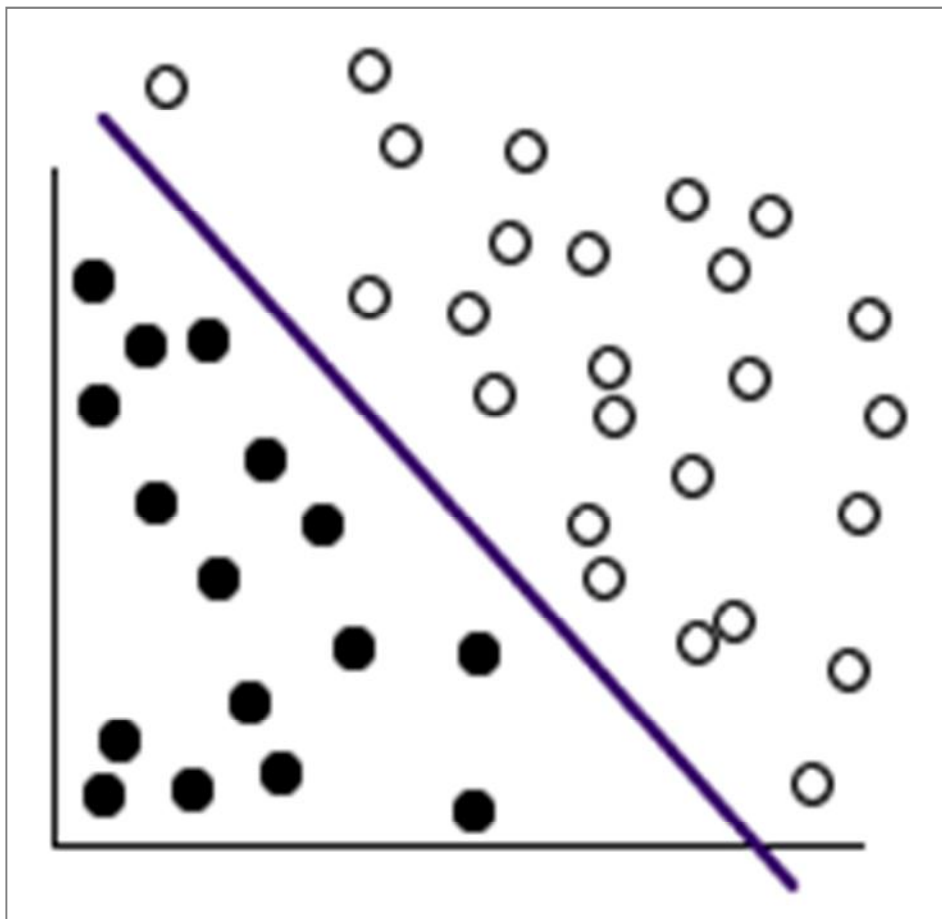


Figure 14 : cas linéairement séparable

Pour décider à quelle catégorie un exemple estimé  $x'$  appartient, il suffit de prendre le signe de la fonction de décision :  $y = \text{sign}(h(x'))$ . La fonction  $\text{sign}()$  est appelée classifieur.

Géométriquement (voir figure15), cela revient à considérer un hyperplan qui est le lieu des points  $x$  satisfaisants  $\langle w, x \rangle + b = 0$

En orientant l'hyperplan, la règle de décision correspond à observer de quel côté de l'hyperplan se trouve l'exemple. On voit que le vecteur  $w$  définit la pente de l'hyperplan ( $w$  est perpendiculaire à l'hyperplan). Le terme  $b$  quant à lui permet de translater l'hyperplan parallèlement à lui-même.

L'objectif de la discrimination linéaire est de trouver la bonne fonction de décision  $h(x)$ . La classe de tous les hyperplans qui en découle sera notée  $H$ .

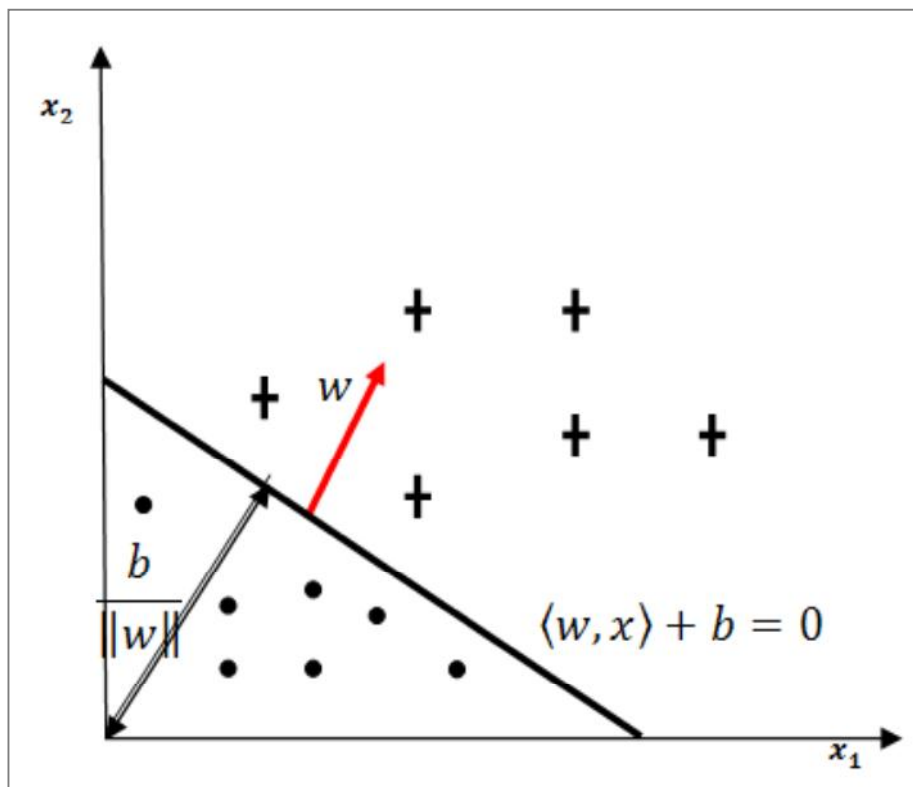


Figure 15 : hyperplan séparateur  $\langle w, x \rangle + b = 0$

**II-9-a) Marge maximale de l'hyperplan**

La marge géométrique représente la distance euclidienne prise perpendiculairement entre l'hyperplan et l'exemple  $x_i$  [10]. En prenant un point quelconque  $x_p$  se trouvant sur l'hyperplan, la marge géométrique peut s'exprimer par :

$$\frac{w}{\|w\|} \cdot x_i - x_p \dots\dots\dots(9)$$

L'hyperplan à marge maximale est le modèle le plus utilisé dans les machines à vecteurs supports. L'estimation des paramètres  $(w^*, b^*)$  de l'hyperplan qui maximise la marge se fait en résolvant le problème d'optimisation suivant :

$$(w^*, b^*) = \mathbf{argmax}_{(w,b)} \min_i \left\{ (y_i(w x_i + b)), \|w\| = 1 \right\} \dots\dots\dots(10)$$

Dire que les deux classes de l'échantillon d'apprentissage  $S$  sont linéairement séparables est équivalent à dire qu'il existe des paramètres  $(w, b)$  tels que l'on a pour tout  $i = 1, 2, \dots, n$

Ce qui est équivalent à [11]:

$$w^* x_i + b^* > 0 \quad \text{si } y_i = 1 \quad \dots\dots\dots(11)$$

$$w^* x_i + b^* < 0 \quad \text{si } y_i = -1 \quad \dots\dots\dots(12)$$

Ce qui est équivalent à :

$$y_i(w^* x_i + b^*) > 0; \quad i = 1, 2, \dots, n \quad \dots\dots\dots(13)$$

La définition consiste à dire qu'il doit exister un hyperplan laissant d'un côté toutes les données positives et de l'autre toutes les données négatives. Dès lors, nous pouvons définir deux plans se trouvant de part et d'autre de l'hyperplan et parallèles à celui-ci, sur lesquels reposent les exemples les plus proches. La figure 16 illustre cette situation. [12]

Dans notre définition de l'hyperplan, il est possible que différentes équations correspondant au même plan géométrique :

$$a(w, x + b) = 0 \quad \dots\dots\dots(14)$$

$a$  est une constante quelconque.

Il est donc possible de redimensionner  $(\mathbf{w}, b)$  de telle sorte que les deux plans parallèles aient respectivement pour équations :

$$(\mathbf{w}^* \mathbf{x}_i + b^*) = 1 \quad \dots\dots\dots(15)$$

$$(\mathbf{w}^* \mathbf{x}_i + b^*) = -1 \quad \dots\dots\dots(16)$$

Ces deux hyperplans sont appelés hyperplans canoniques. Ainsi la marge entre ces deux plans est égale à :

$$y = \frac{2}{\|\mathbf{w}\|} \quad \dots\dots\dots(17)$$

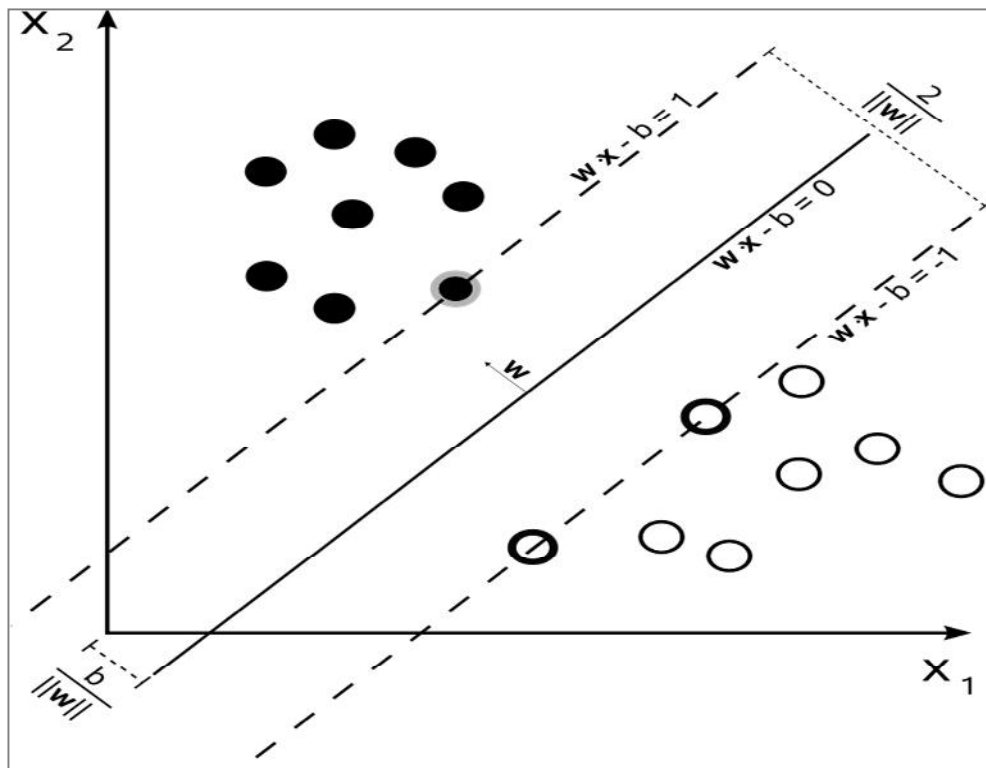


Figure 16 : hyperplans canoniques et marge maximale

**II-11) SVM non-linéaires**

Le paragraphe précédent décrit le principe des SVM dans le cas où les données sont linéairement séparables. Cependant, dans la plupart des problèmes réels, ce n'est pas toujours le cas et il est donc nécessaire de contourner ce problème (difficile de séparer n'importe quel jeu de données par un simple hyperplan). Si par exemple les données des deux classes se chevauchent sévèrement, aucun hyperplan séparateur ne sera satisfaisant. Dans ce but, l'idée est de projeter les points d'apprentissage  $x_i$  dans un espace  $T$  de dimension  $q$ , plus élevée que  $n$  grâce à une fonction non-linéaire qu'on appelle fonction noyau, choisie a priori et d'appliquer la même méthode d'optimisation de la marge dans l'espace  $T$ . L'espace  $T$  ainsi obtenu est appelé espace des caractéristiques ou aussi espace transformé.

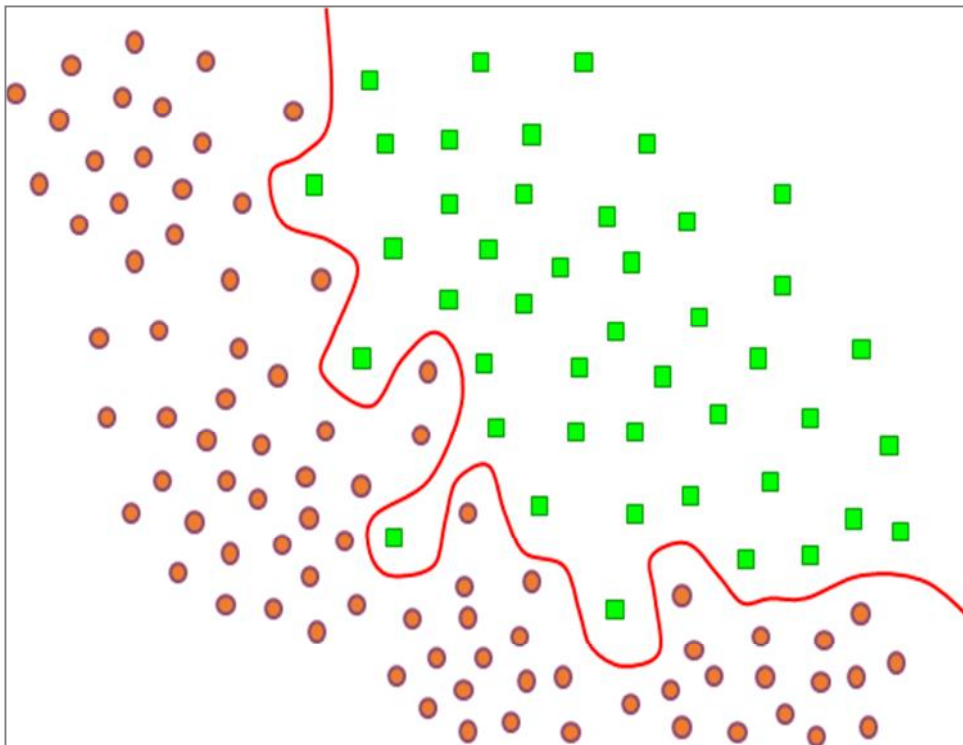


Figure 17 : cas non linéairement séparable

**II- 12) SVM pour le cas multi classes**

La plupart des problèmes ne se contentent pas de deux classes de données. Il existe plusieurs méthodes pour faire la classification multi classes. Citons les plus utilisées : La première méthode, celle que nous utilisons dans notre application, est une méthode dite Un-contre-Un. Au lieu d'apprendre N fonctions de décisions, ici chaque classe est discriminée d'une autre.

La deuxième méthode est appelé Un-Contre-Tous. C'est une approche étendant la notion de marge aux cas multi classes. Cette formulation intéressante permet de poser un problème d'optimisation unique.

**II-13) Conclusion**

Nous avons choisi de travailler avec les LBP comme descripteur de texture car avec les LBP nous prenons en considération tous les pixels de l'image pour la décrire. Comme nous avons choisi les SVM pour la classification. Les SVM possèdent des fondements mathématiques solides, et les exemples de test sont comparés juste avec les supports vecteurs et non pas avec tous les exemples d'apprentissage, et la classification d'un nouvel exemple consiste à voir le signe de la fonction de décision  $f(x)$ , une décision rapide.

# CHAPITRE II

Application et résultats pratiques

### III-1) Préambule

Le système de classification est présenté ici ainsi que les résultats de classification de différentes classes d'images de mammographies obtenus.

L'objectif de ce système est de proposer une méthode de classification des images mammographiques dans le but d'extraction des pathologies mammaires.

Nous débuterons par une présentation du système de classification pour finir par les résultats détaillés.

### III-2) environnement du logiciel

Ce travail a été réalisé avec un ordinateur portable doté du processeur Intel(R) Core (TM) i3-5005U CPU@ 2.00GHz et d'une mémoire vive RAM de 4 Go.

### III-3) Langage adopté pour la programmation

Pour la programmation, notre application a été codée par le langage de programmation Matlab de version 8.0 (R2012b). Notre choix est justifié par la souplesse et la facilité du travail.

MATLAB permet le travail interactif soit en mode commande, soit en mode programmation, tout en ayant toujours la possibilité de faire des visualisations graphiques. Il possède les particularités suivantes :

- Puissance de calcul.
- la continuité parmi les valeurs entières, réelles et complexes.
- l'étendue de gamme des nombres et leurs précisions.
- la compréhension de la bibliothèque mathématique.
- La possibilité de liaison avec les autres langages classiques de programmations.

Notre système est composé essentiellement des étapes montrées dans la figure si dessous :

- Au début nous avons la base de données MIAS qui comporte les images à utiliser.
- Utilisation de l'opérateur lbp pour l'extraction des caractéristiques, ensuite la normalisation des vecteurs caractéristiques.
- Construction de la base d'apprentissage pour avoir le modèle de décision.
- Application des svm

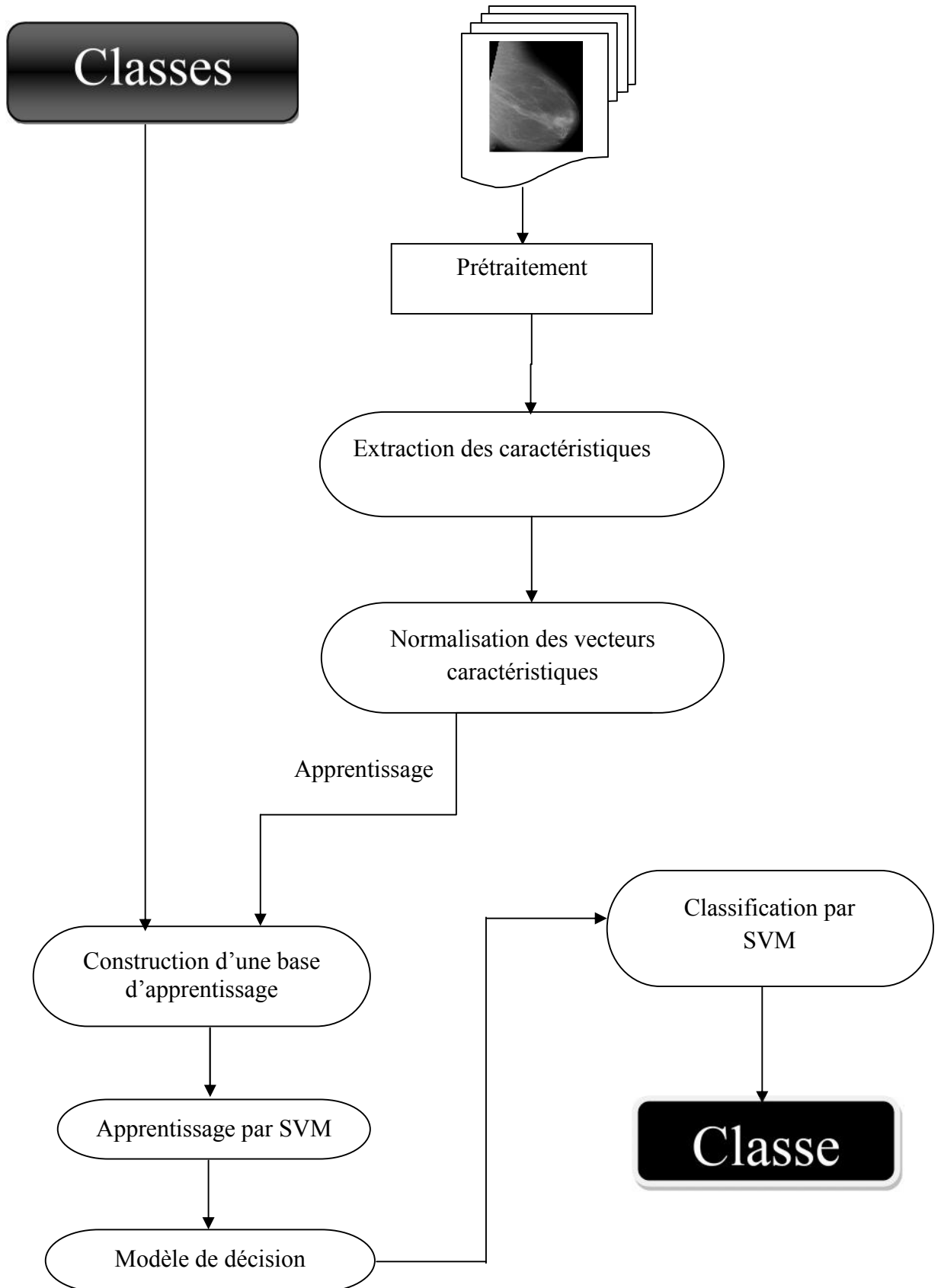


Figure18 : Système proposé pour la classification des images mammographiques.

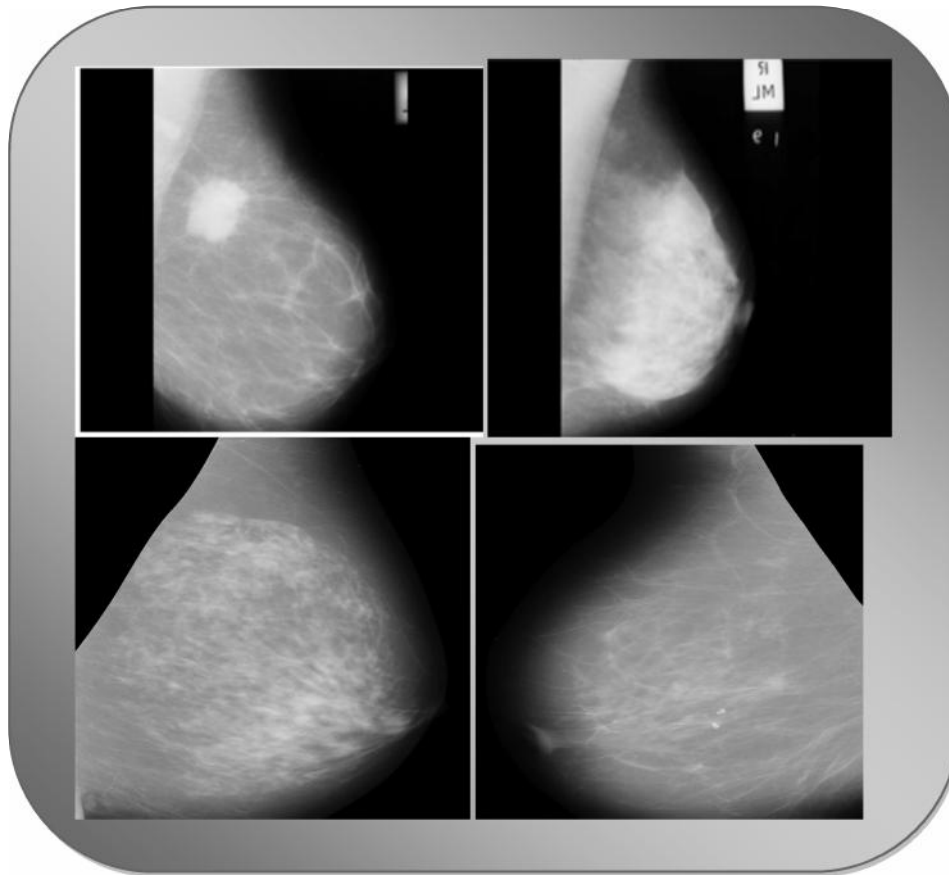
**III-4) Mammography Image Analysis Society (MIAS)**

Figure19: Echantillons des images mammographiques de la base MIAS.

La base de données MIAS [13] est le fruit de travail d'une organisation de groupes de recherche britanniques qui s'intéressent à la mammographie. Les images mammographiques sont obtenues après numérisation de films X-ray soigneusement sélectionnés à partir du programme national de dépistage fait à l'échelle du Royaume-Uni. La numérisation est faite avec un scanner Joyce-Lobel micro densitomètre qui est un dispositif linéaire dans la gamme de densité optique est de 0 à 3,2 et qui a une résolution de 50x50 micromètre. Chaque pixel est codé sur 8 bits. Toutes les images sont découpées de telle sorte qu'elles sont toutes de taille 1024x1024. Tous les noms des mammogrammes sont la forme suivante : mdb XXXBS, où :

- XXX : représentent le nombre de l'image, allant de 001 à 322 ;

- B est le côté du sein, il prend la valeur «L» pour le sein gauche ou «R» pour le sein droit (Left or Right);

- S est la taille de l'image qui peut être «s» pour la petite image (1600x4320 pixels), «m» pour des images de taille moyenne (2048x4320 pixels), «l» pour les grandes images (2600x4320 pixels) et «x» pour les très grandes images (4000x5200 pixels).

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

Pour chaque image, les radiologues ont décrit le type de la pathologie, sa localisation en donnant les coordonnées du centre de la pathologie, le diamètre d'une surface entourant la pathologie, sa taille, l'échelle, et d'autres informations utiles. Les tableaux 1.1 et 1.2 présentent les types de mammogrammes présents dans cette base de données. Sur ce tableau on peut trouver le nombre de mammogrammes contenant des calcifications, des masses, d'autres types d'anomalies, ainsi que le nombre de mammogrammes normaux, tous répartis selon le type du tissu mammaire (gras, glandulaire, ou dense).

Pathologies	Gras	Glandulaire	Dense	Total
Classification	6	9	10	25
Masse	17	15	10	42
Autres	17	15	16	48
Normal	66	65	76	207
Total	106	104	112	322

Tableau 2: Nombre de mammogrammes dans la base de données MIAS

Masses	Gras	Glandulaire	Dense	Total
Circumscribed	12	8	3	23
Spiculated	5	7	7	19
Total	17	15	10	42

Tableau 3: Nombre de mammogrammes contenant des masses dans la base de données MIAS.

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 : Grassex (Fatty)
  - G : Glandes Grassex (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.

### III-5) le système de classification

#### ➤ III-5-1) Prétraitement

Dans la partie prétraitement, nous avons appliquée une égalisation d'histogramme.

L'égalisation d'histogramme est une méthode d'ajustement du contraste d'une image numérique qui utilise l'histogramme. Elle consiste à appliquer une transformation sur chaque pixel de l'image, et donc d'obtenir une nouvelle image à partir d'une opération indépendante sur chacun des pixels. Cette transformation est construite à partir de l'histogramme cumulé de l'image de départ.

L'égalisation d'histogramme permet de mieux répartir les intensités sur l'ensemble de la plage de valeurs possibles, en « étalant » l'histogramme. L'égalisation est intéressante pour les images dont la totalité, ou seulement une partie, est de faible contraste (l'ensemble des pixels sont d'intensité proches). La méthode est rapide, facile d'implémentation, et complètement automatique. [14]

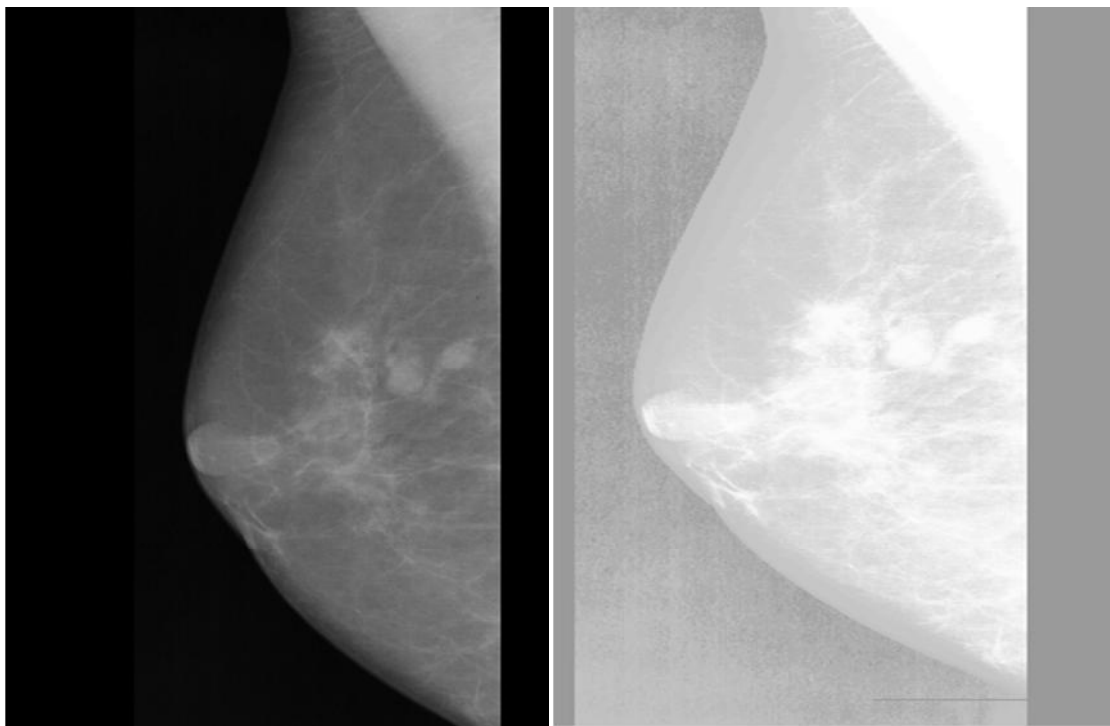


Figure 20 : Exemple de l'égalisation d'histogramme de l'image mdb059.pgm.

### ➤ III-5-2) Extraction des vecteurs caractéristiques

L'application de descripteur LBP nous permet d'extraire les informations contenue dans les images sous forme de matrices que nous avons transformer en un vecteur ligne que nous allons utiliser comme vecteur d'entrée des SVM.

Exemple d'application de LBP sur de images mammaires :

- La figure suivante nous montre le résultat de l'application de l'opérateur lbp sur une image de classe normale de la base de données MIAS,

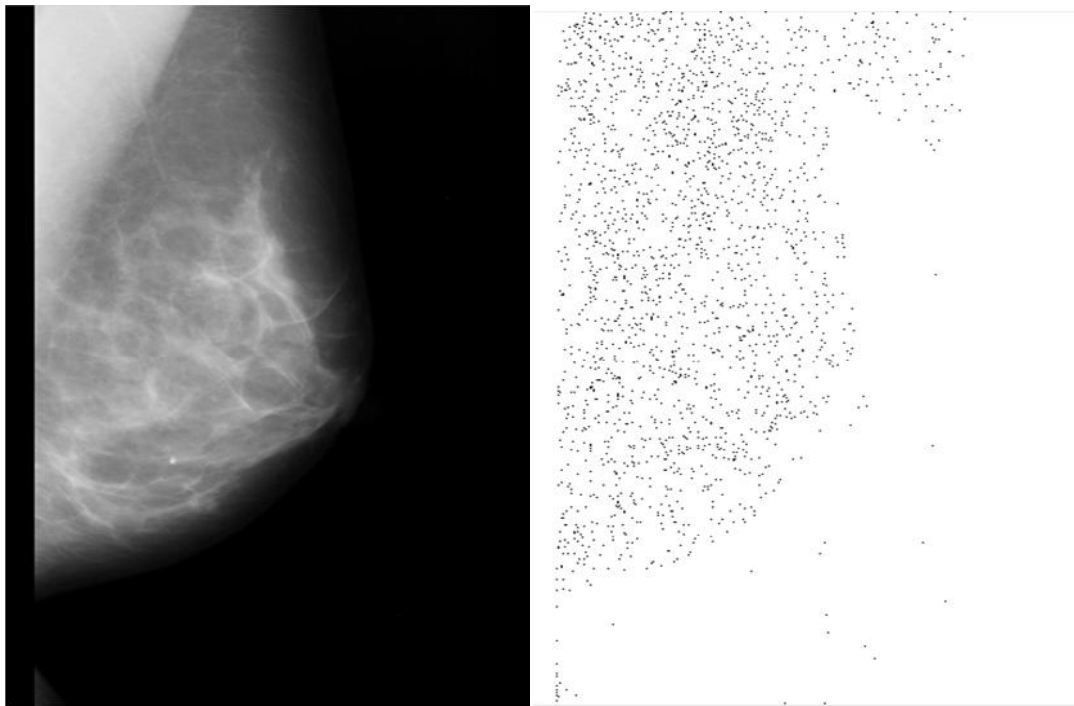


Figure 21 : Application de l'opérateur LBP d'une image normale mdb056.pgm.

- Dans cette figure nous observons le résultat d'une image maligne traitée avec lbp sous matlab.

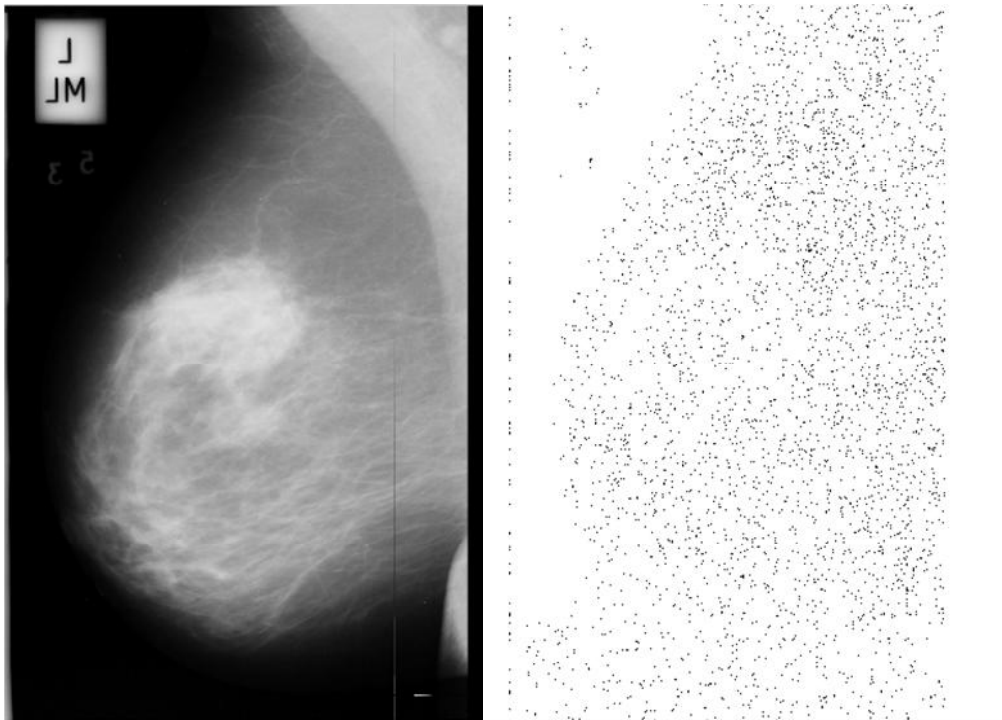


Figure 22 : Application de l'opérateur LBP sur l'image maligne mdb111.pgm.

- La figure suivante nous montre le résultat de l'application de l'opérateur lbp sur une image bénigne.

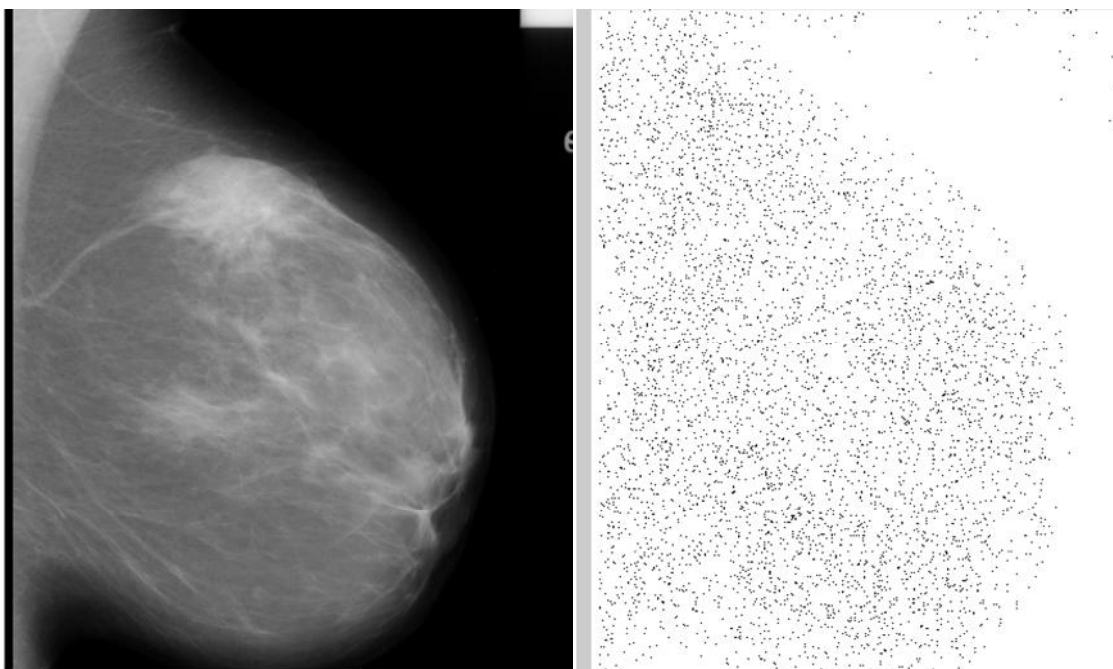


Figure 23 : Application de l'opérateur LBP sur l'image bénigne mdb150.pgm.

### ➤ III-5-3) Application des SVM

Le vecteur caractéristique obtenu est le vecteur d'entrée des SVM. Nous avons défini 3 classes soient :

Normale, bénigne et maligne pour un ensemble de 176 images dont 63 pour bénigne 51 pour maligne et 58 pour normale.

L'apprentissage et l'évaluation se font selon la règle de deux tiers un tiers c'est-à-dire un tiers pour l'évaluation et deux tiers pour l'apprentissage et la génération du modèle SVM.

Résultats d'application de système de classification :

- Pour les deux classes bénigne et maligne :

Classes	<b>b</b>	<b>m</b>
<b>b</b>	19	2
<b>m</b>	0	17

Tableau 4 : La matrice de confusion obtenue pour deux classes

Le taux de classification obtenu ici est de 94,74 % un ensemble de 116 images avec 78 images pour l'apprentissage et 38 pour le test.

- Pour les trois classes bénigne, maligne et normal:

Classe	<b>B</b>	<b>m</b>	<b>n</b>
<b>b</b>	21	0	0
<b>m</b>	4	13	0
<b>n</b>	1	4	15

Tableau 5 : La matrice de confusion obtenue pour trois classes

Le taux de classification obtenu ici est de 84,48% un ensemble de 176 images avec 118 images pour l'apprentissage et 58 pour le test.

### III-6) Interprétation

Dans ce chapitre nous avons appliqué le descripteur LBP sur chaque image. Le résultat de l'étape précédente permet de faire la classification des images de mammographies via les SVM dans un premier temps en deux classe (b : bénigne et m : maligne) ensuite en trois classes (b : bénigne, m : maligne, n : normal). La classification par SVM présente un taux satisfaisant et ouvre beaucoup de perspectives d'amélioration.

Nous avons choisi de travailler avec la base de données MIAS contenant 322 images dont 208 sont normal, 51 sont malignes et 63 sont bénignes, pour la simplicité d'exploitation et la disponibilité par rapport aux autres bases.

La méthode LBP nous a permis de représenter chaque image mammographique par un vecteur caractéristique. A l'aide de ces vecteurs qu'on a normalisé par la suite nous avons construit une règle de décision qui nous a permis de distinguer entre une image normale, une image bénigne et une image maligne.

# CONCLUSION GENERALE

# Conclusion Générale

---

## Conclusion générale

Ce travail s'est intéressé de façon générale à l'étude d'un système de classification des images mammographiques. Le choix de cancer de sein, se justifie par le fait que les statistiques de divers organismes de santé publique ont montrés que le cancer du sein représente une menace considérable pour les femmes à travers le monde et il est la première cause de mortalité par cancer chez les femmes. Malgré l'évolution des nouvelles technologies pour améliorer la qualité des images de mammographies numériques, l'interprétation et l'analyse de ces images restent une tâche difficile même pour un radiologue étant donné que ces images sont fortement texturées. D'où le besoin de développer des systèmes d'analyse et d'interprétation automatiques des images mammographiques numériques comme outil d'aide à la décision et l'interprétation par un expert. Les tests de notre système ont portés sur la base de données MIAS.

Dans la première étape, nous avons commencé par l'extraction du vecteur caractéristique à base des LBP. Dans la deuxième étape, nous avons choisit de faire la classification à travers les SVM qui nous a donné un taux de classification satisfaisant. Néanmoins beaucoup d'améliorations peuvent être apportées tels que :

- La localisation des traitements au niveau des masses, donc une segmentation des zones d'intérêts ainsi qu'un traitement spécifique aux contours afin de permettre une meilleure discrimination des masses et avoir une meilleure classification.
- Extraction des caractéristiques des images par approche LPQ ou une des autres variantes de LBP.
- Développement d'un système de classification basée sur la méthode des réseaux de neurones artificiels.

# BIBLIOGRAPHIE

---

## BIBLIOGRAPHIE

- [1] <http://dspace.univ-tlemcen.dz/handle/112/6223>
- [2] Imane-**CHEIKHROUHO** 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.
- [3] 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
- [4] [https://en.wikipedia.org/wiki/Artificial\\_neural\\_network](https://en.wikipedia.org/wiki/Artificial_neural_network)
- [5] [https://en.wikipedia.org/wiki/K-nearest\\_neighbors\\_algorithm](https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm)
- [6] <https://peipa.essex.ac.uk/pix/mias/all-mias.tar.gz>
- [7] Ojala T ,Pietikinen M & Harwood D (1996) A comparative study of texture measures a with classification based on feature distributions Pattern Recognition 29:51-59
- [8] V. Vapnik, The Nature of Statistical Learning Theory, N-Y, Springer-Verlag, 1995
- [9] Zahia **Zidelmal** épouse **AMIROU**(2012) .Reconnaissance d'arythmies cardiaques par Support Vector Machines (SVMs). Thèse Doctorat. Université Mouloud Mammeri - Tizi-Ouzou
- [10] [www.lsp.ups-tlse.fr/gadat](http://www.lsp.ups-tlse.fr/gadat)

## BIBLIOGRAPHIE

---

[11] Mounir **SEHAD**. Segmentation d'images par une approche basée sur des caractéristiques texturales, temporelles et spectrales : Application aux images MSG. Thèse de doctorat. Université Mouloud Mammeri, Tizi- Ouzou

[12] Chao **ZHU**(2012). Effective and Efficient Visual Description based on Local Binary Patterns and Gradient Distribution for Object Recognition. Doctorat Thesis. Ecole central de Lyon.

[13] <http://peipa.esssex.ac.uk/info/mias.html>

[14] [https://fr.wikipedia.org/wiki/Egalisation\\_d'histogramme](https://fr.wikipedia.org/wiki/Egalisation_d'histogramme)