

République Algérienne Démocratique et Populaire

**Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mouloud MAMMERY, Tizi-Ouzou**



Faculté de Génie Electrique et d'Informatique
Département d'Automatique

MEMOIRE DE FIN D'ETUDES

En vue de l'obtention du diplôme

*de MASTER ACADEMIQUE EN AUTOMATIQUE
OPTION : COMMANDE DES SYSTEMES*

Thème

**Classification Par Réseaux de Neurones
Application Aux Image Médicales
Cérébrales**

Proposé par : Nait Belaid Ouiza

Présenté par : MEZAGUER Smail

Dirigé par : Nait Belaid Ouiza

Soutenu le : 04 / 07 /2013

Promotion 2013

Remerciements

C'est avec grand plaisir que je réserve cette page, en signe de gratitude et de reconnaissance à tous ceux qui m'ont aidé à la réalisation de ce travail.

Je remercie tout d'abord le bon dieu pour m'avoir donné le courage et la santé pour accomplir ce travail.

Je remercie également, ma promotrice Mlle. Naït Belaid Ouïza pour sa patience, pour ses conseils judicieux, sa grande disponibilité et pour m'avoir suivi et orienté.

Je remercie vivement les membres de jury d'avoir accepté de juger mon travail.

Mes remerciements pour mes chers parents, mes chers frères, mes chères sœurs et tous mes amis.

Je n'oublie pas non plus de remercier toutes les personnes que j'ai pu rencontrer et avec lesquelles j'ai pu échanger, et en particulier les enseignants du département d'Automatique.

Je remercie également l'ensemble du personnel du département d'Automatique de l'Université de Mouloud Mammeri Tizi-Ouzou.

Je remercie, enfin, toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce travail.

SOMMAIRE :

Chapitre I : Généralités sur l'imagerie médicale

I.1. Introduction	01
I.2. Techniques utilisées en imagerie médicale	01
I.2.1. Les Ultrasons	01
I.2.1.1 Echographie classique	01
I.2.1.2.2 Effet doppler	02
I.2.2 Les Rayons X	03
I.2.3. La Radioactivité	04
• La scintigraphie	04
• Tomographie par émission de positons (TEP)	04
I.2.4. L'imagerie par résonance magnétique	05
• Le signal IRM	05
• IRM Cérébrales	06
I.3. L'anatomie du cerveau	06
I.4. Observation du cerveau	08
I.5. Séquence d'acquisition	08
I.6. Quelques pathologies cérébrales	10
1. La maladie d'Alzheimer	10
2. La sclérose en plaques	10
3. Les tumeurs	11
Cas 1 : astrocytome de bas grade	11
Cas 2 : glioblastome	12
Cas 3 : L'Oligodendrogliome	12
I.7. Conclusion	13

Chapitre II : Panorama des méthodes de classification

II.1. Introduction	14
II.2. Méthodes de classification	14

SOMMAIRE :

II.2.1. Méthodes supervisées	15
II.2.1.1. K-plus proches voisins	15
II.2.1.2. La classification Bayésienne	15
II.2.1.3. Réseaux de neurones	16
II.2.1.4. Affectation par l'arbre de décision	16
II.2.1.5. Algorithme des machines à support vecteurs	17
II.2.2. Méthodes non supervisée	18
II.2.2.1. Méthode de K-means	19
II.2.2.2. Méthode fuzzy C-means	19
II.2.2.3. Méthodes hiérarchiques	20
II.3. Classification d'images médicales	21
II.4. Conclusion	23

Chapitre III : Introduction aux réseaux de neurones

II.1. Introduction	24
II.2. Historique sur réseaux de neurones	24
II.3. Le neurone biologique	25
II.4. Modélisation d'un neurone	26
II.4.1. Neurone formel	26
II.4.2. Fonctions d'activations	27
II.4.3. Apprentissage des réseaux de neurones	29
• Apprentissage supervisé	29
• Apprentissage non supervisé	30
II.4.4. Topologie des réseaux de neurones	30
II.5. Perceptron multicouches (multi layer perceptron)	31
II.5.1. Architecture	31
II.5.2. Apprentissage des réseaux multicouches	32
II.5.3. Algorithme de rétropropagation du gradient	32
II.5.4. Algorithme de Levenberg-Marquardt	34
II.6. Classification des images médicales par les réseaux de neurones	36

SOMMAIRE :

II.7. Avantages et inconvénients des réseaux de neurones	37
II.7.1. Les avantages des réseaux de neurones	37
II.7.2. Les points faibles des réseaux de neurones	37
II.8. Conclusion	38

Chapitre IV : Tests et résultats

IV .1. Introduction	39
IV .2. Les descripteurs utilisés	39
IV .2.1. Descripteurs de texture	39
IV .2.2. Descripteurs de forme	41
IV .3. Bases d'images	42
IV .4. Conception du réseau de neurones	43
IV .4.1. Architecture et configuration de réseau de neurones	43
IV .4.2. Apprentissage des réseaux de neurones	44
IV .4.3. Phase de test	45
IV .5. Critères de performances	46
IV .5.1 Critères de performances de l'apprentissage	46
IV .5.2 Critères de performances de test	46
IV .6. Résultats de différents apprentissages	47
IV .6.1. Résultats d'apprentissage avec les descripteurs de texture	47
IV .6.2. Résultats de l'apprentissage avec les descripteurs de forme	51
IV.7. Résultats de tests	53
IV .7.1. Résultats de test avec les descripteurs de texture	53
IV .7.2. Résultats de test avec les descripteurs de forme	55
IV .8. Comparaison des résultats obtenus	60
IV .9. Conclusion	61

Liste des figures et des tableaux

Liste des figures et tableaux :

Chapitre I :

Fig.1. Appareil échographe et l'image échographique	02
Fig.2. Échographe doppler et l'image échographique	02
Fig.3. Système de production de rayons X par un tube à rayons X	03
Fig.4. L'image (a) illustre l'encéphale, composée du cerveau, du cervelet et du tronc cérébral. L'image (b) montre une coupe d'un cerveau mettant en évidence les trois matières principales de l'encéphale	07
Fig.5. Plans axial, coronal et sagittal (de gauche à droite) sur une acquisition en T1	08
Fig.6. Séquences d'IRM	09
Fig.7. A droite, IRM d'un cerveau sain. A gauche, cerveau atteint de la maladie d'Alzheimer.....	10
Fig.8. Lésion typique de la sclérose en plaques	11
Fig.9. Un cas d'astrocytome de bas grade	11
Fig.10. Un cas de glioblastome	12
Fig.11.L'Oligodendrogliome	12

Chapitre II :

Fig.1. Méthodes de classification	14
Fig.2. Exemple affectation par l'approche de l'arbre de décision	17

Chapitre III :

Fig.1. Le neurone biologique	25
Fig.2. Représentation d'un neurone artificiel	26
Fig.3. Fonction Heaviside	27
Fig.4. Fonction signe	27
Fig.5 Fonction linéaire	27
Fig.6. Fonction linéaire a seuil	28
Fig.7. Fonction logistique	28
Fig.8. Fonction symétrique	28
Fig.9. La fonction gaussienne	29
Fig.10. Apprentissage supervisé	30
Fig.11. L'apprentissage non supervisé	30

Liste des figures et tableaux :

Fig.12. Définition des couches d'un réseau multicouche	31
Fig.13. Perceptron multicouche	32
Fig.14. Etapes de classification par réseau de neurones	36
Chapitre IV :	
Fig.01. Exemples images de la Base d'Apprentissage	43
Fig.02. Architecture de réseau de neurones	44
Fig.03. Les étapes de l'apprentissage d'un réseau de neurones	45
Fig.04. Histogramme d'erreurs quadratiques en utilisant l'algorithme de Levenberg- Marquardt	48
Fig.05. Histogramme des temps d'apprentissage en utilisant l'algorithme de Levenberg- Marquardt	49
Fig.06. Histogramme d'erreurs en utilisant l'algorithme de descente de gradient	49
Fig.07. Histogramme des temps d'apprentissage en utilisant l'algorithme de descente de gradient	50
Fig.08. Histogramme des performances en utilisant l'algorithme de Levenberg-Marquardt et la fonction Tansig	52
Fig.09. Histogramme des performances en utilisant l'algorithme de descente de gradient et la fonction Tansig	53
Fig.10. Histogramme de meilleur résultat de test	54
Fig.11. Histogramme d'erreurs quadratiques en utilisant l'algorithme de Levenberg- Marquardt et la fonction Tansig	54
Fig.12. Histogramme des temps d'apprentissage en utilisant l'algorithme de Levenberg- Marquardt et la fonction Tansig	55
Fig.13. Histogramme des taux de classification pour les 04 cas	56
Fig.14. Histogramme des sensibilités pour les 04 cas	56
Fig.15. Histogramme des spécificités pour les 04 cas	57
Fig.16. Histogramme des meilleurs taux de classification pour les 04 cas	58
Fig.17. Histogramme des meilleures sensibilités pour les 04 cas	59
Fig.18. Histogramme des meilleures spécificités pour les 04 cas	59
Fig.19. Histogramme des meilleurs résultats de classification	60
Tableau O1. Quelques valeurs d'attributs de 03 images de la base d'apprentissage	48
Tableau.O2.Quelques valeurs de moments de Zernike pour une image IRM cérébrale	51
Tableau.03. les meilleures performances obtenues	60

Chapitre I :

Généralités sur l'imagerie médicale

I.1.Introduction

L'imagerie médicale regroupe les moyens d'acquisition et de restitution d'images du corps humain à partir de différents phénomènes physiques tels que l'absorption des rayons X, la résonance magnétique nucléaire, la réflexion d'ondes ultrasons ou la radioactivité. Ces technologies ont révolutionné la médecine grâce au progrès de l'informatique (programmation et algorithmes) en permettant de visualiser l'anatomie, la physiologie ou le métabolisme du corps humain afin de mieux le diagnostiquer, le pronostiquer, et le soigner.

I.2. Techniques utilisées en imagerie médicale :

Les examens d'imagerie médicale dépendent essentiellement de la technique utilisée, ils permettent d'obtenir des informations sur l'anatomie des organes (la taille, la forme, le volume et localisation de l'organe), dans ce cas on parle de l'imagerie structurelle, ou sur leur fonctionnement (physiologie, métabolisme), dans ce cas on parle de l'imagerie fonctionnelle.

Ci –dessous nous illustrons quelques techniques d'imagerie médicales les plus souvent utilisées et plus particulièrement l'IRM :

I.2.1. Les Ultrasons :

Les ultrasons sont des ondes sonores non audibles par l'oreille humaine à cause de leur fréquence élevée. Ils sont utilisés en médecine notamment dans l'échographie qui permet via l'informatique de générer des images en fonction de la réponse des différents tissus face à l'émission à leur niveau de ces ultrasons. Dans d'autres cas, ils permettent de détruire des substances ou tissus indésirables. Ils sont également utilisés en kinésithérapie dans un but thérapeutique. Le kinésithérapeute utilise un générateur dont l'application permet, grâce à la chaleur créée par ces hautes fréquences, de diminuer les douleurs, de favoriser la circulation, ou de rendre des articulations moins raides [01].

I.2.1.1 Echographie classique :

L'échographie est une technique d'imagerie d'exploration de l'intérieur du corps basée sur les ultrasons. Le faisceau d'ultrasons émis par une sonde se propage au travers de certaines zones de l'organisme et est réfléchi par les différentes structures anatomiques qu'il rencontre selon la densité et la nature du milieu traversé. Le signal recueilli est ensuite traité, ce qui permet la visualisation des organes et des structures anatomiques [02].

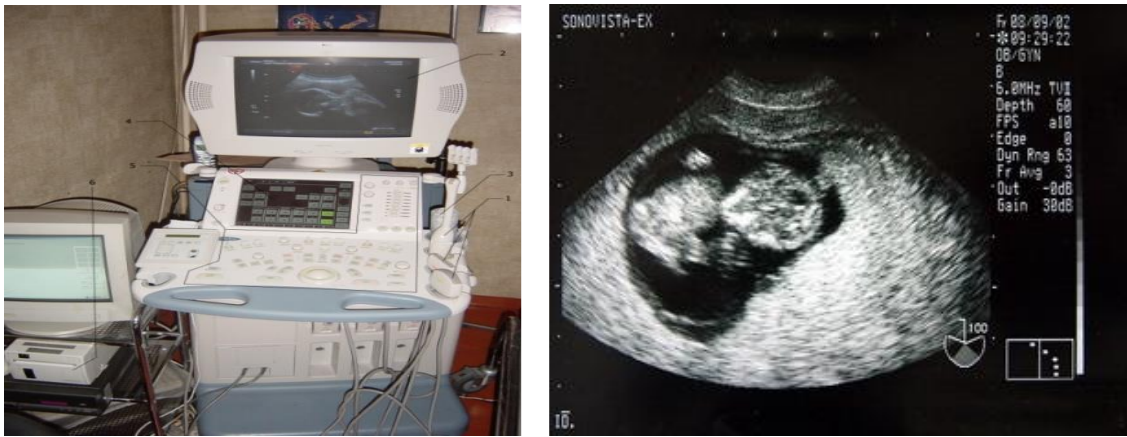


Fig.1. Appareil échographe et l'image échographique

I.2.1.2. Effet doppler :

L'effet Doppler ou effet Doppler-Fizeau est le décalage de fréquence d'une onde acoustique ou électromagnétique entre la mesure à l'émission et la mesure à la réception lorsque la distance entre l'émetteur et le récepteur varie au cours du temps. Si on désigne de façon générale ce phénomène physique sous le nom d'effet Doppler, on réserve le terme d'« effet Doppler-Fizeau » aux ondes lumineuses.

L'échographe Doppler est constitué d'un échographe classique couplée à une sonde Doppler. Lorsqu'un faisceau d'ultrasons traverse les cavités cardiaques ou les vaisseaux, l'écho renvoyé par les éléments figurés du sang (qui deviennent des émetteurs) aura une longueur d'onde plus longue (son plus grave) s'ils s'éloignent du capteur ; et une longueur d'onde plus courte (son plus aigu) s'ils se dirigent vers le capteur. La sonde Doppler recueillera donc un écho dont la fréquence (longueur d'onde) sera différente de la fréquence d'émission. La différence des fréquences d'émission et de réception : $F - F'$ ou $F' - F$ permet de calculer la vitesse et la direction des globules rouges [03].



Fig.2 Échographe doppler et l'image échographique

I.2.2 Les Rayons X :

Les rayons X sont une forme de rayonnement électromagnétique à haute fréquence constitué de photons dont la longueur d'onde est comprise approximativement entre 0,01 nanomètre et 10 nanomètres, correspondant à des fréquences de 30 péta-hertz à 30 exa-hertz. L'énergie de ces photons va de quelques électronvolts, à plusieurs dizaines de Mégaélectronvolts. C'est un rayonnement ionisant utilisé dans de nombreuses applications dont l'imagerie médicale tel que la radiographie conventionnelle. Les rayons X peuvent être produits de deux manières, soit par des transitions électroniques ou par accélération d'électrons. Le système de production des rayons X est représenté dans la figure ci-dessous. Il est basé sur le freinage des électrons sur une cible dans un tube à rayon X [04].

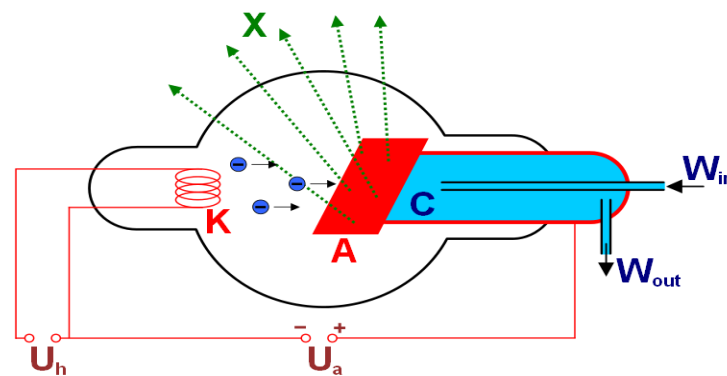


Fig.3. Système de production de rayons X par un tube à rayons X (ici à anode fixe).

X : Rayons X, K : Cathode (et radiateur), A : Anode, W_{in} : Entrée d'eau

W_{out} : Sortie d'eau, C : Réfrigérant à eau, U_h : Cathode-haute-tension, U_a : Anode-tension

Le premier point à retenir est qu'un tube radiogène ne fonctionne pas tout seul, il doit être relié à un générateur de haute tension (environ 100 kV), à un générateur secondaire de basse tension et à un système de refroidissement. Le tube radiogène est constitué d'une cathode et d'une anode entourée par des enveloppes de protection. La cathode est la source des électrons. Il s'agit d'un filament en forme de spirale, composé généralement de tungstène, qui s'échauffe lors de la mise en route du tube pour laisser s'échapper les électrons. Les électrons sont accélérés entre la cathode et l'anode par une forte différence de potentiel, délivré par le générateur. L'anode est la cible des électrons et le lieu de production des rayons X. La surface de bombardement des électrons sur l'anode s'appelle le foyer. La surface de l'anode est oblique par rapport à la direction du faisceau d'électron de manière à permettre d'avantage aux rayons X de pouvoir sortir du tube. Le tube radiogène, mis sous vide, est

entouré de plusieurs enveloppes de protection permettant d'assurer une protection thermique, électrique et mécanique.

I.2.3. La Radioactivité :

La radioactivité est un phénomène physique naturel au cours duquel des noyaux atomiques instables, dits radio-isotopes, se transforment spontanément (désintégration), en dégageant de l'énergie sous forme de rayonnements divers, en des noyaux atomiques plus stables ayant perdu une partie de leur masse. Les rayonnements ainsi émis sont appelés, selon le cas, des rayons alpha, des rayons bêta ou des rayons γ .

Deux techniques d'imagerie médicale font appel au phénomène radioactivité : la scintigraphie et la tomographie par émissions de positons [05].

A. La scintigraphie :

La scintigraphie est une méthode d'imagerie médicale qui procède par l'administration, dans l'organisme, d'isotopes radioactifs afin de produire une image médicale par la détection des rayonnements émis par ces isotopes après captation par les organes à examiner. Le patient reçoit des molécules radioactives. Elles vont être absorbées par les organes. Ensuite, on analyse les rayons émis par le corps. Enfin, on reconstruit l'image de l'organe. La dose est contrôlée de manière à minimiser les risques liés à l'irradiation.

Par exemple, une scintigraphie osseuse est égale à une radio du bassin, si on regarde la dose de radiations reçue. Elle peut être utilisée pour visualiser plusieurs parties du corps, selon le traceur utilisé et ce que l'on veut voir. Elle permet de voir la forme et le fonctionnement de l'organe, la présence de tumeurs primaires ou secondaires, l'origine de douleurs ou de maladies.

B. Tomographie par émission de positons (TEP) :

La TEP repose sur le principe général de la Scintigraphie qui consiste à injecter un traceur dont on connaît le comportement et les propriétés biologiques pour obtenir une image du fonctionnement d'un organe. Ce traceur est marqué par un atome radioactif (carbone, fluor, azote, oxygène...) qui émet des positons dont l'annihilation produit elle-même deux photons. C'est la détection en coïncidence de ces derniers qui permet la localisation du lieu de leur émission et donc la concentration du traceur en chaque point de l'organe. C'est cette

information quantitative que l'on représente sous la forme d'une image faisant apparaître en couleurs les zones de forte concentration du traceur.

I.2.4.L'imagerie par résonance magnétique :

L'imagerie par résonance magnétique (IRM) est une technique d'imagerie médicale permettant d'obtenir des vues 2D ou 3D de l'intérieur du corps de façon non-invasive avec une résolution relativement élevée. L'IRM repose sur le principe de la résonance magnétique nucléaire (RMN) qui utilise les propriétés quantiques des noyaux atomiques pour la spectroscopie en chimie, elle nécessite un champ magnétique puissant et stable produit par un aimant supraconducteur qui crée une magnétisation des tissus par alignement des moments magnétiques de spin.

Des champs magnétiques oscillants plus faibles, dits radiofréquence, sont alors appliqués de façon légère pour modifier cet alignement et produire un phénomène de précession qui donne lieu à un signal électromagnétique mesurable. La spécificité de l'IRM consiste à localiser précisément dans l'espace l'origine de ce signal RMN en appliquant des champs magnétiques non-uniformes, des « gradients », qui vont induire des fréquences de précession légèrement différentes en fonction de la position des atomes dans ces gradients. Il est alors possible de reconstruire une image en deux dimensions puis en trois dimensions de la composition chimique et donc de la nature des tissus biologiques explorés [06].

A. Le signal IRM :

Le corps humain contient un grand nombre d'atomes d'hydrogène et, à l'heure actuelle, l'imagerie par résonance magnétique du corps humain est basée sur les propriétés magnétiques de ces atomes d'hydrogène. Le noyau de cet atome d'hydrogène est composé d'un unique proton (chargé positivement) qui est en mouvement. Tous ces protons peuvent être considérés comme des petits aimants. En plaçant le patient à l'intérieur du tunnel où règne un champ magnétique, tous ces protons vont s'aligner. Le système est en équilibre et la production d'images nécessite la survenue d'un élément perturbateur, excitateur : l'émission d'ondes radio. Lorsque celles-ci ont une fréquence particulière, les noyaux d'hydrogène vont être placés dans un état dit de «résonance». Cette mise en résonance produit une rotation des petits aimants du patient. Le retour à l'état normal produit une variation du champ magnétique. Dans une antenne correctement placée, cette variation du champ magnétique induit la formation d'un courant électrique : c'est le signal IRM.

B. *IRM Cérébrales* :

L'étude structurelle et fonctionnelle du cerveau a connu un développement fulgurant ces dernières années grâce aux méthodes d'imagerie, dont notamment l'imagerie par résonance magnétique (IRM).

L'imagerie cérébrale, en particulier l'I.R.M. cérébrale, a été un progrès majeur dans le diagnostic et la prise en charge des maladies neurologiques. L'I.R.M. permet de détecter des lésions, par exemple en cas de crises d'épilepsie, de perte de l'usage de la parole ou d'un membre (tumeur cérébrale, malformation vasculaire, hématome, anomalie acquise ou congénitale des structures cérébrales). Une véritable cartographie anatomique et fonctionnelle du cerveau a ainsi pu être dressée en visualisant ses zones actives lorsqu'elles sont sollicitées au cours de l'exécution d'une tâche donnée. Pour voir le cerveau penser et mieux comprendre les dysfonctionnements cérébraux, une des méthodes consiste à suivre en direct les modulations du débit sanguin liées à l'activité cérébrale. Il est également possible de mesurer, à l'échelle microscopique, les mouvements des molécules d'eau du cerveau et établir ainsi l'architecture fine du tissu neuronal et de ses variations. D'autres molécules d'intérêt biologique impliquées dans le métabolisme renseignent également sur les pathologies cérébrales. Le développement d'IRM à champ magnétique intense permettra d'observer le cerveau et ses pathologies avec une précision encore plus fine, à une échelle plus représentative des phénomènes qui l'animent.

I.3. L'anatomie du cerveau :

Bien que représentant seulement 2% du poids total du corps humain (soit environ 1,4 kilogrammes), le cerveau gère directement ou indirectement 98 % de ses fonctions. Il est responsable des fonctions humaines les plus complexes comme la pensée, la résolution de problèmes, les émotions, la conscience et les comportements sociaux, et régit les fonctions essentielles du corps comme la respiration, le processus d'alimentation, le sommeil, les mouvements et les cinq sens. En dépit de son extrême complexité, le cerveau n'est composé que de deux types de cellules : les neurones et les cellules gliales [07].

- Les neurones sont des cellules nerveuses capables de recevoir et de transmettre l'information. Ils sont constitués d'un corps cellulaire, de plusieurs prolongements afférents appelés dendrites et d'un prolongement efférent appelé axone. Chaque neurone peut posséder jusqu'à 10.000 connexions avec d'autres neurones, ce qui conduit à un nombre très élevé de réseaux interconnectés.

- Les cellules gliales sont des cellules de soutien qui contribuent à assurer le bon fonctionnement des neurones, sans participer directement au transfert de l'information. Le cerveau contient plus de 100.000 milliards de neurones et encore davantage de cellules gliales.

Le cerveau est la partie la plus volumineuse du système nerveux central. Il est placé dans la boîte crânienne. Il comprend le cerveau, le cervelet et le tronc cérébral.

-*Le cerveau* : occupe la majeure partie de l'encéphale, et est divisé en 2 hémisphères cérébraux : l'hémisphère droit et l'hémisphère gauche. La surface du cerveau présente de nombreux replis, avec un grand nombre de circonvolutions limitées par des sillons.

-*Le cervelet* : est situé en parallèle sur deux grandes voies nerveuses : celles qui amènent les messages sensoriels vers les zones du cortex qui en font l'analyse, et celles qui partent du cortex et descendent vers les muscles pour les faire contracter. Le cervelet est en partie impliqué dans la coordination des mouvements du corps.

- *Le tronc cérébral* : représente un carrefour majeur du système nerveux central entre le cerveau et le cervelet. Il est responsable en particulier de la régulation de la respiration et du rythme cardiaque. C'est également un centre de passage des voies motrices et sensitives, ainsi qu'un centre de contrôle de la douleur.

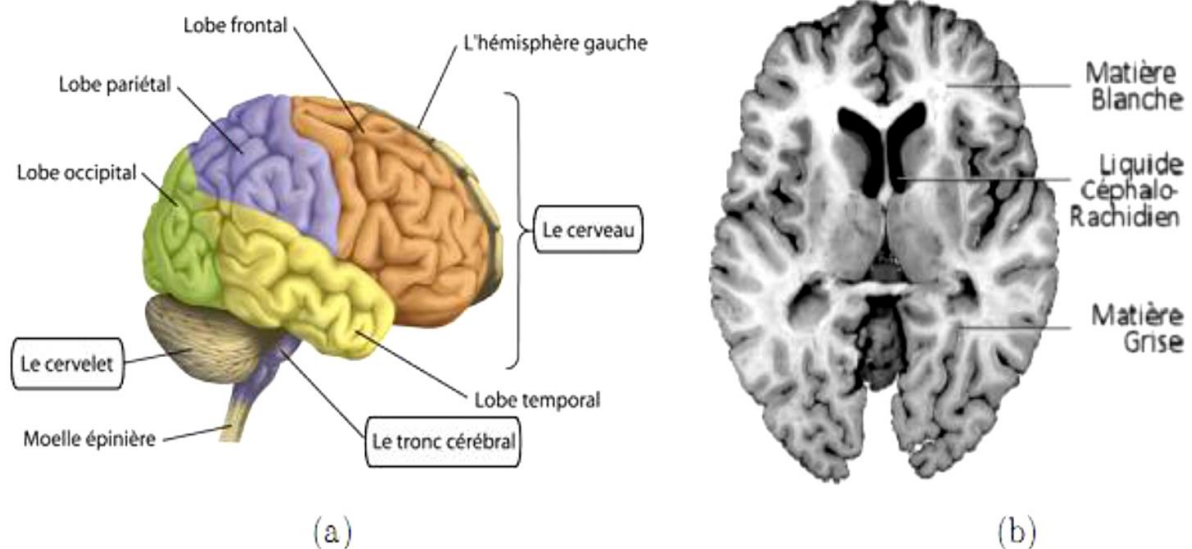


Fig.4.L'image (a) illustre l'encéphale, composée du cerveau, du cervelet et du tronc cérébral.

L'image (b) montre une coupe d'un cerveau mettant en évidence les trois matières principales de l'encéphale.

I.4. Observation du cerveau :

Le cerveau peut être observé par des coupes bidimensionnelles selon plusieurs angles de vue. Il existe principalement trois axes anatomiques qui permettent de réaliser les coupes : axiale, sagittale et frontale [07].

- *Coupe axiale (ou transverse)* : cette coupe représente une vue de dessus du cerveau. Elle correspond à un plan perpendiculaire au champ magnétique statique.
- *Coupe sagittale* : cette coupe est relative à un plan parallèle au plan inter-hémisphérique. Il s'agit de vues latérales du cerveau.
- *Coupe frontale (ou coronale)* : cette coupe est une vue de face du cerveau. Elle est acquise sur le plan perpendiculaire aux coupes axiale et sagittale.

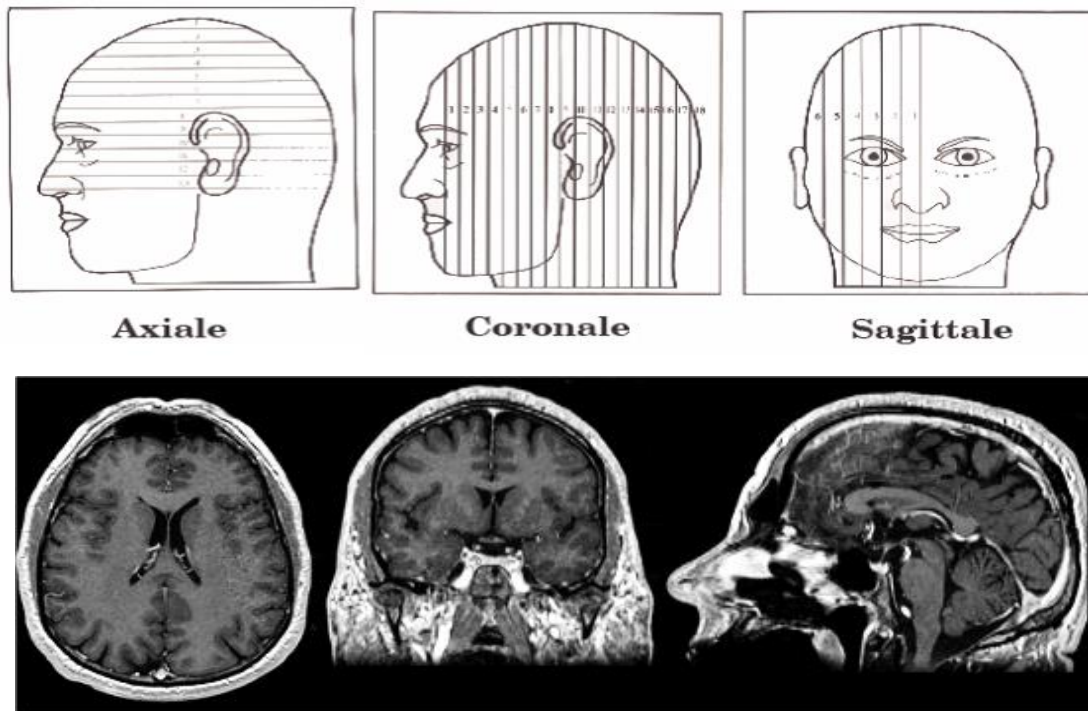


Fig.5. Plans axial, coronal et sagittal (de gauche à droite) sur une acquisition en T1

I.5. Séquence d'acquisition :

Dans l'IRM classique, on peut pondérer l'image en T1, T2 ou en densité de protons [07], suivant certains paramètres d'acquisition :

- le temps d'écho TE, c'est-à-dire le temps qui sépare l'impulsion RF et l'acquisition du signal.
- le temps de répétition TR, c'est à dire le temps entre deux impulsions RF consécutives.

Les tissus ayant des temps T1 et T2 différents en fonction de leur richesse en atome d'hydrogène et en fonction du milieu dans lequel ces derniers évoluent, peuvent renvoyer des signaux différents si l'on arrive à mettre en évidence ces différences de temps. Trois types de contrastes peuvent donc être obtenus :

- En utilisant un temps de répétition court (TR) et un temps d'écho (TE) court (neutralise les différences de temps T2), on obtient un contraste d'image pondérée en T1, pondération dite « anatomique ».
- En utilisant un temps de répétition long (neutralise les différences de temps T1) et un temps d'écho long, on obtient un contraste d'image dite pondérée en T2, dite aussi pondération « tissulaire ».
- Pour les images pondérées en densité de protons, le contraste est obtenu lorsque l'intensité de l'image dépend essentiellement de la densité locale de protons et beaucoup moins des constantes de relaxation T1 et T2. Il est obtenu en utilisant un TR long et un TE court.

Chaque modalité contient des informations spécifiques ne se retrouvant pas dans les autres. À partir des images pondérées en T1, on peut par exemple distinguer les différents tissus cérébraux tels que la matière blanche, la matière grise et le liquide céphalo-rachidien, tandis que les images pondérées en T2 mettent plus facilement en évidence certaines anomalies comme les lésions dues à la sclérose en plaque (SEP).

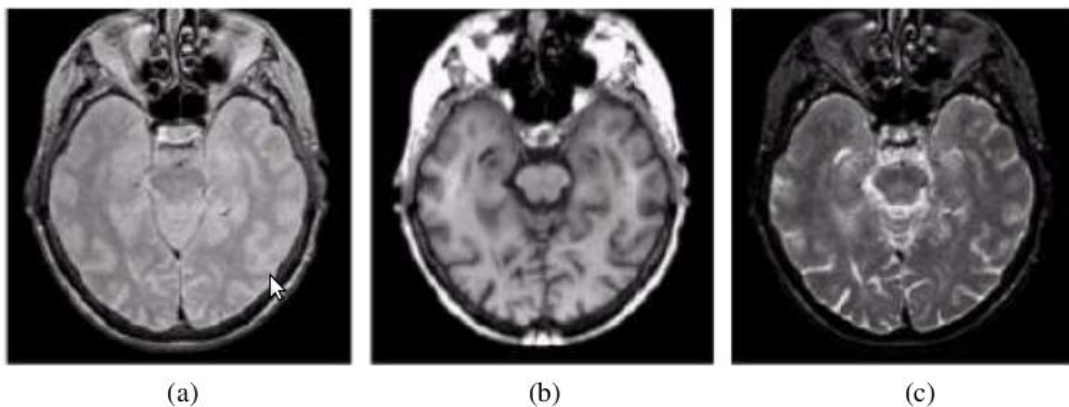


Fig.6.Séquences d'IRM.

Images d'une même coupe du cerveau humain, pondérées en :

(a) densité de protons, (b) en T1 et (c) en T2.

I.6. Quelques pathologies cérébrales :

I.6.1. La maladie d'Alzheimer :

Le meilleur exemple des pathologies cérébrales est donné par la plus fréquente des maladies neurodégénératives : la maladie d'Alzheimer [08]. Au cours de cette pathologie, on observe une perte diffuse des neurones de l'encéphale. Cette dégénérescence bien que diffuse touche préférentiellement 2 zones : le cortex cérébral et l'hippocampe. Cette perte neuronale est responsable d'une atrophie cérébrale c'est-à-dire une diminution du volume de tissu cérébral.

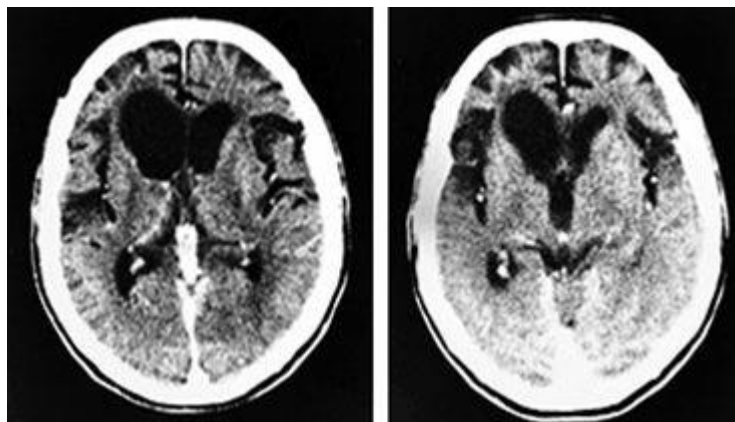


Fig.7.A droite, IRM d'un cerveau sain. A gauche, cerveau atteint de la maladie d'Alzheimer.

Ce vide tissulaire est comblé par une augmentation parallèle du volume liquidien occupé par le liquide céphalorachidien. On observe ainsi augmentation de la taille des ventricules cérébraux et des sillons corticaux.

I.6.2. La sclérose en plaques :

La sclérose en plaques (SEP) [09] est une maladie neurologique auto immune (dus à une hyperactivité) chronique du système nerveux central. Ses manifestations cliniques sont liées à une démyélinisation (disparition ou la destruction) des fibres nerveux du cerveau, de la moelle épinière et du nerf optique. Les causes restent inconnues. Des théories évoquent des facteurs génétiques et environnementaux comme les infections.

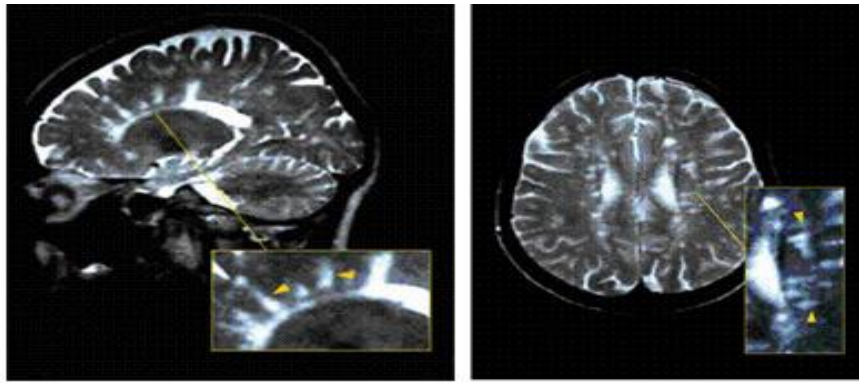


Fig.8.Lésion typique de la sclérose en plaques

I.6.3. Les tumeurs :

Une tumeur au cerveau est une masse intracrânienne due à une croissance incontrôlée des cellules on la trouve généralement dans les neurones, les tissus lymphatiques, les cellules gliales etc. [08].Les Tumeurs cérébrales peuvent être primitives ou des métastases.

Cas 1 : astrocytome de bas grade

Le cas présenté sur la figure 1.13 correspond à un astrocytome de bas grade. Ce type de tumeur représente environ 5% des tumeurs intracrâniennes et se caractérise généralement par une masse hémisphérique non encapsulée et infiltrante. Souvent non accompagnés d'œdème, les astrocytomes de bas grade sont hyper-intenses en pondération T2 et peu ou moyennement contrastés en pondération T1.

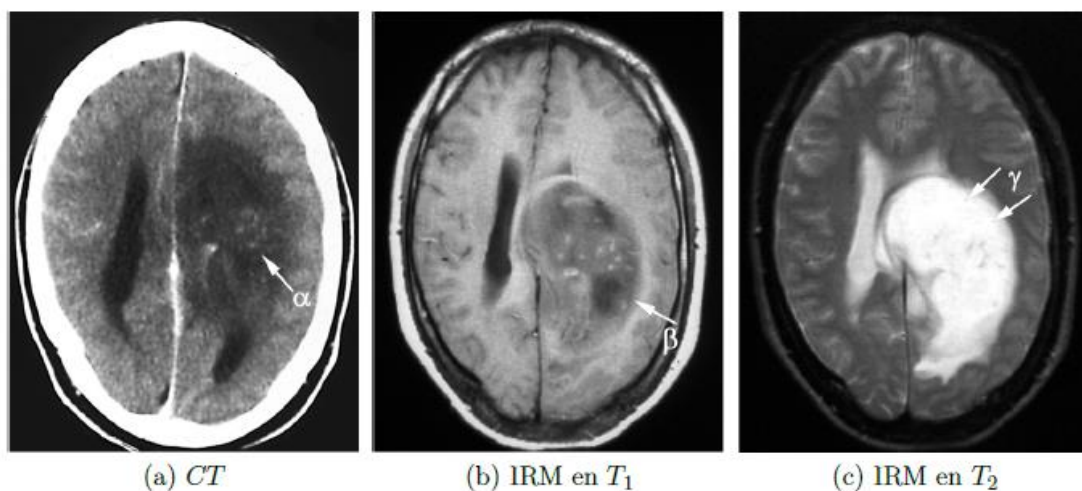


Fig.9. Un cas d'astrocytome de bas grade

Cas 2 : glioblastome :

La lésion présentée sur la figure 1.14 correspond à un glioblastome. Ce type de tumeur, de grade IV, est relativement fréquent puisqu'il représente 20% de toutes les tumeurs intracrâniennes. Survenant le plus souvent chez des patients entre 50 et 70 ans, les glioblastomes forment des masses molles et riches en vaisseaux sanguins. Souvent hétérogènes, ils présentent des nécroses et sont accompagnés d'œdèmes.

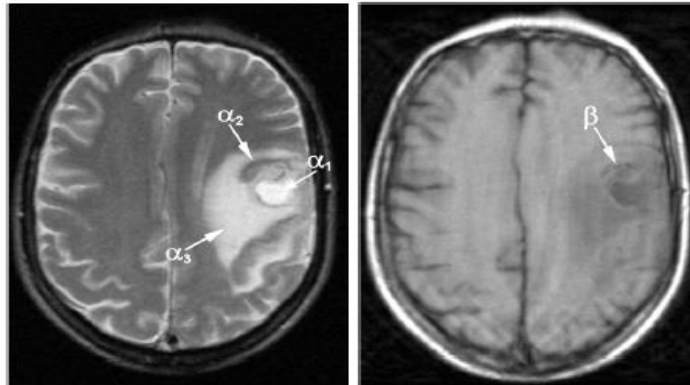
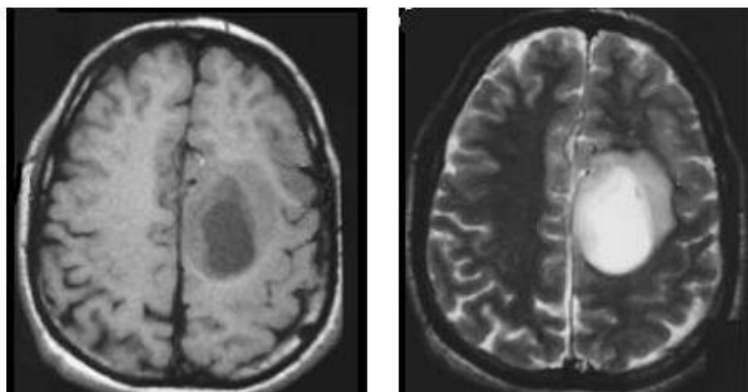


Fig.10. Un cas de glioblastome

Cas 3 : L'Oligodendrogliome :

Il s'agit d'une tumeur bénigne évoluant lentement. Les oligodendrogliomes représentent de 2% à 5% des tumeurs cérébrales primitives et 4% à 15% des gliomes. Ils sont généralement situés dans le lobe frontal, temporal ou les lobes pariétaux. Une dégénérescence kystique est courante, mais l'hémorragie et l'œdème sont rares. L'IRM est souvent moins caractéristique que le scanner. Ces tumeurs peuvent apparaître comme un "œuf sur un plat".



(a)

(b)

*Fig.11. L'Oligodendrogliome
a) coupe axiale pondéré en T1.*

b) le même coupe pondéré en T2 qui montre le kyste au centre.

I.7. Conclusion :

Nous avons présenté dans ce chapitre quelques notions générales portant sur l'imagerie médicale. Cette dernière permet d'analyser les tissus par des médias extrêmement divers, leurs exploitations et leurs interprétations permettent d'établir plus finement le diagnostic médical. Les techniques d'imagerie médicale sont multiples basées sur différents types de rayonnements et phénomènes physiques, avec le développement technologique, les appareils médicaux fournissent la plus part de temps des images numériques qui peuvent donc subir des traitements afin de faciliter le diagnostic du médecin.

Nous avons aussi détaillé la technique d'acquisition d'image par la résonance magnétique (IRM) en passant par l'anatomie cérébrale. Enfin nous avons cité quelques pathologies cérébrales. Dans notre travail, nous allons effectuer une classification d'images médicales cérébrales en deux classes d'images saines et pathologiques.

Chapitre II :
Panorama des méthodes de classification

II.1. Introduction :

Les méthodes de classification ont pour but d'identifier les classes auxquelles appartiennent des objets à partir de certains traits descriptifs. On parle de classification d'images lorsque les objets correspondent à des images où chaque image peut être caractérisée par un ensemble d'attributs tels que les attributs de couleur, de texture, ou de forme. La classification des images intervient comme étape très importante dans de différentes applications et surtout dans le domaine médical, elle permet d'aider les médecins à établir un diagnostic médical ou encore contrôler une action thérapeutique en lui offrant de nouveaux outils avec plus de précision et de performance.

Nous présentons dans ce chapitre un panorama des méthodes de classification les plus connues et qui font référence à l'existence de groupes ou classes de données.

II.2. Méthodes de classification :

Les méthodes utilisées pour la classification sont nombreuses, elles peuvent être divisées en deux catégories selon l'information disponible a priori sur les objets. La figure suivante montre quelques méthodes de classification.

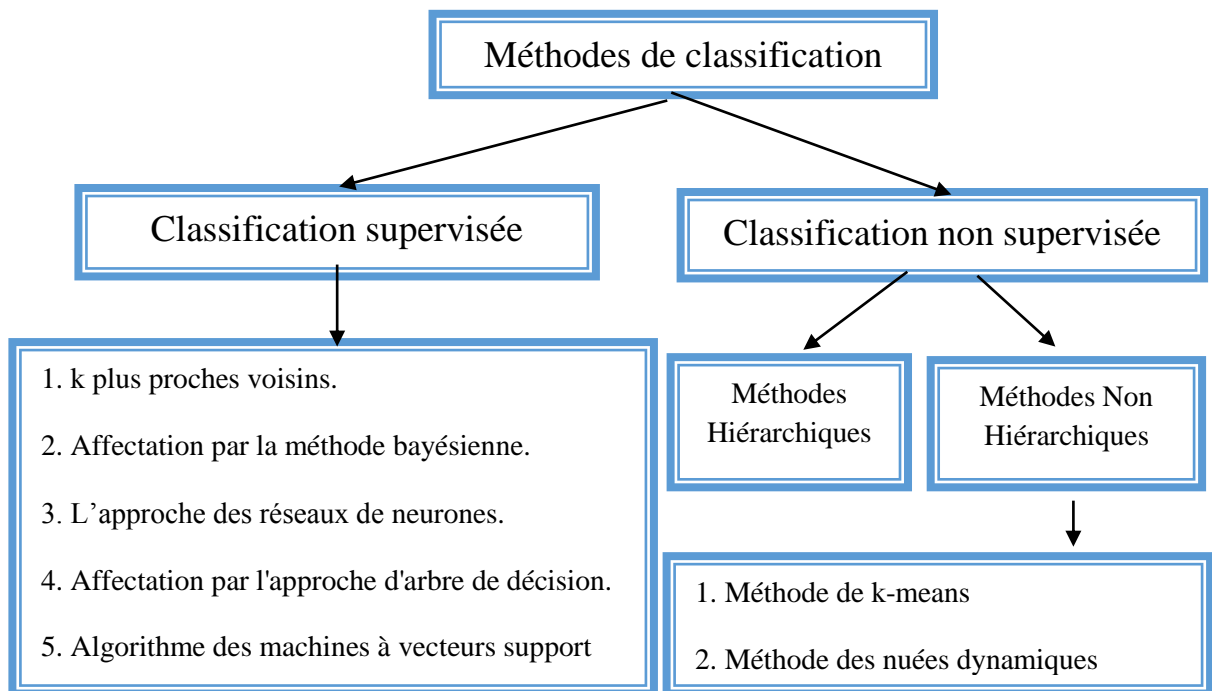


Fig.1.Méthodes de classification

II.2.1. Méthodes supervisées :

Les méthodes d'affectation (aussi appelées «classificateurs») basées sur la notion d'apprentissage supervisé : méthodes utilisant un ensemble d'exemples où les classes d'appartenance sont connues au préalable. À partir de cet ensemble, des normes (ou règles) d'affectation seront définies.

II.2.1.1. K-plus proches voisins :

Plus connus en anglais sous le nom K-Nearest Neighbor (K-NN), ou encore Memory Based Reasoning. La méthode des plus proches voisins consiste à déterminer pour chaque nouvel objet que l'on veut classer, la liste des plus proches voisins parmi les objets déjà classés. L'objet est affecté à la classe qui contient le plus d'objets parmi ces plus proches voisins. Cette méthode nécessite de choisir une distance, la plus classique est la distance euclidienne, et le nombre K de voisins à prendre en compte.

L'avantage de la méthode des K-plus proches voisins qu'elle ne nécessite pas d'apprentissage. La méthode peut s'appliquer dès qu'il est possible de définir une distance sur les champs. Ainsi cette méthode permet de traiter des problèmes avec un très grand nombre d'attributs. Les inconvénients de cette méthode sont le temps de classification qui est assez grand et le choix de K et la distance [10].

II.2.1.2. La classification Bayésienne :

Les techniques se basant sur les lois statistiques sont les premières qui ont été utilisées pour l'analyse de données. Elles consistent à prendre un sous ensemble d'une population et essayer d'arriver à des conclusions concernant toute la population. Ce sont des méthodes qui reposent sur la théorie de Bayes représentant une référence théorique pour les approches statistiques de résolution des problèmes de classification [11]. Le principe de cette théorie est le suivant :

Soit un ensemble C de K classes : $C = \{C_r\}$ tel que : $(r = 1, 2, \dots, K)$

Soit un vecteur X aléatoire dont on connaît la densité de probabilité conditionnelle : $P(X/C_r)$

La règle de Bayes permet de déterminer la meilleure classe (la classe la plus probable) $C_r(X)$ en appliquant la formule suivante :

décider C_j si :

$$P(X/C_j).P(C_j) \geq P(X/C_r).P(C_r) \quad \text{II. 02}$$

$$\forall r \neq j, \quad 1 \leq r \leq K$$

La probabilité qu'une observation X appartient à la classe C_j est donnée par la loi suivante :

$$P(C_j/X) = \frac{P(X/C_j)P(C_j)}{P(x)} \quad \text{II. 03}$$

Cette règle est optimale si le coût moyen des erreurs commises lors de la classification est minimal. La minimisation de coût moyen revient à la maximisation de $P(X/C_j)$.

II.2.1.3. Réseaux de neurones :

Le réseau de neurones est un classifieur non linéaire se basant sur le fonctionnement des neurones humains de manière à imiter le fonctionnement de notre cerveau. Un réseau de neurones est constitué de neurones artificiels interconnectés dont le rôle est de mimer les propriétés des neurones biologiques. Un neurone permet de définir une fonction discriminante dans l'espace d'entrée en réalisant une combinaison linéaire de son vecteur de caractéristiques.

Afin de traiter les problèmes de classification, on utilise les réseaux à couches. Les réseaux à couches sont connus sous le nom de perceptron multicouche. Ce sont des réseaux où les neurones sont regroupés en couches connectées entre elles. On distingue trois types de couches : la couche d'entrée, la couche de sortie et les couches cachées. Les perceptrons multicouches utilisent le principe de rétropropagation. L'avantage majeur d'utilisation des réseaux de neurones pour la classification est la facilité de mise en œuvre. Les inconvénients de cette méthode qu'elle nécessite une phase d'apprentissage et les résultats fournis à la sortie de réseau de neurones ne sont pas justifiés [12]. Les réseaux de neurones sont utilisés dans notre travail, ils seront vus en détail dans le prochain chapitre.

II.2.1.4. Affectation par l'arbre de décision :

La méthode des arbres de décision est une méthode de classification supervisée représentée sous la forme d'une arborescence dont le principe repose sur un partitionnement récursif des données. Elle est composée d'une racine (nœud initial) qui est le point de départ

de l'arbre de décision, des nœuds internes représentant des questions sur les attributs, un lien (arc) qui correspond à une valeur d'attribut ou bien la réponse à la question, le résultat indique le chemin à suivre dans l'arbre. En bref, pour déterminer à quelle classe appartient un nouvel individu, l'arbre de décision test ses attributs (d'où l'appellation d'arbre de décision)[13].

La figure suivante illustre l'affectation par l'approche de l'arbre de décision.

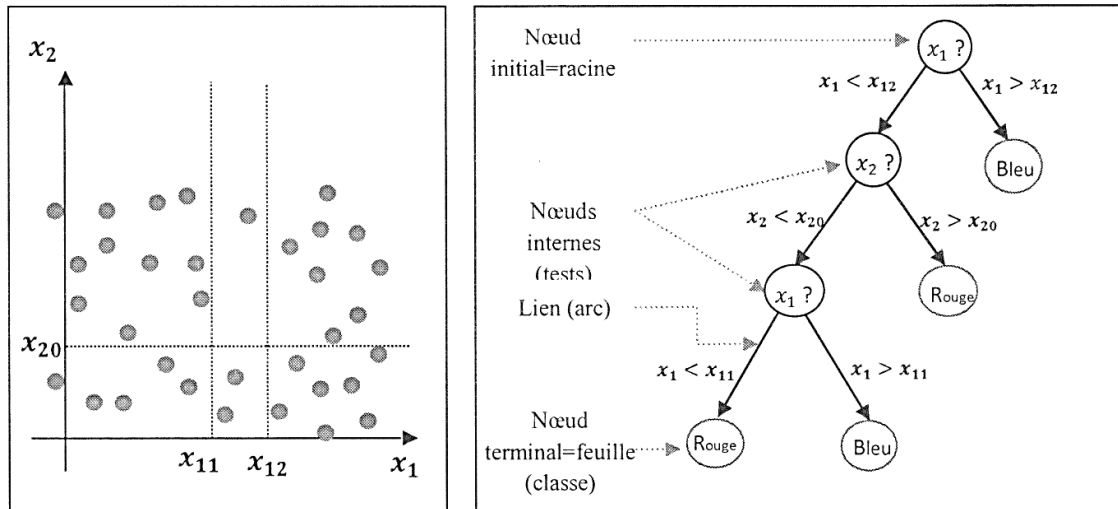


Fig.2.Exemple affectation par l'approche de l'arbre de décision.

Les avantages de cette méthode sont la facilité et la rapidité de construction et d'utilisation, ainsi l'interprétation directe de l'arbre vers une base de règles. La clarté des règles de décision rend possible le dialogue homme machine. En outre, elles ne font aucune hypothèse sur les données (méthodes non paramétriques). Les inconvénients de cette méthode sont la sensibilité aux bruits et la difficulté d'optimisation (le choix de meilleur arbre) et les arbres de décision sont extrêmement complexes si le nombre d'attributs et de classes augmentent.

II.2.1.5. Algorithme des machines à vecteurs support :

Les méthodes des Machines à support de vecteurs (Support Vector Machines : SVM) ont reçu une attention considérable ces dernières années et un grand nombre d'applications sont décrites dans la littérature. L'objectif des SVM est de maximiser la marge de séparation entre les classes. Les SVM sont des algorithmes d'apprentissage faisant l'hypothèse que les éléments sont linéairement séparables s'ils sont préalablement projetés dans un espace à grande dimension adéquat. La projection est réalisée par l'utilisation de la fonction Kernel

$k(x_i, x_j)$. La fonction la plus utilisée est le noyau gaussien qui réalise cette projection de façon implicite [14].

$$k(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|}{2\sigma^2}\right) \quad \text{II. 04}$$

La recherche du meilleur hyperplan de séparation peut alors se formaliser comme un problème de programmation quadratique sous contrainte linéaire. L'hyperplan fournit par les SVM est ainsi représenté comme une combinaison linéaire de certains exemples de la base d'apprentissage appelés vecteurs de support. Il existe plusieurs algorithmes pouvant résoudre le problème cité ci-dessus. Le plus compétitif étant actuellement l'algorithme Séquentiel Minimal Optimisation (Optimisation Minimum Séquentiel). Il reste deux hyper-paramètres (C et σ) qui doivent être choisis afin de minimiser le risque d'erreur déterminé à partir d'une base de tests. L'algorithme produit alors une fonction de décision pouvant être formalisée comme ci-dessous.

$$\varphi(x) = \sum_{i \in SV} \alpha_i y_i k(x_i, x) + b \quad \text{II. 05}$$

Où : $0 \leq \alpha_i \leq C$ et (x_i, y_i) est un exemple de la base de tests.

Les avantages de cette méthode sont l'absence d'optimum local et le contrôle explicite du compromis entre la complexité du classifieur et l'erreur ainsi cette méthode traite des données à grandes dimensions. Les inconvénients de cette méthode sont le besoin d'une bonne fonction Kernel et les problèmes de stabilité des calculs dans la résolution de certains programme quadratique a contraintes.

II.2.2. Méthodes non supervisée :

Les méthodes de classification automatique (aussi appelées méthodes de clustering) méthodes basées sur la notion d'apprentissage non supervisé, elles consistent à regrouper des objets appartenant à un ensemble C en classes restreintes de telle sorte que les objets d'une même classe soient le moins dispersés possible. Contrairement aux méthodes supervisées, on ne possède pas des connaissances a priori sur les classes prédéfinies des éléments. L'objectif des méthodes du Clustering est de grouper des éléments proches dans un même groupe de manière à ce que deux données d'un même groupe soient le plus similaires

possible et que deux éléments de deux groupes différents soient le plus dissemblables possible.

II.2.2.1. Méthode de K-means :

L'algorithme des k-means mis au point par McQueen en 1967 est l'un des algorithmes de clustering les plus connus [15]. Il est basé sur la méthode des centroïdes (ou centres de gravité), dans ce type de méthodes, la classe est représentée par son centre de gravité.

On se donne pour commencer, k centres arbitraires L_1, L_2, \dots, L_k où chaque L_i représente le centre d'une classe C_i . Chaque classe C_i est représentée par un ensemble d'individus plus proches de L_i que de tout autre centre. Après cette initialisation, on effectue une deuxième partition en regroupant les objets autour des M_j qui prennent alors la place des L_i (M_j est le centre de gravité de la classe C_j , calculé en utilisant les nouvelles classes obtenues). Le processus est ainsi réitéré jusqu'à atteindre un état de stabilité où aucune amélioration n'est possible.

Cette méthode est convergente et surtout avantageuse du point de vue calcul mais elle dépend essentiellement de la partition initiale. Il existe donc un risque d'obtenir une partition qui ne soit pas optimale mais seulement meilleure que la partition initiale. De plus, la définition de la classe se fait à partir de son centre, qui pourrait ne pas être un individu de l'ensemble à classer, d'où le risque d'obtenir des classes vides.

II.2.2.2. Méthode fuzzy C-means :

Fuzzy C-means (FCM) est une méthode de clustering qui permet à un objet de données d'appartenir à deux ou plusieurs clusters. Cette méthode a été développée par Dunn en 1973 et améliorée par Bezdek en 1981, repose sur la recherche des valeurs prises par les centroïdes des classes et par les degrés d'appartenance des objets à ces classes [16].

Si μ_{ij} est le degré d'appartenance de X_j à la classe i , la matrice $U_{C \times N} = [\mu_{ij}]$ est appelée matrice de C-partition floue si et seulement si elle satisfait :

$$(\forall i \in \{1 \dots C\})(\forall j \in \{1 \dots N\})\mu_{ij} \in [0,1]$$

$$0 < \sum_{j=1}^N \mu_{ij} < N$$

II. 06

$$(\forall j \in \{1 \dots N\}) \sum_{i=1}^C \mu_{ij} = 1 \quad \text{II. 07}$$

Bezdek a montré que le problème de partition de X en C classes floues pouvait être formulé comme la minimisation d'une fonctionnelle $J(B, U, X)$ définie par :

$$J(B, U, X) = \sum_{i=1}^C \sum_{j=1}^N \mu_{ij}^m d^2(X_j, b_i) \quad \text{II. 08}$$

Tel que :

$$b_i = \frac{\sum_{k=1}^N \mu_{ik}^m X_k}{\sum_{k=1}^N \mu_{ik}^m} \quad \text{II. 09}$$

$$\mu_{ij} = \left\{ \sum_{k=1}^C \left(\frac{d^2(X_j, b_i)}{d^2(X_j, b_k)} \right)^{2/(m-1)} \right\}^{-1} \quad \text{II. 10}$$

L'algorithme des FCM consiste alors à l'application itérée de l'équation **II.07** et **II.08** jusqu'à stabilité des solutions. Le critère d'arrêt des itérations, définissant cette stabilité, peut par exemple consister à l'étude de la norme de la matrice U ou à la stabilité des centres de classe sur deux itérations successives.

Avec :

- C : le nombre de classes, connu a priori.
- N : la taille du vecteur de données (nombre de pixels à classer).
- U_{ij} : le degré d'appartenance du pixel X_k à la classe i connue par son centroïde b_i .
- d : le degré de similitude, il peut être la distance Euclidienne.
- m : un réel >1 appelé degré de flou. Plus m tend vers 1, plus la classification devient dure et U_{ij} se rapproche de 0 ou de 1.

Les paramètres de la méthode sont choisis d'une façon arbitraire, ces paramètres ont une grande influence sur le résultat attendu. Ce qu'il nous oblige de faire une étude approprié sur les données en entrée et le regroupement que l'on souhaite obtenir.

II.2.2.3. Méthodes hiérarchiques :

La classification hiérarchique consiste à effectuer une suite de regroupements en classes de moins en moins fines en agrégeant à chaque étape les objets ou les groupes d'objets les plus proches. Elle fournit ainsi un ensemble de partitions de l'ensemble d'objets [17]. Cette

approche utilise la notion de distance, qui permet de refléter l'homogénéité ou l'hétérogénéité des classes. Ainsi, on considère qu'un élément appartient à une classe s'il est plus proche de cette classe que de toutes les autres.

Son principe est simple, initialement chaque objet forme une classe, soit n classes, donc on cherche à réduire ce nombre de classe $T < n$ itérativement de sorte que dans chaque étape on fusionne deux classes ensemble (Les deux classes choisies pour être fusionnées sont celles qui sont les plus "proches" en fonction de leur dissimilarité) ou ajouter un nouveau élément à une classe (un élément appartient à une classe s'il est plus proche de cette classe que de toutes les autres). La valeur de dissimilarité est appelée indice d'agrégation. Qui commence dans la première itération faible, et croîtra d'itération en itération. Parmi les algorithmes plus connus de ce type :

- ✓ *La classification ascendante hiérarchique (CHA) :* Où le mot ascendante est utilisé pour désigner qu'elle part d'une situation dont tous les objets représentent des clusters à part entière, puis on cherche les rassembler en classes de plus en plus grandes.
- ✓ *La classification descendante hiérarchique :* Elle consiste à fractionner l'ensemble des données en sous-ensemble et chaque sous-ensemble est à son tour fractionnée en d'autres classes jusqu'à avoir de petites classes. Donc, on démarre d'une classe mère et on aura comme résultat un ensemble de sous-ensembles.

L'avantage principal de cette méthode est la facilité d'utilisation. La principale difficulté présentée par cette méthode est la définition du critère de regroupement de deux classes, c'est-à-dire la détermination d'une distance entre les classes. Il en existe un bon nombre d'algorithmes pour les méthodes hiérarchiques ainsi que cette méthode est coûteuse en temps de calcul.

II.3. Classification d'images médicales :

La classification d'images est devenue très importante surtout dans le domaine médical, elle permet d'aider les médecins à établir un diagnostic médical ou encore contrôler une action thérapeutique en lui offrant de nouveaux outils avec plus de précision et de performance. Le but d'une classification médicale de pathologies est de rassembler les cas qui

ont des similitudes biologiques et qui sont susceptibles de partager certains facteurs étiopathologiques. L'identification des classes est importante car elle permet, d'une part de comprendre le processus de la maladie et d'autre part d'instaurer l'approche thérapeutique adéquate. En outre, elle permet de dégager le pronostic global de la maladie.

Plusieurs méthodes de classification comprenant les statistiques, la reconnaissance des formes, l'intelligence artificielle et les réseaux de neurones ont été utilisées pour l'aide au diagnostic médical.

De nombreux articles de revue résument l'utilisation des différents classifieurs dans le domaine de la mammographie et de l'imagerie TDM. Le choix du classifieur et de ces caractéristiques associées varie d'un article à l'autre et les performances de chaque classifieur ne sont pas toujours comparables. Il est donc quasiment impossible d'ordonner les classifieurs selon leur capacité à être ou pas les plus adaptés aux problématiques ciblées. Cheng et al [18] décrit un ensemble de méthodes de classification utilisées pour le développement de systèmes d'aide à la détection en mammographie, tandis que Li [19] fait de même pour l'imagerie TDM.

Campadelli et al [20] analyse la robustesse des SVM à un apprentissage à partir de classes de taille très inégale pour la détection de tumeurs pulmonaires en imagerie TDM. Des coefficients de régularisation sont ainsi ajoutés de manière à obtenir des performances sensiblement identiques à des cas de classes correctement réparties. Boroczky et al [21], Lu et al [22] utilisent aussi les SVM sur des images TDM pulmonaires.

En ce qui concerne les réseaux de neurones, Gletsos et al [23] ont proposé un système d'aide au diagnostic dédié aux tumeurs du foie en TDM sur une combinaison séquentielle de trois réseaux de neurones permettant de classifier les nodules pré segmentés manuellement successivement selon leur malignité et le type de maladies qu'ils représentent. Cheng and Cui[24] adapte un réseau de neurones à quatre couches à l'aide de règles de logique floue permettant de détecter toutes les micro-calcifications présentes dans 100 mammographies au prix d'un faible nombre de faux positifs par image. En mammographie, Sajda et al [25] génère un réseau de neurones hiérarchique pyramidal où chaque entrée reçoit des informations qui lui sont propres ou qui proviennent de la sortie d'un réseau de neurones de hiérarchie inférieure. De nombreuses études similaires portent sur l'adaptation des classifieurs de base vers des algorithmes plus rapides et efficaces : Wu et al [26], Zhang and Jiang [27], Gil-Pita and Yao [28], Leung et al [29].

II.4. Conclusion :

Nous avons évoqué dans ce chapitre les différentes méthodes de classification supervisées et non supervisées en soulignant quelques avantages et inconvénients de chaque méthode.

Plusieurs méthodes sont proposées pour le problème général de classification. Ils diffèrent par les mesures de proximités qu'ils utilisent, la nature des données qu'ils traitent et l'objectif final de la classification. C'est donc, le choix d'une méthode approprié dépend fortement de l'application, la nature des données et les ressources disponibles. Une analyse attentive des données aide à bien choisir le meilleur algorithme. Il n'existe pas un algorithme qui peut répondre à toutes les demandes.

Malgré le nombre important de méthodes existants, plusieurs problématiques restent encore ouvertes dans le cadre de la classification. Un problème très souvent rencontré concerne la difficulté de fixer les paramètres en entrées des méthodes par l'utilisateur. De même, la présentation des résultats sous une forme facilement interprétable par l'utilisateur est rarement abordée malgré son intérêt. La complexité des méthodes mises en œuvre est aussi problématique dans certains cas. La prise en compte du contexte dans lequel les classes sont créées constitue un sujet d'étude récent. Dans la suite de cette thèse, nous présentons les réseaux de neurones.

Chapitre III :
Introduction aux réseaux de neurones

II.1.Introduction :

La classification est un problème sur lequel les réseaux de neurones donnent de bons résultats et s'appliquent avec succès. Les réseaux de neurones, fabriqués de structures cellulaires artificielles, constituent une approche permettant d'aborder sous des angles nouveaux les problèmes de perception, de mémoire, d'apprentissage et de raisonnement. Ils s'avèrent aussi des alternatives très prometteuses pour contourner certaines des limitations des ordinateurs classiques. Grâce à leur traitement parallèle de l'information et à leurs mécanismes inspirés des cellules nerveuses (neurones), ils infèrent des propriétés émergentes permettant de solutionner des problèmes jadis qualifiés de complexes.

Ce chapitre présente une brève généralité sur les réseaux de neurones ainsi que l'architecture des réseaux de neurones à apprentissage supervisé utilisé dans notre travail pour la classification.

II.2.Historique sur réseaux de neurones :

Pour faire un bref historique, les réseaux de neurone ont connu leurs débuts dans les années 1943 avec les travaux de Warren Mc Culloch & Walter Pitt sur le « neurone formel ». En 1949, D. Hebb présente dans son ouvrage « The Organization of Behavior » une règle d'apprentissage. De nombreux modèles de réseaux aujourd'hui s'inspirent encore de la règle de Hebb. En 1958 les travaux de Franck Rosenblatt sur « le perceptron » proposent le premier algorithme d'apprentissage permettant d'ajuster les paramètres d'un neurone. Il est à présent communément admis que le perceptron, comme classifieur linéaire, est le réseau de neurones le plus simple. En 1969, Minsky et Papert publient le livre Perceptrons dans lequel ils utilisent une solide argumentation mathématique pour démontrer les limitations des réseaux de neurones à une seule couche. Ce livre aura une influence telle que la plupart des chercheurs quitteront le champ de recherche sur les réseaux de neurones. En 1982, Hopfield propose des réseaux de neurones associatifs et l'intérêt pour les réseaux de neurones renaît chez les scientifiques. En 1986, Rumelhart, Hinton et Williams publient l'algorithme de la rétropropagation de l'erreur, qui permet d'optimiser les paramètres d'un réseau de neurones à plusieurs couches. À partir de ce moment, la recherche sur les réseaux de neurones connaît un essor fulgurant et les applications commerciales de ce succès académique suivent au cours des années 90.

II.3. Le neurone biologique :

Le neurone biologique est une cellule vivante spécialisée dans le traitement des signaux électriques. Les neurones sont reliés entre eux par des liaisons appelées axones. Ces axones vont eux-mêmes jouer un rôle important dans le comportement logique de l'ensemble. Ces axones conduisent les signaux électriques de la sortie d'un neurone vers l'entrée (synapse) d'un autre neurone. Les neurones font une sommation des signaux reçus en entrée et en fonction du résultat obtenu vont fournir un courant en sortie [30].

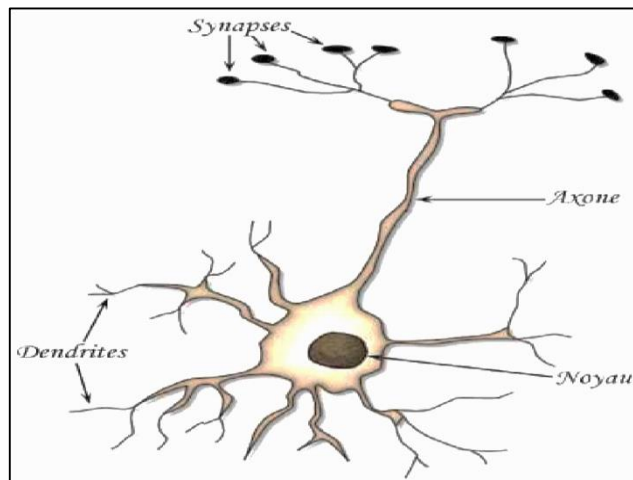


Fig.1. Le neurone biologique

Un neurone biologique est composé de trois parties essentielles :

- ✓ Le corps cellulaire effectue les transformations biochimiques nécessaires à la vie du neurone.
- ✓ Les dendrites forment une arborescence autour du corps cellulaire et permettent au neurone de capter les signaux qui parviennent de l'extérieur.
- ✓ L'axone est une fibre nerveuse qui transporte les signaux émis par le neurone, il se ramifie en extrémité là où il communique avec les autres neurones à travers des synapses.

L'influx nerveux se propage toujours de la dendrite vers le corps cellulaire et de celui-ci vers l'axone. Chaque neurone reçoit des signaux excitateurs ou inhibiteurs par ses dendrites. Ces signaux pondérés sont combinés dans le corps cellulaire. Le potentiel résultant est

comparé au seuil interne, s'il est supérieur à celui-ci, le neurone provoque l'émission dans l'axone d'un train d'impulsions vers les synapses, dans le cas contraire il reste inactif.

II.4. Modélisation d'un neurone :

II.4.1. Neurone formel :

Le neurone est un modèle mathématique très simple, dérivé d'une analyse de la réalité biologique, ce modèle imite quelques fonctions comme la mémorisation et l'apprentissage, reste de même que le neurone artificiel est lion de posséder toutes les capacités du neurone biologique. Le fonctionnement de ce neurone formel est basé sur une règle de calcul assez simple : on cherche à évaluer la valeur d'une sortie O_j , à partir de plusieurs entrées X_i , qui elles-mêmes sont pondérées par des coefficients appelés synapses ou poids synaptiques W_{ij} . Selon cette description, chaque neurone est relié à d'autres par des connexions. L'activation du neurone se produit lorsque celui-ci atteint un certain seuil d'activation. Cette activation est générée par les connexions qui ont pour propriété d'être excitatrices ou inhibitrices. La sommation des données reçues à l'entrée du neurone est transformée par une fonction d'activation ou fonction de sortie Φ non linéaire. La figure suivante montre le modèle mathématique d'un neurone artificiel [31].

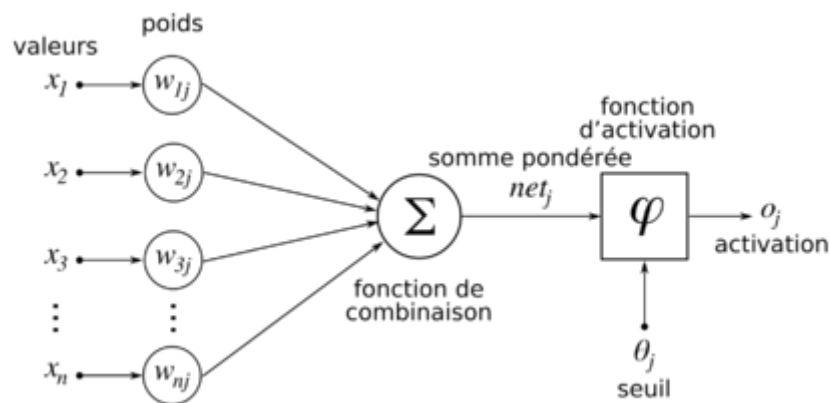


Fig.2. Représentation d'un neurone artificiel

Tel que :

X_1, X_2, \dots, X_n : représentent les entrées du neurone.

$W_{1j}, W_{2j}, \dots, W_{nj}$: représentent les poids synaptiques.

Σ : la sommation au biais θ_j tel que : $\Sigma = W_{1j} * X_1 + W_{2j} * X_2 + \dots + W_{nj} * X_n + \theta_j$

φ : la fonction d'activation

O_j : la sortie de neurone tel que : $O_j = \varphi(\Sigma)$

II.4.2. Fonctions d'activations :

La fonction d'activation (ou fonction de seuillage, ou encore fonction de transfert) joue un rôle très important dans le comportement du neurone. Elle retourne une valeur représentative de l'activation du neurone, cette fonction a comme paramètre la somme pondérée des entrées ainsi que le seuil d'activation. La nature de cette fonction diffère selon le réseau. On peut citer quelques-unes :

A-Fonction binaire a seuil :

Fonction Heaviside définie par : $h(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$ III. 01

Fonction Signe définie par : $h(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ -1 & \text{sinon} \end{cases}$ III. 02

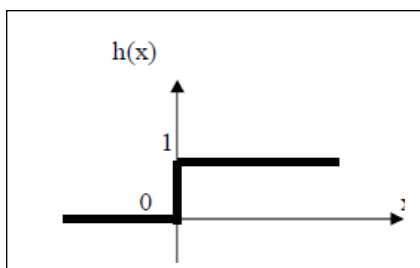


Fig.3.Fonction Heaviside

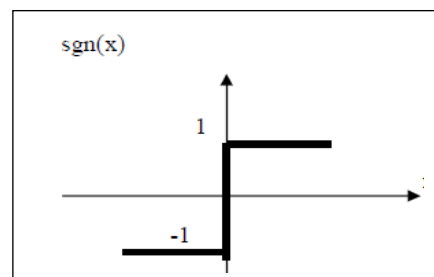


Fig.4.Fonction signe

Le seuil introduit une non-linéarité dans le comportement du neurone, cependant il limite la gamme des réponses possibles à deux valeurs.

B-Fonction linéaire :

C'est l'une des fonctions d'activations les plus simples, sa fonction est définie par : $F(x)=x$

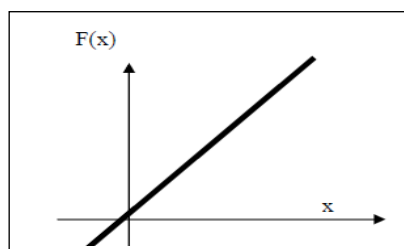


Fig.5 Fonction linéaire

C-Fonction linéaire à seuil ou multi-seuils :

On peut la définir comme suit :

$$F(x) = \begin{cases} X & \text{si } x \in [V, U] \\ V & \text{si } x \geq V \\ U & \text{si } x \leq U \end{cases} \quad \text{III. 03}$$

Cette fonction représente un compromis entre la fonction linéaire et la fonction seuil : entre ses deux barres de saturation, elle confère au neurone une gamme de réponses possibles. En modulant la pente de la linéarité, on affecte la plage de réponse du neurone.

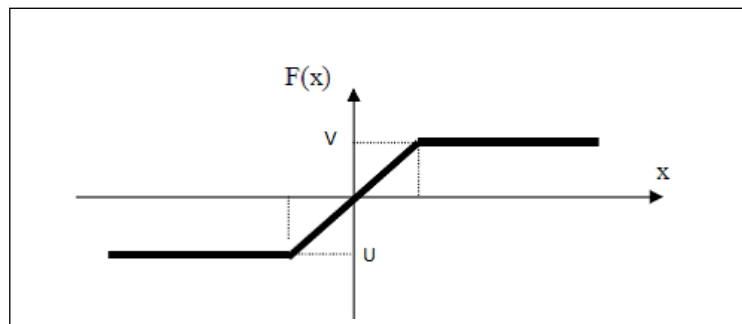


Fig.6. Fonction linéaire à seuil

D-Fonction sigmoïde :

La fonction sigmoïde (dite aussi courbe en **S**) est définie par :

$$f(x) = \frac{1}{1 + e^{-bx}} \quad \text{III. 04}$$

C'est la plus utilisée car elle introduit de la non-linéarité, mais c'est aussi une fonction continue, différentiable. En outre possède une propriété simple permet d'accélérer le calcul de sa dérivée, ce qui réduit le temps de calcul nécessaire à l'apprentissage d'un réseau de neurones.

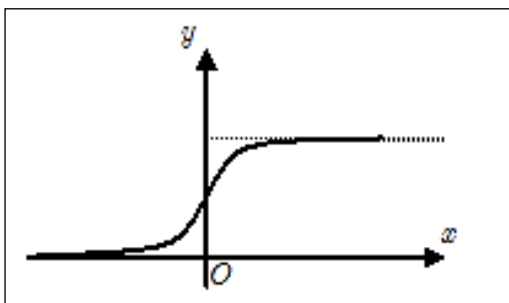


Fig.7. Fonction logistique

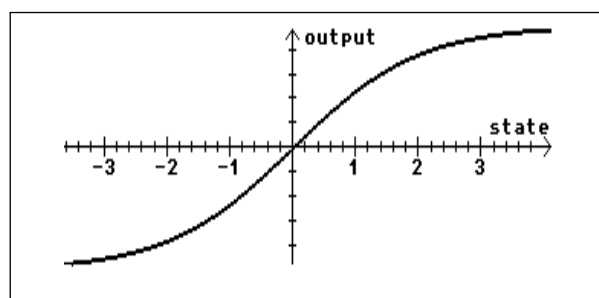


Fig.8. Fonction symétrique

E-La fonction radiale (gaussienne) :

La fonction gaussienne est définie comme suit :

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-u)^2}{2\sigma^2}} \quad \text{III. 05}$$

C'est une fonction locale qui fonctionne dans des intervalles restreints, elle est continue et bornée sur l'intervalle de fonctionnement.

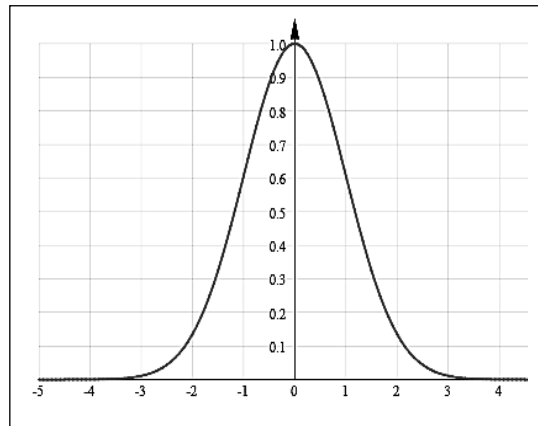


Fig.9.La fonction gaussienne

II.4.3. Apprentissage des réseaux de neurones :

L'apprentissage est la propriété la plus intéressante des réseaux neuronaux. Durant cette phase, le comportement du réseau est modifié à travers les poids en fonction d'un jeu de données présentées en entrée, jusqu'à l'obtention du comportement désiré.

L'apprentissage peut être supervisé ou non supervisé selon la présence ou l'absence de la réponse désirée $d(n)$ [32].

- *Apprentissage supervisé :*

L'apprentissage est dit supervisé lorsque le réseau est forcé à converger vers un état final précis, ce qui nécessite la connaissance a priori de la réponse désirée $d(n)$. La méthode la plus utilisée est la rétro propagation du gradient. Elle consiste à présenter des exemples au réseau, calculer sa sortie, ajuster les poids de façon à réduire l'écart entre cette sortie et la réponse désirée pour satisfaire un certain critère de performance.

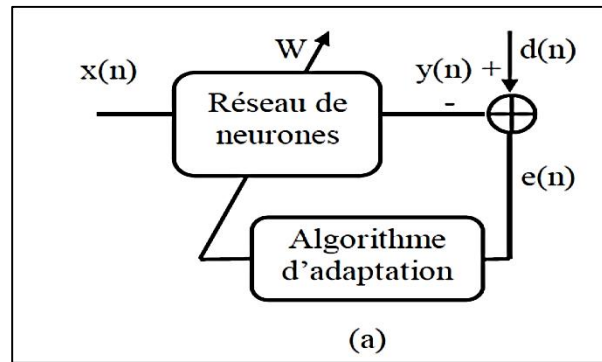


Fig.10. Apprentissage supervisé

- *Apprentissage non supervisé :*

Dans l'apprentissage non supervisé, seules les valeurs d'entrée sont disponibles et le réseau est laissé libre de converger vers n'importe quel état final. La connaissance à priori de la sortie désirée n'est pas nécessaire, la procédure d'apprentissage est basée uniquement sur les valeurs d'entrées. Le réseau s'auto-organise de façon à optimiser une certaine fonction de coût. Le modèle de *Kohonen* est parmi les réseaux qui s'adaptent à cet apprentissage.

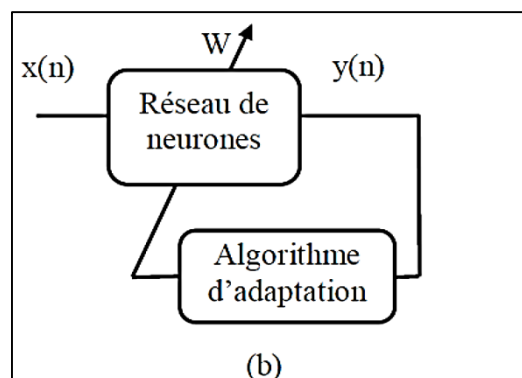


Fig.11. L'apprentissage non supervisé

II.4.4. Topologie des réseaux de neurones :

La topologie d'un réseau de neurone est la façon dont les neurones sont reliés (d'une manière générale l'architecture locale du réseau). Il n'existe pas une méthode formelle pour déterminer le nombre de neurones ou de couches ni les connexions nécessaires pour résoudre un problème donné (sauf le nombre de neurones de la couche d'entrée et de sortie) [33]. Nous présentons les différentes topologies [34] :

Réseau multicouche : Les neurones sont arrangés par couche. Il n'y a pas de connexion entre neurones d'une même couche et les connexions ne se font qu'avec les neurones des couches avales.

Réseau à connexions locales : Il s'agit d'une structure multicouche, mais qui à l'image de la rétine, conserve une certaine topologie. Chaque neurone entretient des relations avec un nombre réduit et localisé de neurones de la couche avale.

Réseau à connexions récurrentes : Les connexions récurrentes ramènent l'information en arrière par rapport au sens de propagation défini dans un réseau multicouche. Ces connexions sont le plus souvent locales.

Réseau à connexion complète : C'est la structure d'interconnexion la plus générale. Chaque neurone est connecté à tous les neurones du réseau (et à lui-même).

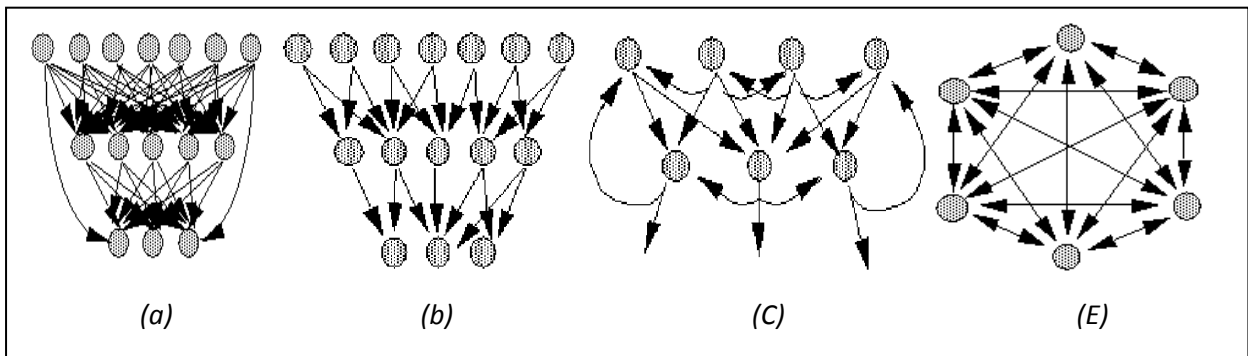


Fig.12. Définition des couches d'un réseau multicouche.

(a) Réseaux multicouches, (b) Réseaux à connexions local, (c) Réseaux à connexions récurrentes,,
(E) Réseaux à connexion complète

Nous nous intéressons aux réseaux multicouches qui sont les plus utilisés dans les problèmes de classification supervisé.

II.5. Perceptron multicouches (multi layer perceptron) :

II.5.1. Architecture

Le perceptron multicouche est un réseau orienté de neurones artificiels organisé en couches ou l'information voyage dans un seul sens, de la couche d'entrée vers la couche de sortie. La couche d'entrée représente toujours une couche virtuelle associée aux entrées du système. Elle ne contient aucun neurone. Les couches suivantes sont des couches de neurones et les sorties des neurones de la dernière couche correspondent toujours aux sorties du système.

Les Perceptrons Multicouches (**PMC**) sont des réseaux de neurones pour lesquels les neurones sont organisés en couches successives, les connexions sont toujours dirigées des couches inférieures vers les couches supérieures et les neurones d'une même couche ne sont

pas interconnectées. Un neurone ne peut donc transmettre son état qu'à un neurone situé dans une couche postérieure à la sienne. Les neurones sont reliés entre eux par des connexions pondérées. Ce sont les poids de ces connexions qui gouvernent le fonctionnement du réseau et programment une application de l'espace des entrées vers l'espace des sorties à l'aide d'une transformation non linéaire. La création d'un perceptron multicouche pour résoudre un problème donné passe donc par l'inférence de la meilleure application possible telle que définie par un ensemble de données d'apprentissage constituées de paires de vecteurs d'entrées et de sorties désirées. Cette inférence peut se faire, entre autre, par l'algorithme dit de rétropropagation [35].

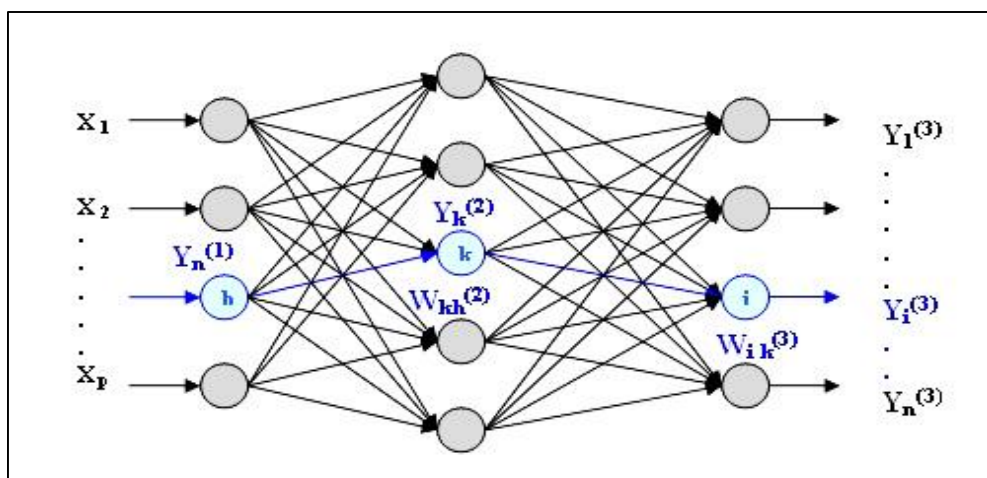


Fig.13. Perceptron multicouche.

II.5.2. Apprentissage des réseaux multicouches :

L'apprentissage d'un perceptron multicouches se fait par plusieurs algorithmes citons par exemples la méthode de Gauss-Newton, méthode des moindres carrés non-linéaire, méthode du gradient conjugué et la méthode Quasi-Newton. Dans notre travail, on s'intéresse à deux algorithmes de rétropropagation qui sont l'algorithme de descente de gradient et l'algorithme de Levenberg-Marquardt.

II.5.3. Algorithme de rétropropagation du gradient :

L'algorithme d'apprentissage par rétropropagation du gradient de l'erreur est un algorithme itératif qui a pour objectif de trouver les poids des connexions minimisant l'erreur quadratique moyenne commise par le réseau sur l'ensemble d'apprentissage. Cette minimisation par une méthode du gradient conduit à l'algorithme d'apprentissage de rétropropagation.

Cet algorithme, qui présente l'avantage d'exister, reste discutable dans la mesure où sa convergence n'est pas prouvée. Son utilisation peut conduire à des blocages dans un minimum local de la surface d'erreur. Son efficacité dépend, en effet, d'un grand nombre de paramètres que doit fixer l'utilisateur : le pas du gradient, les paramètres des fonctions d'activation, l'architecture du réseau, le nombre de couches, le nombre de neurones par couche, l'initialisation des poids...etc.

• **Etapes de l'algorithme de rétropropagation :** (voir la figure précédente)

W_{ik} : La connexion entre le neurone i et le neurone k.

W_{kh} : La connexion entre le neurone k et le neurone h

net_i : L'entrée totale du neurone i de la couche (3).

net_k : L'entrée totale du neurone k de la couche (2).

δ_i : L'erreur quadratique

α : Pas d'apprentissage

Y : la sortie

θ : le seuil (le biais)

Dans ce qui suit nous allons résumer les différentes étapes de la règle de la rétropropagation :

- 1) Appliquer un vecteur d'entrée, aux unités d'entrée :

$$X = (X_1, X_2, X_3, \dots, X_N)^P \quad \text{III. 06}$$

- 2) Calculer les entrées totales des neurones cachés :

$$net_k^{(2)} = \sum_{j=1}^P W_{kj} X_j + \theta_k^{(2)} \quad \text{III. 07}$$

- 3) Calculer les sorties des neurones cachés :

$$Y_k^{(2)} = F(net_k^{(2)}) \quad \text{III. 08}$$

- 4) Calculer les entrées totales des neurones de la couche de sortie :

$$net_i^{(3)} = \sum_{k=1}^n W_{ik} Y_k^{(2)} + \theta_i^{(3)} \quad \text{III. 09}$$

- 5) Calculer les sorties réelles des unités de la 3ème couche (couche de sortie) :

$$Y_i^{(3)} = F(\text{net}_i^{(3)}) \quad \text{III. 10}$$

- 6) Calculer les termes d'erreur sur chaque unité de sortie :

$$\delta_i^{(3)} = Y_i^{(3)} - Yd_i^{(3)} \quad \text{III. 11}$$

- 7) Calculer les termes d'erreur sur chaque unité cachée :

$$\delta_k^{(2)} = F(\text{net}_k^{(2)}) \sum_{i=1}^n \delta_i^{(3)} W_{kh}^{(3)} \quad \text{III. 12}$$

Notons que les termes d'erreur pour les unités de la couche cachée sont calculés avant que les connexions des poids aux unités de la couche cachée n'aient été mises à jour. Ainsi, l'erreur est rétropropagée de la sortie vers l'entrée, d'où le nom rétropropagation.

- 8) Mettre à jour les poids de la couche de sortie :

$$W_{ik}^{(3)}(t+1) = W_{ik}^{(3)}(t) - \alpha \delta_i^{(3)} Y_k^{(2)} \quad \text{III. 13}$$

- 9) Mettre à jour les poids de la couche cachée :

$$W_{kh}^{(2)}(t+1) = W_{kh}^{(2)}(t) - \alpha \delta_k^{(2)} Y_h^{(1)} \quad \text{III. 14}$$

Répéter ce processus jusqu'à ce que l'erreur soit atteinte pour tous les exemples, c'est-à-dire lorsqu'elle devienne suffisamment petite pour chaque paire d'apprentissage.

II.5.4. Algorithme de Levenberg-Marquardt :

La méthode de LEVENBERG-MARQUARDT [37] est un standard pour l'optimisation de l'erreur quadratique due à ses propriétés de convergence rapide et de robustesse.

On définit :

J : la fonction coût de notre réseau de neurones.

W : désigne la matrice de l'ensemble des poids de réseau.

K : désigne la K -ème itération.

S_k : la variation $W(t+1) - W(t)$

g : étant le gradient et G_K le Hessien

H_k : Le Hessien modifié par la variable μ_k (μ_k : pas d'apprentissage)

$$H_k = G_k + \mu_k I_n \quad \text{III. 15}$$

$$R_k = \frac{\Delta J_k}{\Delta Q_k} \quad \text{avec : } \begin{cases} \Delta J_k = J(W_k) - J(W_k + S_k) \\ Q_k = Q(0) - Q(S_k) \end{cases} \quad \text{III. 16}$$

Avec I_n matrice identité de rang

$$Q(s) = J_k(W_k) + g(W_k)^T \cdot S + \frac{1}{2} S^T H_k(W_k) \cdot S \quad \text{III. 17}$$

$$g_k = J a_k^T r_k \quad \text{Avec : } J_{a_{i,j}} = \frac{\partial r_i}{\partial w_j} \quad \text{III. 18}$$

$$G_k \approx J a_k^T J a_k \quad \text{III. 19}$$

- **Les étapes de l'algorithme de Levenberg-Marquardt :**

Initialisation des poids W_0 par une distribution uniforme et $\mu_0 > 1$

La solution est donnée par l'algorithme suivant la K -ème itération :

- 1) Dédurre de W_k et μ_k , calculer g_k et H_k suivant les relations de H_k et g_k
- 2) Résoudre - $g_k = H_k S_k$
- 3) Évaluer de $J(W_k + S_k)$ et R_k

- 4) **Calcul de paramètre μ_k :**

Si $R_k \leq 0,25$ alors $\mu_{k+1} = 4\mu_k$

Si $R_k \geq 0,75$ alors $\mu_{k+1} = \frac{\mu_k}{2}$ sinon $\mu_{k+1} = \mu_k$

- 5) **Calcul de la correction des poids W_k :**

Si $R_k \leq 0$ alors $W_{k+1} = W_k$

Sinon $W_{k+1} = W_k + S_k$

II.6. Classification des images médicales par les réseaux de neurones :

La classification par réseau de neurone passe par plusieurs étapes comme le montre la figure 14 .Le vecteur d'entrée de perceptron multicouche correspond aux descripteurs des images IRM cérébrales et sa sortie correspond aux classes fixées au préalable. Le réseau cherche à atteindre une sortie désirée suivant plusieurs itérations de l'algorithme de rétropropagation. On dit que l'algorithme est convergé s'il atteint une erreur quadratique désirée.

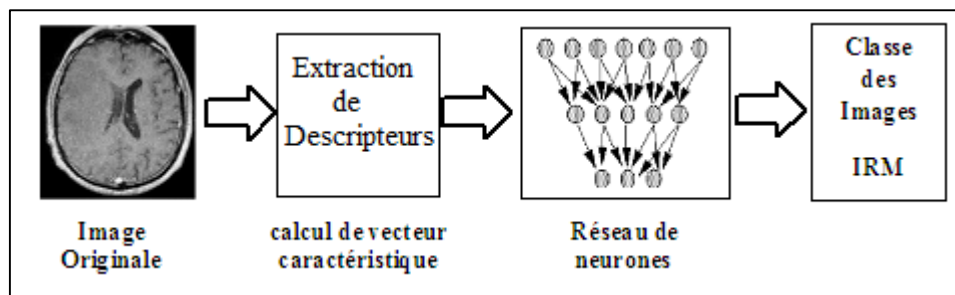


Fig.14. Etapes de classification par réseau de neurones

Dans ce qui suit quelques travaux de classification d'images médicales effectués par réseaux de neurones.

Lalit P. Bhaiya , Virendra Kumar Verma [38] Ont utilisé un perceptron multicouche pour développer un système d'aide à la décision médicale avec 02 classes, une classe saine et une autre pathologique. Le système de prise de décision médicale a été conçu par la transformation en ondelettes. Ce classifieur neuronal a donné des résultats très prometteurs en classifiant les images cérébrales en image saine et pathologique. Le taux de classification était de 95.7 %, tandis que la sensibilité et la spécificité sont 92.5% et 90.4 respectivement.

K.V.Ramana, Raghu.B.Korrapati [39] Dans leur étude ont considéré un perceptron multicouches avec l'algorithme de rétropropagation pour la classification des hémorragies de cerveau. La classification et le diagnostic des hémorragies de cerveau ont une grande importance dans le dépistage précoce des hémorragies qui réduisent les taux de mortalité. Le but de cette recherche était de détecter des hémorragies de cerveau et de les classifier et de fournir au patient le diagnostic le plus correct. Le taux de classification obtenu était de 90%.

Amir Ehsan Lashkari [40] a proposé une méthode de classification automatique des images cérébrale en pondération T1 et T2 pour détecter les anomalies dans les tissus de cerveau. Les images IRM cérébral sont classer en deux classes, tissus normales (images saines) et tissus anormale (images pathologique) en utilisant les descripteurs de texture : énergie, entropie, contraste, moyenne...et les ondelettes de Gabor. Ces descripteurs sont par la suite introduits dans un réseau de neurones multicouches en vue de classification. Les résultats de la classification sont encourageants tels que le taux de classification était de 98,28%, la sensibilité de réseau égale à 97,17% et spécificité était de 98%.

Anchana Khemphila , Veera Boonjing [41] ont utilisé un perceptron multicouche pour la classification des images IRM cérébrale dans l'optique de détecter les personnes suspects de la maladie de Parkinson. Le taux de classification était de 83,33%.

II.7. Avantages et inconvénients des réseaux de neurones :

II.7.1. Les avantages des réseaux de neurones :

- a. Souplesse : les réseaux de neurones sont capables de traiter une gamme très étendue de problèmes. Leur résultat peut être une régression, une classification ou encore une analyse de clusters (clustering neuronal).
- b. Bonne résolution : ils donnent de bons résultats même dans des domaines complexes car ils sont beaucoup plus puissants que les statistiques ou les arbres de décisions.
- c. Bonne adaptation : une fois que les données sont codées, ils traitent aussi bien des variables continues qu'énumératives.
- d. Parallélisme massif : les réseaux de neurones sont constitués d'unités de calcul qui peuvent opérer d'une manière parallèle. La plupart des implémentations des réseaux de neurones peuvent être facilement converties d'une version séquentielle à une version parallèle.
- e. Outils disponibles : il existe de nombreux produits sur le marché intégrant la technique des réseaux de neurones.

II.7.2. Les points faibles des réseaux de neurones :

- a. Codage des entrées : toutes les entrées d'un réseau de neurone doivent se trouver dans un intervalle défini en général entre 0 et 1, ce qui entraîne des transformations qui impliquent des traitements supplémentaires, et risque de fausser les résultats.

- b. Lisibilité : les réseaux de neurones ne fournissent pas l'explication de leur résultat. Ce qui peut être gênant si on cherche à comprendre un phénomène.
- c. Détermination de la taille : afin que l'échantillon fournisse de bons résultats, sa taille doit être calculée en fonction du nombre d'entrée, du nombre des couches et du taux de connexion ce qui entraîne une augmentation du nombre d'exemples qui ne sont pas toujours disponibles.
- d. Performance : le nombre de calculs à effectuer pour définir un réseau optimal peut être très consommateur de puissance, ce qui peut donner de mauvaises performances à cette technique.

II.8.Conclusion :

Nous avons présenté dans ce chapitre les notions de base sur le neurone biologique et le neurone formel, ceci dans le but de bien connaître les caractéristiques d'un neurone qui nous intéresse, de même le principe de son fonctionnement. Par la suite nous avons exposé les différentes topologies et les différents modes d'apprentissage existants des réseaux de neurones. Nous avons vu que pour réaliser l'apprentissage d'un réseau de neurones, il existe plusieurs algorithmes de rétropropagation d'erreur quadratique dont on a détaillé l'algorithme de descente de gradient et la technique de Levenberg-Marquardt.

Nous avons montré que le perceptron multicouche est un outil puissant pour la classification des images médicales dans plusieurs travaux effectués. On s'intéresse dans le chapitre suivant à une classification des images IRM cérébrale en deux classes, une classe d'images saines et une classe d'images pathologiques.

Chapitre IV :

Tests et résultats

IV .1. Introduction :

Le diagnostic médical est un processus de classification. L'utilisation de techniques dite intelligentes pour la réalisation de cette classification devient de plus en plus importante. Même si la décision de l'expert est le facteur le plus important lors du diagnostic, les systèmes de classification fournissent une aide substantielle, car ils réduisent les erreurs dues à la fatigue et le temps nécessaire pour le diagnostic.

On s'intéresse dans cette section à l'étude d'un réseau de neurones multicouches conçu pour des problèmes de classification d'images IRM cérébrales. Les images utilisées sont des images saines (normale) et pathologiques (qui présentent des anomalies). Les images sont divisées en deux groupes, un groupe pour la phase d'apprentissage et les autres pour la phase de test où chaque image une fois présentée à notre système de classification elle est classée dans la classe normale ou la classe pathologique. Il est très important de noter que le choix des entrées du système et dans notre cas le choix des caractéristiques descriptives des images contribue dans la discrimination des anomalies par rapport au tissu normal. Pour cela, deux descripteurs sont utilisés, les descripteurs de texture des images et les moments de ZERNIKE.

IV .2. Les descripteurs utilisés :

Un système de classification ne traite que très rarement l'image brute telle quelle. Différentes caractéristiques descriptives de l'image ou du cas clinique (patient) sont généralement préférées. Celles-ci sont en effet porteuses d'un nombre plus important d'informations facilitant leur interprétation. Ces caractéristiques sont assimilées à ce que l'œil humain analyse avant de poser son diagnostic sur la présence ou pas de tumeurs. Elles sont l'interprétation de l'image d'un point de vue numérique et leur choix est primordial. Elles doivent en effet permettre au classifieur de discriminer les éléments pathologiques des éléments sains et être caractéristiques de chacune des classes.

IV .2.1. Descripteurs de texture

La texture est une des propriétés de base de l'image. Elle représente l'ordonnement visuel du changement des niveaux de gris dans une surface de l'espace ou une structure d'un objet. Haralick [41] a proposé un ensemble de 14 caractéristiques qui peuvent être utilisées afin de caractériser les textures. Ces mesures proviennent de la matrice de cooccurrence de l'image. Celle-ci représente l'estimation de la probabilité de transition d'un niveau de gris à un autre. En d'autres termes, elle détermine la fréquence d'apparition d'un motif formé par deux pixels de niveau de gris différents. Chaque élément $P(d, \theta)$ de la matrice de

cooccurrence représente le nombre de pixels de niveau de gris a ayant pour voisin un pixel de niveau de gris b situé à une distance d selon la direction ($\Theta=0, \Theta=45, \Theta=90, \Theta=135$) par rapport à l'horizontale.

Dans notre cas on s'intéresse aux 06 descripteurs de texture suivants :

- a. **Le contraste** : il mesure les variations locales des niveaux de gris, il permet aussi de caractériser la dispersion des valeurs de la matrice par rapport à sa diagonale principale :

$$CONT = \sum_i \sum_j (i - j)^2 \cdot P(i, j)$$

- b. **L'homogénéité** : On définit l'homogénéité par la loi suivante :

$$HOM = \sum_{i,j} \frac{1}{1 + (i - j)^2} P(i, j)$$

- c. **L'Energie** : ce paramètre mesure l'homogénéité des niveaux de gris de l'image, il a une valeur numérique faible quand les $P(i, j)$ de la matrice de cooccurrence ont des valeurs très proches et une valeur forte quand les niveaux de gris ne sont plus équiprobable.

$$E = \sum_{i,j} [P(i, j)]^2$$

- d. **La corrélation** : la corrélation à une forte valeur quand les valeurs sont uniformément distribuées dans la matrice de cooccurrence. Elle mesure la dépendance linéaire des niveaux de gris dans l'image.

$$COR = \sum_i \sum_j (i - U_x)(j - U_y)P(i, j)$$

Avec : U_x et U_y sont respectivement les moyennes des lignes et des colonnes

- e. **L'entropie** : elle caractérise le désordre dans l'image et rend compte de la complexité de la texture de celle-ci, il en résulte alors que l'entropie est maximum si la distribution spatiale des pixels est équiprobable et faible si les couples de pixels sont inégalement répartis dans l'image. De plus elle permet aussi de caractériser le degré de granulation de l'image : plus l'entropie est élevée et plus la granulation est grossière.

$$ENT = - \sum_i \sum_j P(i,j) \log[P(i,j)]$$

f. **La moyenne :** La moyenne dans une image est calculée comme suit :

$$MOY = \sum_i \sum_j P(i,j)$$

Le calcul des descripteurs est les quatre directions ($\Theta=0, \Theta=45, \Theta=90, \Theta=135$) ce qui donne un vecteur de descripteurs de taille 24 pour chaque image.

IV .2.2. Descripteurs de forme :

La forme est une caractéristique visuelle importante. Elle présente un des attributs de base pour décrire le contenu visuel d'une image. Pour extraire l'information pertinente de la forme, nous avons utilisé les moments. Ces moments ont été introduits par F. Zernike en 1934.

Dans le domaine du traitement de l'information, les moments de Zernike ont beaucoup été utilisés pour leur propriété d'orthogonalité qui permet la génération de descripteurs non redondants et leurs propriétés d'invariance en translation, en échelle et en rotation. La formulation des moments de ZERNIKE pour une fonction discrète $f(x,y)$ est donnée par (cas d'images par exemple) [42] :

$$Z_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x,y) \cdot [V_{nm}(x,y)]^*$$

tel que: $x^2 + y^2 = r \leq 1$

[...]* : est utilisé pour indiquer la valeur complexe conjuguée.

n : représente ici l'ordre de décomposition ($n=0, 1, 2, \dots, \infty$), dit aussi ordre radial.

m : le nombre de répétitions de la décomposition pour un ordre n donné. L'ordre et la répétition sont liés par les deux conditions suivantes : ($n - m$) toujours pair et ($m \leq n$).

$V_{nm}(x, y)$: représente les polynômes de Zernike constituant la base orthogonale de projection.

Ils s'écrivent en général en représentation polaire sous la forme suivante :

$$V_{nm}(r, \theta) = R_{nm}(r) \cdot e^{jm\theta}$$

Afin de conserver l'orthogonalité de la base au niveau des moments calculés, la fonction $f(x,y)$ doit être recalculée à l'intérieur du cercle unité en la transférant en représentation polaire (r,θ) tel que le centre de l'image soit le centre du cercle unité. Les relations de transformation s'écrivent alors :

$$x = r \cdot \cos \theta \quad \text{et} \quad y = r \cdot \sin \theta$$

Avec :

$$r = \sqrt{x^2 + y^2} \quad \text{et} \quad \theta = \tan^{-1} \left(\frac{y}{x} \right).$$

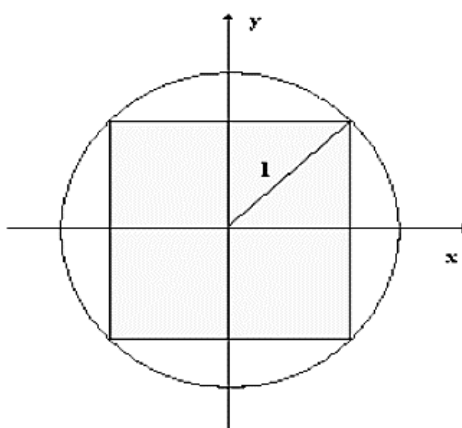


Fig.01. Façon de recalculer de la fonction $f(x,y)$ (rectangle gris) dans le cercle unité

Ainsi le vecteur caractéristique extrait d'une image pour un ordre donné n est de taille :

$$g = (n + 2)^2 / 2 \quad \text{pour } n \text{ pair}$$

$$g = [(n + 1) (n + 3)] / 2 \quad \text{pour } n \text{ impair.}$$

En considérant uniquement les moments de facteurs de répétition positif, le nombre de moments est réduit à :

$$g = (n + 2)^2 / 4 \quad \text{pour } n \text{ pair}$$

$$g = [(n + 1) \cdot (n + 3)] / 4 \quad \text{pour } n \text{ impair.}$$

Dans notre cas on fixe $n=12$ ce qui fait l'extraction de $g = (12 + 2)^2 / 4 = 49$ descripteurs pour chaque image sachant que nous avons considéré uniquement les nombre de répétition positifs. Cependant, ces nombres complexes sont généralement convertis en plusieurs nombres réels ce qui apporte une certaine robustesse aux variations de luminosités et aux instabilités de normalisation. Les moments de Zernike Z_{nm} peuvent être exprimés comme suit, suite à une conversion non linéaire [33] :

- 1) Un nombre complexe à une dimension est converti en nombre réel à deux dimensions.

- 2) L'amplitude qui est la valeur absolue du nombre complexe.
- 3) La phase ou l'argument du nombre complexe.
- 4) L'amplitude et l'argument du nombre complexe.

IV .3. Bases d'images :

Nous avons travaillé sur des images téléchargées de l'internet [43], la Base d'apprentissage contient 150 images IRM cérébrales en pondération T1 et de taille 256*256, divisées en images saines et images pathologiques. Cette base sert à l'apprentissage de notre réseau de neurones. La Base de test contient 30 images IRM cérébrales en pondération T1 et de taille 256*256, divisées en images saines et images pathologiques. Cette base sert à tester la capacité de notre réseau de neurones à classer les images IRM cérébrales.

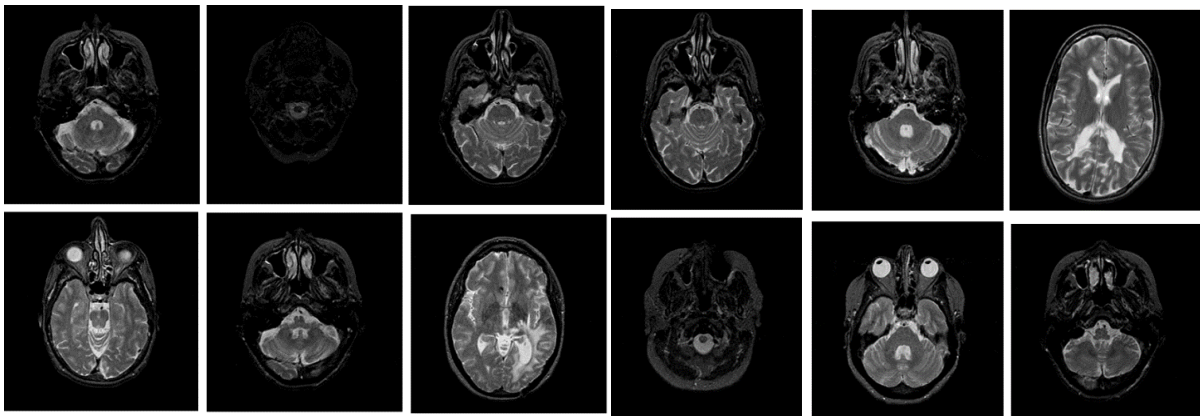


Fig.01. Exemples images de la Base d'Apprentissage et de test

IV .4. Conception du réseau de neurones :

IV .4.1. Architecture et configuration de réseau de neurones :

On a adopté un réseau de neurones MLP (Multi Layer Perceptron) qui représente les caractéristiques suivantes :

- ✓ Une couche d'entrée avec un nombre 'n' d'entrées égales au nombre de descripteurs caractérisant l'image à classer. Le vecteur d'entrée de réseau de neurones doit être normalisé. Toutes les valeurs de vecteur d'entrée doit appartenir à l'intervalle [0 1]. Cela pour éviter l'influence d'une entrée par rapport aux autres.
- ✓ Une couche cachée avec un nombre « k » variable de neurones pour avoir le taux de classification le plus élevé
- ✓ Une couche cachée avec 2 sorties qui représente les classes des images, Dans notre cas, l'image peut être soit saine ou pathologique.

- ✓ On a utilisé deux fonctions d'activation sigmoïde :
 1. Une de type logistique (Logsig)
 2. L'autre de type symétrique (Tansig).
- ✓ Le nombre d'itérations est fixé à **50.000** afin de limiter le temps écoulé lors de l'apprentissage, et pour réaliser un apprentissage plus ou moins satisfaisant. En plus que s'il était trop grand, il y aurait une situation de sur apprentissage, et par contre s'il était trop petit, un arrêt précoce de l'apprentissage est survenu.
- ✓ L'erreur d'apprentissage est fixée à 0,00001 pour la première partie et 0,000001 pour deuxième partie.
- ✓ Le pas d'apprentissage est fixé à 0.001 : un pas d'apprentissage plus petit produira des oscillations, tandis qu'un pas plus élevé grand ne donne pas des résultats satisfaisants.

L'image suivante illustre l'architecture de notre réseau :

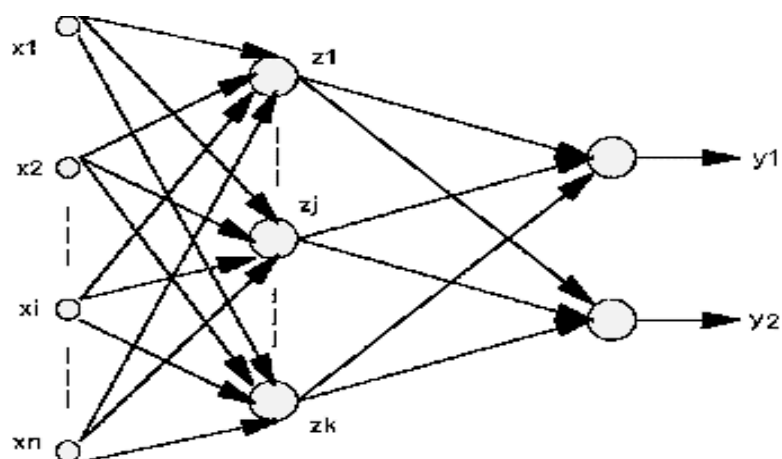


Fig.02. Architecture de réseau de neurones

IV .4.2. Apprentissage des réseaux de neurones :

L'apprentissage consiste à calculer les pondérations optimales des différentes connexions du réseau de neurones en utilisant un échantillon de données. A un même type du réseau de neurones, plusieurs scénarios d'apprentissage peuvent être envisagés. La vitesse de convergence (temps d'apprentissage) et l'erreur d'apprentissage sont les deux critères de choix les plus utilisés. L'apprentissage permet une comparaison entre le résultat obtenu en fonction des entrées fournies, et la sortie désirée. Ainsi, le réseau continue à se modifier jusqu'à obtention des résultats satisfaisants.

Pour l'apprentissage de notre réseau de neurones, on a choisi 02 algorithmes prédéfinis dans MATLAB qui sont l'algorithme de descente de gradient et l'algorithme de Levenberg-Marquardt basés sur la minimisation de l'erreur quadratique. Cela afin de comparer la robustesse des 02 algorithmes vis-à-vis de l'apprentissage. Pour plus de détail de ces 02 algorithmes, voir le chapitre III.

La figure suivante illustre les étapes de l'apprentissage d'un réseau de neurones :

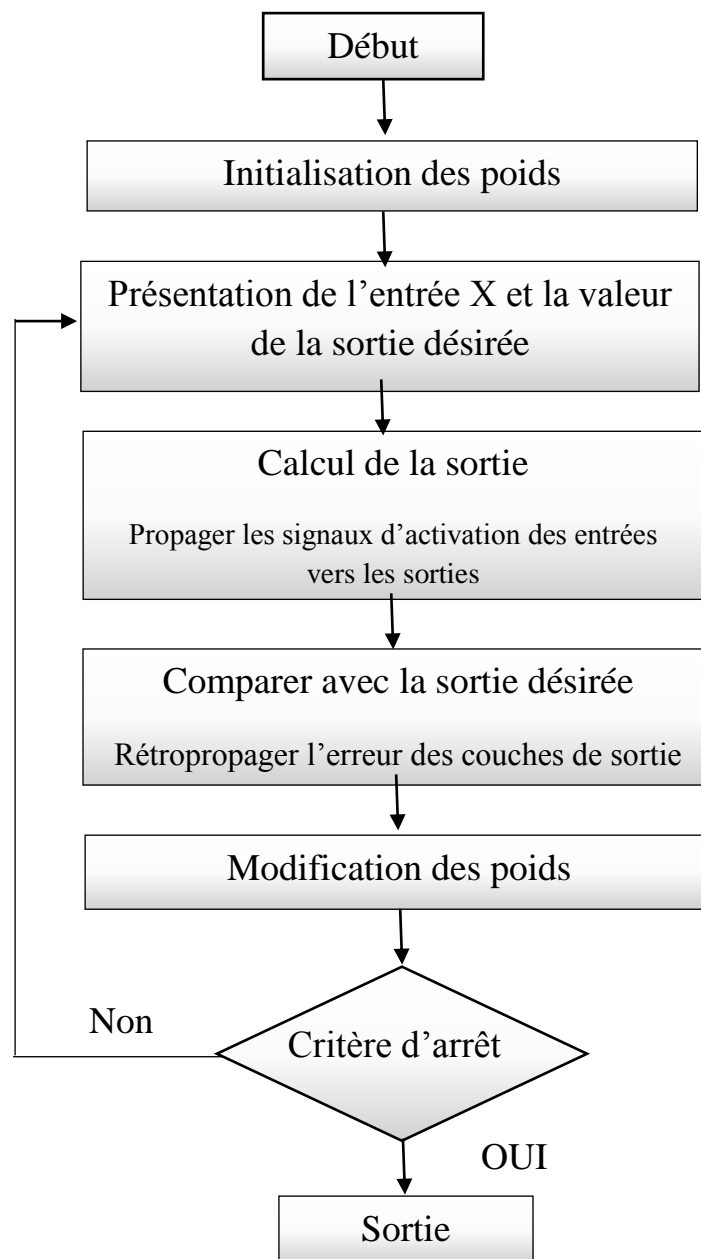


Fig.03. Les étapes de l'apprentissage d'un réseau de neurones

IV .4.3. Phase de test :

La phase de test s'effectue après entraînement du réseau de neurones (après l'apprentissage). Il s'agit de décider à quelle classe affecter une nouvelle entrée, l'algorithme de cette phase est très simple dans le sens où il fait qu'une propagation de l'entrée vers l'avant. Pour que le système soit performant la base de données utilisée doit être différente de celle utilisée en phase d'apprentissage.

IV .5. Critères de performances :

IV .5.1 Critères de performances de l'apprentissage :

- *Le temps d'apprentissage* : Le temps d'apprentissage définit le temps nécessaire à la convergence de l'algorithme de rétropropagation vers une erreur quadratique désirée.
- *L'erreur quadratique* : On définit l'erreur quadratique par l'écart qui existe entre la sortie réelle et la sortie désirée. Elle est définie par la loi suivante :

$$e = \frac{1}{2} \sum_{k=1}^m (Y_i - Y_d)^2$$

IV .5.2 Critères de performances de test :

La performance d'un réseau de neurones conçu pour des problèmes de classification repose sur quatre notions essentielles que sont les vrais positifs (VP), faux positifs (FP), vrais négatifs (VN) et faux négatifs (FN).

FN (Faux Négatif) : désigne un test négatif dans un individu atteint d'une maladie.

FP (Faux Positif) : désigne un test positif dans un individu sain.

VN (Vrai Négatif) : désigne un test négatif dans un individu sain.

VP (Vrai Positif) : désigne un test positif dans un individu atteint d'une maladie.

Un bon test de diagnostic minimise le nombre de FN et de FP.

Ces notions sont utilisées pour calculer des critères classiques comme le taux de classification, la sensibilité et la Spécificité comme suit :

- **Taux de classification :**

Le taux de classification permet d'évaluer la qualité du classifieur par rapport au problème pour lequel il a été conçu. Les performances en termes de taux de classification sont

alors déterminées en présentant au classifieur chacun des exemples de la base de test et en comparant la classe donnée en résultat à la vraie classe. Donc le taux de classification représente le rapport des images bien classées sur le nombre total d'images. Il est défini par la loi :

$$Tc(\%) = \frac{VP + VN}{VP + VN + FP + FN} \times 100$$

- **Sensibilité :**

On appelle sensibilité (Se) du test sa capacité de donner un résultat positif quand la maladie est présente. Dans le langage des probabilités, la sensibilité mesure la probabilité conventionnelle que le test soit positif lorsque la maladie est présente. La sensibilité est estimée par la proportion de résultats positifs par suite de l'application du test à un groupe d'individus reconnus comme ayant la maladie. Elle est défini par :

$$Se (\%) = \frac{VP}{VP + FN} \times 100$$

- **Spécificité :**

On appelle spécificité (Sp) du test cette capacité de donner un résultat négatif quand la maladie est absente. Dans le langage des probabilités, la spécificité mesure la probabilité conventionnelle que le test soit négatif lorsque la maladie est absente. La spécificité est estimée par la proportion de résultats négatifs conséquemment à l'application du test à un groupe d'individus reconnus comme n'ayant pas la maladie. Elle est définie par la loi :

$$Sp (\%) = \frac{VN}{VN + FP} \times 100$$

IV .6. Résultats de différents apprentissages :

Nous avons travaillé avec **MATLAB** version **7.9 (R2009b)** sous l'environnement Windows7, qui contient la toolbox pour l'analyse d'images (image processing toolbox) et la toolbox (neural network toolbox) pour l'apprentissage des réseaux de neurones.

IV .6.1. Résultats d'apprentissage avec les descripteurs de texture :

Nous avons utilisé 06 descripteurs mentionnés ci-dessus, calculés à partir de la matrice de cooccurrence dans les quatre directions. Ce qui fait l'extraction de 24 descripteurs.

Le tableau suivant donne quelques valeurs d'attributs de 03 images de la base d'apprentissage.

Descripteurs Images	Moyenne	Contraste	Homogénéité	Energie	Corrélation	Entropie
Image 1	42.9191	96.2575	0.8320	0.6010	0.6455	0.9368
Image 2	47.5485	96.9729	0.8122	0.5582	0.6496	0.9904
Image 3	42.9191	106.1307	0.8091	0.5551	0.6373	0.9934

Tableau 01. Quelques valeurs d'attributs de 03 images de la base d'apprentissage

Les critères de performances de l'apprentissage d'un réseau de neurones sont l'erreur quadratique et le temps d'apprentissage. Les 02 figures suivantes illustrent ces 02 critères selon le nombre de neurones de la couche cachée et l'algorithme d'apprentissage.

- *Résultats expérimentaux :*

Algorithme d'apprentissage : Levenberg-Marquardt

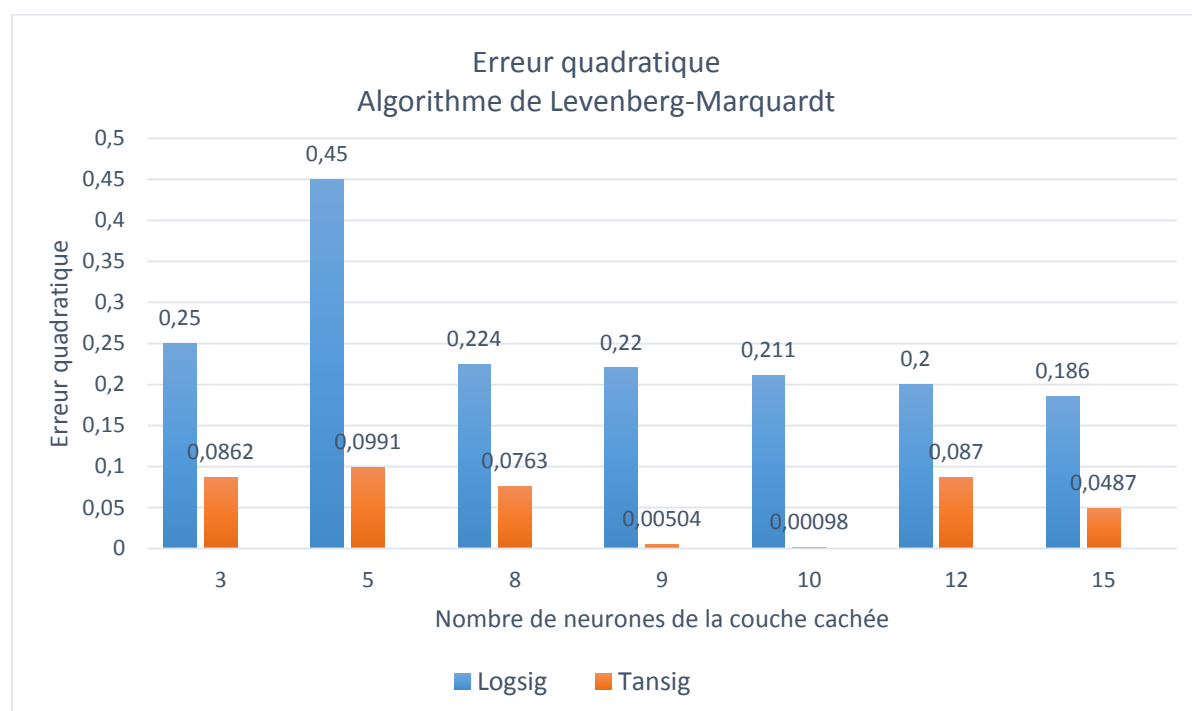


Fig.04. Histogramme d'erreurs quadratiques en utilisant l'algorithme de Levenberg-Marquardt

On constate que l'erreur quadratique varie selon le nombre de neurones de la couche cachée et la fonction d'activation. La fonction symétrique (Tansig) donne de meilleurs résultats par rapport à la fonction Logsig. L'erreur est minimale ($e = 0,0098$) pour $N=10$ neurones et la fonction d'activation symétrique (Tansig).

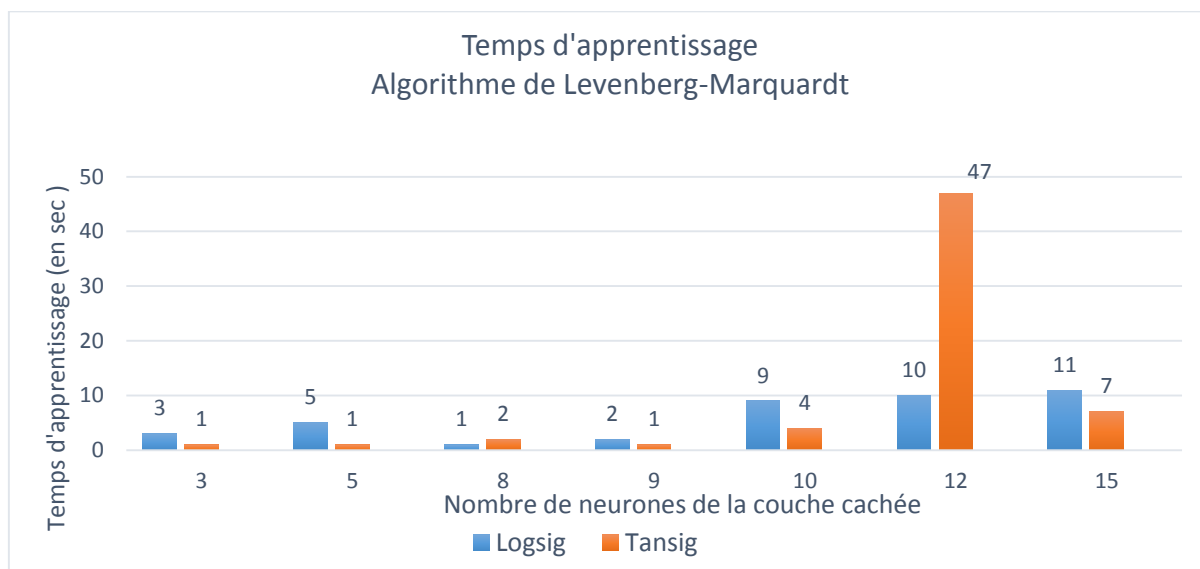


Fig.05. Histogramme des temps d'apprentissage en utilisant l'algorithme de Levenberg-Marquardt

Le temps d'apprentissage varie selon le nombre de neurones de la couche cachée. Le temps d'apprentissage augmente en augmentant le nombre de neurones cachés. Ainsi on constate que l'algorithme de Levenberg-Marquardt assure une convergence rapide vers l'erreur désirée.

Algorithme d'apprentissage : Descente de gradient

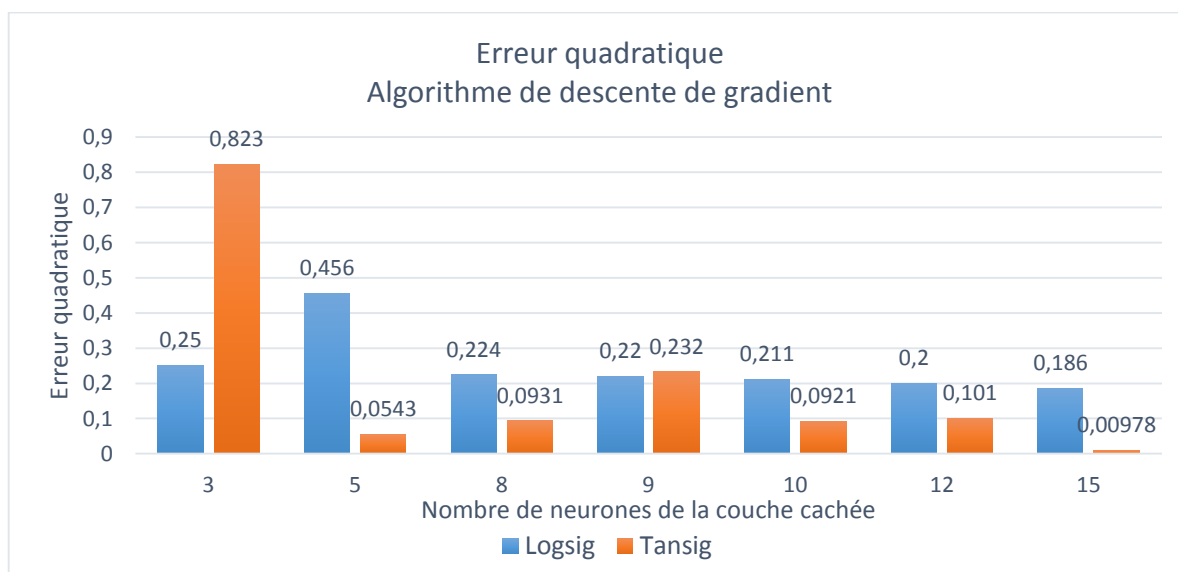


Fig.06. Histogramme d'erreurs quadratiques en utilisant l'algorithme de descente de gradient

On constate que l'erreur quadratique varie selon le nombre de neurones cachés. Ainsi la fonction symétrique (Tansig) permet d'avoir un écart plus au moins petit par rapport à la fonction logistique. Dans notre cas, l'erreur la plus minime ($e = 0,00978$) est obtenue pour $N=15$ en utilisant la fonction symétrique.

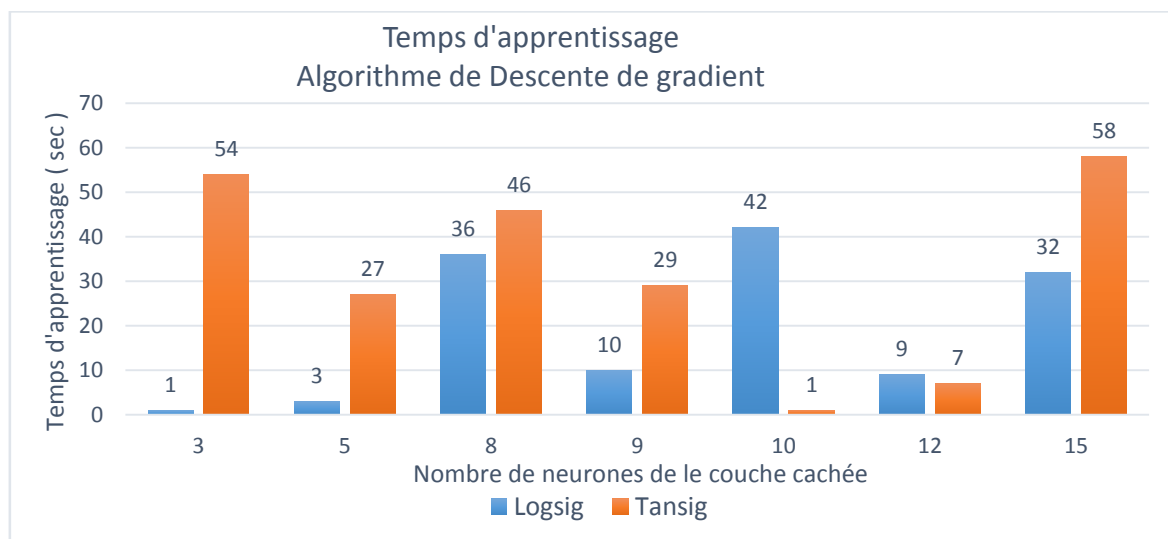


Fig.07. Histogramme des temps d'apprentissage en utilisant l'algorithme de descente de gradient

On constate que l'algorithme de descente de gradient est plus lent par rapport à l'algorithme de Levenberg-Marquardt et il n'y a pas de règle qui lie le temps d'apprentissage et le nombre de neurones cachés.

- Critère de l'erreur quadratique :

D'après les résultats obtenus, nous remarquons que l'erreur d'apprentissage varie selon l'algorithme d'apprentissage utilisé, la fonction d'activation et le nombre de neurones de la couche cachée.

- La fonction symétrique (Tansig) donne des meilleurs résultats avec un nombre assez grand de neurones de la couche cachée.
- L'algorithme de Levenberg-Marquardt assure une erreur minimale par rapport à celui de descente de gradient.
- L'erreur la plus petite est $e = 0.000980$, avec les paramètres : 10 neurones cachés, algorithmes de Levenberg-Marquardt et la fonction d'activation symétrique (Tansig).
- Critère de temps d'apprentissage :

Nous notons que le temps d'apprentissage est le temps nécessaire à la convergence de l'algorithme d'apprentissage.

- L'algorithme d'apprentissage de Levenberg-Marquardt assure un temps d'apprentissage minimal par rapport à celui de descente de gradient.
- Contrairement à l'erreur d'apprentissage, le temps d'apprentissage atteint sa valeur minimale avec peu de neurones cachés.

IV.6.2. Résultats d'apprentissage avec les descripteurs de forme (Moments de ZERNIKE) :

Le calcul des moments de Zernike se fait directement à partir des images IRM. Un seul paramètre à fixer, l'ordre de décomposition.

Le tableau suivant donne quelques valeurs de moments de Zernike d'une image IRM cérébrale jusqu'à l'ordre de $n=7$. Le nombre de moments est $g = (7+1).(7+3)/4=20$ moments.

Ordre	Les moments de Zernike			
0	638.53			
1	-510.37 + 318.36i			
2	-29.13	116.85 - 148.74i		
3	153.77 - 153.90i	-54.46 - 170.02i		
4	64.01	21.59 + 95.18i	146.43 + 146.54i	
5	-103.24 - 57.95i	40.95 + 65.83i	7.65 + 2.95i	
6	-269.88	84.68 + 49.60i	-91.98 + 37.32i	-146.20 + 36.56i
7	147.72 - 27.57i	-88.79 - 05.56i	-108.44 - 113.62i	51.31 - 106.78i

Tableau.O2.Quelques valeurs de moments de Zernike pour une image IRM cérébrale

Pour la phase d'apprentissage en utilisant les descripteurs de forme on a utilisé uniquement la fonction symétrique (Tansig) et l'algorithme de Levenberg-Marquardt comme algorithme d'apprentissage

On fait varier le nombre de neurones de la couche cachée de 5 à 28 jusqu'à l'obtention des résultats satisfaisants et on calcul à chaque fois l'erreur quadratique et le temps d'apprentissage.

Les moments de ZERNIKE sont introduits dans un réseau de neurones en quatre cas possibles Soit on introduit :

Cas 1 : l'amplitude des moments de Zernike

Cas 2 : la phase des moments de Zernike

Cas 3 : l'amplitude et la phase en même temps

Cas 4 : Convertir le nombre complexe en un nombre réel à deux dimensions.

- *Résultats expérimentaux :*

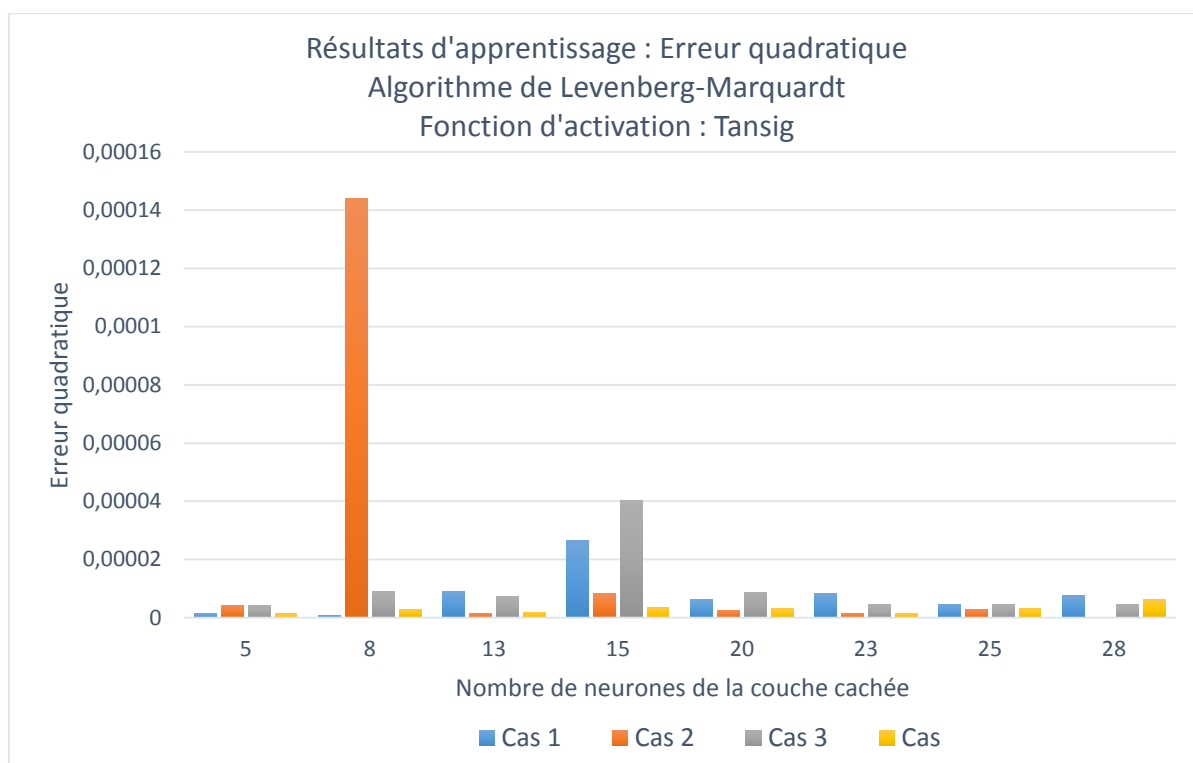


Fig.11. Histogramme d'erreurs quadratiques en utilisant l'algorithme de Levenberg-Marquardt et la fonction Tansig.

Nous constatons que l'erreur quadratique varie selon le nombre de neurones cachés et le types de vecteur d'entrée de réseau de neurones. L'erreur minimale ($e=0,0000019$) est obtenue pour $N=23$ dont le vecteur caractéristique est constitué des parties réelles et imaginaires des moments de Zernike.

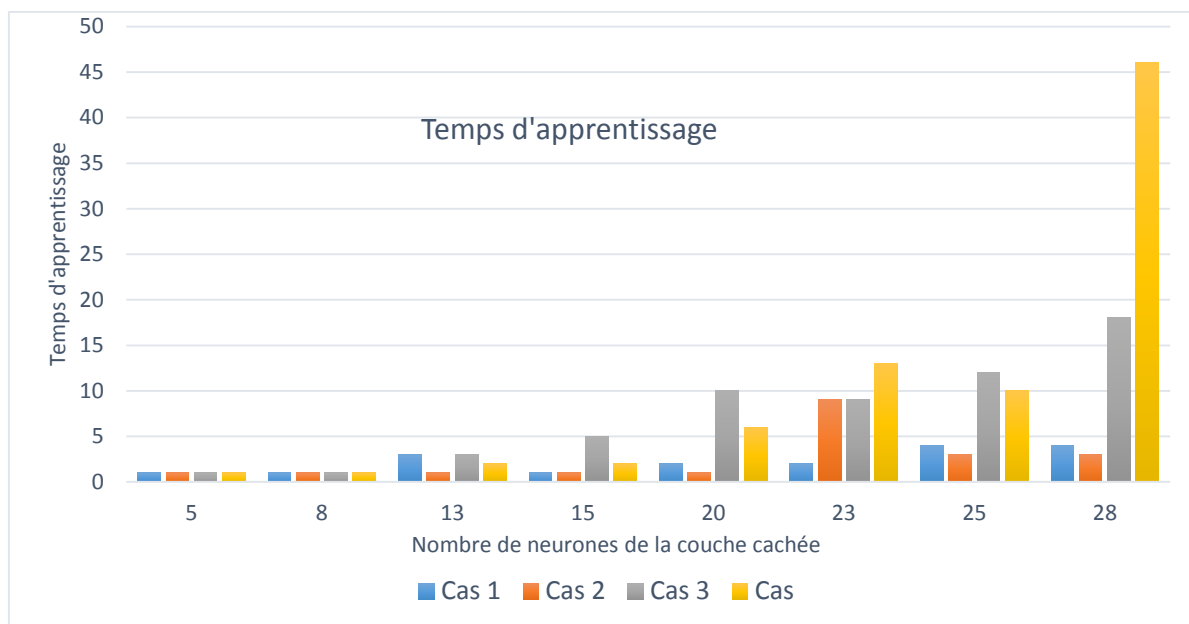


Fig.12. Histogramme des temps d'apprentissage en utilisant l'algorithme de Levenberg-Marquardt et la fonction Tansig.

Le temps d'apprentissage est proportionnellement relative au nombre de neurones de la couche cachée. On augmentant le nombre de neurones, le temps d'apprentissage augmente et l'erreur quadratique, en général, diminue. Donc pour le choix de la meilleure configuration, on doit satisfaire le compromis erreur quadratique-temps d'apprentissage.

Discussion des résultats de l'apprentissage :

- Erreur quadratique : On constate que l'erreur quadratique est minimale pour la majorité des configurations des réseaux de neurones.
- Temps d'apprentissage : l'algorithme de Levenberg-Marquardt assure une convergence rapide vers l'erreur quadratique désirée.
- On constate aussi que le meilleur cas d'apprentissage est celui dont le vecteur caractéristique est constitué par la partie réelle et imaginaire des moments de Zernike.

IV.7. Résultats de tests :

IV.7.1. Résultats de test avec les descripteurs de texture :

Pour tester la performance du réseau de neurones, on choisit les meilleures configurations prises dans la phase d'apprentissage. On a choisi des configurations qui minimisent l'erreur quadratique et le temps de convergence de l'algorithme. La base de test contient des images saines et pathologiques. Notons que les images de la base de test sont

différentes de la base d'apprentissage. Dans cette phase, on teste la capacité de réseau de neurones à classer ces images en deux classes, saine ou pathologique.

Les figures suivantes donnent les résultats de test : le taux de classification, la sensibilité et la spécificité pour les différentes configurations de réseau de neurones :

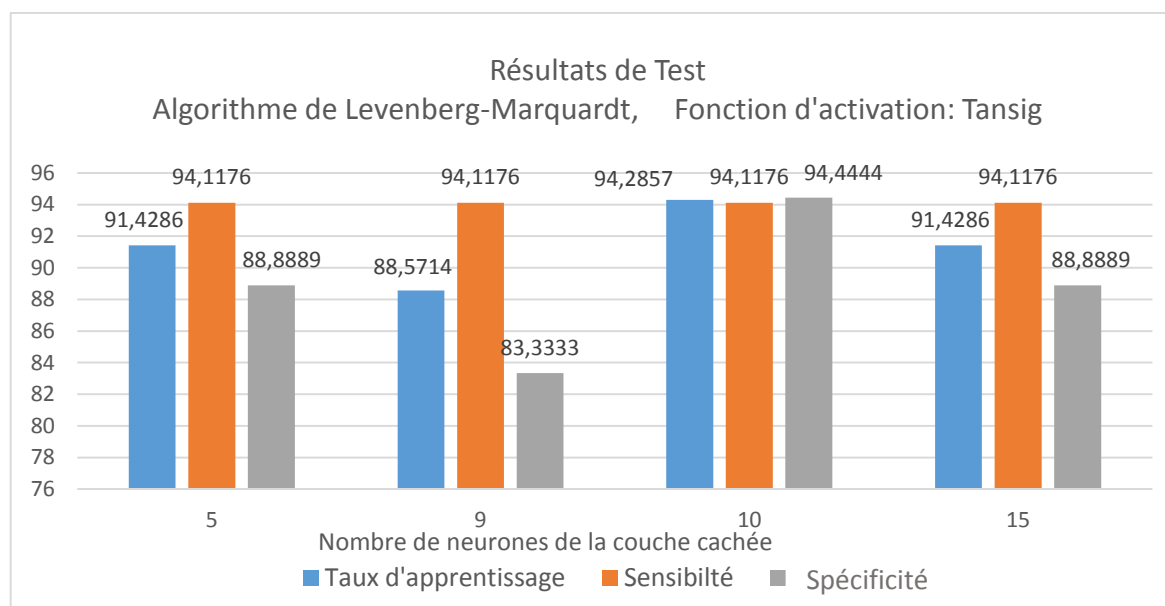


Fig.08. Histogramme des performances en utilisant l'algorithme de Levenberg-Marquardt et la fonction Tansig.

Le meilleur résultat obtenu est pour le nombre de neurones $N=10$ neurones. Le taux de classification, la sensibilité et la spécificité les plus élevés sont : $T_c=94,2857$, $S_e=94,1176$, $S_p=94,1176$.

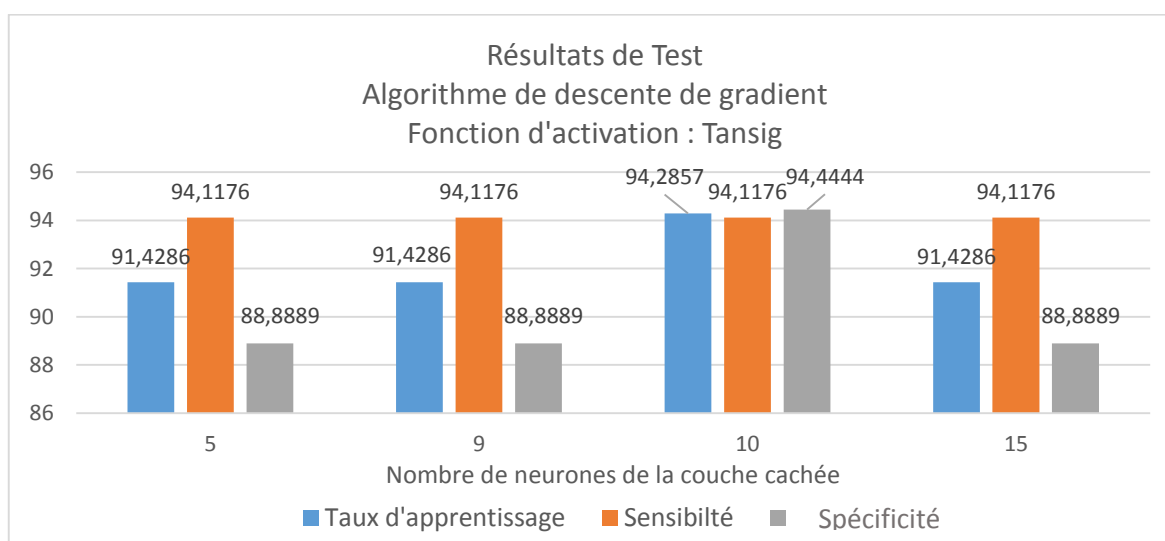


Fig.09. Histogramme des performances en utilisant l'algorithme de descente de gradient et la fonction Tansig.

On constate que pour un nombre de neurones $N=10$, on aura avec l'algorithme de descente de gradient le même résultat de classification que précédemment : $T_c=94,2857$, $Se=94,1176$, $Sp=94,1176$.

- *Discussion des résultats de test :*

D'après les figures précédentes on constate que le meilleur résultat de test est :

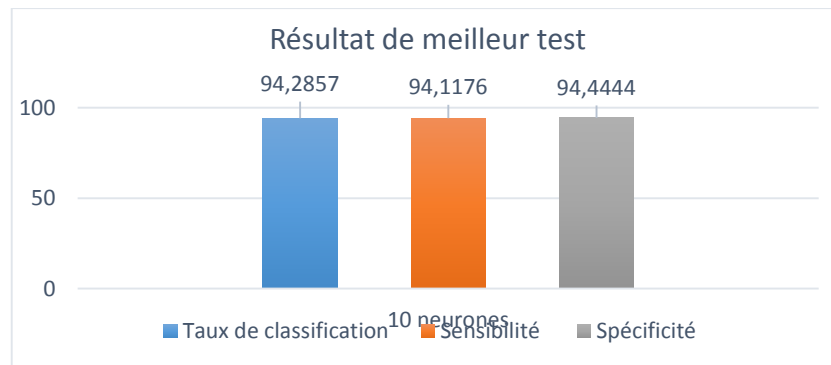


Fig.10. Histogramme de meilleur résultat de test

Les meilleurs résultats de test sont obtenus pour $N = 10$ neurones cachés et la même configuration de réseau de neurone qui a donné l'erreur quadratique la plus petite. Donc, on déduit que la phase de test dépend de la phase d'apprentissage. Pour augmenter le taux de classification, la sensibilité et la spécificité d'un réseau de neurones, on doit prendre les configurations qui minimisent l'erreur quadratique dans la phase d'apprentissage.

IV .7.2. Résultats de test avec les descripteurs de forme :

Dans cette phase, on teste notre réseau à classer les images IRM cérébrales. Les performances de réseau sont calculées en termes de taux de classification, de sensibilité et de spécificité pour différentes configurations.

1. *Le taux de classification :*

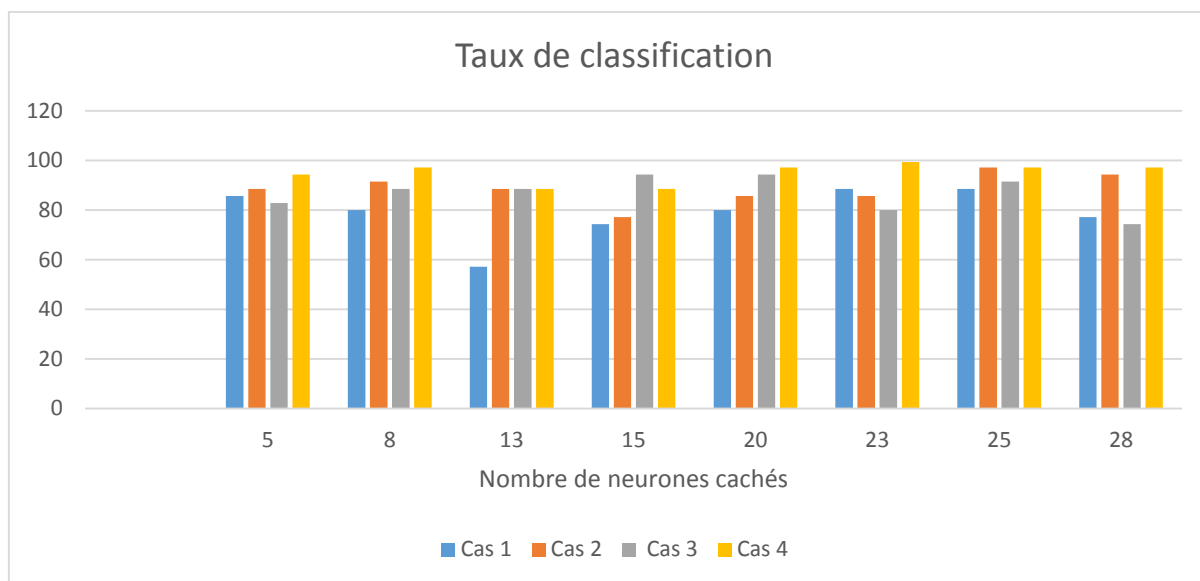


Fig.13. Histogramme des taux de classification pour les 04 cas

L’histogramme représente les réseaux de neurones (pour chaque nombre de neurones cachés utilisé) qui ont donné des taux de reconnaissance élevés. D’après cette figure, on constate que les taux de classification les plus élevés sont ceux de cas 3 et le cas 4. Cela revient à la taille de vecteur d’entrée (qui est le double par rapport aux cas 1 et le cas 2). Le meilleur taux de classification est $T_c=99,3\%$ pour $N=23$ neurones cachés, algorithme de Levenberg-Marquardt et la fonction symétrique (Tansig).

2. La sensibilité :

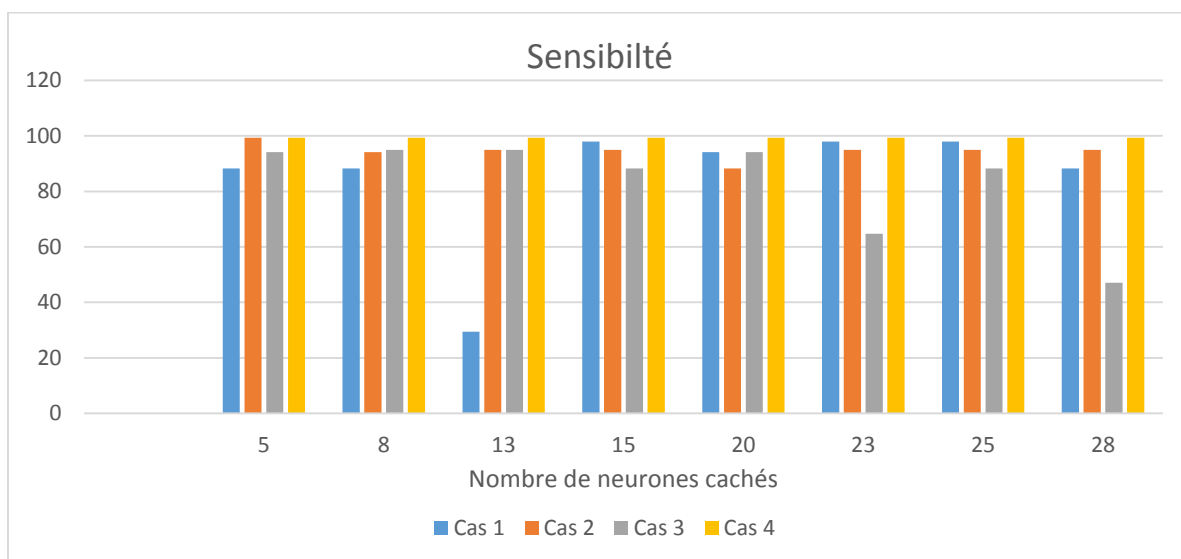


Fig.14. Histogramme des sensibilités pour les 04 cas

L'histogramme montre que la sensibilité des réseaux de neurones est meilleure pour les réseaux qui possèdent un nombre de neurones cachés plus de 5 neurones ($Se=99,3\%$) dont le vecteur caractéristique est composé des parties réelle et imaginaire des moments de Zernike. le nombre des vrais positifs est relativement élevé.

Nous remarquons que les réseaux de neurones possédant un nombre petit de neurones cachés, ne sont pas assez sensibles à la reconnaissance des pathologies. Et par la suite le nombre de neurones dans la couche cachée influe fortement sur la sensibilité du réseau.

3. La spécificité :

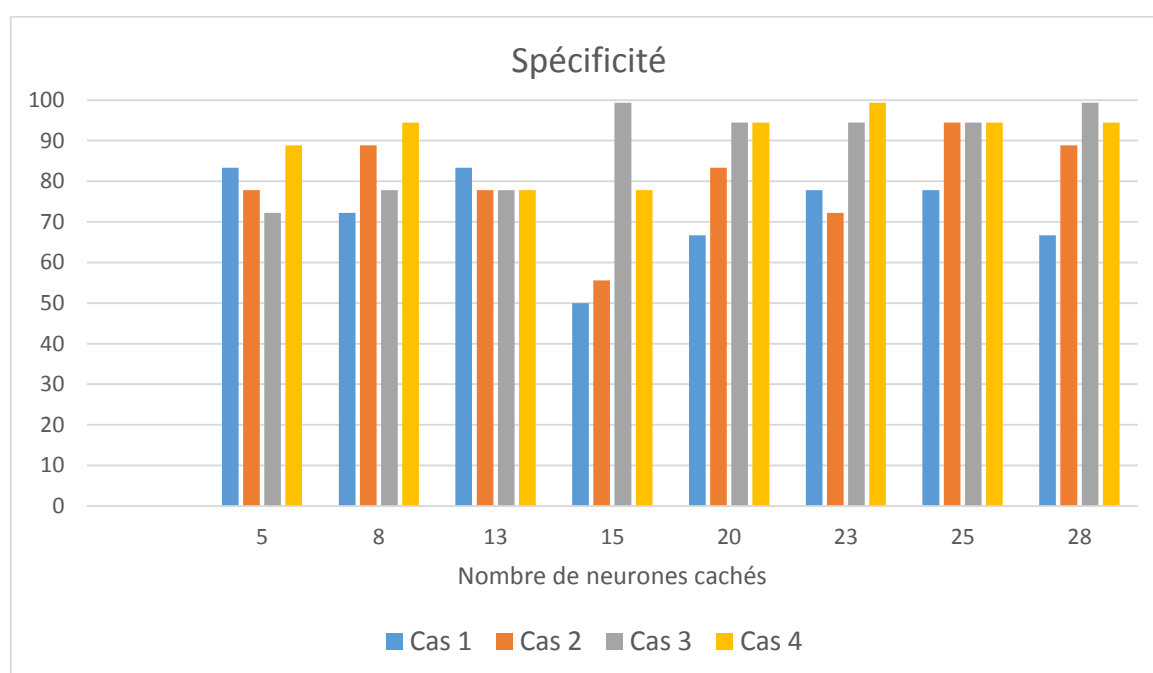


Fig.15. Histogramme des spécificités pour les 04 cas

Nous rappelons que la spécificité est estimée par la proportion de résultats négatifs conséquemment à l'application du test à un groupe d'individus reconnus comme n'ayant pas la maladie. Nous remarquons que la spécificité varie selon le nombre de neurones cachés. Le meilleur résultat est $Sp=99,3\%$ pour $N=23$ neurones.

Discussion des résultats de l'analyse de forme :

- *Le taux de classification :*

D'après les résultats précédents, on déduit que le taux de classification dépend :

- De nombre de neurones de la couche cachée. En général, en augmentant le nombre de neurones cachés, on diminue l'erreur quadratique ce qui permet d'augmenter le taux de classification.
- De la taille de vecteur caractéristique. En augmentant la taille de vecteur caractéristique, le taux de classification augmente.

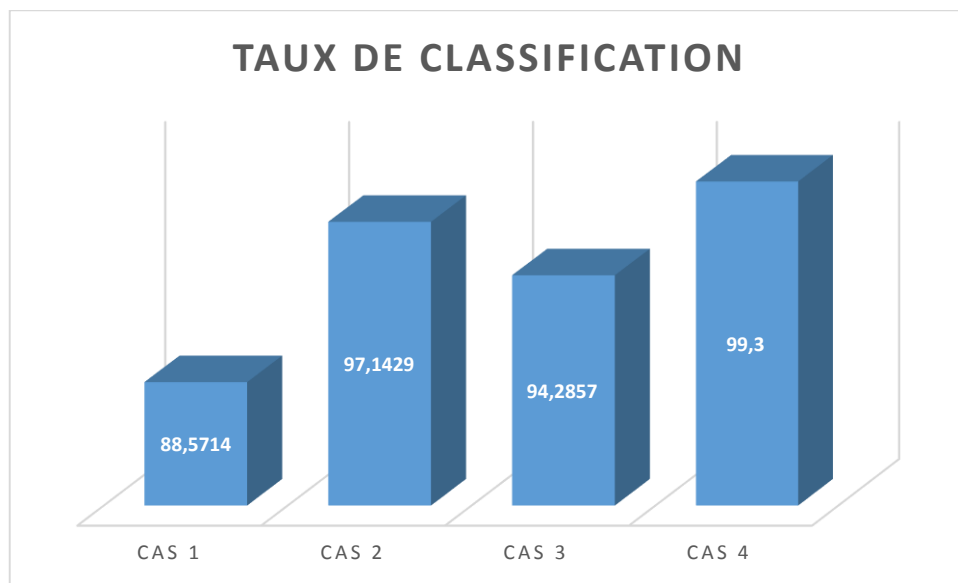


Fig.16. Histogramme des meilleurs taux de classification pour les 04 cas

- *La sensibilité :*

La sensibilité représente la capacité de réseau de neurones à détecter les pathologies. Le meilleur résultat trouvé est $Se=99,3\%$ pour $N=23$ neurones. On a vu que la sensibilité varie selon le nombre de neurones cachés. Pour améliorer la sensibilité d'un réseau de neurones, on doit choisir un nombre suffisant de neurones de la couche cachée.

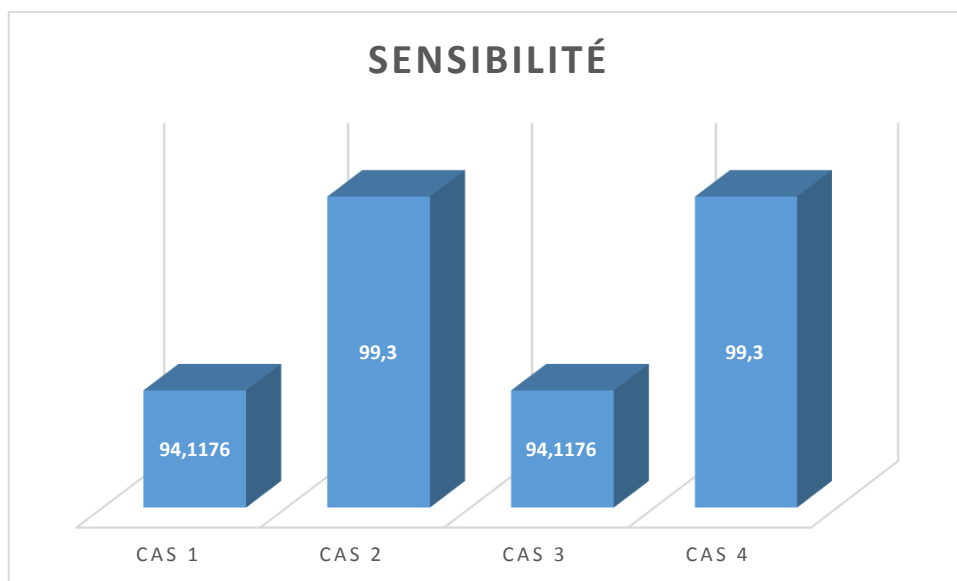


Fig.17. Histogramme des meilleures sensibilités pour les 04 cas

- la spécificité :

La spécificité mesure la probabilité conventionnelle que le test soit négatif lorsque la maladie est absente. Autrement dit, la capacité de réseau de neurones à classer les images saines comme étant des images saines. Dans notre travail, la spécificité $Sp=99,3\%$ (cas 2, cas 4).

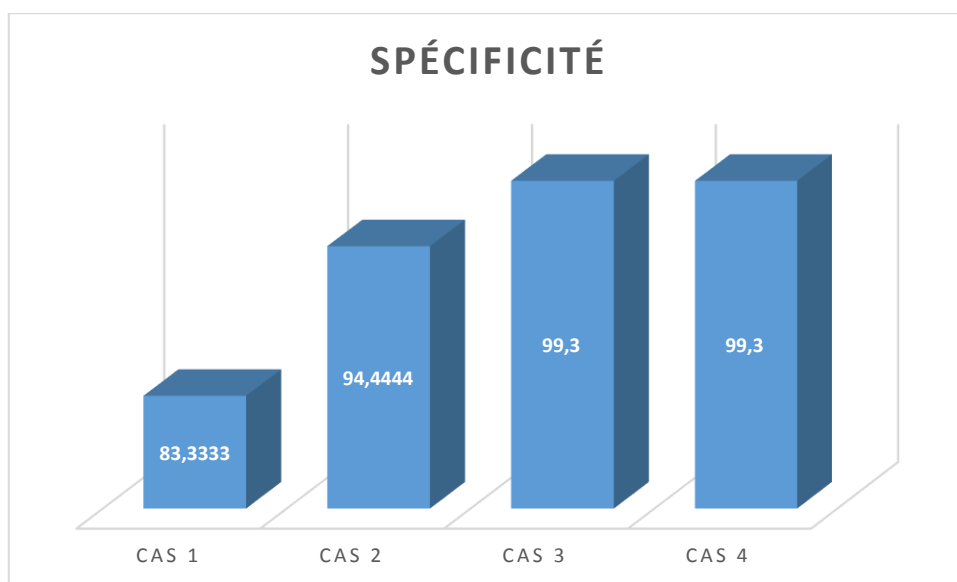


Fig.18. Histogramme des meilleures spécificités pour les 04 cas

Notons aussi que le taux de classification, la sensibilité et la spécificité de réseau de neurones dépend essentiellement de nombre d'exemples de pathologies présentés lors de

l'apprentissage de réseau de neurones. Pour augmenter les performances de réseau de neurones, vis-à-vis la classification, on doit avoir une base d'apprentissage assez importante pour entrainer le réseau à détecter toutes pathologies.

IV .8. Comparaison des résultats obtenus :

Les meilleurs résultats obtenus dans notre étude de classification d'images IRM cérébrales par les réseaux de neurones sont représentés ci-dessous :

Performances		Taille de vecteur d'entrée	Nombre de neurones cachés	Taux de classification (%)	Sensibilité (%)	Spécificité (%)
Descripteurs						
Analyse de texture (Matrice de cooccurrence)		24	10	94,2857	94,1176	94,1176
Analyse de forme (Moments de Zernike)	Cas 1	49	23	88,5714	99,3	83,3333
	Cas 2	49	23	94,2857	99,3	99,3
	Cas 3	98	23	97 ,1429	99,3	94,4444
	Cas 4	98	23	99,3	99,3	99,3

Tableau.03. les meilleures performances obtenues

Le meilleur résultat de classification en utilisant les descripteurs de texture et de forme est représenté dans les deux histogrammes suivants :

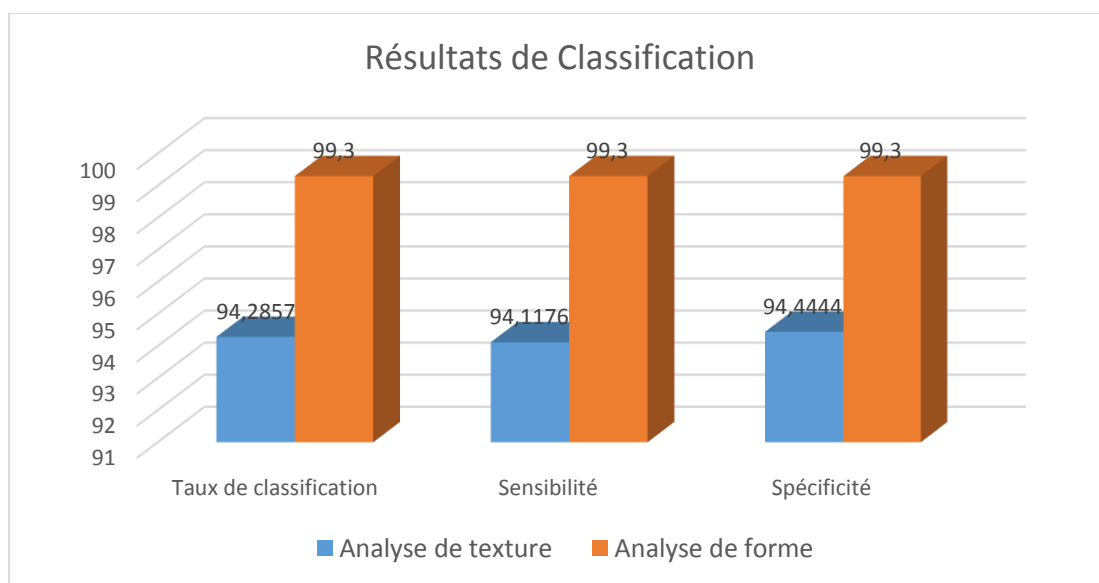


Fig.19. Histogramme des meilleurs résultats de classification

D'après ces résultats, on déduit que les caractéristiques de forme ont donné de meilleurs résultats. Malgré que les caractéristiques de texture permettent de faire la distinction entre les tissus sains et tissus pathologiques, les moments de Zernike constituent un descripteur robuste à fort pouvoir descriptif. Ils ont donné de meilleurs résultats avec un taux de classification de 99,3%.

IV .9. Conclusion :

L'objectif de ce chapitre était d'appliquer les principes théoriques des réseaux de neurones afin de réaliser un classifieur neuronal des images IRM cérébrales à deux classes (classe d'images saines et classe d'images pathologiques). Le classifieur a été conçu avec des architectures différentes et une variété de paramètres d'apprentissage à savoir les algorithmes d'apprentissage et les fonctions d'activation.

Une étude de ce classifieur neuronal a été faite concernant les performances d'apprentissage (erreur et temps d'apprentissage), ainsi qu'une évaluation pour trouver les meilleurs réseaux de neurones, ce qui reste purement empirique. Nous notons que la faisabilité d'une amélioration des performances en enrichissant la base d'apprentissage, et puis en élargissant le vecteur d'entrée avec d'autres paramètres pertinents pour pouvoir augmenter le taux de reconnaissance.

Finalement, Il ressort de cette étude que l'utilisation des moments de Zernike comme descripteurs d'images IRM cérébrales présentées pour la classification donne des résultats meilleurs par rapport aux descripteurs de texture.

Conclusion générale

Conclusion générale

Le travail présenté dans ce mémoire concerne le domaine du traitement d'images et plus précisément celui de la classification d'images médicales. Nous avons développé des classifieurs neuronaux différents, comme des outils d'aide au diagnostic médical.

Une étude comparative des différents classifieurs basée sur l'analyse de leurs performances (erreurs quadratiques et temps d'apprentissage) a montré que leurs capacités sont influencées par le choix des paramètres d'apprentissage : fonctions d'activation et l'algorithme d'apprentissage. Les classifieurs neuronaux se caractérisent par une grande capacité d'apprentissage en termes de précision et d'adaptation (facile à implémenter).

Pour l'évaluation expérimentale de notre système, nous avons prévu deux cas pour les entrées du réseau, le 1^{er} cas les entrées sont les caractéristiques de texture et le deuxième cas les entrées sont les moments de Zernike, Chacun de ces cas a permis d'avoir un taux de classification particulier. En effet, en utilisant notre classifieur, le taux de classification est évalué respectivement à 94.28 %, et 99,3% pour les différents cas cités ci-dessus. Les moments de Zernike donnent de meilleurs résultats par rapport aux descripteurs de texture et ces résultats sont très encourageants et très prometteurs.

En perspective de ce travail, nous pouvons améliorer les performances des classifieurs neuronaux, on faisant appel à des techniques d'hybridation à savoir la logique floue et les algorithmes génétiques, la logique floue contribue par sa capacité de représentation des connaissances imprécises.

Un modèle hybride de l'approche neuronale et l'approche floue fusionne la capacité de modélisation des réseaux de neurones avec la facilité d'interprétation et d'intégration des connaissances des systèmes flous. La principale propriété du modèle neuro flou est sa capacité de traiter en même temps des connaissances numériques et symboliques.

Les algorithmes génétiques ont une large capacité d'adaptation d'où l'apparition des systèmes neuro génétiques pour contourner le problème de choix d'architecture des réseaux de neurones ou encore le choix des descripteurs optimaux.

Références bibliographiques

Références bibliographiques :

- [01] : Support de cours « LES ULTRASONS EN MÉDECINE », Biophysique et Imagerie Médicale, Faculté de Médecine de LILLE.
- [02] : <http://www.em-premium.com>
- [03] : M. L'ABBE MOINGO « Analyse complète des travaux modernes relatifs aux phénomènes de la lumière » Imprimerie de Wittersheim, Paris.
- [04] : Olivier Ernst « Les rayons X » support de cours, Université Lille
- [05] : Y. Arnoud « Cours de Radioactivité », Cours de physique nucléaire.
- [06] : Support de cours en ligne « http://www.med.univ-rennes1.fr/cerf/edicerf/BASES/BA004_idx.html », Université de Rennes.
- [07] : Pr L. BOUMGHAR « Déploiement sur une plateforme de visualisation, d'un algorithme coopératif pour la segmentation d'images IRM basé sur les systèmes multi-agents », Département Informatique, Université des Sciences et de la Technologie Houari Boumediene.
- [08] : Anne-Sophie Capelle « Segmentation d'images IRM multi-échos tridimensionnelles pour la détection des tumeurs cérébrales par la théorie de l'évidence », l'université de Poitiers
- [09] : J.Pélissier / P. Labauge « La Sclérose en plaques » Editions Masson, ISBN-13 : 9782294012679
- [10] : Sylvain Arlot « Classification supervisée : des algorithmes et leur calibration automatique », Cnrs – École Normale Supérieure – Inria, Équipe-projet Willow.
- [11] : Florence Tupin « Analyse bayésienne en traitement d'images », Ecole Nationale Supérieure des télécommunications.

Références bibliographiques :

- [12] : LAMRI LAOUAMER « Approche exploratoire sur la classification appliquée aux images », université du Québec.
- [13] : Sandrine TOMEI LE DIGARCHER « Détection automatique de fixations radioactives dans les images TEP oncologiques », L'Institut National des Sciences Appliquées de Lyon
- [14] : Cyril Meurie « Segmentation d'images couleur par classification pixellaire et hiérarchie de partitions », l'UNIVERSITE de CAEN.
- [15] : Laseldi Loria « la méthode de classification non supervisé K-means axiales », Université de Franche Comité.
- [16] : Nesma Settouti « Renforcement de l'Apprentissage Structurel pour la Reconnaissance du Diabète », Département Génie Électrique et Électronique, Université Abou Bekr Belkaid – Tlemcen.
- [17] : Cel eux .G, Diday .E, Govaert .G, « Classification automatique de données environnement statistique et informatique ». Dunod, Informatique, 1989.
- [18] Cheng, H. D., Shi, X. J., Min, R., Hu, L. M., Cai, X. P. and Du, H. N. (2006). « Approaches for automated detection and classification of masses in mammograms » , *Pattern Recognition* vol. 39, pp. 646 – 668.
- [19] Li, Q. (2007). « Recent progress in computer-aided diagnosis of lung nodules on thin-section CT », *Computerized Medical Imaging and Graphics* vol. 31, n° 4-5, pp. 248-257.
- [20] Campadelli, P., Casiraghi, E. and Valentini, G. (2005). « Support vector machines for candidate nodules classification », *Neurocomputing* vol. 68, n° 1-4, pp. 281-288.
- [21] Boroczky, L., Zhao, L. and Lee, K. P. (2006). « Feature subset selection for improving the performance of false positive reduction in lung nodule CAD », *IEEE Transactions on Information Technology in Biomedicine* vol. 10, n° 3, pp. 504-511.

Références bibliographiques :

- [22] Lu, L., Wanyu, L. and Xiaoming, S. (2009). « Automated detection of pulmonary nodules in CT images with support vector machines ». *Proceedings of SPIE - The International Society for Optical Engineering*, Shenyang, China, SPIE.
- [23] Gletsos, M., Mougiakakou, S. G., Matsopoulos, G. K., Nikita, K. S., Nikita, A. S. and Kelekis, D. (2003). « A Computer-Aided Diagnostic System to Characterize CT Focal Liver Lesions: Design and Optimization of a Neural Network Classifier. » *IEEE Transactions on Information Technology in Biomedicine* vol. 7, n° 3, pp. 153-162.
- [24] Cheng, H. D. and Cui, M. (2004). « Mass lesion detection with a fuzzy neural network », *Pattern Recognition* vol. 37, n° 6, pp. 1189–1200.
- [25] Sajda, P., Spence, C. and Pearson, J. (2002). « Learning contextual relationships in mammograms using a hierarchical pyramid neural network », *IEEE Transactions on Medical Imaging* vol. 21, pp. 239–250.
- [26] Wu, Y., Ianakiev, K. and Govindaraju, V. (2002). « Improved k-nearest neighbor classification », *Pattern Recognition* vol. 35 pp. 2311–2318.
- [27] Zhang, X.-L. and Jiang, C. (2007). « Improved SVM for Learning Multi-Class Domains with ROC Evaluation », *Proc. International Conference on Machine Learning and Cybernetics*, vol. 5, pp. 2891-2896.
- [28] Gil-Pita, R. and Yao, X. (2008). « Evolving edited k-nearest neighbor classifiers », *International Journal of Neural Systems* vol. 18, n° 6, pp. 459-467.
- [29] Leung, F. H. F., Ling, S. H. and Lam, H. K. (2008). « An improved genetic-algorithm-based neural-tuned neural network », *International Journal of Computational Intelligence and Applications* vol. 7, n° 4, pp. 469-492.
- [30] : Zahra Zerdoumi « Application des réseaux de neurones artificiels à la poursuite des son linéarités fluctuantes des systèmes satellitaires » Thèse de Magister : Université Mohamed Boudiaf de M'Sila.2006.

Références bibliographiques :

- [31] Claude Touzet, Les réseaux de neurones artificiels «Introduction au connexionnisme». Laboratoire d'Etudes et Recherche en Informatique', ISBN 2 - 906 899 - 78X. Juillet 1992.
- [32] : Hervé.Abdi, « Les Réseaux de Neurones », Presses Universitaire de Grenoble, 1994.
- [33] : Claude Touzet, « Les réseaux de neurones artificiels, introduction aux connexionnismes,cours, exercices et travaux pratiques », Juillet 1992.
- [34] Idiou Ghania « Régression et modélisation par les réseaux de neurones », Thèse de Magister, Constantine 2009.
- [35] : J.A.Freeman et D.M.Skapura « Neural Network : Algorithm, Application and Programming Techniques », Edition Addison –Westly Publication Company1992.
- [36] : Aude RONDEPIERRE & Pierre WEISS « méthodes standards en optimisation non linéaire déterministe », département génie mathématique et modélisation, Université de Toulouse.
- [37] : Lalit P. Bhaiya , Virendra Kumar Verma « Classification of MRI Brain Images Using Neural Network », Department of Electronics & Tele-Communication, RCET, Bhilai, C.G. India.
- [38] : K.V.Ramana, Raghu.B.Korrapati « Neural Network Based Classification and Diagnosis of Brain Hemorrhages », CSE Department, Jawaharlal Nehru Technological University.
- [39] : AmirEhsan Lashkari « A Neural Network based Method for Brain Abnormality Detection in MR Images Using Gabor Wavelets », University of Tehran, School of Electrical and Computer engineering.
- [40] : Anchana Khemphila, Veera Boonjing « Parkinsons Disease Classification using Neural Network and Feature selection », World Academy of Science, Engineering and Technology

Références bibliographiques :

- [41] : [HARA79] R.M.Haralick, « Statistical and structural approaches to texture », Proceedings of the IEEE, mai 1979, number 5, vol.67.
- [42] : Anis CHAARI « Nouvelle approche d'identification dans les bases de données biométriques basée sur une classification non supervisée », UFR Sciences et technologies, Université d'Evry Val d'Essonne.
- [43] : Reference <http://www.med.harvard.edu/AANLIB/home.html>,