

Introduction générale

Depuis l'aube des temps, l'homme pratique la classification dans sa vie quotidienne, quand il essaie de répondre aux problèmes et questions sur la catégorie des objets, c'est-à-dire d'affectation d'objets à leur classe.

Avec l'apparition des écoles de pensées scientifiques la classification a constitué l'un des thèmes majeurs, et a été utilisée dans plusieurs domaines, commençant par la taxinomie en laquelle on classe les organismes vivants en les regroupant en entités appelées taxons. Ainsi, qu'en biologie, plus précisément à l'étude des espèces vivants ce qu'on a appelle : les systèmes de classification phylogénétique, jusqu'à le grand développement des ordinateurs et des outils d'analyse des grandes données qui a permis de rendre la classification un outil primordial de nombreux domaines comme : la reconnaissance des formes, l'apprentissage et la recherche opérationnelle.

Récemment, pour des besoins en intelligence artificielle, on a eu la naissance de la classification automatique et ses applications dans le domaine médicale, en particulier la détection et l'identification des maladies tels que le cancer et l'Alzheimer en observant seulement des cellules microscopiques, en fait, pour que le spécialiste puisse donner un bon diagnostic, il faut qu'il y ait une bonne lisibilité et une bonne reconnaissance des cellules anormales des autres cellules qui composent une image médicale.

L'imagerie médicale est certainement l'un des domaines de la médecine qui a le plus progressé ces vingt dernières années. Ces récentes découvertes permettent non seulement un meilleur diagnostic mais offrent aussi de nouveaux espoirs de traitement pour de nombreuses maladies. Cancer, épilepsie... l'identification précise de la lésion facilite déjà le recours à la chirurgie, seule solution thérapeutique pour certains malades. De telles techniques permettent également de mieux comprendre le fonctionnement de certains organes encore mystérieux, comme le cerveau.

La présence des images médicale avec une bonne qualité rend l'utilisation des techniques et méthodes de l'intelligence artificielle plus intéressante, surtout pour l'aide dans la tâche du médecin spécialiste dans le diagnostic. Les systèmes informatiques d'aide au diagnostic (CAD) sont en amélioration constante et sont capables de détecter et de caractériser des images mammographiques suspectes, foyers de microcalcifications, masses ou plus difficilement, distorsion architecturale. L'analyse numérique et informatique des clichés de mammographie

peut alors servir de deuxième lecture informatique pour le radiologue, le but de ces systèmes experts étant d'augmenter la sensibilité de la mammographie tout en réduisant le nombre de faux positifs. Plusieurs études récentes ont montré que ces systèmes (CAD pouvaient augmenter le taux de cancers détectés en dépistage de 15 à 20 %.

Notre objectif est d'élaborer un système de classification automatique des images de mammographie en trois classes ; normales, bénignes ou malignes. Ce système sert à assister le médecin spécialiste à la prise de décision pour la phase de diagnostic du cancer du sein. Le système élaboré dans notre projet de fin d'études comporte trois étapes essentielles :

- Prétraitements des images mammographiques pour l'élimination de tous ce que n'en a pas besoin dans le diagnostic,
- Extraction et sélection des caractéristiques de l'image à l'aide de la transformée en cosinus discrète (DCT),
- La classification des images par un réseau de neurones artificiel.

Pour cela nous avons réparti notre travail en 3 chapitres, comme suit :

Le premier chapitre : contient des généralités sur les différents types des images médicales ainsi que les tumeurs et ses classifications.

Le deuxième chapitre : montre les systèmes d'aide au diagnostic (CAD) avec les différentes étapes nécessaires pour la conception de tels systèmes.

Le troisième chapitre : contient les résultats expérimentaux du système implémenté pour la classification du cancer du sein en utilisant le langage MATLAB [1].

1. L'imagerie médicale

1.1. Le scanner

Le scanner, également appelé tomographie axial assisté par ordinateur (TAO), est un dispositif de radiographie associant des rayons X et des traitements informatiques, permettant d'obtenir une image des plans de coupe d'un objet, en particulier du corps humain, par des mesures de densité. Le scanographe appelé également Scanner en raison de l'analyse par balayage qu'il effectue (de l'anglais scanning) reconstitue l'image du corps en mesurant la densité des rayonnements X à travers le corps humain sous différents angles (tomodensitométrie).

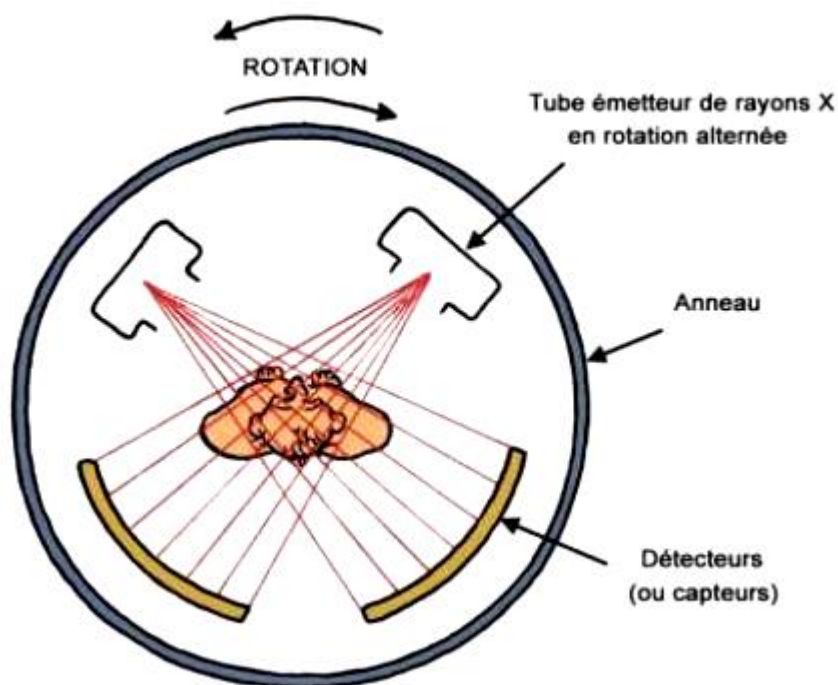


Figure 1.1 : Principe de scanner

Le tube à rayons X et les détecteurs tournent autour de l'objet à examiner de multiples profils d'atténuation obtenus à des angles de rotation différents transformés en niveaux de gris formant ainsi l'image analogique.

La tomodensitométrie (TDM) repose sur deux propriétés essentielles :

- La mesure de la densité radiologique des volumes élémentaires d'une coupe à partir de l'absorption d'un faisceau de rayons X, ce qui permet d'accroître la résolution en densité de l'image d'une coupe de corps d'un facteur 100 par rapport à ce qu'elle est en radiologie conventionnelle.
- La reconstitution d'une image en coupe du corps humain à partir des différentes projections transversales obtenue par le système constituant l'appareillage.

1.2. L'imagerie par résonance Magnétique IRM

L'IRM ou imagerie par résonance magnétique est une technique non invasive, d'une totale innocuité, basée sur le principe de la résonance des atomes de certaines molécules (l'hydrogène en imagerie diagnostique) sous l'action de certaines ondes de radiofréquence. L'appareil est constitué d'un tunnel formé d'un aimant très puissant (0.1 et 1,5 Tesla) entourant le lit d'examen sur lequel s'allonge le patient. Des antennes spécifiques y sont connectées. Certaines émettent une onde radiofréquence qui stimule les noyaux d'hydrogène contenus dans l'eau composant nos cellules.

Après simulation (qui dure à peine quelques millisecondes), les atomes d'hydrogène restituent cette énergie qui se dissipe dans différents plan de l'espace sous l'action du champ magnétique de l'aimant. L'énergie est alors captée par d'autres antennes dites antennes réceptrices puis analysée par un puissant ordinateur qui construit alors une véritable carte énergétique de la partie du corps étudiée.

Suivant la composante en eau des tissus analysés, leurs vascularisations et leurs pathologies éventuelles, les images seront différentes et l'ordinateur réalise des images en noir et blanc d'une très grande sensibilité qui est très précieuse pour le diagnostic, notamment en matière de pathologies tumorales ou infectieuses. Il est ainsi possible de réaliser des coupes de n'importe quelle partie du corps, dans tous les plans de l'espace. Cette technique permet d'obtenir des résultats similaires à une étude à ciel ouvert.

L'inconvénient majeur réside dans le fait que le patient doit respecter une stricte immobilité pendant toute la durée de l'examen (environ une quinzaine de minutes). Par ailleurs, le fait d'être enfermé dans un tunnel est parfois difficile à supporter pour certaines personnes. De plus du fait

du champ magnétique puissant pouvant déplacer les objets dits ferromagnétiques, il existe certaines contre-indications : les valves cardiaques artificielles métalliques, les stimulateurs cardiaques, certains clips cérébraux ou certaines prothèses ou objets métalliques. Il va de soi que tous les objets métalliques ou magnétiques doivent être déposés à l'extérieur de la salle d'examen (clefs, pièces de monnaie, carte bancaires...). La salle d'examen est par ailleurs protégée (blindée) par une protection spécifique qui la met à l'abri des perturbations ou pollutions magnétiques extérieures.



Figure 1.2 : Image par IRM.

1.3. Echotomographie

L'échotomographie utilise les ultrasons, qui sont des ondes acoustiques inaudibles par l'être humain. Il s'agit donc d'une technique qui ne comporte pas les risques des explorations accomplies par des rayons X, d'où son intérêt en particulier pour les explorations chez la femme enceinte (Figure 1.3).



Figure 1.3 : Echotomographie d'une femme enceinte

1.4. La mammographie

Il est devenu de plus en plus évident ces dernières années que l'obtention de clichés mammographiques de très haute qualité n'est pas seulement souhaitable, mais est devenu indispensable pour assurer le progrès des programmes de dépistage. Le développement de tels programmes, destinés à détecter les petits cancers du sein dans des populations asymptomatiques, a donné un énorme élan à la recherche pour améliorer la qualité des clichés. La qualité de l'image en mammographie est à la fois difficile à définir et à mesurer, comme le montre l'absence d'un test de qualité de l'image (Figure 1.4).

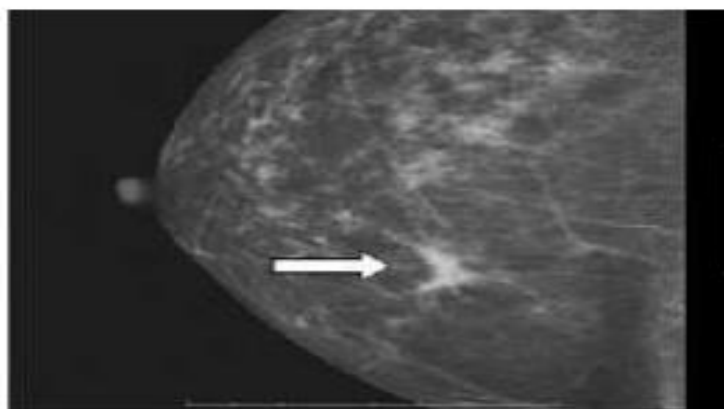


Figure 1.4 : Cliché du sein d'une femme

L'objet-cible consiste en une matrice de résine dont la densité radiologique est intermédiaire entre celle de l'eau de la graisse. Plusieurs objets sont placés au sein de cette matrice afin de tester, soit séparément, soit ensemble, la résolution et la définition à faible contraste. La taille et la densité de ces objets s'échelonnent de celles visibles sur tous les films à celles qui n'ont pu être visualisées sur aucun film existant, et inclut également une série d'objets intermédiaire qui deviennent visibles ou invisibles selon la qualité de l'image. Lorsque ces tests sont élaborés, on vérifie leur fiabilité en faisant varier la tension et le filtrage. Un score de qualité d'image, défini à partir des films obtenue avec ce test, a montré sans surprise l'importance de la tension et le filtrage : ce score chute à peu près d'un facteur 1 l'accroissement de la tension de 2KV, et d'un facteur 1 pour une épaisseur 0.14 mm d'aluminium supplémentaire de filtration. Lors du test, différents films sont exposés dans des conditions standard, afin de déterminer une densité de basse dans une partie [2].

2. Le cancer du sein : Il existe un certain nombre de cancers différents, ayant chacun leurs propres symptômes et pouvant toucher diverses parties du corps. Parmi les cancers les plus fréquents, on retrouve :

- Le cancer du sein
- Le cancer du colon
- Le cancer de la prostate
- Le cancer de la peau
- Le cancer du poumon
- Le cancer du foie
- Le cancer de l'estomac
- Le cancer de la vessie.

Le cancer du sein est la forme de cancer la plus fréquemment diagnostiquée chez les femmes, représentant environ 30% de tous les nouveaux cas de cancer chaque année. En 2003, plusieurs cas ont été diagnostiqués et 5 300 femmes sont décédées de cette maladie. En 1995, le cancer du sein représentait 97 000 de vies perdues pour les femmes canadiennes. Depuis 1984, les taux d'incidence du cancer pour tous les âges ont augmentés progressivement alors que les taux de mortalité sont restés relativement stable. Une femme sur neuf devrait développer un cancer du

sein au cours de sa vie, et un sur 25 devrait en mourir. La mammographie joue un rôle central de détecter des anomalies dans le cancer du sein en train de dépistage. Bien que la mammographie soit considérée comme le moyen le plus fiable de détection du cancer du sein, entre 10% et 30% des femmes chez lesquelles un cancer du sein a ensuite été diagnostiqué ont des mammographies faussement négatives. La classification des tumeurs est fondée sur leur organe ou tissu d'origine, leur type histologique, et leur degré de malignité. Le diagnostic est fondé sur l'histologie, mais fait de plus en plus souvent appel à des techniques complémentaires telles que l'immunohistochimie, la cytogénétique et la biologie moléculaire. Le terme « tumeur » désignait autrefois toute augmentation de volume localisé déformant un organe ou une partie du corps. Elle réunissait des lésions différentes, notamment :

- des collections liquidiennes collectées dans une cavité préformée ;
- des tuméfactions d'origine inflammatoire ;
- des hypertrophies tissulaires d'origine dystrophique (goitre) ;
- des lésions liées à des désordres d'origine embryologique (dysembryoplasies). La définition actuelle est plus restrictive et repose sur la notion d'homéostasie des tissus.

Le développement et la croissance d'un tissu ou d'un organe sont conditionnés par des processus complexes permettant la régulation des différentes étapes de la vie d'une cellule : la prolifération, la différenciation, la sénescence et la mort cellulaire programmée. Tous ces phénomènes sont aussi mis en jeu lors du renouvellement des cellules dont la durée de vie est limitée. Au sein d'un tissu, l'équilibre entre ces processus est à l'origine de l'homéostasie tissulaire. Les anomalies de cette homéostasie, par augmentation de la prolifération et/ou diminution de la mort cellulaire, sont à l'origine de l'accumulation des cellules aboutissant à la formation d'une tumeur macroscopiquement visible. Le terme de tumeur (synonyme : « néoplasme » ou « néoplasie ») désigne actuellement une prolifération cellulaire excessive aboutissant à une masse tissulaire ressemblant plus ou moins au tissu normal homologue (adulte ou embryonnaire), ayant tendance à persister et à croître, témoignant de son autonomie biologique.

3. La tumeur

3.1. Caractères d'une tumeur

a) *Prolifération cellulaire excessive*

La prolifération est liée à la multiplication des descendants d'une ou plusieurs cellules anormales. C'est la notion de clonalité. Un clone est un ensemble de cellules dérivées d'une seule cellule initiale. Une tumeur est dite poly-oligo- ou monoclonale, selon qu'elle se développe à partir de plusieurs, quelques ou une seule cellule.

b) *Masse tissulaire ressemblant plus ou moins à un tissu normal :*

Les caractères cytologiques et architecturaux de ce nouveau tissu réalisant un aspect plus ou moins proche de celui du tissu normal homologue adulte ou embryonnaire. Cette ressemblance définit une notion fondamentale : la différenciation tumorale. Plus la fonction et la structure tumorales se rapprochent de la fonction et de la structure du tissu normal, plus la tumeur n'est dite différenciée (Figure 1.5).

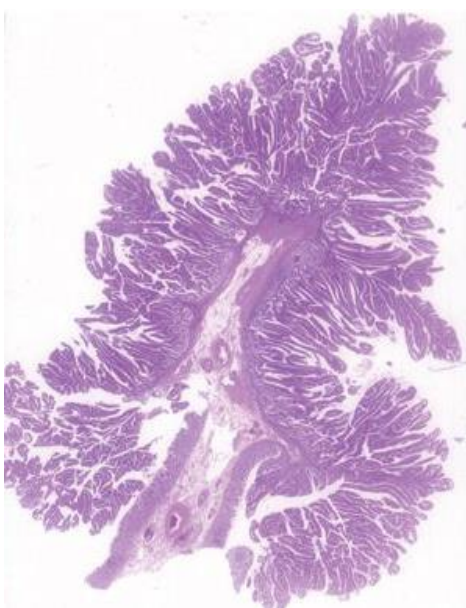


Figure 1.5 : Tumeur bénigne

c) Tendance à persister et à croître

La prolifération tumorale se poursuit après la disparition du « stimulus » qui lui a donné naissance. La prolifération tumorale est biologiquement autonome.

d) Succession d'événements génétiques

Ces anomalies génétiques s'accumulent généralement en plusieurs années. Au cours de ce processus en plusieurs étapes, le génome des cellules tumorales acquiert des allèles mutants de proto-oncogènes, de gènes suppresseurs de tumeur et de gènes contrôlant directement ou indirectement l'intégrité de l'ADN.

La conséquence de ces anomalies génétiques est l'acquisition de nouvelles propriétés, dont :

- la capacité de générer leurs propres signaux mitogènes ;
- de résister aux signaux externes d'inhibition de la croissance ;
- de proliférer sans limite (immortalisation) ;
- d'infiltrer les tissus adjacents ;
- de constituer une néo-vascularisation (angiogénèse).

3.2. Catégories de tumeurs

Il existe deux catégories de tumeurs :

1. Les tumeurs bénignes.
2. Les tumeurs malignes ou cancers.

Ces deux catégories de tumeurs s'opposent par leurs caractères :

- Macroscopiques.
- Histologiques.
- Evolutifs [3].

3.2.1. Les tumeurs bénignes

Les tumeurs bénignes du sein sont des formations nodulaires de la glande mammaire. La glande mammaire est constituée de lobules et de canaux galactophores entourés de tissu conjonctif dense dans lequel cheminent les vaisseaux. Les canaux principaux s'ouvrent à l'extérieur au niveau du mamelon et en profondeur. Après s'être divisés en canaux plus étroits, de moyens et petits

calibres, ils se terminent par l'unité terminale ductulo-lobulaire de la glande mammaire. Une tumeur maligne a les caractéristiques générales suivantes :

- *Caractères évolutifs*

Les tumeurs bénignes se développent localement et restent cantonnées au tissu dans lequel elles ont pris naissance. Leur croissance est lente. Toutefois, elles peuvent atteindre un volume et un poids importants. Elles ne récidivent pas après ablation chirurgicale, à condition que l'exérèse soit complète. Ces tumeurs ne métastasent jamais. Leur évolution est généralement favorable. Toutefois dans certains cas, elles peuvent être la cause de complications graves voire mortelles, en raison de leur siège ou de désordres métaboliques. Exemples : 1- un méningiome du trou occipital, situé dans un orifice non expansible, peut avoir une évolution mortelle en provoquant un engagement du tronc cérébral à travers l'orifice occipital. 2- un adénome parathyroïdien est responsable d'une hyperparathyroïdie et par conséquent d'une hypercalcémie parfois dangereuse.

- *Caractères Macroscopiques*

Il s'agit de tumeurs circonscrites, bien limitées, nettement séparées des tissus avoisinants, parfois même entourées par une capsule (coque faite de tissu conjonctif). Cette limitation explique la facilité de l'exérèse chirurgicale et la possibilité d'une exérèse limitée à la seule tumeur. (Exemple : adénofibrome du sein, léiomyome de l'utérus).

- *Caractères histologiques*

Le tissu tumoral reproduit de très près la structure du tissu initial (tumeur différenciée). Les cellules ont une morphologie normale et ne présentent aucun caractère de malignité. Il n'y a pas d'envahissement des tissus voisins. Les tumeurs bénignes refoulent sans les détruire les tissus sains de voisinage : elles sont expansives. (Exemple : adénome du foie).

3.2.2. Les tumeurs malignes

Une tumeur maligne est une prolifération indéfinie d'une lignée cellulaire dont l'évolution spontanée est la mort de l'individu porteur, habituellement liée à l'extension de la tumeur à tout l'organisme. Une tumeur maligne a les caractéristiques générales suivantes :

- *Caractères évolutifs*

Les tumeurs malignes ont habituellement une croissance rapide. Elles donnent naissance à une dissémination tumorale à distance (surtout par voie lymphatique et sanguine) avec éclosion et développement de tumeurs secondaires dans d'autres viscères : les métastases. Les tumeurs malignes ont tendance à récidiver après éradication locale. L'évolution, en dehors des traitements se fait spontanément vers la mort.

- *Caractères Macroscopiques*

Les tumeurs malignes sont mal limitées, non encapsulées ; elles détruisent et envahissent l'organe dans lequel elles ont pris naissance, ainsi que les organes de voisinage. Leurs contours sont irréguliers. Les foyers de nécrose et d'hémorragie sont habituels.

- *Caractères histologiques*

Les cellules tumorales malignes présentent habituellement des caractères anormaux (caractères cytologiques de malignité). Le tissu tumoral est plus ou moins différencié. Il « caricature » le tissu normal orthologue [4].

4. Conclusion

Le but de ce chapitre est de présenter différents types d'images médicales et spécialement le cancer du sein. Alors, nous avons donné un aperçu sur quelques techniques d'acquisition des images médicales. Ces techniques permettent de visualiser une image d'une partie d'un corps humain ou d'un organe donné pour apporter plus de précision à un diagnostic médical. Ensuite nous avons intéressé au problème du cancer du sein ou nous avons montrés les deux types des tumeurs du sein : bénignes et malignes.

D'un point de vue informatique, le traitement et l'analyse des images ont donné lieu à un certain nombre d'applications automatiques dont le but est d'aider le médecin à améliorer sa lecture et par voie de conséquence son diagnostic médical. Le chapitre suivant est consacré à la présentation de systèmes informatiques automatisé d'aide au diagnostic [5].

Introduction

Le cancer du sein est le cancer le plus répandu et le principal cause de morbidité et de mortalité chez les femmes entre 50 et 74 ans à travers le monde. Les statistiques récentes ont montré qu'une femme sur huit aux États-Unis et une femme sur 10 les femmes européennes développent un cancer du sein au cours de leur vie. Le cancer du sein est donc un problème majeur de santé publique, et la meilleure stratégie pour lutter contre le cancer du sein est la précoce détection. Pour cette raison, la mammographie reste l'outil le plus précis et le meilleur pour la détection précoce du cancer du sein.

La détection assistée par ordinateur (*CADe*) est le processus d'identification des anomalies potentielles avec une mammographie, classant les régions d'une mammographie comme positif ou négatif. Les systèmes *CADe* utilisent les algorithmes de traitement d'images pour analyser les possibles anomalies de mammographies, y compris masses, micro calcifications, distorsions et densités symétriques dans l'architecture du sein. Dans un examen initial des images mammographies, un radiologue examine les régions suspectes suggérées par le Système *CADe* et détermine s'ils méritent d'être poursuivis. Le système *CADe* assiste le radiologue en confirmant la détection de régions suspectes ou identifier ceux qui auraient autrement été manqués.

Face à l'augmentation du nombre de mammographies au cours des dernières décennies et la difficulté de lecture et d'interprétation des mammographies, différentes recherches font l'effort, soit pour détecter automatiquement les lésions mammaires à l'aide de systèmes de détection assistée par ordinateur (Communément appelé *CADe*), soit interpréter automatiquement les mammographies par des systèmes de diagnostic assistés par ordinateur (communément appelé *CADx*). Ces systèmes sont utilisés comme un complément à l'évaluation des radiologues.

1. Architecture d'un système CAD

En règle générale, la procédure de développement d'un système de diagnostic assisté par ordinateur (*CAD*) permettant de diagnostiquer des régions suspectes dans les mammographies se déroule en trois étapes fondamentales (Figure 2.1) :

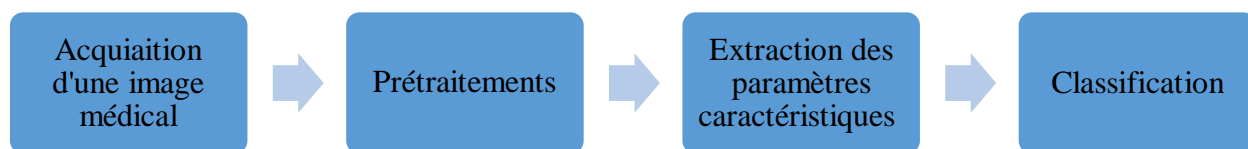


Figure 2.1 : Processus de classification des images médicales

- Etape des prétraitements : elle consiste à préparer les mammographies pour les prochaines étapes ou opérations. Dans cette étape, on peut utiliser des algorithmes pour la détection de régions d'intérêt, qui consistent à analyser la mammographie et extraire les informations nécessaires, par exemple, la segmentation.
- Extraction de caractéristiques : dans cette étape, nous pouvons extraire des paramètres caractéristiques ou primitifs discriminantes pour la classification.
- Classification des ROI : le but de cette étape est de classer les mammographies comme normales ou anormal et malin ou bénin [6].

2. Les prétraitements :

Le but de cette étape est d'éliminer, le plus possible, les informations non pertinentes dues au bruit résultant de l'acquisition de l'image, et par conséquent, faciliter l'extraction des informations utiles à l'analyse. Plusieurs méthodes de base pour le traitement des signaux et des images peuvent être utilisées dans cette étape.

3.1. Rehaussement de contraste

Un traitement de rehaussement de contraste peut être appliqué aux images rayon X. Ce traitement consiste à accentuer les caractéristiques d'une image afin de rendre son affichage plus convenable à l'analyse. L'objectif est d'augmenter le contraste d'images afin d'accroître la séparabilité des régions [7].

3.2. Filtrage

Pour améliorer la qualité visuelle de l'image, on doit éliminer les effets de bruit (parasites) en lui faisant subir un traitement appelé filtrage. Le principe du filtrage est de modifier la valeur des pixels d'une image, généralement dans le but d'améliorer son aspect. En pratique, il s'agit de créer une nouvelle image en se servant des valeurs des pixels de l'image d'origine. Parmi les filtres linéaires et non linéaires de base on peut trouver :

a) Filtre passe-bas (lissage)

Ce filtre n'affecte pas les composantes de basse fréquence dans les données d'une image, mais doit atténuer les composantes de haute fréquence.

b) Filtre passe-haut (accentuation)

Le filtre passe-haut a les caractéristiques inverses du filtre passe-bas. Ce filtre n'affecte pas les composantes de haute fréquence d'un signal, mais doit atténuer les composantes de basse fréquence [16]. Par exemple, le renforcement des contours et leur extraction s'obtiennent par l'application d'un filtre passe-haut.

c) Filtres médians

C'est un filtre non-linéaire utilisé pour éliminer le bruit de type poivre et sel. Un filtre passe-bas a tendance à rendre l'image floue et donc perdre de l'information sur les contours caractérisés par des fortes variations d'intensité. Pour diminuer cet effet, on prend la valeur médiane sur un voisinage du pixel, c'est le filtre médian.

3.3. Les histogrammes

Un histogramme est un graphique statistique permettant de représenter la distribution des intensités des pixels d'une image, c'est à dire le nombre de pixels pour chaque intensité lumineuse. Par convention un histogramme représente le niveau d'intensité en abscisse en allant du plus foncé (à gauche) au plus clair (à droite). Ainsi, l'histogramme d'une image en 256 niveaux de gris sera représenté par un graphique possédant 256 valeurs en abscisses, et le nombre de pixels de l'image en ordonnées.

Un histogramme est utile pour le choix du seuil lors de la séparation de deux régions. Par exemple, sur la figure 2.2, on distingue clairement trois classes différentes, classe de fond, classe du sein et classe regroupe les annotations, bande d'intensité et les calcifications :

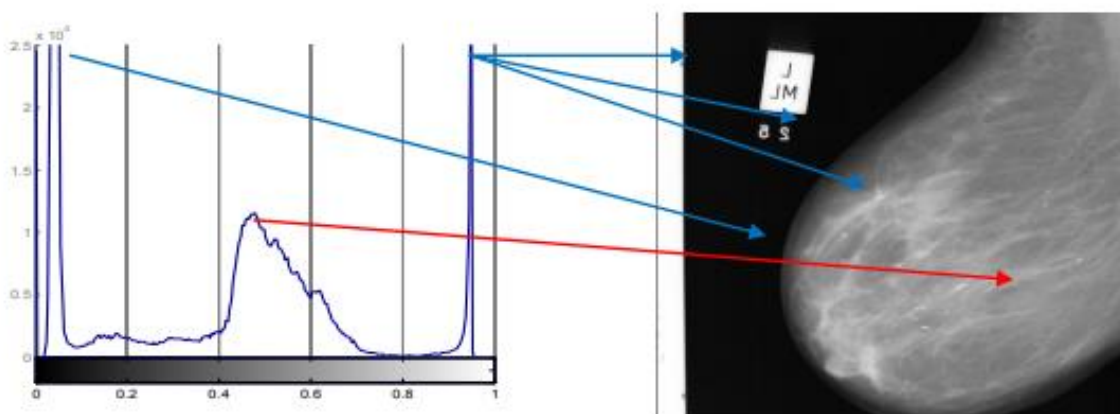


Figure 2.2 : L'histogramme typique d'une mammographie.

- Classe 1 : (à gauche) elle regroupe tous les pixels ayant des valeurs de (faible d'intensités), le grand pic correspondant aux pixels de fond.
- Classe 2 : (au milieu) elle regroupe tous les pixels ayant des valeurs grises correspondant au sein lui-même.
- Classe 3 : (à droite) elle regroupe tous les pixels de haute intensité, le grand pic correspondant au muscle pectoral, aux annotations et peut être à des lésions mammaires [8].

3.4. La segmentation d'image

La segmentation d'image est une opération de traitement d'images qui a pour but de rassembler des pixels entre eux suivant des critères prédéfinis. Les pixels sont ainsi regroupés en régions, qui constituent un pavage ou une partition de l'image. Il peut s'agir par exemple de séparer les objets du fond. Si le nombre de classes est égal à deux, elle est appelée aussi binarisation ou seuillage (*Thresholding*).

La segmentation est une étape primordiale en traitement d'image. Si l'homme sait naturellement séparer des objets dans une image c'est grâce à des connaissances de haut niveau (compréhension

des objets et de la scène). Mettre au point des algorithmes de segmentation de haut niveau (chaque région est un objet sémantique) est encore un des thèmes de recherche les plus courants en traitement d'images.

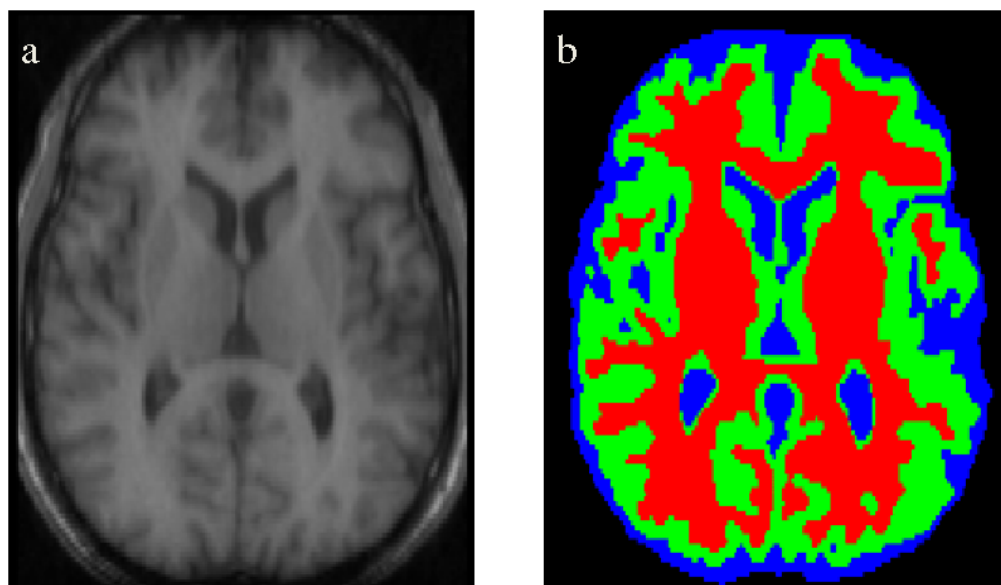


Figure 2.3 : Exemple de segmentation d'image IRM en plusieurs régions

- **Extraction des paramètres caractéristiques**

Les caractéristiques ne sont que des motifs observables dans l'image qui donnent des informations sur l'image. La précision de la classification dépend de l'étape d'extraction des caractéristiques. Plus spécifiquement, les caractéristiques peuvent faire référence au résultat d'une opération de voisinage général appliquée à l'image, à des structures spécifiques dans l'image même, allant de structures simples telles que des points ou des arêtes à des structures plus complexes telles que des objets.

4.1. Caractéristiques du domaine spatial

De nombreuses caractéristiques peuvent être extraites dans le domaine spatial pour la détection des anomalies des mammographies. Par exemple, les méthodes d'extraction des caractéristiques

de texture jouent un rôle très important dans la détection des anomalies des mammographies en raison de la nature des mammographies. Les caractéristiques de texture se sont avérées utiles pour différencier les masses et les tissus mammaires normaux. Les caractéristiques de texture permettent d'isoler les lésions normales et anormales avec des masses et une micro-calcification. Les caractéristiques de texture les plus connues sont les suivantes :

- Périmètre
- Surface
- La compacité
- Épaisseur
- Orientation
- Direction
- Ligne
- Fond
- premier plan
- Distance [9].

4.2. Caractéristiques du domaine fréquentiel :

Ces caractéristiques sont extraites dans le domaine fréquentiel de l'image par l'utilisation d'une transformation en cosinus discrète (DCT) qui donne une autre représentation de l'image.

La transformation en cosinus discrète (DCT) exploite les fonctions de cosinus, elle transforme un signal de la représentation spatiale en domaine de fréquence. La DCT représente une image comme une somme de sinusoides de magnitudes et de fréquences variables. La transformée F d'une image M de dimension $N \times N$ est calculée par la formule suivante :

$$F(i, j) = \frac{1}{\sqrt{2}} C(i) C(j) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} M(x, y) \cos\left(\frac{2x+1}{2N} i\pi\right) \cos\left(\frac{2y+1}{2N} j\pi\right)$$

Avec :

$$C(x) = \begin{cases} \frac{1}{\sqrt{2}} & \text{si } x = 0 \\ 1 & \text{ailleurs} \end{cases}$$

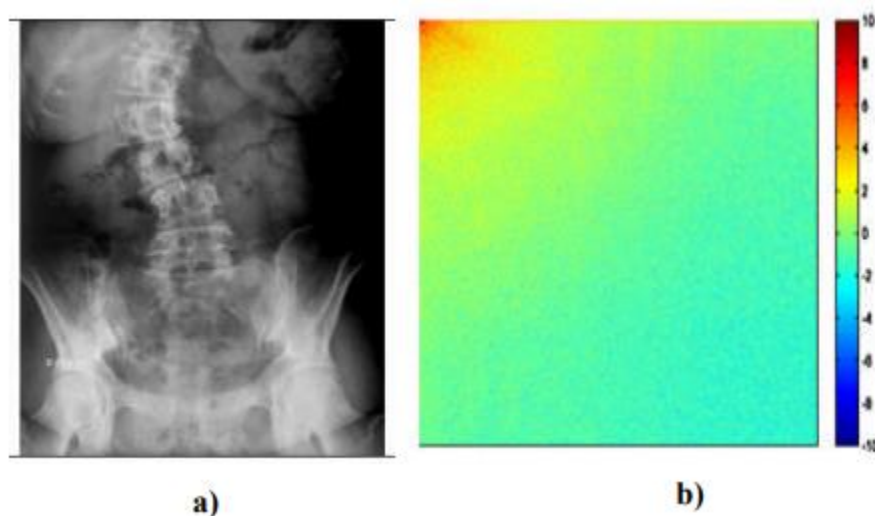


Figure 2.4 : a) Image rachis_scoliose b) DCT de l'image

La DCT a la propriété que, pour une image typique, la plupart des informations visuellement significatives sur une image sont concentrées uniquement dans quelques coefficients de DCT.

Un exemple de la transformée DCT est montré sur la figure 2.4. Les coefficients DCT à haute amplitude sont représentés par des points rouges, les coefficients DCT à très faible amplitude sont représentés par des points verts et les coefficients DCT à haute amplitude mais de signe négative sont représentés par des points bleus. On voit bien que l'énergie de l'image rachis scoliose est concentrée en haut de l'image sur le côté gauche dans les faibles fréquences. Alors, nous avons une variation faible de l'intensité sur une large surface et le détail spatial est assez faible [10].

- **La classification**

C'est la dernière étape dans un système de classification automatique des images médicales. Dans notre projet nous avons utilisé les réseaux de neurones artificiels (*Artificial Neural Networks*) comme classificateur.

5.1. K-Plus proches voisins (KNN)

La méthode des k plus proches voisins est une méthode classique d'apprentissage supervisé non paramétrique où une nouvelle observation est classée dans la classe d'appartenance de l'observation de l'échantillon d'apprentissage qui lui est la plus proche. En abrégé k -NN ou KNN, de l'anglais *k-nearest neighbors*. Son fonctionnement peut être assimilé à l'analogie suivante “*dis-moi qui sont tes voisins, je te dirais qui tu es...*”.

Le principe de cette méthode consiste à calculer la distance d'un nouvel objet par rapport à ceux dont la classe est déjà connue ; le nouvel objet appartiendra à la classe dont il est le plus proche. Dans un contexte de classification d'une nouvelle observation x , l'idée fondatrice est de faire voter les plus proches voisins de cette observation. La classe de x est déterminée en fonction de la classe majoritaire parmi les k plus proches voisins de l'observation x . La méthode k -NN est donc une méthode basée sur le voisinage, non-paramétrique.

La méthode k -NN est une approche basée sur des mesures de distances. Il n'a pas besoin d'apprentissage. Lors de la phase de décision, l'algorithme évalue la classe majoritaire des k objets les plus proches de l'exemple à tester.

5.2. Les réseau de neurones

Un réseau de neurones artificiels (NN : Neural Network) peut apprendre et donc être formé pour trouver des solutions, reconnaître des modèles, classer les données et prévoir les événements futurs. Les gens utilisent souvent les NN pour résoudre des problèmes plus complexes. Problèmes tels que la reconnaissance des caractères, la prévision boursière ...etc.

Un réseau neuronal est l'association, en un graphe plus ou moins complexe, d'objets élémentaires, les neurones formels. Les principaux réseaux se distinguent par l'organisation du graphe (en couches, complets. . .), c'est-à-dire leur architecture, son niveau de complexité (le nombre de neurones, présence ou non de boucles de rétroaction dans le réseau), par le type des neurones (leurs fonctions de transition ou d'activation) et enfin par l'objectif visé : apprentissage supervisé ou non, optimisation, systèmes dynamiques.

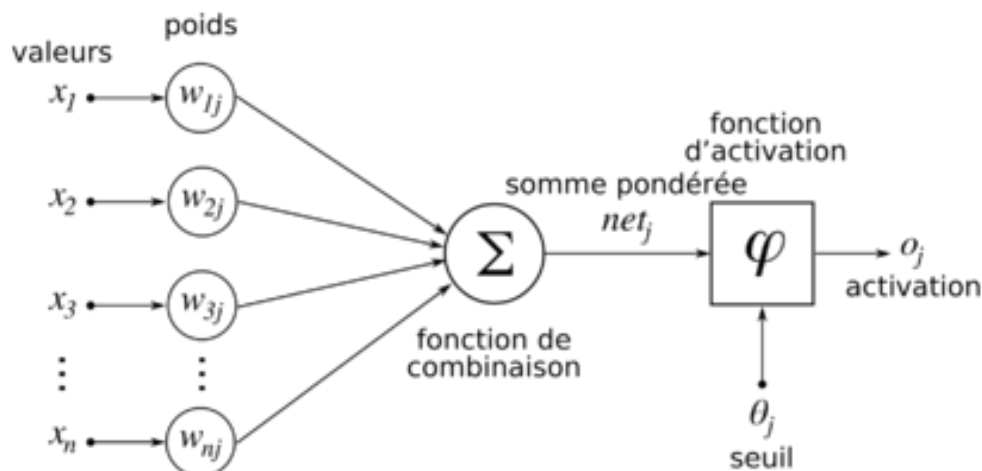


Figure 2.5 : Exemple de perceptron multicouche élémentaire avec une couche cachée et une couche de sortie.

Le perceptron multicouche (PMC) est un réseau composé de couches successives. Une couche est un ensemble de neurones n'ayant pas de connexion entre eux. Une couche d'entrée lit les signaux entrant, un neurone par entrée x_j , une couche en sortie fournit la réponse du système. Selon les auteurs, la couche d'entrée qui n'introduit aucune modification n'est pas comptabilisée. Une ou plusieurs couches cachées participent au transfert.

Dans un perceptron, un neurone d'une couche cachée est connecté en entrée à chacun des neurones de la couche précédente et en sortie à chaque neurone de la couche suivante.

Le comportement d'un réseau de neurones est défini par la manière dont ses éléments informatiques sont connectés et par les forces de ces connexions, ou pondérations. Les poids sont ajustés automatiquement en formant le réseau selon une règle d'apprentissage spécifiée jusqu'à ce qu'il accomplisse la tâche souhaitée correctement. Pour les utilisateurs expérimentés, les NN sont parfaits pour la modélisation de données non linéaires avec un grand nombre d'entités en entrée. Utilisés correctement, les NN peuvent résoudre des problèmes trop difficiles à résoudre avec un algorithme simple.

L'apprentissage est l'estimation des poids des vecteurs entre les couches par minimisation de la fonction perte quadratique ou de celle d'une fonction d'entropie en classification. Différents algorithmes d'optimisation sont proposés, ils sont généralement basés sur une évaluation du gradient par rétro-propagation.

5.3. Machine à vecteurs support (SVM)

Le "*Support Vector Machine*" ou (SVM) est un classifieur discriminant paramétrique établissant un hyperplan séparateur de marge maximale entre les exemples représentants de chacune des classes des données d'apprentissage.

Sur la Figure 2.5, nous illustrons un problème de classification binaire en 2 dimensions. Il existe une infinité de droites séparatrices des deux classes en présence mais une seule maximise la distance aux exemples aux frontières des classes (i.e. maximise la distance z_2 à l'hyperplan). Ces exemples frontières sont appelés vecteurs de support.

Le SVM est l'un des algorithmes les plus cités dans la littérature en raison de ses performances, de sa généralité et de ses fondations théoriques

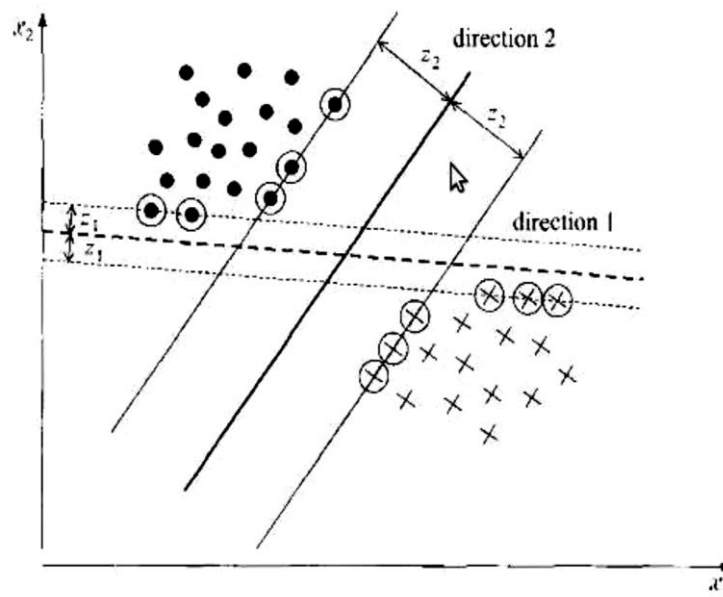


Figure 2.5 : Hyperplan séparateur de marge maximale.

De manière simple, il s'agit de construire, dans l'espace des attributs, un hyperplan séparant les points-images représentatifs des prototypes de chacune des deux classes. Cette technique, également appelée méthode des séparateurs à vastes marges, repose sur deux idées clés : la notion de marge maximale et la notion de noyau.

La marge est la distance entre la frontière de séparation (l'hyperplan) et les prototypes les plus proches de cette frontière. Ces prototypes sont appelés vecteurs supports. Dans les SVM, la frontière de séparation est choisie comme étant celle qui maximise la marge [11].

- **Conclusion**

Dans ce chapitre nous avons présentés l'architecture générale d'un système de classification des images médicales ainsi que les différentes étapes nécessaires pour aboutir à la décision finale. Le chapitre suivant contient les résultats d'implémentation de l'algorithme de classification des images médicales par l'utilisation des caractéristiques fréquentiels.

1. Introduction

La mammographie est l'une des méthodes disponibles pour la détection précoce du cancer. Le défi consiste à détecter avec précision le cancer du sein à son stade précoce. Les caractéristiques de la mammographie peuvent fournir des preuves des signes précoces de tumeurs suspectes. De nombreuses méthodes de traitement d'images ont été développées au cours des deux dernières décennies, pour aider les radiologues à diagnostiquer le cancer du sein.

Dans notre projet de fin d'étude, nous avons utilisé une technique de classification automatique pour le diagnostic des lésions suspectes d'une mammographie.

La méthode utilisée comprend 6 blocs principaux :

- Prétraitements : elles contiennent les étapes suivantes :
 - Filtrer l'image.
 - Supprimer les étiquettes et les objets supplémentaires dans les mammographies.
 - Supprimer le fond et le muscle pectoral.
- Segmentation et détection des ROI.
- Transformation d'images mammographie par DCT.
- Extraction et sélection de caractéristiques pour chaque ROI.
- Classification par réseaux de neurones.

L'organigramme de la méthode que nous avons utilisée est illustré par la figure 3.1.

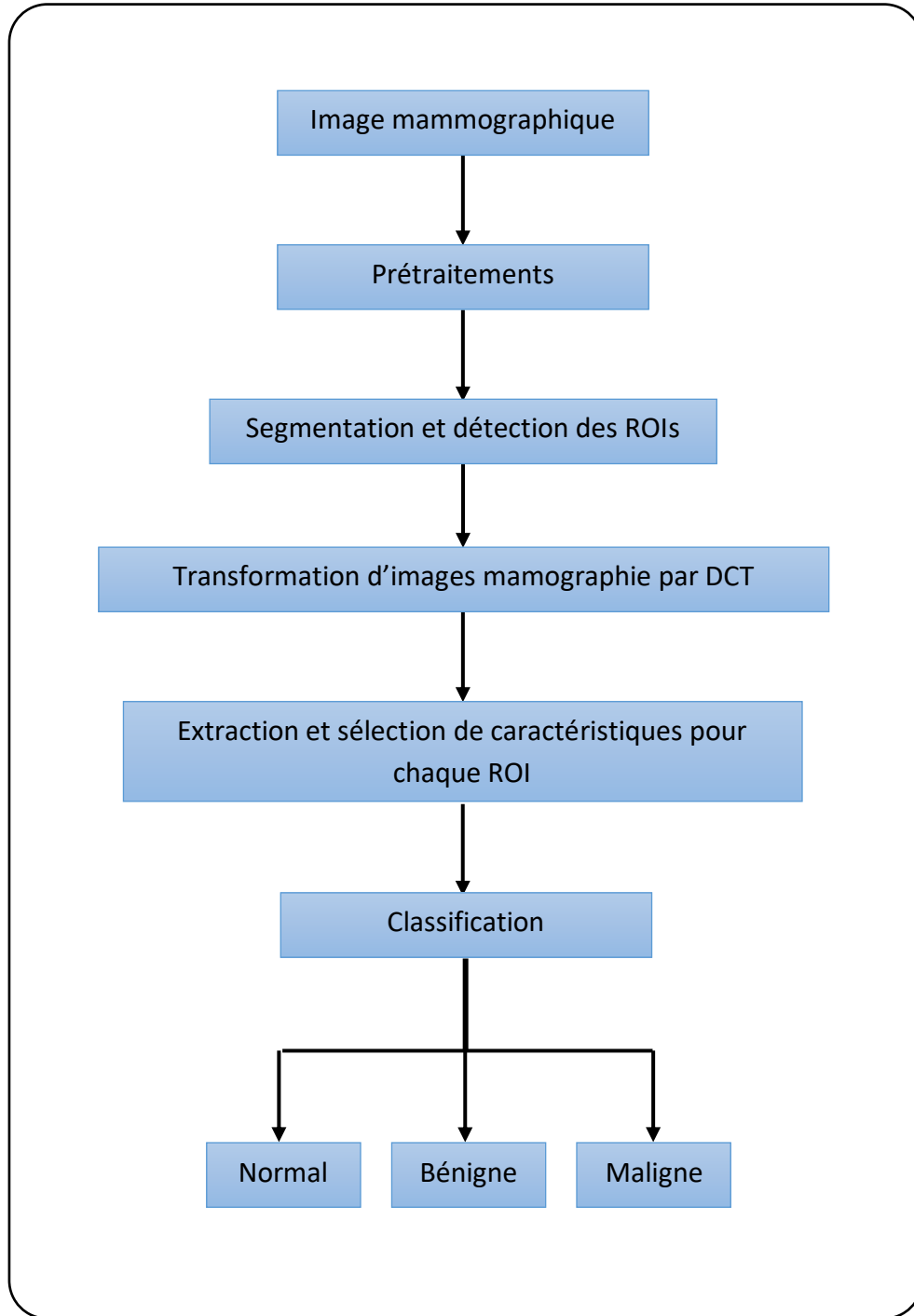


Figure 3.1 : Les étapes de la méthode proposée

2. Résultats des prétraitements

Les mammographies sont des images médicales difficiles à interpréter. Une phase de préparation est donc nécessaire pour améliorer la qualité de l'image et rendre les résultats de la segmentation plus précis. L'objectif principal des prétraitements de la mammographie est d'améliorer la qualité de l'image et de la préparer au traitement ultérieur, c.-à-d. pour la segmentation et l'extraction de caractéristiques. L'extraction de la bordure mammaire et la suppression des muscles pectoraux font également partie du prétraitement. Les mammographies ont des étiquettes rectangulaires de haute intensité, des étiquettes de faible intensité, des artefacts en bande. Les types de bruits présents dans la mammographie sont représentés à la figure 3.2, un prétraitement de la mammographie est nécessaire pour détecter et éliminer ces bruits [12].

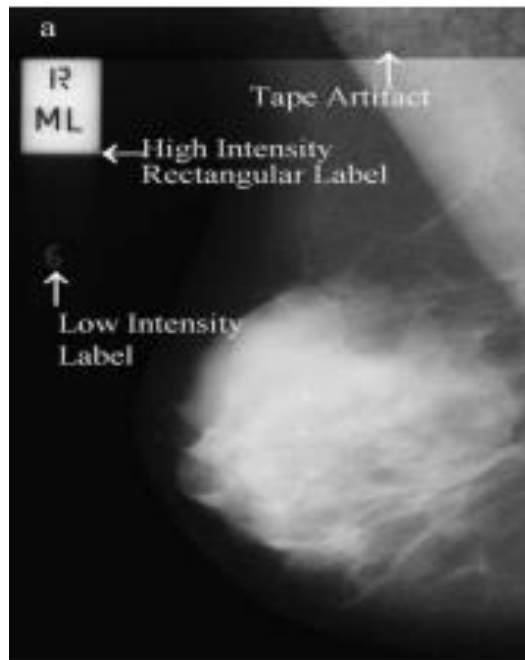


Figure 3.2 : Mammographie avec divers bruits

Dans notre méthode proposée, nous utilisons les images de la base de données MIAS (Mammography Image Analysis Society : MIAS), c'est une organisation du Royaume-Uni, un groupes de recherche qui a produit une base de données de mammographie numérique. Les images sont au format de fichier en niveaux de gris (PGM - Portable Grey Map). La base de données MIAS original (numérisée au bord de 50 microns) a été réduite à 200 microns, chaque image est represent en 1024×1024 pixels. Nous avons utilisé la base de données mini-MIAS

car elle contient des informations sur les anomalies de chaque image mammographique Cette base de données composée de 322 mammographies du sein droite et gauche, 52 mammographies malignes, 63 bénignes et 207 normal.

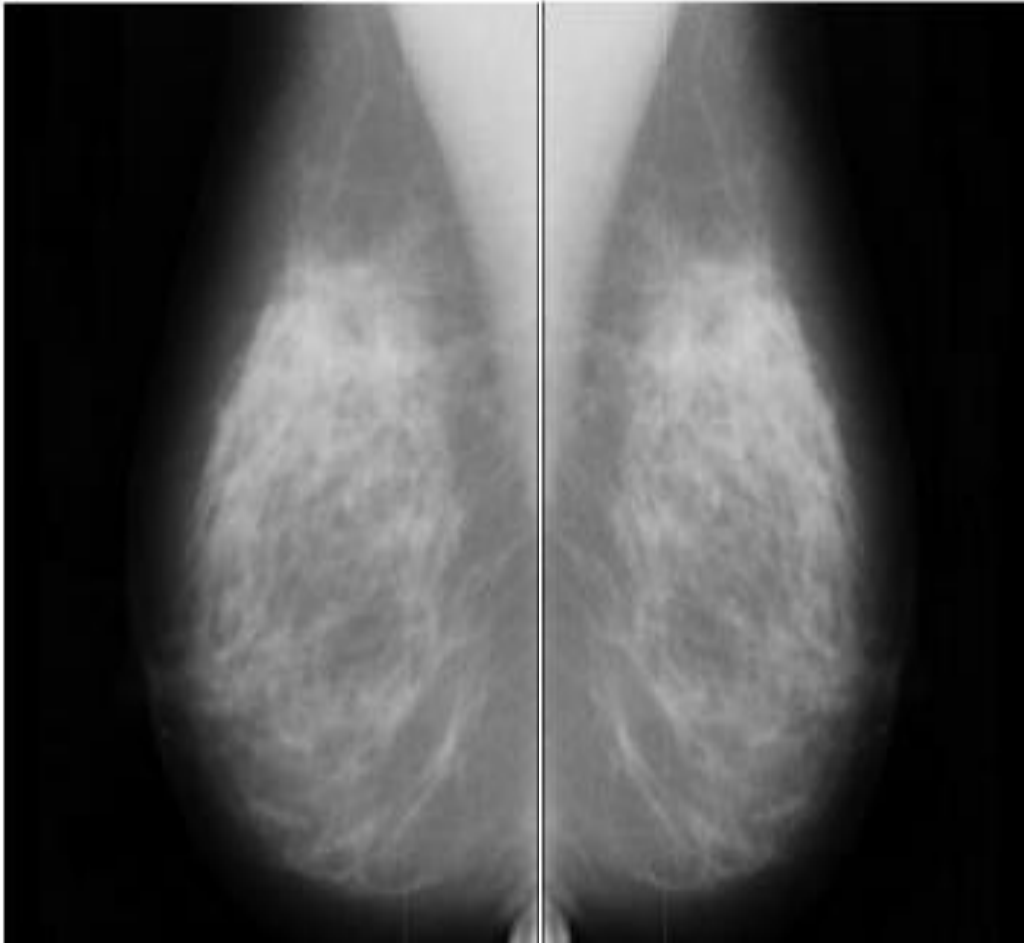


Figure 3.3 : Mammographie droite et mammographie inversée (côté gauche).

La première étape dans notre projet est l'élimination du bruit de l'image. Nous avons effectué un filtrage par le filtre médian (3×3). Il est principalement utilisé pour réduire ou éliminer le bruit présent dans l'image. Le filtre médian réduit non seulement le bruit mais préserve également les informations utiles dans l'image. Un filtre médian est un filtre non linéaire qui élimine efficacement le bruit du sel et du poivre.

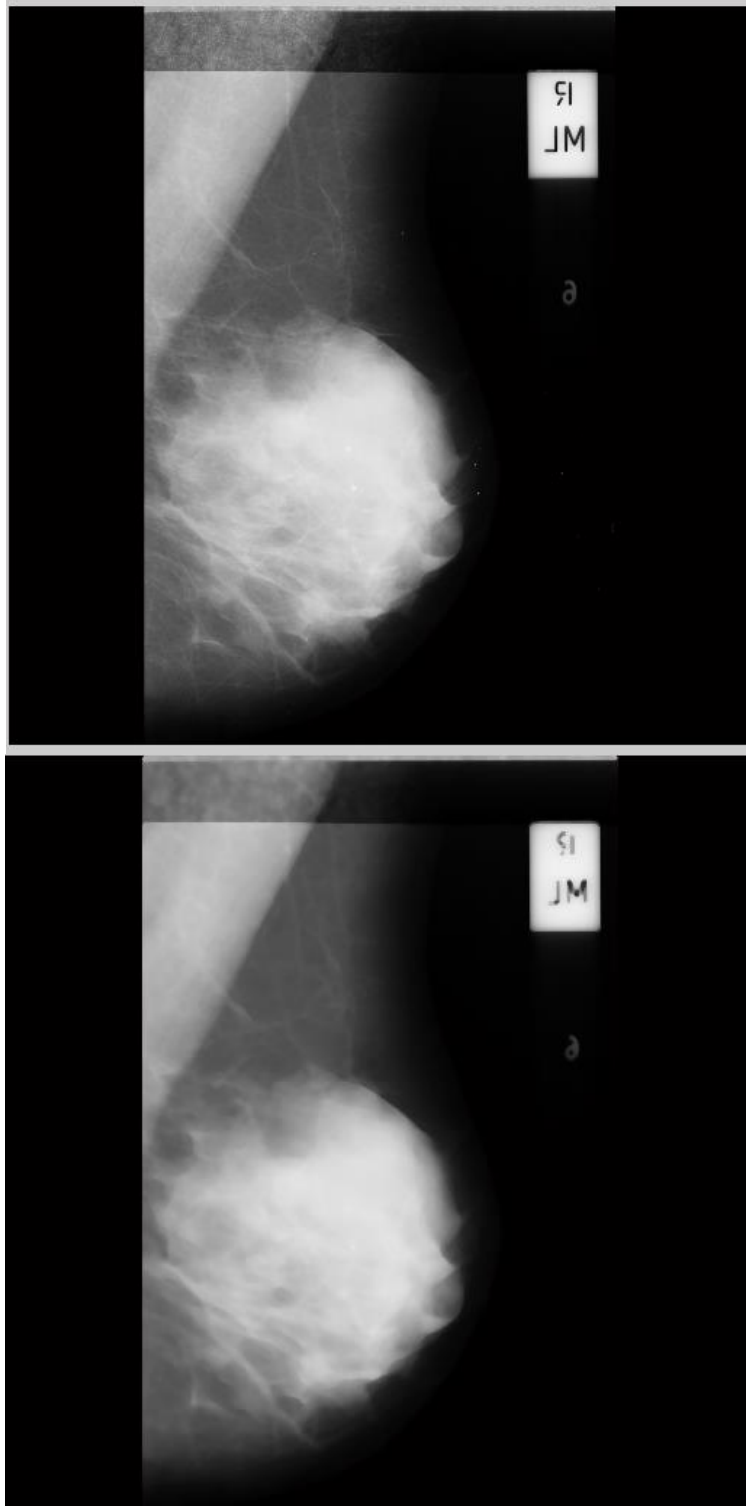


Figure 3.4 : images de mammographie originelle et filtrée

Pour supprimer les étiquettes dans l'image nous agissons comme suit;

- Convertir l'image en noir et blanc en utilisant un seuillage avec la valeur 0.0706.
- Rechercher l'objet de plus grande surface dans l'image noir et blanc,
- Utiliser l'objet trouvé comme un masque de l'image originelle.

Ce processus peut supprimer tous les objets de l'image, à l'exception du plus gros objet. Ainsi nous obtenons une image comme le montre la figure 3.5.

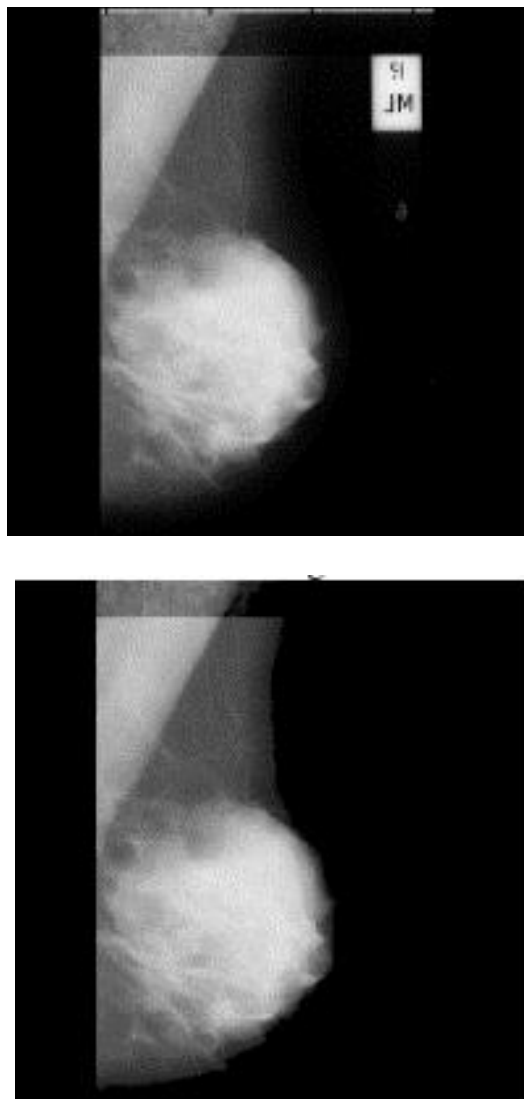


Figure 3.5 : Mammographie originelle et traitée.

Ensuite, il est également important de détecter la région d'intérêt (ROI) pour une analyse ultérieure. nous avons supprimé les parties inutiles de l'image. Les parties redondantes sont des pièces de fond qui n'appartiennent pas à la région de couverture du sein. Les fonds sont sombres et leur niveau de gris est zéro. Pour supprimer les parties noires, nous faisons collecté toutes les lignes et les colonnes de valeur noir (niveaux de gris = 0) dans un vecteur, puis nous les supprimons.

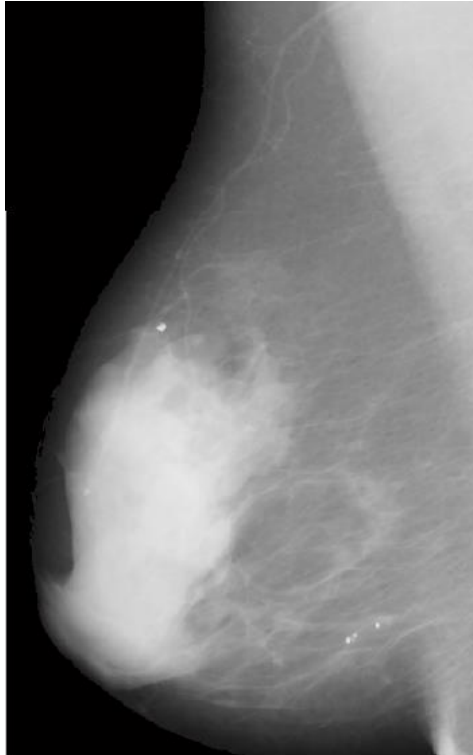


Figure 3.6 : Mammographie sans fond

Nous avons aussi essayé de supprimer le muscle pectoral, c'est une tâche importante et difficile au même temps. Ceci est fait en sélectionnant un pixel dans le coin haut gauche de l'image et ensuite nous appliquons un seuillage de tel sorte que la région détectée soit connecter au pixel choisi et en plus elle doit avoir une surface très petite par rapport à la surface du sein. Cette méthode très simple m'a réussi uniquement avec un groupe d'images, la figure 3.7 illustre une image de mammographie sans muscle pectoral [13].

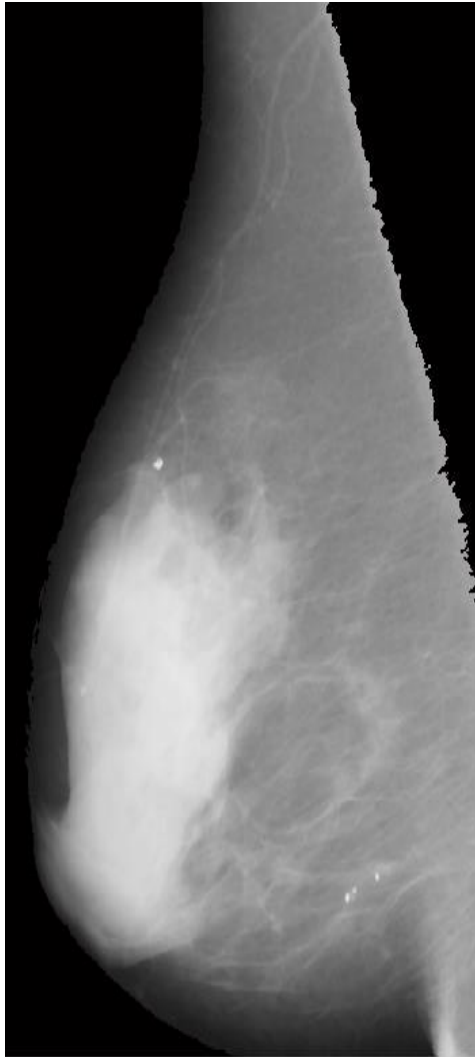


Figure 3.7 : Mammographie sans muscle pectoral

La dernière étape de prétraitement est de faire en sorte que toutes les images aient la même taille, soit 512×512 en utilisant la fonction '*imresize*' dans MATLAB, puis en retourne tous les images du sein gauche a droite.

3. Extraction et sélection de caractéristiques pour chaque ROI

Pour extraire les caractéristiques des images, nous faisons la transformation des images du domaine spatiale au domaine fréquentiel par la DCT en utilisons l'instruction "*dct2*" de MATLAB.

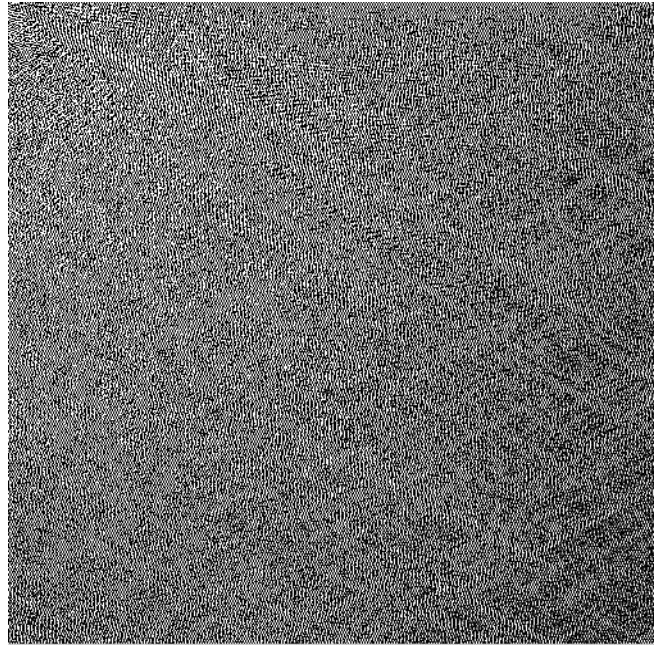


Figure 3.8 : Mammographie par DCT

Nous prenons ensuite les valeurs de coefficients de la DCT dans la zone $[0 - 128, 0 - 128]$, c'est la zone qui contient la grande partie de l'information contenue dans l'image (figure 3.9).

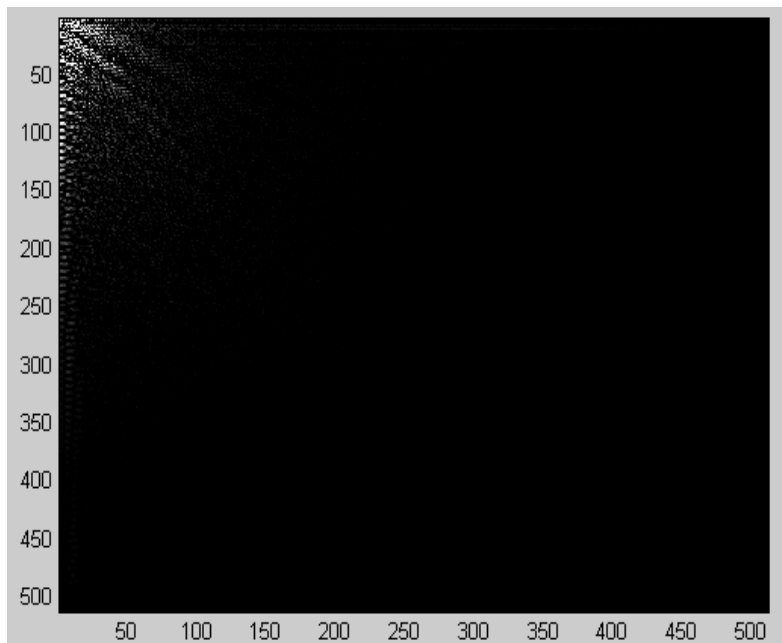


Figure 3.9 : Mammographie par DCT partie d'information

Les coefficients de la zone choisie pour chaque image sont placées dans un vecteur colonne, c'est-à-dire que nous obtenons 150 vecteurs que nous sauvegardons et les mettons dans une matrice. Les 50 colonnes de 1 à 50 pour la classe N (Normal), et les colonnes de 51 à 100 pour la classe B (bénigne) et les derniers 50 colonnes pour la classe M (Maligne).

3.1. Selection des caractéristiques

Le nombre des coefficients dans la zone choisie est très grand, ainsi nous allons choisir un nombre réduit de coefficients par l'utilisation des statistiques des ces valeurs dans chaque classe par rapport à tous les classes. Alors, on procède comme suit :

- On calcule la variance de chaque coefficient dans chaque classe : V_B , V_M et V_N .
- On fait la moyenne des 3 variances (V_{moy}).
- On calcule la variance totale de chaque coefficient V_T .
- On prend le résultats de division de V_T/V_{moy} .

Le but de cette étape est de voir les paramètres qui changent constamment, elles correspondent aux valeurs les plus élevées du résultat de la division. Les valeurs trouvées sont classées du plus grand au plus petit. Les grandes valeurs sont celles sur lesquelles nous nous appuyons car elles changent constamment et grâce auxquelles nous pouvons obtenir des informations pour déterminer le type de tumeur et sa classification.

4. Classification

Dans cette étape, nous utilisons le classifieur réseaux de neurones NN (neural network). Il contient 3 couches :

- La couche 1 d'entrée : les entrées (les caractéristiques extraits).
- Une couche cachée
- La couche de sortie (3 neurones pour 3 classes).

Alors, pour les images de la première classe (N), la couche de sortie est comme suit :

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Pour les images de la deuxième classe (B), la couche de sortie est comme suit :

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

et pour les images de la deuxième classe (M), la couche de sortie est comme suit :

$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Nous avons utilisé l'interface graphique du langage MATLAB pour l'apprentissage et les tests sur un réseau de neurones.

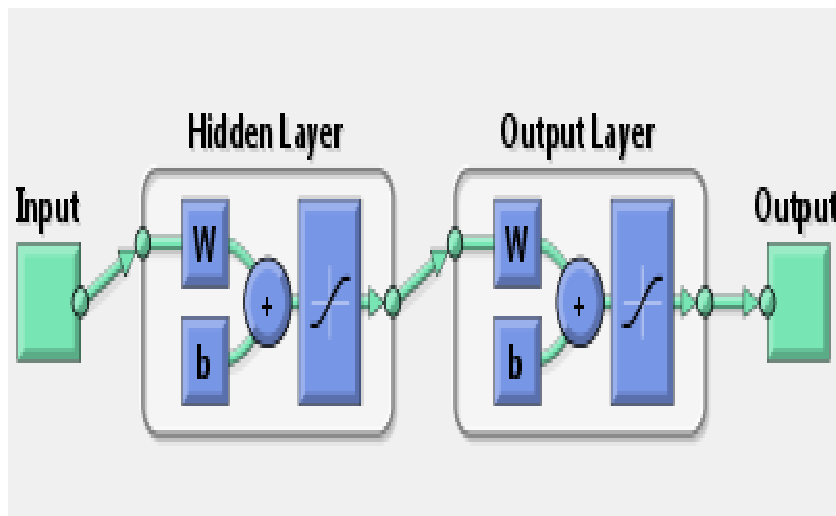


Figure 3.10 : Le classificateur NN

5. Les résultats

Nous avons collectés les paramètres sous forme d'une matrice avec un autre vecteur contenant les sorties correspondante de chaque vecteur de la matrice.

Nous devons donner des pourcentages pour chacune des étapes précédentes. Nous avons donné les pourcentages suivants apprentissages 70%, validation 15% et test 15%. Le nombre de neurones utilisé est fixé à 20. Enfin le résultat final est donné par la figure 3.11.

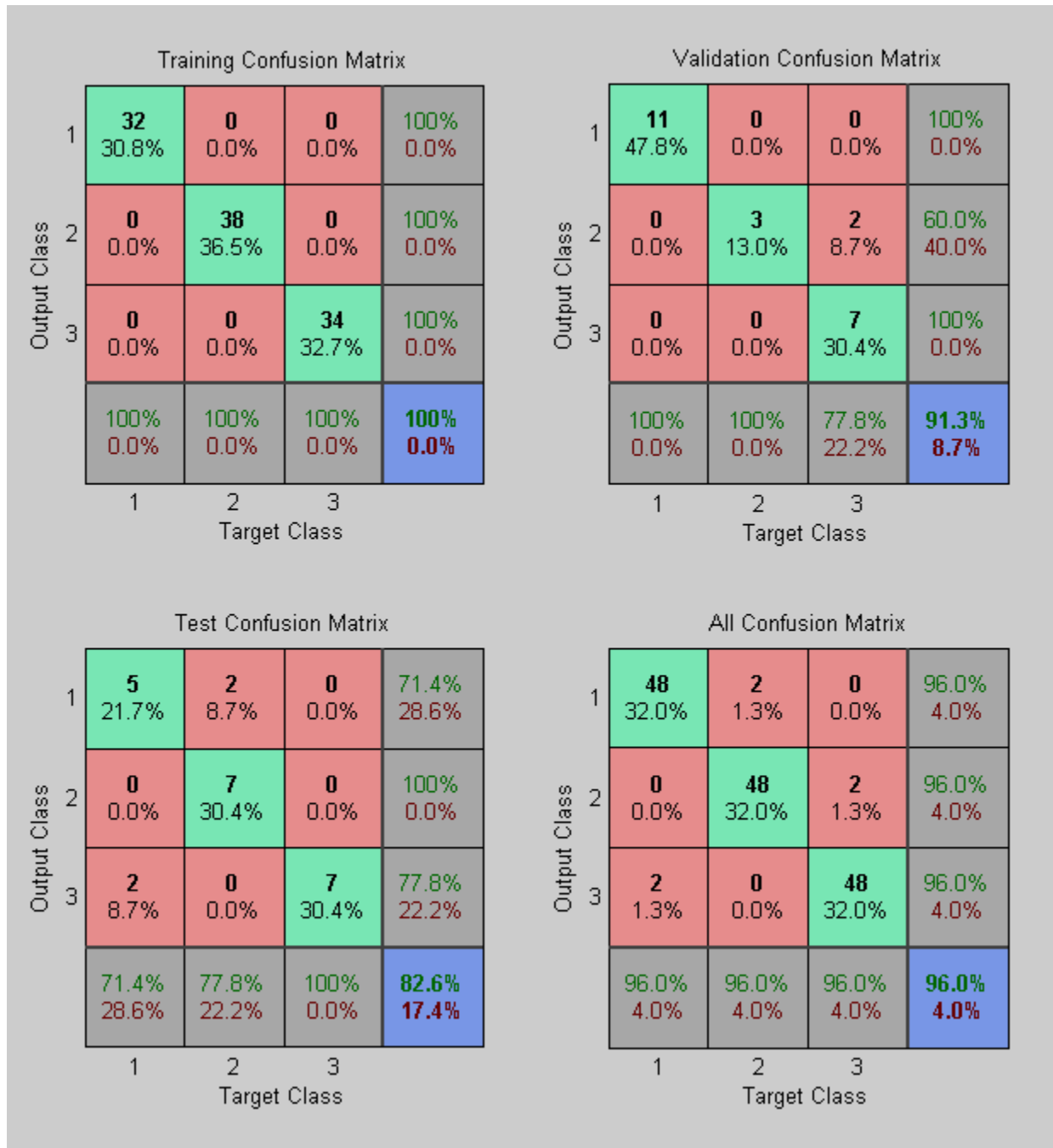


Figure 3.11 : Le résultat de classification.

Sur l'exemple de la figure 3.11, on peut voir les résultats de test sur la partie « *Test confusion Matrix* » :

La colonne n° : 01 :

- 21.5 % (5 images) les images de classe N classées correctement.
- 8.7% (2 images) les images de classe N mal classées.

La colonne n° : 02 :

- 30.4% (7 images) les images de classe B classées correctement.
- 8.7% (2 images) les images de classe B mal classées

La colonne n° : 03 :

- 30.4% (7 images) les images de classe M classé correctement.
- 0% (0 images) les images mal classées.

Au total, le taux de classification est 82.6%, soit 19 parmi 23 images de test avec un taux d'erreur de 17.4 % (4 images mal classées).

Nous avons effectués une évaluation du système par le taux de classification en fonction du nombre de neurones dans la couche cachée du réseau (Tableau 3.1). Nous avons effectué 10 exécutions et le résultat final est la moyenne. On remarque qu'un nombre de neurones entre 20 et 60 donne de meilleurs résultats. D'autre part, nous avons fait une autre évaluation pour le choix du nombre des caractéristiques et les résultats sont montrés dans le tableau 3.2. On remarque qu'avec un nombre de caractéristiques égale à 200 le taux de classification est le meilleur.

Tableau 3.1 : Résultats de la classification en fonction du nombre de neurones de la couche cachée.

N° : de Neurones de la couche cachée	Test n° :										Moyenne
	1	2	3	4	5	6	7	8	9	10	
20	91.7	91.3	95.7	95.7	95.7	91.3	95.7	95.7	100	87.0	94.00
40	91.3	91.3	100	87.0	95.7	91.3	100	100	91.3	95.7	94.36
60	95.7	95.7	100	91.3	100	91.3	91.3	100	91.3	91.3	94.79
80	82.6	95.7	100	95.7	87	91.3	95.7	82.6	82.6	87.0	90.20
100	91.3	95.7	82.6	91.3	95.7	87.0	87.0	100	91.3	91.3	90.87
120	87.0	87.0	82.6	87.0	87.0	82.6	95.7	87.0	95.7	95.7	88.73
140	91.3	91.3	95.7	91.3	82.6	95.7	91.3	91.3	91.3	100	92.18
160	91.3	91.3	87.0	91.3	91.3	95.7	95.7	91.3	87.0	95.7	91.76
180	82.6	95.7	91.3	95.7	87.0	82.6	95.7	91.3	91.3	95.7	91.00
200	91.3	95.7	91.3	91.3	91.3	95.7	91.3	95.7	82.6	91.3	91.75

Tableau 3.2 : Résultats de la classification en fonction du nombre de caractéristiques

N° : de caractéristiques	Test n° :										Moyenne
	1	2	3	4	5	6	7	8	9	10	
20	56.5	60.9	69.9	56.5	65.2	60.9	69.6	69.6	69.6	69.6	64.83
40	69.6	60.9	78.3	87	52.2	82.6	73.9	73.9	69.6	69.6	71.76
60	91.3	91.3	78.3	78.3	73.9	82.6	65.2	78.3	91.3	78.3	80.88
80	82.6	73.9	91.3	87	87	78.3	78.3	82.3	69.6	82.6	81.29
100	78.3	91.3	82.6	95.7	95.7	100	82.6	87	69.6	73.9	84.36
120	82.6	73.9	82.6	95.7	82.6	82.6	78.3	87	87	91.3	84.36
140	91.3	52.2	82.6	82.6	91.3	87	91.3	95.7	78.3	87	83.93
160	87	100	91.3	91.3	87	91.3	87	78.3	91.3	69.6	88.74
180	95.7	73.9	91.3	91.3	95.7	87	91.3	100	91.3	95.7	91.08
200	91.7	91.3	95.7	95.7	95.7	91.3	95.7	95.7	100	87.	94.00

6. Conclusion

Dans ce chapitre un algorithme de classification automatique pour le diagnostic du cancer du sein est montré. La classification et le diagnostic sont mis en œuvre dans l'environnement MATLAB. Les résultats obtenus montrent l'efficacité de cette méthode pour avoir une décision sur la mammographie du sein an tant que normale, bénigne ou maligne.

L'algorithme utilisé peut contribuer à la détection et le diagnostic du cancer du sein. Il est possible de l'utiliser dans un système de diagnostic assisté par ordinateur pour l'aide du médecin.

Conclusion générale

La classification d'images est une tâche importante dans le domaine de la vision par ordinateur, la reconnaissance d'objets et l'apprentissage automatique. Bien que les capacités des activités réalisées dans le domaine de classification des images soient nombreuses, aucune méthode n'est jugée fiable à 100%, mais au fur et à mesure les nouveaux travaux essaient d'améliorer les scores pour des meilleurs résultats.

Dans le cadre de notre projet nous avons choisie l'implémentation d'un système d'aide de diagnostique par ordinateur, qui a pour objectif d'implémenter une application qui réalise une classification d'une base d'images en un ensemble de classes (normales, bénignes ou malignes). Pour réaliser notre travail de classification on a utilisé les caractéristiques de la DCT pour faire la classification. Nous avons choisi le classificateur à base de réseaux de neurones, ce choix est justifié par la simplicité et l'efficacité de ce classificateur.

D'après les résultats trouvés et montrés dans ce mémoire, nous pouvons conclure que les caractéristiques extraites dans le domaine fréquentiel par DCT sont efficaces pour la classification des images de mammographies en 3 classes (Normales, bénignes ou malignes). D'autre part, le classificateur à base de réseaux de neurones montre ces performances dans l'utilisation des caractéristiques de la DCT pour aboutir à la décision finale [14].