
Classification automatique de la densité des tissus mammaires

Mohamed El Habib Boukhobza, Malika Mimi

*Laboratoire signaux et systèmes, Université Abdelhamid Ibn Badis de Mostaganem l'UABM, Algérie
elhabibmohamed1@gmail.com*

RÉSUMÉ. *Le cancer du sein est un problème de santé publique. L'imagerie médicale est l'un des éléments clés dans le diagnostic. Cependant, la qualité d'interprétation d'une mammographie reste variable. Une des caractéristiques de l'anatomie et de la physiologie du sein est la densité du tissu mammaire qui est importante pour deux raisons principales : (1) la densité mammaire accrue est associée à une diminution de sensibilité de la mammographie pour la détection du cancer du sein (Schetter, 2014), (2) la densité du sein est l'un des plus importants facteurs de risque connus pour le cancer du sein (Prevrhal et al., 2002 ; Boyd et al., 1995). Le classement automatique de la densité des tissus est donc un processus important dans le diagnostic. De plus, le système de classification BI-RADS identifie quatre niveaux de densité du sein, mais la base de données mini-MIAS est divisée en trois catégories de densité. Dans cet article, nous décrivons une méthode pour la classification de la densité globale du sein en utilisant les réseaux de neurones artificiels. Cette approche présente l'avantage de ne pas nécessiter d'étape de prétraitement et de s'adapter aux différentes bases de données de mammographies. La validité de notre méthode est démontrée en utilisant 240 mammographies de la base de données DDSM et 180 mammographies de la base de données mini-MIAS, avec un taux de classification correcte de 87,50 % et 86,11 %, respectivement.*

ABSTRACT. *Breast cancer is an international public health concern. Medical imaging is one of the key elements in diagnosis. However, the quality of the interpretation of mammograms remains variable. One of the important characteristics in breast anatomy and physiology is breast tissue density. Density is important for two main reasons: first, increased breast density is associated with decreased mammographic sensitivity for the detection of breast cancer (Schetter, 2014). Second, breast density is one of the strongest known risk factors for breast cancer (Prevrhal et al., 2002; Boyd et al., 1995). For these reasons, automatic tissue density classification is an important process in diagnosis. Moreover, the BI-RADS (Breast Imaging-Reporting And Data System) classification system identifies four levels of breast density, but the mini-MIAS (Mammographic Image Analysis Society) database is divided into three density categories. In this article we describe a method for overall breast density classification using artificial neural networks. This approach has the advantages of not requiring a preprocessing step and the ability to be adapted to different mammography databases. The validation of our method is demonstrated using 240 mammograms from the DDSM database and 180 mammograms from mini-MIAS database, with the correct classification rate of 87.50% and 86.11%, respectively.*

MOTS CLÉS : *classification automatique, réseau de neurones artificiel, histogramme.*

KEYWORDS: *automatic classification, artificial neural networks, histogram.*

DOI: 10.3166/ts.2017.00001 © 2016 Lavoisier

1. Introduction

Le cancer du sein est l'un des types de cancer les plus fréquents chez les femmes, la mammographie est la méthode la plus efficace pour la détection précoce des maladies du sein (Liyakathunisa et Kumar, 2011). La densité mammaire a un impact majeur sur la fiabilité de l'interprétation des mammographies (Stines, 2004). La densité mammaire est un terme radiologique, définie par l'atténuation du faisceau de rayons X au travers des constituants du parenchyme mammaire (tissu graisseux, tissu glandulaire et stroma conjonctif). La graisse radiotransparente apparaît noire sur les clichés de mammographie. Le tissu glandulaire et fibro-conjonctif radio-opaque apparaît blanc (Seror, 2005). En 1976, Wolfe a décrit une association entre une classification qualitative des densités mammaires et le risque de cancer du sein (Kerlikowske, 2007 ; McCormack et Dos Santos, 2006). Les tissus mammaires à densité élevée sont beaucoup plus susceptibles de présenter un cancer du sein (Boyd *et al.*, 2007). Les femmes dont le tissu mammaire est plus dense doivent bénéficier d'exams plus avancés. Différentes méthodes de classification ont été appliquées pour classer les tissus mammaires, parmi lesquelles l'analyse topographique de la mammographie (Chen *et al.*, 2013), celle basée sur l'algorithme des k plus proches voisins (Oliver *et al.*, 2005a, 2005b ; Mustra *et al.*, 2012) et celle basée sur l'arbre de décision (Oliver *et al.*, 2005a, 2005b). Par ailleurs, une technique de classification utilisant un histogramme multi-résolution a été proposée (Muhimmah et Zwiggelaar, 2006), de même qu'une méthode utilisant une approche statique (Petroudi *et al.*, 2003). La propriété d'approximation universelle parcimonieuse justifie, dans une large mesure, l'utilisation des réseaux de neurones dans le domaine des sciences de l'ingénieur (Rivals *et al.*, 1995). La classification est un des problèmes récurrents en vision par ordinateur, problème sur lequel les réseaux de neurones donnent de bons résultats. Grâce aux deux propriétés suivantes :

- la propriété d'« approximateurs universels » : un perceptron multicouche doté d'une unique couche cachée, avec un nombre suffisant de neurones, peut approximer n'importe quelle fonction avec la précision voulue (Hornik, 1993) ;

- la propriété d'« approximateurs parcimonieux » : les réseaux à une couche cachée forment une famille d'approximateurs parcimonieux, ce qui signifie qu'à nombre égal de paramètres, il est possible d'approximer correctement davantage de fonctions qu'avec des polynômes (Barron, 1993).

Notre objectif est de développer un système de classification automatique de la densité des tissus mammaires à base des réseaux de neurones artificiels. Le vecteur d'entrée de chaque réseau de neurones est calculé à partir de l'histogramme de la mammographie. L'originalité de l'approche proposée dans cet article est qu'elle ne nécessite pas de traitement initial ni pré-segmentation des images. La méthode est appliquée sur deux bases d'images et a pour objectif de classer les mammographies de la base DDSM en quatre types selon le lexique BI-RADS, mais les mammographies de la base mini-MIAS, seront classées en trois catégories de tissu mammaire : graisseux ; graisseux-glandulaire ; dense. D'autres études ont réalisé la

classification des tissus mammaires en deux classes (Mustra *et al.*, 2012 ; Petroudi *et al.*, 2003).

La deuxième section présente les étapes d'élaboration du système de classification de tissus mammaires, nous exposons ensuite une comparaison de nos résultats avec ceux obtenus par d'autres méthodes.

2. Système de classification automatique de la densité des tissus mammaires

La densité mammaire peut être quantifiée à partir de l'histogramme d'une mammographie (Byng *et al.*, 1994). Dans l'approche proposée, les valeurs déduites de l'histogramme seront traitées par des réseaux de neurones artificiels (RNA). Depuis une dizaine d'années, l'utilisation de réseaux de neurones artificiels s'est développée dans de nombreuses disciplines (sciences économiques, écologie, environnement, biologie et médecine...) (Zribi et Boujelbene, 2012). Ils sont notamment appliqués pour résoudre des problèmes de classification (Drew et Monson, 2000). Les réseaux de neurones artificiels (RNA) sont inspirés de la biologie, représentent un modèle mathématique (Jodouin, 1994). En général le réseau de neurones possède :

- une couche d'entrée qui contient les neurones qui reçoivent les données ;
- une couche de sortie ;
- une ou plusieurs couches de neurones cachés dont les signaux demeurent dans le système.

Il n'existe pas de règle théorique permettant de choisir la meilleure configuration d'un réseau de neurones artificiels, c'est la raison pour laquelle la conception d'un réseau de neurones se fait de manière expérimentale. Les approches utilisant les réseaux de neurones posent souvent des problèmes de convergence, d'initialisation des poids synaptiques et de choix de la topologie du réseau de neurones. Pour résoudre ces problèmes, on doit appliquer les considérations pratiques suivantes :

- initialiser tous les poids du réseau à des petites valeurs aléatoires (Cherif, 2010) ;
- un taux d'apprentissage compris entre 0 et 1, ce taux détermine l'importance de la modification des poids durant l'apprentissage (Stéphane, 2012) ;
- l'utilisation d'une base d'apprentissage comprenant les exemples d'entrée ainsi que les sorties désirées associées, permet de réaliser un apprentissage supervisé. Les paramètres du modèle vont ainsi s'adapter en comparant à chaque fois les sorties obtenues et les sorties désirées (Cherif, 2013 ; Mitcell, 1997).

Un problème de classification multiclasse peut être résolu par un réseau de neurones à plusieurs sorties ou bien un système composé de plusieurs réseaux de neurones, chaque réseau possédant une seule sortie binaire. Notre choix d'utiliser un système de classification composé de plusieurs réseaux de neurones est justifié par le fait que plusieurs chercheurs (Ahmad et Roslin, 2007 ; Nguyen et Chan, 2004 ; Patan, 2008) affirment que le système multi-réseaux de neurones est plus performant que le réseau de neurones unique à plusieurs sorties.

L'élaboration de notre système de classification s'effectue, selon les étapes suivantes :

- établissement de l'histogramme de la mammographie ;
- extraction des valeurs d'entrée du système à partir de l'histogramme ;
- normalisation des valeurs ;
- envoyer les valeurs normalisées à l'entrée des réseaux de neurones.

La méthode proposée est appliquée sur les deux bases suivantes :

- la base DDSM (*Digital Database for Screening Mammography*) (Heath *et al.*, 1998 ; Oliveira *et al.*, 2008, 2010) de l'université de la Floride du Sud ;
- la base mini-MIAS (*Mammographic Image Analysis Society*) (Suckling *et al.*, 1994) réalisée par un groupe de chercheurs britanniques qui s'intéressent à la compréhension de la mammographie.

2.1. Classification automatique de la densité des tissus mammaires de la base DDSM

La base DDSM contient des mammographies recueillies auprès de l'hôpital Massachusetts General Hospital (MGH), l'université Wake Forest University (WFU), l'hôpital Sacred Heart Hospital (SH) et Washington University of St. Louis Medical Center (WU). Plusieurs scanners ont été utilisés avec des résolutions différentes, comme l'indique le tableau 1 (Heath *et al.*, 1998).

La version IRMA de la base DDSM nous offre des images en niveaux de gris codés sur 16 bits (noir — 0, blanc — 65535) (Oliveira *et al.*, 2008, 2010). De nombreuses classifications ont été proposées pour évaluer la densité mammaire. Celle qui est actuellement utilisée dans la base DDSM provient du lexique BI-RADS de l'American College of Radiology (ACR) et consiste à apprécier visuellement le pourcentage de tissu radio-opaque. Quatre types sont ainsi décrits (figure 1) :

Tableau 1. Les scanners utilisés pour la numérisation des mammographies de la base DDSM

Source	Scanner utilisé	Résolution en micron par pixel
MGH	DBA M2100 Image Clear	42
	Howtek 960	43,5
WFU	Lumisys 200 Laser	50
SH		
WU	Howtek MultiRad850	43,5

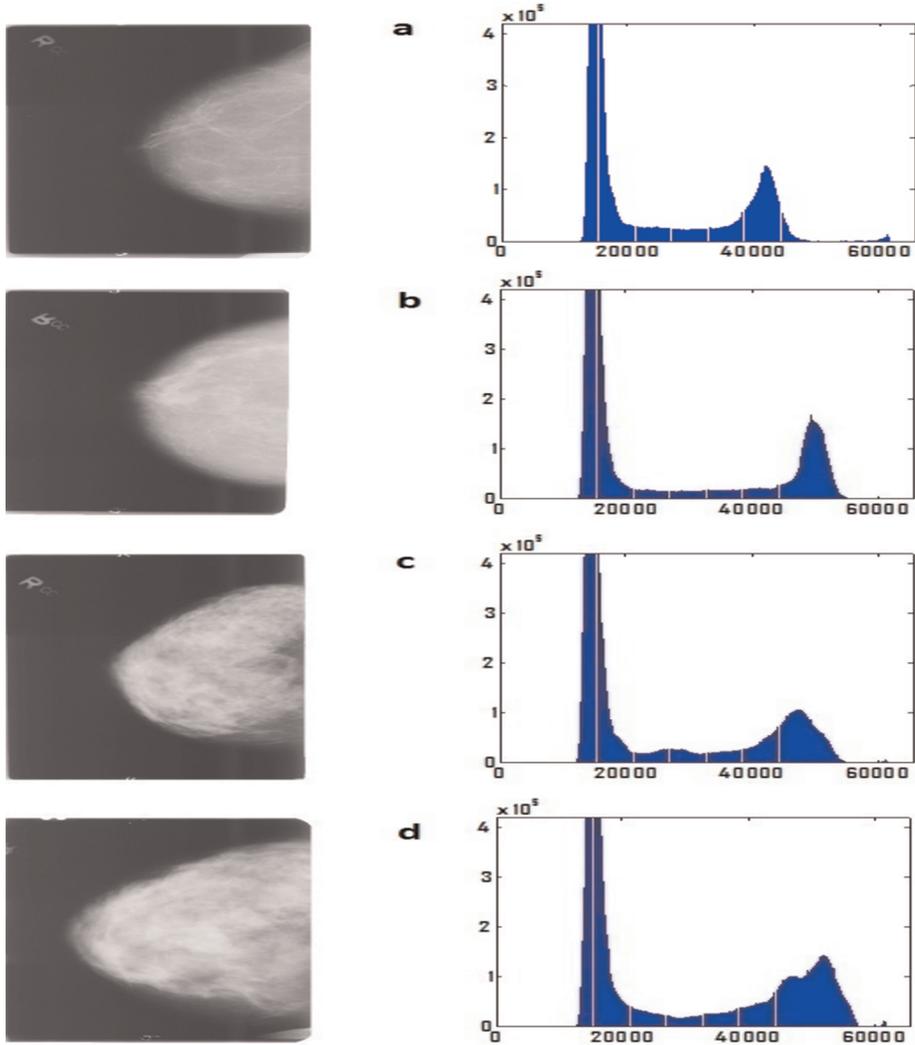


Figure 1. Quatre mammographies de la base DDSM et leurs histogrammes :
 (a) correspond à un tissu du type BI-RADS I, (b) correspond à un tissu
 du type BI-RADS II, (c) correspond à un tissu du type BI-RADS III,
 (d) correspond à un tissu du type BI-RADS IV

- BI-RADS I : sein presque entièrement grasseux (< 25 % de glande mammaire) ;
- BI-RADS II : présence d'opacités fibro-glandulaires éparses (approximativement 25 à 50 % de glande mammaire) ;
- BI-RADS III : tissu mammaire dense et hétérogène, ce qui peut compliquer la détection de petites masses (approximativement 51 à 75 % de glande mammaire) ;

– BI-RADS IV : tissu mammaire très dense, ce qui peut diminuer la sensibilité de la mammographie (> 75 % de glande mammaire) (Tardivon, 2008).

Chaque histogramme de la figure 1 présente deux pics, celui de gauche correspond au fond de l'image, donnant peu d'informations, celui de droite représente la zone d'intérêt et doit être traité puisqu'il correspond aux tissus mammaires.

L'utilisation d'un nombre limité de valeurs fournies par l'histogramme permet de réduire le nombre d'entrées, ceci afin de diminuer le temps de calcul et accélérer l'apprentissage du système de classification de tissus mammaires. Le nombre de valeurs à utiliser est choisi de manière arbitraire. La zone d'intérêt est limitée par les niveaux de gris 33 000 et 57 500. À partir de cet intervalle, on prend 50 valeurs régulièrement réparties pour former un vecteur d'entrée « E ». Chacune de ces valeurs représente le nombre de pixels pour chaque niveau de gris.

En pratique, il est recommandé d'effectuer une normalisation, pour chaque vecteur d'entrée « E » (Dreyfus *et al.*, 2002). On calcule la valeur moyenne « m » et l'écart-type « σ » par rapport aux composantes du vecteur d'entrée « E », ensuite une normalisation du vecteur « E » est réalisée pour obtenir le vecteur normalisé « E' » suivant la formule ci-dessous (Dreyfus *et al.*, 2002) :

$$E' = \frac{E - m}{\sigma} \quad (1)$$

De sorte que le vecteur normalisé « E' » a une moyenne voisine de zéro et un écart-type de l'ordre de 1 (Dreyfus *et al.*, 2002). Les valeurs du vecteur « E' » sont introduites dans la couche d'entrée des réseaux de neurones.

Notre système de classification de la densité des tissus mammaires de la base DDSM est composé de trois étapes successives, chaque étape utilise un réseau de neurones :

- le premier réseau permet d'identifier les tissus mammaires BI-RADS I ;
- le deuxième réseau permet d'identifier les tissus mammaires BI-RADS II ;
- le troisième réseau permet d'identifier les tissus mammaires de type BI-RADS III et IV, tel qu'illustré sur la figure 2.

Quatre-vingts pour cent des mammographies traitées seront utilisées pour l'apprentissage et 20 % pour le test, pour chaque réseau de neurones.

La première étape consiste à calculer des indicateurs à partir de l'histogramme de l'image en niveaux de gris sans prétraitement. À partir de la base DDSM, on a choisi 240 mammographies réparties en 60 images pour chaque type de tissu, afin de concevoir un système capable de classifier les quatre types de tissus. Les réseaux de neurones de notre application sont composés de trois couches :

- une couche d'entrée, elle contient autant de neurones que le nombre d'éléments du vecteur d'entrée E' : E1'.....E50' ;

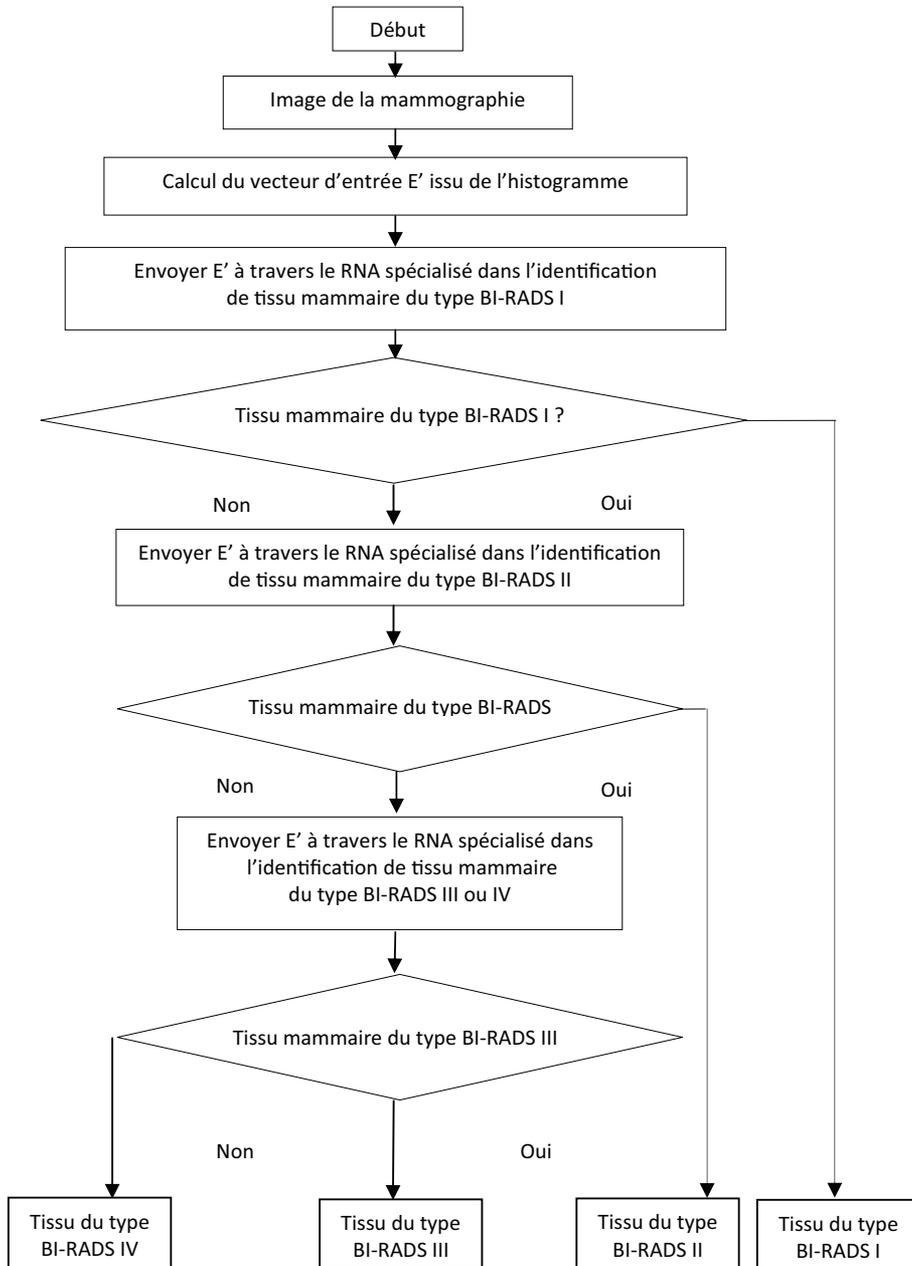


Figure 2. Organigramme du système de classification de la densité des tissus mammaires conformément à la classification BI-RADS

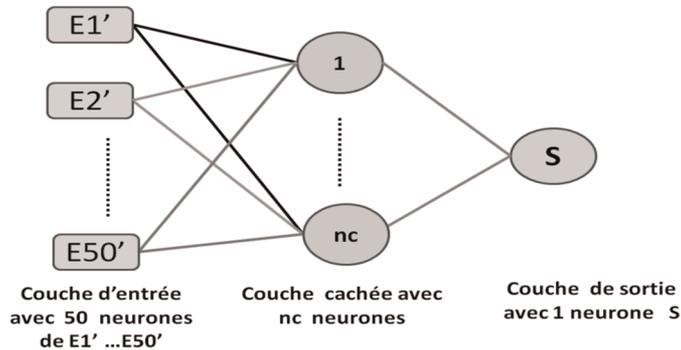


Figure 3. RNA à trois couches, nc neurones cachés, un neurone de sortie et 50 d'entrées

– pour ne pas augmenter la complexité du réseau, nous avons opté pour une seule couche cachée ;

– une couche de sortie avec un neurone « S », cette sortie permet d'identifier le type du tissu, « S » peut prendre la valeur « 0 » ou « 1 ».

La fonction d'activation (ou fonction de transfert) sert à convertir le résultat de la somme pondérée des entrées d'un neurone en une valeur de sortie, cette conversion s'effectue par un calcul de l'état du neurone et permet d'introduire une non-linéarité dans le fonctionnement du neurone (Cybenko, 1989). Les fonctions d'activation et le nombre de neurones dans la couche cachée ont été choisis expérimentalement ; le modèle le plus performant sera sélectionné (figure 3).

L'expérimentation du réseau de neurones, nécessite l'application des tâches suivantes :

- faire varier nc de 1 à 50, nc étant le nombre de neurones en couche cachée ;
- changer fc , fc désignant la fonction de transfert de la couche cachée ;
- changer fs , fs désignant la fonction de transfert de la couche de sortie.

Pour simplifier, on considère un seul type de fonction par couches de neurones, trois fonctions de transfert seront envisagées :

- fonction linéaire : « purelin » (figure 4a) ;
- la tangente hyperbolique : « tansig » (figure 4b) ;
- fonction sigmoïde : « logsig » (figure 4c).

L'utilisation des fonctions « purelin », « tansig » et « logsig » au niveau de la couche cachée et de la couche de sortie, permet d'obtenir neuf couples de fonctions d'activation :

- tansig-tansig ; tansig-logsig ; tansig-purelin ;

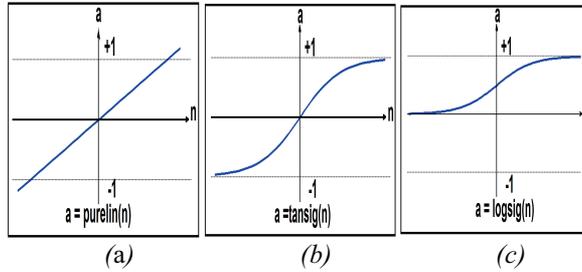


Figure 4. Les trois fonctions de transfert

- logsig-logsig ; logsig-tansig ; logsig-purelin ;
- purelin-purelin ; purelin-logsig ; purelin-tansig.

L'apprentissage est effectué pour chaque configuration, avec les paramètres suivants :

- nombre maximum d'itérations (Epochs) = 1000 ;
- gradient minimum = 1 e-10.

L'apprentissage est arrêté si l'une de ces conditions est satisfaite.

Critères de performance. — Les performances de chaque réseau de neurones sont observées par comparaison des résultats estimés par le réseau avec les états réels. Le taux de reconnaissance est défini comme étant le rapport entre le nombre des cas reconnus correctement et le nombre des cas présentés pour le test du système (Tremblay, 2004).

$$\text{Taux de reconnaissance} = \frac{\text{Bonnes décisions}}{\text{Nombre total d'exemples}} \times 100 \quad (2)$$

Le tableau 2 est une présentation non exhaustive des taux de reconnaissance obtenue pour la base DDSM, en variant le nombre de neurones en couche cachée et les couples des fonctions de transfert.

Tableau 2. Exemples de taux de reconnaissance des tissus mammaires selon la classification BI-RADS pour plusieurs configurations de réseaux de neurones

Type de tissus identifié par le réseau de neurones	Nombre de neurones cachés nc	fc = logsig fs = purelin	fc = purelin fs = logsig	fc = logsig fs = tansig
RNA spécialisé dans l'identification de tissu mammaire du type BI-RADS I	3	25,00 %	16,66 %	00,00 %
	31	00,00 %	91,66 %	33,33 %
	50	08,33 %	41,66 %	00,00 %

Tableau 2 – (suite)

Type de tissus identifié par le réseau de neurones		Nombre de neurones cachés nc	fc = logsig fs = purelin	fc = purelin fs = logsig	fc = logsig fs = tansig
RNA spécialisé dans l'identification de tissu mammaire du type BI-RADS II		3	08,33 %	08,33 %	00,00 %
		43	00,00 %	83,33 %	00,00 %
		50	08,33 %	08,33 %	00,00 %
RNA spécialisé dans l'identification de tissu mammaire du type BI-RADS III et BI-RADS IV	BI-RADS III	3	00,00 %	00,00 %	00,00 %
		42	00,00 %	83,33 %	08,33 %
		50	00,00 %	08,33 %	00,00 %
	BI-RADS IV	3	00,00 %	66,66 %	00,00 %
		42	00,00 %	91,66 %	75,00 %
		50	00,00 %	58,33 %	00,00 %

Le tableau 3 indique les paramètres de chaque réseau de neurones sélectionné constituant le système de classification de la base DDSM.

Tableau 3. Les paramètres des réseaux de neurones sélectionnés constituant le système de classification de la base DDSM

N° d'étape	Type de tissu identifié par le réseau de neurones	Les paramètres du réseau de neurones sélectionné
Étape 1	Réseau de neurones spécialisé dans l'identification de tissu mammaire du type BI-RADS I	<ul style="list-style-type: none"> - nombre d'images utilisées NI = 60*4 types de tissus - nombre d'images d'apprentissage NA = 192 - nombre d'images testées NT = 48 - nombre de neurones d'entrée NE = 50 - nombre de neurones cachés NC = 31 - fonction de transfert de la couche cachée : Purelin - fonction de transfert de la couche de sortie : Logsig - un neurone de sortie : S1, telle que : <ul style="list-style-type: none"> * si S1 = 0 ⇒ tissu du type BI-RADS I ; * si S1 = 1 ⇒ autre type de tissu.

Tableau 3 – (suite)

N° d'étape	Type de tissu identifié par le réseau de neurones	Les paramètres du réseau de neurones sélectionné
Étape 2	Réseau de neurones spécialisé dans l'identification de tissu mammaire du type BI-RADS II	<ul style="list-style-type: none"> - nombre d'images utilisées NI = 60*3 types de tissus - nombre d'images d'apprentissage NA = 144 - nombre d'images testées NT = 36 - nombre de neurones d'entrée NE = 50 - nombre de neurones cachés NC = 43 - fonction de transfert de la couche cachée : Purelin - fonction de transfert de la couche de sortie : Logsig - un neurone de sortie : S2, telle que : <ul style="list-style-type: none"> * si $S2 = 0 \Rightarrow$ tissu du type BI-RADS II ; * si $S2 = 1 \Rightarrow$ autre type de tissu.
Étape 3	Réseau de neurones spécialisé dans l'identification des tissus mammaires du type BI-RADS III et IV	<ul style="list-style-type: none"> - nombre d'images utilisées NI = 60*2 types de tissus - nombre d'images d'apprentissage NA = 96 - nombre d'images testées NT = 24 - nombre de neurones d'entrée NE = 50 - nombre de neurones cachés NC = 42 - fonction de transfert de la couche cachée : Purelin - fonction de transfert de la couche de sortie : Logsig - un neurone de sortie : S3, telle que : <ul style="list-style-type: none"> * si $S3 = 0 \Rightarrow$ tissu du type BI-RADS III ; * si $S3 = 1 \Rightarrow$ tissu du type BI-RADS IV.

2.2. Classification automatique de la densité des tissus mammaires de la base mini-MIAS

La base mini-MIAS (Mammography Image Analysis Society) (Suckling *et al.*, 1994) nous offre des mammographies en niveaux de gris codées sur 8 bits (noir — 0, blanc — 255) avec une résolution de 200 μm par pixel. Pour la conception d'un système de classification de tissu mammaire de la base mini-MIAS, on applique la méthode décrite précédemment pour la conception d'un système de classification de la base DDSM. La classification de la base mini-MIAS s'effectuera par l'utilisation de deux réseaux de neurones, puisque dans cette base d'images, il y a seulement trois catégories de tissus mammaires :

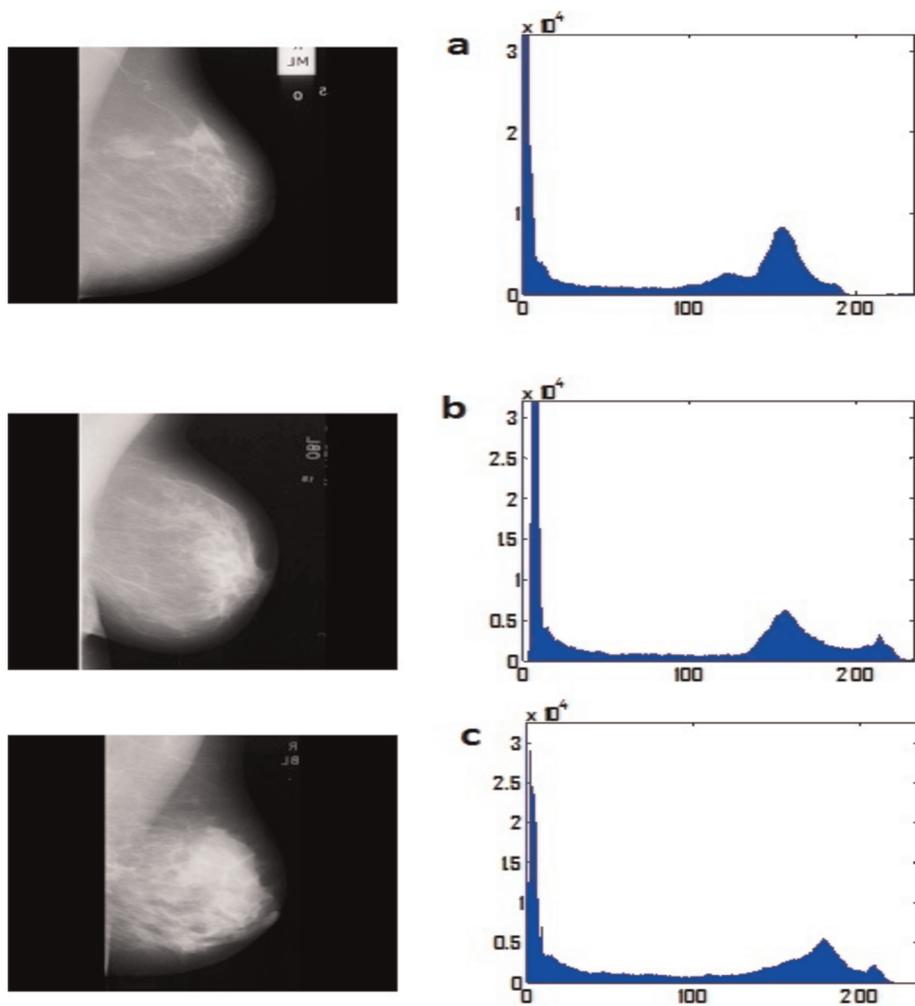


Figure 5. Trois mammographies de la base MIAS et leurs histogrammes : (a) correspond à un tissu gras, (b) correspond à un tissu gras-glandulaire (c) correspond à un tissu dense

- gras (Fatty F) ;
- gras-glandulaire (Fatty glandular G) ;
- dense (Dense D).

Les histogrammes de la figure 5 montrent des variations sur la plage de niveaux de gris entre 100 et 255, cet intervalle correspond aux tissus mammaires. À partir de cet intervalle, on cherche à utiliser un nombre limité de valeurs. Afin de

Tableau 4. Des exemples de taux de reconnaissance des tissus mammaires conformément à la classification de la base mini-MIAS réalisé pour plusieurs configurations des réseaux de neurones

Type de tissus identifié par le réseau de neurones		Nombre de neurones cachés nc	fc = logsig fs = purelin	fc = purelin fs = logsig	fc = logsig fs = tansig
RNA spécialisé dans l'identification de tissu mammaire du type grasieux (F)		3	66,66 %	00,00 %	08,33 %
		16	00,00 %	91,66 %	00,00 %
		50	00,00 %	25,00 %	25,00 %
RNA spécialisé dans l'identification de tissu mammaire du type grasieux-glandulaire (G) et dense (D)	Grasieux-glandulaire	3	08,33 %	00,00 %	08,33 %
		23	00,00 %	83,33 %	00,00 %
		50	00,00 %	08,33 %	08,33 %
	Dense	3	00,00 %	50,00 %	00,00 %
		23	08,33 %	83,33 %	25,00 %
		50	00,00 %	33,33 %	08,33 %

* fc désigne la fonction de transfert de la couche cachée.

* fs désigne la fonction de transfert de la couche de sortie.

réduire le nombre d'entrées, cela nous amène à déduire 32 valeurs régulièrement réparties.

Le nombre de valeurs à utiliser est choisi de manière arbitraire. Chacune de ces valeurs représente le nombre de pixels normalisés pour chaque niveau de gris, ces valeurs servent à former le vecteur d'entrée : $E1' \dots E32'$. Pour la conception d'un système de classification de tissu mammaire de la base mini-MIAS, on a choisi 180 mammographies réparties en trois sous-groupes de 60 images pour chaque type de tissu. Le système de classification est formé de deux réseaux de neurones. Le premier réseau permet d'identifier les tissus mammaires grasieux (F), ce même réseau de neurones peut indiquer que le tissu traité et non grasieux dans ce cas, le deuxième réseau a pour rôle d'identifier les tissus grasieux-glandulaires (G) et les tissus denses (D). Le tableau 4 est une présentation non exhaustive des taux de reconnaissance obtenue pour la base mini-MIAS en variant le nombre de neurones en couche cachée et les fonctions de transfert de la couche cachée et la couche de sortie.

Le tableau 5 indique les paramètres de chaque réseau de neurones sélectionné constituant le système de classification de la base mini-MIAS.

Tableau 5. Les paramètres des réseaux de neurones sélectionnés constituant le système de classification de la base mini-MIAS

N° d'étape	Type de tissu identifié par le réseau de neurones	Les paramètres du réseau de neurones sélectionné
Étape 1	Réseau de neurones spécialisé dans l'identification de tissu mammaire graisseux	<ul style="list-style-type: none"> - nombre d'images utilisées NI = 60*3 types de tissus - nombre d'images d'apprentissage NA = 144 - nombre d'images testées NT = 36 - nombre de neurones d'entrée NE = 32 - nombre de neurones cachés NC = 16 - fonction de transfert de la couche cachée : Purelin - fonction de transfert de la couche de sortie : Logsig - un neurone de sortie : S1, telle que : <ul style="list-style-type: none"> * si S1 = 0 ⇒ tissu graisseux ; * si S1 = 1 ⇒ autre type de tissu.
Étape 2	Réseau de neurones spécialisé dans l'identification des tissus graisseux-glandulaires et les tissus denses	<ul style="list-style-type: none"> - nombre d'images utilisées NI = 60*2 types de tissus - nombre d'images d'apprentissage NA = 96 - nombre d'images testées NT = 24 - nombre de neurones d'entrée NE = 32 - nombre de neurones cachés NC = 23 - fonction de transfert de la couche cachée : Purelin - fonction de transfert de la couche de sortie : Logsig - un neurone de sortie : S2, telle que : <ul style="list-style-type: none"> * si S2 = 0 ⇒ tissu graisseux-glandulaire ; * si S2 = 1 ⇒ tissu dense.

3. Résultats et comparaison

Les mammographies utilisées pour évaluer notre méthode et les autres méthodes proviennent de la base DDSM et/ou de la base mini-MIAS.

Le taux de reconnaissance (TR) global n'est pas toujours la moyenne des taux de chaque type de tissu, puisque le taux global dépend du nombre de tissus de chaque type utilisé dans le test. Le système de classification de la densité des tissus mammaires est simple à élaborer, puisqu'il utilise des réseaux de neurones à une seule couche cachée. L'entrée de chaque réseau utilise seulement les données issues des histogrammes.

De plus, la classification de la densité des tissus mammaires s'applique sur des bases engendrées par des campagnes de dépistage du cancer du sein, ce qui entraîne la production d'une masse importante d'informations à traiter. Ceci implique que la simplicité de la méthode est un facteur important puisqu'il permet de réduire le temps de calcul.

L'approche proposée ne nécessite pas de traitement initial ni pré-segmentation des images. Par contre, les autres méthodes nécessitent un prétraitement des images, ce qui engendre une perte d'information. Cette perte d'information explique la différence des résultats entre la méthode proposée et les autres méthodes.

Notre système de classification est composé de plusieurs réseaux de neurones et chaque réseau est spécialisé dans l'identification d'un seul type de tissu mammaire ou deux types. Cette caractéristique offre une possibilité d'adaptation de notre approche aux différents types de tissus mammaires.

L'observation des tableaux 6 et 7 nous permet de constater la performance de notre méthode pour chaque type de tissu mammaire. Cette performance est vérifiée pour la classification BI-RADS et mini-MIAS.

Tableau 6. Comparaison des méthodes conformément à la classification BI-RADS

Réf. – année				
Type de tissu	(Petroudi <i>et al.</i> , 2003)	(Oliver <i>et al.</i> , 2005b)	(Chen <i>et al.</i> , 2013)	Méthode Proposée
BI-RADS I	91,00 %	46,00 %	88,00 %	91,66 %
BI-RADS II	64,00 %	48,00 %	85,00 %	83,33 %
BI-RADS III	70,00 %	52,00 %	78,00 %	83,33 %
BI-RADS IV	78,00 %	36,00 %	72,00 %	91,66 %
TR global	76,00 %	47,00 %	81,22 %	87,50 %

Tableau 7. Comparaison des méthodes conformément à la classification de la base mini-MIAS

Réf. – année				
Type de tissu	(Oliver <i>et al.</i> , 2005a)	(Muhimmah et Zwiggelaar, 2006)	(Mustra <i>et al.</i> , 2012)	Méthode proposée
F	63,33 %	80,16 %	87,73 %	91,67 %
G	58,33 %	74,71 %	75,00 %	83,33 %
D	66,66 %	77,00 %	83,50 %	83,33 %
TR global	62,77 %	77,57 %	82,50 %	86,11 %

4. Conclusion

Cette étude a été consacrée à l'application des réseaux de neurones artificiels, dans la classification automatique de la densité des tissus mammaires.

Le système de classification est développé grâce à l'exploitation des indicateurs issus des histogrammes par des réseaux de neurones artificiels. Ces indicateurs ont l'avantage d'être d'une grande simplicité d'emploi.

Notre système de classification est composé de plusieurs réseaux de neurones artificiels. Chaque réseau du système de classification est spécialisé dans la reconnaissance soit d'un seul type de tissu mammaire, soit de deux types conformément à la classification BI-RADS ou mini-MIAS.

Les résultats expérimentaux montrent l'efficacité de notre système proposé. En perspective d'autres bases de données de mammographie seront testées, pour approuver la fiabilité du système de classification de tissu mammaire à base de réseaux de neurones artificiels.

Remerciement

Nous tenons à exprimer nos remerciements à M. Abdelmalik Taleb Ahmed, Professeur, à l'IUT de Valenciennes et chercheur au LAMIH, pour son aide, sa disponibilité et pour l'attention portée à notre travail.

Bibliographie

- Ahmad Z. et Roslin F. (2007). Modeling of real pH Neutralization Process using Multiple Neural Networks (MNN) Combination Technique. *International Conference on Control, Instrumentation and Mechatronics Engineering 2007*, Johor Bahru, Johor, Malaysia.
- Barron A.R. (1993). Universal approximation bounds for superpositions of a sigmoidal function, *IEEE Transactions on Information theory*, vol. 39, n° 3, p. 930-945.
- Boyd N.F., Byng J.W., Jong R.A., Fishell E.K., Little L.E., Miller A.B. et Yaffe M.J. (1995). Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study, *Journal of the National Cancer Institute*, vol. 87, n° 9, p. 670-675.
- Boyd N.F., Guo H., Martin L.J., Sun L., Stone J., Fishell E. et Yaffe M.J. (2007). Mammographic density and the risk and detection of breast cancer, *New England Journal of Medicine*, vol. 356, n° 3, p. 227-236.
- Byng J.W., Boyd N.F., Fishell E., Jong R.A. et Yaffe M.J. (1994). The quantitative analysis of mammographic densities, *Physics in Medicine and Biology*, vol. 39, n° 10, p. 1629.
- Cherif M. (2010). *Capacité d'une mémoire associative à fonction de sortie chaotique*. Mémoire en informatique, Université du Québec à Montréal.

- Cherif A. (2013). Réseaux de neurones, SVM et approches locales pour la prévision de séries temporelles. Thèse de doctorat. Université François-Rabelais de Tours.
- Chen Z., Oliver A., Denton E. et Zwigelaar R. (2013). Automated mammographic risk classification based on breast density estimation, *Iberian Conference on Pattern Recognition and Image Analysis 2013*, Madeira, Portugal.
- Cybenko G. (1989). Approximation by superpositions of a sigmoidal function: *Math. Control, Signals and Systems*, vol. 2, n° 4, p. 303-314.
- Drew P.J. et Monson J.R. (2000). Artificial neural networks, *Surgery*, vol. 127, n° 1, p. 3-11.
- Dreyfus G., Martinez J.M., Samuelides M., Gordon M.B., Badran F., Thiria S. et Hérault L. (2002). *Réseau de neurone méthodologie et application*, Édition Eyrolles, Paris.
- Heath M., Bowyer K., Kopans D., Kegelmeyer J.P., Moore R., Chang K. et Munishkumaran S. (1998). Current status of the digital database for screening mammography. *In: Digital mammography*, Springer Netherlands, p. 457-460.
- Hornik K. (1993). Some new results on neural network approximation, *Neural Networks*, vol. 6, n° 8, p. 1069-1072.
- Jodouin J.F. (1994). *Les Réseaux de neurones : principes et définitions*, L'Hermès.
- Kerlikowske K. (2007). The mammogram that cried Wolfe, *New England Journal of Medicine*, vol. 356, n° 3, p. 297-300.
- Kolb T., Lichy J. et Newhouse J.H. (2002). Comparison of the performance of screening mammography, physical examination, and breast us and evaluation of factors that influence them: an analysis of 27,825 patient evaluations 1, *Radiology*, vol. 225, n° 1, p. 165-175.
- Liyakathunisa et Ravi Kumar C.N. (2011). A Novel and Efficient Lifting Scheme based Super Resolution Reconstruction for Early Detection of Cancer in Low Resolution Mammogram Images, *International Journal of Biometrics and Bioinformatics*, vol. 5, n° 2, p. 53-75.
- McCormack V.A. et Dos Santos Silva I. (2006). Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis, *Cancer Epidemiology Biomarkers & Prevention*, vol. 15, n° 6, p. 1159-1169.
- Muhimmah I. et Zwigelaar R. (2006). Mammographic density classification using multiresolution histogram information, *Proceedings of the International Special Topic Conference on Information Technology in Biomedicine, ITAB 2006*, Ioannina, Greece.
- Mustra M., Grgi M. et Delač V.K. (2012). Breast density classification using multiple feature selection, *AUTOMATIKA*, vol. 53, n° 4, p. 362-372.
- Nguyen H. et Chan C. (2004). Multiple neural networks for a long term time series forecast, *Neural Computing & Applications*, vol. 13, n° 1, p. 90-98.
- Oliveira J.E., Gueld M.O., Araújo A.D.A., Ott B. et Deserno T.M. (2008). Toward a standard reference database for computer-aided mammography. *In: Medical Imaging. International Society for Optics and Photonics*. p. 69151Y-69151Y-9.

- Oliveira J.E., Machado A.M., Chavez G.C., Lopes A.P.B., Deserno T.M. et Araújo A.D.A. (2010). MammoSys: a content-based image retrieval system using breast density patterns, *Computer Methods and Programs in Biomedicine*, vol. 99, n° 3, p. 289-297.
- Oliver A., Freixenet J., Bosch A., Raba D. et Zwigelaar R. (2005a). Automatic classification of breast tissue. In: *Iberian Conference on Pattern Recognition and Image Analysis*, p. 431-438.
- Oliver A., Freixenet J., Bosch A. et Zwigelaar R. (2005b). Automatic classification of breast density. In: *IEEE International Conference on Image Processing*, vol. 2, p. 1258-1261.
- Patan K. (2008). *Artificial neural networks for the modelling and fault diagnosis of technical processes* Springer Science & Business Media.
- Petroudi S., Kadir T. et Brady M. (2003). Automatic classification of mammographic parenchymal patterns: A statistical approach. In: *Engineering in Medicine and Biology Society. Proceedings of the 25th Annual International Conference of the IEEE. IEEE, 2003*, Cancun, Mexico, p. 798-801.
- Pevrhál S., Shepherd J.A., Smith-Bindman R., Cummings S.R. et Kerlikowske K. (2002). Accuracy of mammographic breast density analysis: results of formal operator training, *Cancer Epidemiology Biomarkers & Prevention*, vol. 11, n° 11, p. 1389-1393.
- Rivals I., Personnaz L., Dreyfus G. et Ploix J.L. (1995). Modélisation, classification et commande par réseaux de neurones : principes fondamentaux, méthodologie de conception et illustrations industrielles. Les réseaux de neurones pour la modélisation et la commande de procédés, JP Corriou ed. Lavoisier Tec & Doc.
- Schetter S.E. (2014). Breast density as an independent risk factor for cancer, *J Am Osteopath Coll Radiol*, vol. 3, p. 10-19.
- Seror J.Y. (2005). Quelles explorations chez une patiente à densité mammaire élevée ? *Centre de Radiologie Duroc*.
- Stéphane T. (2012). *Data Mining et statistique décisionnelle : l'intelligence des données*, Éditions Technip.
- Stines J. (2004). La densité mammaire : un concept radiologique, *26^{es} journées de la SFSPM 2004*, Nancy, p. 150-158.
- Suckling J., Parker J., Dance D.R., Astley S., Hutt I., Boggis C. et Savage J. (1994). The Mammographic image analysis society digital mammogram database. In: *Excerpta Medica. International Congress Series 1994*, p. 375-378.
- Tardivon A. (2008). Quels risques, pour quelles femmes ? Densité mammaire et cancer du sein, *30^{es} journées de la SFSPM 2008*, La Baule, p. 93-101.
- Tremblay G. (2004). *Optimisation d'ensembles de classifieurs non paramétriques avec apprentissage par représentation partielle de l'information*. Thèse de doctorat. École de technologie supérieure. Université du Québec.
- Wolfe J.N. (1976a). Breast patterns as an index of risk for developing breast cancer, *American Journal of Roentgenology*, vol. 126, n° 6, p. 1130-1137.

Wolfe J.N. (1976b). Risk for breast cancer development determined by mammographic parenchymal pattern, *Cancer*, vol. 37, n° 5, p. 2486-2492.

Zribi M. et Boujelbene Y. (2012). Les réseaux de neurones un outil de sélection de variables : le cas des facteurs de risque de la maladie du cancer du sein, *Éthique et économique/Ethics and Economics*, vol. 9, n° 1, p. 173-182.

Article reçu le : 14 février 2013

Accepté le : 20 juillet 2016

