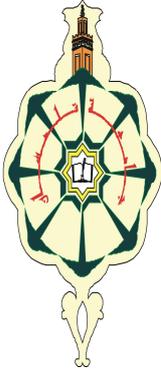


République Algérienne Démocratique et Populaire



Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Abou Baker Belkaid – Tlemcen

Faculté de Technologie

Département de Génie Electrique et Electronique

Filière : électronique biomédicale

Projet de Fin d'Etudes

Magister en électronique biomédicale

Intitulé : classification non supervisée du cancer du sein

Présenté par : youbi Mohammed Ridha

Soutenu le : 15/12/2013, devant le jury composés de :

Président : Pr. besaidabdlhafid

Examineur : Dr. hadj slimanezinedin

Encadreur : Pr. chikh Mohammed amine

Remerciements

L'acheminement d'un mémoire est un parcours de quelques années de notre vie au cours duquel on apprend, on évolue, on croise de nouvelles connaissances, on arrive parfois au bout de la patience,...

Un mémoire est en effet le fruit d'un savoir-faire, d'un savoir vivre...

On ne peut pas passer par un mémoire sans remercier tous ceux qui étaient là pour nous épauler...

J'aimerais avant tout exprimer ma gratitude à mon directeur de mémoire,

Mr. CHIKH Mohammed Amin, professeur à l'université Abou Baker Belkaid. Ce fut un réel plaisir de l'avoir en tant que directeur de mémoire de Magistère. En effet, il a su me faire entrer dans le domaine réel de la recherche, me pousser a toujours cherché le mieux. il m'a permis aussi d'acquérir énormément de connaissances.

Je le remercie finalement pour tout le temps qu'il m'a consacré.

Chaque discussion avec lui est une vraie motivation et un vrai encouragement vers l'avant.

Je le remercie chaleureusement pour ses conseils avisés.

J'exprime ma gratitude à Mr. BESSAID, professeur à l'université Abou Baker Belkaid ainsi qu'à Mr. HADJ Selimane, Maître de conférence à l'université Abou Bakr Belkaid pour avoir accepté de juger mes travaux en tant que rapporteurs. Je souhaite aussi les remercier pour leur implication Fructivie dans l'évaluation de mes travaux. Certes, leurs remarques pertinentes ont permis l'amélioration de mon mémoire de magistère.

Finalement, je remercie tous ceux qui ont participé d'une manière ou d'une autre au bon déroulement de cette mémoire: amis, collègues, famille, voisins... Les citer n'exprime pas réellement ma reconnaissance mais c'est le minimum que je peux leur offrir.

Abstract:

- An artificial neural network (ANN) is an information-processing paradigm inspired by the way the densely interconnected, parallel structure of the mammalian brain processes information. The key element of the ANN paradigm is the novel structure of the information processing system. Learning in ANN typically occurs by example through training, or exposure to a set of input/output data where the training algorithm iteratively adjusts the connection weights. The Kohonen self-organizing map neural network performs a mapping from a continuous input space to a discrete output space, preserving the topological properties of the input. For training and testing the neural network various databases available on the Internet as well as gathered information from hospitals is used.

The International Agency for Research on Cancer has reported breast cancer to be by far the most frequent cancer and the leading cause of death in women. It has been estimated that 719,100 new cases (19% of all new cancers in females) occurred worldwide in 1985. With annual totals expected to be around 340,000 deaths in 1990 and 420,000 by 2000.

Currently, breast imaging for the detection and characterization of suspicious breast lesions relies upon mammography and ultrasound. Mammography is the modality of choice for early diagnosis of breast cancer. Only 20% of currently biopsy cases actually reveal cancer. The remainders are all benign cases, which underwent a potentially unnecessary surgical procedure.

Preventing benign biopsies is the most important way to improve the efficacy of mammography screening, especially as screening becomes more widespread. This clearly demonstrates a need for efficient breast cancer diagnosis techniques. Some of the works done in this direction include linear programming approach (Mangasarian, 1995), machine-learning approach (Wolberg, 1994).

It is proposed to develop an efficient neural network models to provide accurate diagnosis while being completely noninvasive. These models will utilize existing, available information such as mammography findings and patient history data. Artificial Neural Network (ANN) is collections of mathematical models that emulate some of the observed properties of biological nervous systems and draw on the

analogies of adaptive biological learning. The key element of ANN paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements that are analogous to synapses.

Computer based diagnostic system hold promising means the challenges of the clinical situations.

INTRODUCTION GENERALE

L'automatisation du diagnostic médical du cancer du sein est devenu une Nécessité absolue vu la croissance exponentielle du nombre de mammographies effectuées chaque année. En particulier, le diagnostic à l'identification des masses mammaires et leur classification suscitent actuellement un grand intérêt.

L'avancement de la technologie informatique a encouragé les chercheurs à développer des logiciels pour aider les médecins dans la prise de décision sans consulter directement les spécialistes. Le développement de logiciels exploite le potentiel de l'intelligence humaine telle que le raisonnement, la prise de décision, l'apprentissage (par l'expérience) et bien d'autres.

En effet, la complexité des données traitées et la difficulté rencontrée pour la reconnaissance des différentes pathologies nécessitent l'utilisation de plusieurs approches directement.

Dans ce travail, nous abordons quelques techniques de la classification non supervisée, sachant que dans le plus part des cas les données ne sont pas annotées.

La classification non supervisée est alors sollicitée lorsqu' on ne connaît pas les classes des données. A la fin, de processus de classification non supervisée les données doit apparaître à l'une des classes créés par la procédure de la classification.

Une étude concernant les différentes techniques de classification non supervisé est réalisée. Nous avons fait appel à l'approche Fuzzy C-Means (FCM) est la carte de Kohonen. Fuzzy C- Means est un algorithme de classification non supervisée floue. Issu de l'algorithme des C-moyennes (C-means), et il a introduit la notion d'ensemble flou dans la définition des classes.

Pour le deuxième modèle, La carte de Kohonen a effectué une compression sur les données d'entrée représentés par des vecteurs de dimension N dans un espace de dimension M ($M < N$). La propriété la plus intéressante de cette carte est la faculté de préserver les relations topologiques des vecteurs d'entrée. Il est intéressant de noter que des fonctions similaires sont présentes dans les cerveaux des êtres vivant, ce qui est un premier point en faveur de la préservation de la topologie comme un mécanisme

essentiel dans le traitement des données .les résultats obtenus ont montré l'intérêt de la 1^{ère} approche (FCM) par rapport à la carte de Kohonen.

Mots-clefs: Cancer du sein, Diagnostic Assisté par Ordinateur (DAO), Classification

Chapitre I

Le Cancer du sein

Introduction

Le cancer du sein est un problème majeur de santé publique. Il constitue, en France et dans les pays occidentaux, l'affection tumorale la plus fréquente de la femme.[1]

Parkin a estimé qu'en 1980 le cancer du sein représentait 18 % des cancers de la femme dans le monde. La même année, il était le cancer féminin le plus fréquent, suivi par le cancer du col de l'utérus, du colon et rectum, de l'estomac puis des poumons à égalité avec le cancer du corps utérin. [1]

Entre 1978 et 1982, en Europe, il y eut 135 000 nouveaux cas de cancers du sein découverts chaque année, représentant 24 % de tous les cancers féminins. Le taux d'incidence standardisé sur l'âge en Europe est de 50,3 pour la France et de 56,8 pour l'Europe. [2]

A l'intérieur de l'Europe et pour la période 1978-1982, le taux d'incidence a varié entre 39 pour 100 000 en Espagne et 76 pour 100 000 au Royaume-Uni. Dans le monde, la variation est encore plus importante. Les taux les plus bas se retrouvent en Asie avec 22 pour 100 000 habitants au Japon, et les plus élevés sont aux États-Unis avec des taux qui varient entre 87 et 90 pour 100 000 suivant les régions. [2]

Tableau 1 - Taux d'incidence annuelle brut suivant l'âge*

Age (ans)	0- 44	45- 64	> 65	Tot al
France	14. 8	138. 4	210. 4	73. 8
Europe	19. 3	156. 5	224. 5	83. 1

**pour 100 000 habitants*

L'incidence du cancer du sein augmente régulièrement avec l'âge bien que dans les pays occidentaux, cette croissance soit moins importante après 50 ans qu'avant.

L'incidence des cancers du sein augmente en moyenne de 1,5 % par an dans le monde. Cette augmentation est variable d'un pays à l'autre mais aussi, parfois, dans un même pays : ainsi aux États-Unis, l'augmentation est plus importante chez les femmes de race noire que chez celles de race blanche. [1]

Durant les années 80, l'augmentation de l'incidence s'est portée surtout sur les petites lésions : les tumeurs inférieures à 1 cm de diamètre sont passées de 9 à 36 pour 100 000 de 1982 à 1988 et les tumeurs qui font entre 1,0 et 1,9 cm sont passées, durant la même période, de 40 à 84 pour 100 000 habitants. En même temps, les tumeurs de 2,0 à 2,9 cm avaient une incidence annuelle stable et celles qui mesuraient plus de 3 cm diminuaient de 33 à 24 pour 100 000 personnes. [3]

En 1990 il y avait 19 % des femmes décès par cancer de sein et se place donc comme la cause la plus fréquente des décès féminins par cancer. [3]

II. Facteur de risque

En dépit de la multitude des enquêtes entreprises pour tenter de découvrir des facteurs de risque susceptibles de jouer un rôle dans la genèse de cette pathologie, nous verrons que de nombreux points restent à éclaircir. [1]

II.1 Facteurs hormonaux

Il est actuellement bien établi que l'âge des premières règles, de la ménopause, la parité et l'âge à la première grossesse a une influence sur l'apparition d'un cancer du sein. Les femmes ayant une vie génitale supérieure à 30 ans présentent des facteurs de risque plus importants que lorsque celle-ci est inférieure à 30 ans. Parmi les femmes ayant eu une ménopause naturelle, le risque encouru par celles ayant été ménopausées après 55 ans est le double de celui encouru par les femmes ménopausées avant 44 ans. Le nombre de grossesses est important mais le facteur majeur est l'âge à la première grossesse. Ainsi une femme ayant mené une grossesse à terme avant 19 ans voit son risque diminuer de 50 % par rapport à une nullipare. Ces observations ont conduit au développement de plusieurs hypothèses étio-pathogéniques regroupées sous le terme de

"hypothèse de la fenêtre oestrogénique". Cependant, pour l'instant, le rôle de ces modifications hormonales et, en particulier, oestrogéniques reste mystérieux. En particulier, on ignore encore si celles-ci ont un rôle étiologique propre au cancer du sein ou simplement une action promotrice. [4]

En ce qui concerne la contraception orale, un méta analyse publiée en 1990 a porté sur 28 études cas témoins. Les conclusions montraient qu'il n'existait aucune augmentation du risque de cancer du sein chez les femmes utilisatrices de contraception orale qu'elle que soit la durée. Toutefois, ce risque est apparu légèrement mais significativement augmenté (risque relatif : 1,72) chez les femmes qui avaient utilisé la contraception orale pendant au moins quatre ans avant la première grossesse à terme. [4]

II.2 Facteurs familiaux et génétiques

Il est maintenant couramment admis qu'un antécédent de cancer du sein dans une famille augmente le risque de chaque apparenté de développer un cancer du sein au cours de sa vie.

La fréquence des formes familiales de cancer du sein, leur apparition chez des patientes jeunes, les formes pluri focales ou bilatérales conduit à évoquer le caractère héréditaire de ce cancer. Certains auteurs pensent que les concentrations familiales de cancer du sein sont la conséquence d'une susceptibilité génétique à la maladie.

Récemment ont été identifiés deux gènes liés aux cancers du sein familiaux : BRCA1 (pour BReastCAncer 1) sur le chromosome 17 (pour les familles où l'on observe des cancers du sein et des cancers de l'ovaire) et BRCA2 sur le chromosome 13 (pour les familles où l'on observe des cancers du sein dans les deux sexes) ; de plus, les femmes ayant une hétérozygotie pour le gène de l'ataxie télangiectasie (ATM) seraient particulièrement prédisposées au cancer du sein. La connaissance de ces anomalies génétiques pourrait permettre, dans l'avenir, un diagnostic précoce des tumeurs chez les patientes qui en seraient porteuses. [3]

Une forme particulière de cancer du sein est connue pour être associée à la perte ou la mutation du gène suppresseur de tumeur p53 sur le chromosome 17. Cette anomalie

génétique caractérise le syndrome de Li-Fraumeni dans lequel six types différents de cancers dont le cancer du sein peuvent affecter certaines familles. Cependant, il ne faut pas considérer les formes familiales de cancer du sein comme de plus mauvais pronostic que les cas isolés. [3]

II.3 Alimentation

Une alimentation riche en graisses favoriserait l'apparition d'un cancer du sein. Des études de laboratoire ont montré que des rats ayant un régime riche en graisses avaient une incidence plus forte de tumeurs mammaires. Les études cas contrôle effectuées sur ce sujet ont montré des résultats contradictoires. Certaines ont montré une augmentation de l'incidence des cancers du sein chez les femmes ayant un régime riche en graisses et d'autres aucune différence. [4]

La consommation d'alcool est une des plus fortes associations montrées avec le cancer du sein. Une méta analyse de 4 études de cohortes a montré une forte corrélation entre le risque de cancer du sein et la consommation quotidienne d'alcool. [4]

II.4 Mastopathies bénignes

Un grand nombre d'études ont montré que les femmes ayant eu une ou plusieurs biopsies du sein pour mastopathie sont plus à risque de développer un cancer. Trois mille biopsies du sein ont été analysées en fonction de l'histologie. En tant que groupe comparé à un échantillon de la population générale, le risque de développer un cancer du sein est augmenté. Cependant, lorsque l'on analyse les résultats en fonction du caractère prolifératif ou non, seul le groupe présentant des lésions prolifératives présente un risque augmenté. [1]

En conclusion, de nombreux facteurs de risque ont été mis en évidence dans cette pathologie. Une étude effectuée par l'American Cancer Society, par l'intermédiaire d'un questionnaire, a montré que seulement 21 % des cancers du sein diagnostiqués entre 30 et 54 ans et 29 % de ceux diagnostiqués entre 55 et 84 ans pourraient être attribués à un ou plusieurs facteurs de risque. [1]

III. classification du cancer de sein

La classification des cancers du sein repose sur une étude du type histologique de ces cancers. [2]

L'O.M.S. a proposé en 1981 une classification des tumeurs malignes du sein qui a été adoptée par la plupart des équipes dans le monde. Le pathologiste a à sa disposition plusieurs types de prélèvement. Il peut s'agir d'une biopsie par "truc ut" effectuée, en général, dans les tumeurs non opérables d'emblée, ou d'une ponction cytologique. L'exérèse chirurgicale d'une tumeur comporte un examen extemporané sur tissu frais qui permet un diagnostic immédiat dans 99 % des cas. La pièce opératoire est par la suite fixée et examinée. [2]

Tableau 2 - Classification histologique des tumeurs du sein (OMS, 1981)

<i>Tumeurs épithéliales</i>	<ul style="list-style-type: none">- Papillome intracanalair- Adénome du mamelon- Adénome tubuleux- Adénome lactant- Autres
<i>* Bénignes</i>	<ul style="list-style-type: none">. carcinome intracanalair. carcinome lobulaire in situ
<i>* Malignes</i>	<ul style="list-style-type: none">- Infiltrantes. carcinome canalaire infiltrant. carcinome canalaire infiltrant avec composante intracanalair prédominante. carcinome lobulaire infiltrant

	<ul style="list-style-type: none"> . carcinome médullaire . carcinome papillaire . carcinome tubuleux . carcinome sécrétant (juvénile) . carcinome apocrine . carcinome métaplasique . autre - <i>Maladie de Paget du mamelon</i>
<i>Tumeurs épithéliales et conjonctives mixtes</i>	<ul style="list-style-type: none"> - Fibroadénome - Tumeur phyllode - Carcinosarcome
<i>Tumeurs diverses</i>	<ul style="list-style-type: none"> - Tumeur des parties molles - Tumeurs cutanées - Tumeur des tissus hématopoïétiques et lymphoïdes

VI. Diagnostic des cancers du sein

A- DIAGNOSTIC CLINIQUE

L'étape diagnostique est essentielle puisque la précocité du diagnostic améliore le pronostic de ces cancers. L'examen clinique doit être mené de façon rigoureuse. La palpation des seins devrait être intégrée dans tout examen clinique réalisé par le praticien puisque aucun matériel n'est exigé.

a) Interrogatoire

Dans près de 90 % des cas, c'est la patiente elle-même qui consulte pour l'apparition d'un nodule mammaire découvert de façon fortuite (toilette, traumatisme...). Plus rares sont les autres signes d'appel : douleur, rougeur, rétraction du mamelon, écoulement mamelonner, adénopathie axillaire. [4]

Il est essentiel de rechercher lors de l'interrogatoire les antécédents familiaux de cancer et, plus particulièrement, du sein. Il sera nécessaire de noter les antécédents gynécologiques, les dates des premières et dernières règles, la chronologie des grossesses et des allaitements, les traitements hormonaux et la contraception. Il faudra également préciser le contexte médical psychologique (consommation d'alcool). [4]

b) Examen clinique

Il se fait en position assise et couchée. Il comporte l'examen des seins quadrant par quadrant ainsi que celui des aires ganglionnaires axillaires et sus claviculaires. [4]

La tumeur maligne du sein est caractérisée par un nodule dur, irrégulier, indolore. Il peut exister une rétraction de la peau en regard, ou de la plaque aréole- mamelonner. L'adduction contrariée permet de rechercher la fixation de la tumeur sur le pectoral contracté. Les adénopathies, lorsqu'elles existent, intéressent le plus fréquemment le groupe inférieur du creux axillaire homolatéral. [4]

Certains carcinomes mammaires, en particulier les cancers colloïde et médullaire, ont toutes les apparences de la bénignité : tumeur ronde, régulière, mobile, sans modification des téguments. Certaines tumeurs ont une topographie atypique : sus ou sous mammaire, para mammaire interne ou externe, prolongement axillaire. La maladie de Paget du mamelon est typiquement une lésion eczématiserforme, prurigineuse, rouge, suintante avec érosion du mamelon. Cette lésion implique l'existence d'un cancer sous-jacent parfois très à distance du mamelon. [4]

Un problème particulier est posé par le cancer à évolution rapide. Il est défini par une tumeur d'apparition récente dont le temps de doublement est inférieur à 6 mois. Environ

10 % des cancers présentent des signes inflammatoires : tumeur chaude, douloureuse, érythème cutané en regard et accentuation des pores cutanés (peau d'orange). Dans 3 % des cas, l'ensemble du sein est inflammatoire : il s'agit d'une mastite carcinomateuse. [4]

Chez la femme âgée qui souvent se néglige, une tumeur du sein peut se présenter sous une forme très évoluée envahissant les plans musculaires et cutanés et accompagnée d'une rétraction cutanée réalisant dans la forme le plus typique, le squirrhe.

c) Diagnostic différentiel

Certaines affections bénignes peuvent simuler un cancer du sein mais la fréquence des tumeurs malignes doit inciter à avoir recours à l'histologie au moindre doute. [2]

** Diagnostic différentiel d'une tumeur mammaire*

. Placard de mastose parfois de contours imprécis;

. adénofibrome souvent ferme et lobulé;

. Tumeur phyllode;

. Granulome inflammatoire chronique : l'origine la plus fréquente est l'ectasie canalaire sécrétante. Il s'agit d'une tumeur riche en plasmocytes;

. La cytotéatonécrose post-traumatique : il s'agit d'un granulome lipophagique sur stéatonécrose d'origine traumatique. [2]

** Diagnostic d'un sein inflammatoire*

. les mastites aiguës bénignes : le plus souvent d'origine infectieuse : mastite puerpérale, surinfection de kyste, de galactocèle, d'ectasie canalaire sécrétante. [1]

** Diagnostic d'une rétraction cutanée*

. la maladie de Mondor : phlébite superficielle de la paroi thoracique qui se présente, lorsque le cordon phlébitique n'est plus perçu, sous la forme d'une dépression cutanée linéaire. [1]

. L'ombilication congénitale du mamelon souvent bilatérale. Le mamelon est rétracté plat ou fendu horizontalement. [1]

Diagnostic d'un écoulement mamelonner

. L'ectasie canalaire sécrétante : écoulement uni-orificiel spontané et récidivant parfois hématique;

. Les papillomes intracanaux.

B- EXAMENS COMPLÉMENTAIRES

Le diagnostic sénologique est actuellement basé sur un trépied diagnostique associant examen clinique, mammographie et cytologie. L'échographie a, actuellement, une place privilégiée dans cet arsenal. D'autres techniques (Doppler couleur, IRM) doivent trouver leur place dans la stratégie diagnostique. [4]

a) Mammographie

Elle doit toujours être bilatérale. Le sein est exploré par trois incidences fondamentales : de face ou crânio-podale, de profil, et en oblique dégageant le quadrant supéro-externe et le prolongement axillaire. En fonction des besoins cliniques, des incidences complémentaires pourront être centrées sur une masse. Il n'est pas possible de définir radiologiquement un sein normal. Le rapport des composantes glandulaires et adipeuses varie au cours de l'existence et, en particulier, en fonction de l'imprégnation hormonale de la femme. La lecture comparative des clichés des deux seins est donc indispensable. [2]

Signes classiques du cancer du sein :

- Les micro calcifications : elles permettent, le plus souvent, de détecter des petits cancers du sein. Certains critères sont en faveur de la malignité : punctiformes, irrégulières ou vermiculaires, en amas de plus de 30 éléments.

- Les nodules à contours flous : il s'agit de petites opacités inférieures à 10 mm, à contours spiculaires mal définis. [3]

- Les distorsions de l'architecture glandulaire : ce sont des désorganisations localisées de l'architecture du sein sous forme d'images linéaires divergentes sans opacité tumorale identifiable.

- La multicentricité des lésions. [3]

Tous stades confondus, on peut estimer que le diagnostic mammographique isolé est exact dans 85 à 90 % des cas. Certaines zones restent difficiles à explorer : la partie haute et interne du quadrant supéro-interne, le sillon sous-mammaire, la partie externe du quadrant inféro-externe, la partie postérieure de la glande mammaire. [3]

b) Échographie

L'échographie est une exploration complémentaire souvent indispensable en sénologie. Elle complète les renseignements fournis par la mammographie. C'est dans les zones de densité accrue que l'échographie apporte le plus de renseignements. L'expression échographique typique d'un cancer du sein est celle d'une lésion solide, hypoéchogène, plus ou moins hétérogène, à contours irréguliers et dont le grand axe est, en général, vertical par rapport à la peau. Une ombre postérieure est présente dans environ un tiers des cancers. Cependant, l'échographie dépend beaucoup de l'opérateur, elle ne visualise pas les foyers isolés de microcalcifications, elle est de spécificité médiocre dans les seins adipeux. Elle permet souvent de réaliser une ponction échoguidée. [3]

c) Doppler couleur

L'échotomographie doppler à codage couleur allie la rapidité du doppler couleur à la haute précision du doppler pulsé. Actuellement, cette méthode est toujours en phase d'évaluation. Les résultats obtenus sont très dépendants de l'appareil utilisé. Le doppler pulsé permet de calculer des index de résistance qui sont théoriquement effondrés en pathologie maligne. Le doppler couleur permet de détecter les vaisseaux. Leur présence au centre de la tumeur est évocatrice de malignité. [3]

d) Tomodensitométrie

Par son absence totale de spécificité, elle ne présente aucune indication pour le diagnostic de la lésion elle-même. Son rôle est limité aux rares cas d'atteinte pariétale lorsque la lésion se situe dans le sillon sous mammaire. Elle est intéressante pour les radiothérapeutes qui peuvent réaliser des dosimétries prévisionnelles (scanner hélicoïdal et reconstruction 3D à l'étude). [4]

e) IRM

Les micros calcifications sont invisibles. La lésion elle-même est souvent de contours irréguliers à limites imprécises, polycycliques, spicules. L'IRM, en particulier après injection de gadolinium, a certainement une place actuellement dans l'étude des récidives et l'évaluation du suivi des plasties mammaires. [4]

f) Galactotrophie

C'est l'opacification du galactophore par un produit iodé. Elle est indiquée devant tout écoulement uni-orificiel spontané ou provoqué. L'intérêt essentiel de la galactographie est le repérage pré chirurgical du secteur drainé par le galactophore dans sa topographie et son étendue. [4]

C- CYTOLOGIE MAMMAIRE

L'examen histologique ou cytologique est indispensable pour confirmer le diagnostic. La ponction aspiration d'une tumeur mammaire peut être réalisée de façon simple et rapide. La fiabilité globale varie de 77 à 93 %. Le taux de faux positifs varie entre 3 et 12 % et celui de faux négatifs de 2 à 10 %. Une adénopathie axillaire peut également être ponctionnée. De même un écoulement mammaire doit être prélevé et étalé sur lame.

Actuellement, le plus souvent, on a recours à une biopsie par "trucut". Elle permet une étude histologique de la tumeur avant chirurgie, en particulier la tumeur peut être gradée selon la classification de Scarff et Bloom de même qu'il peut être réalisé un dosage des récepteurs hormonaux. [4]

Si la tumeur est infra clinique, il est nécessaire d'avoir recours soit à une ponction ou une micro biopsie guidée par échographie ou après repérage mammographie par stéréotaxie. De même, ces deux méthodes permettent un repérage préopératoire de la tumeur par mise en place d'un harpon au niveau du site tumoral. [4]

D- DÉPISTAGE

Il est classique de considérer le cancer du sein comme une maladie générale et ce dès le diagnostic. Ceci est dû à la fréquence et à la précocité des micro métastases. On estime que la moitié des patientes est porteuse de métastases occultes au moment où le diagnostic est établi. On conçoit donc l'importance du gain de temps obtenu par le dépistage précoce des tumeurs du sein. On entend par dépistage l'identification, par un test de réalisation facile et peu coûteuse, de sujets atteints d'une affection donnée et qui sont apparemment en bonne santé. Le dépistage n'a pas pour but de poser un diagnostic. Il se différencie en cela du diagnostic précoce. Les sujets qui présentent un test de dépistage positif doivent subir des examens à visée diagnostique. [1]

Les études qui se sont attachées à décrire l'histoire naturelle du cancer du sein montrent que pour chaque tumeur il existe une taille critique à partir de laquelle survient la dissémination métastatique. Si la patiente est traitée avant cette taille tumorale critique, elle n'aura pas de métastase ; en revanche, si le traitement est débuté au-delà de cette taille critique, des métastases se déclareront cliniquement après un délai plus ou moins long. Les calculs mathématiques montrent qu'une avance de douze mois au moment du traitement diminue de 30 % le nombre de patientes qui auront des métastases. [1]

a) Moyens du dépistage

Ils se résument à deux méthodes cliniques, l'auto-examen des seins et l'examen médical, et à un seul moyen para clinique, la mammographie.

** Auto examen des seins (AES)*

Cette méthode consiste à apprendre aux femmes l'auto-inspection et l'autopalpation des seins et, ce, de manière mensuelle. Les femmes apprennent ainsi à connaître parfaitement leurs seins afin d'y détecter ultérieurement une anomalie. Cette méthode

est simple, peu coûteuse et traumatique. Cependant, un sondage de la SOFRES montre que si 86 % des femmes sont conscientes de l'importance de cet examen, 47 % le font une fois par an, et seulement 10 à 20 % des femmes le pratiquent de façon régulière. L'AES entraîne une réduction de 10 % du taux des cancers évolués, une augmentation de 20 % du taux de petits cancers et une réduction de 7 à 15 % de l'envahissement ganglionnaire sur des séries rétrospectives. Cependant, il n'existe actuellement pas de publication qui montre que l'AES ait amélioré la détection du cancer du sein à un stade pré métastatique.

** Examen médical*

Son efficacité sur la réduction de la mortalité par cancer du sein est de l'ordre de 25 %. Cet examen peut s'intégrer dans le cadre d'une visite médicale annuelle. Cependant l'examen médical des seins est souvent mal accepté lors d'une consultation pour un autre motif. L'acceptabilité de l'examen médical dépend, en fait, essentiellement de la valeur de l'information donnée par le médecin.

** La mammographie*

L'European Group for Breast Cancer Screening, qui rassemble la plupart des équipes européennes qui travaillent sur le dépistage du cancer du sein, a émis les recommandations suivantes :

Les essais contrôlés ont montré que la mortalité par cancer du sein peut être réduite grâce à un dépistage mammographie parmi les femmes âgées de plus de 50 ans. Des programmes de dépistage doivent donc être offerts aux femmes de 50 ans et plus.

Actuellement, la mammographie est le test de dépistage de choix. Deux clichés sont préférables pour le premier examen de dépistage. Les données actuelles indiquent qu'un intervalle de deux à trois ans est raisonnable. L'avantage d'un dépistage systématique chez les femmes de moins de 50 ans n'a pas été démontré de façon convaincante. Des recherches concernant le dépistage chez les femmes âgées de 40 à 50 ans doivent être poursuivies.

La mammographie est actuellement la seule méthode efficace de dépistage du cancer du sein. Sa sensibilité et sa spécificité sont supérieures à 90 %. Les essais randomisés ont montré que la mortalité par cancer du sein peut être réduite par le dépistage mammographie chez les femmes de plus de 50 ans avec un intervalle entre les vagues de

deux à trois ans. Cet effet favorable de réduction de la mortalité peut atteindre 40 %. Le risque de cancer du sein radio induit par la mammographie a fait, au cours des dernières années, l'objet d'un débat souvent passionné et parfois anxiogène pour le grand public. Un contrôle de qualité très strict, en particulier de dosimétrie, permet de vérifier que la dose administrée est minime et que le risque lié à l'irradiation est négligeable par rapport aux taux d'incidence de ce cancer dans la population générale. Cette dose est de l'ordre de 0,17 rad par cliché si bien que le risque de cancer induit pour les femmes âgées de plus de 35 ans serait d'environ 3,5 pour 1 million de femmes par an et par rad.

Les cancers du sein familiaux posent un problème particulier. Ces cancers sont assez peu nombreux (de 5 à 10 % des cancers du sein) mais le risque de cancer parmi les femmes appartenant à ces familles est élevé, voisin de 40 % et ces cancers touchent souvent des femmes jeunes. Ces femmes sont actuellement définies par l'existence de trois cancers en ligne directe ou de deux cancers dont un bilatéral. Pour ces femmes, des modalités particulières de dépistage sont à l'étude.

V. Bilan pré thérapeutique

Au cours de cette étape, l'équipe thérapeutique doit recueillir tous les éléments cliniques et para cliniques qui caractérisent le cancer et son hôte. [1]

L'examen clinique de la tumeur permet de déterminer sa taille, son siège (quadrant), la présence d'adénopathies axillaires ou sus claviculaires et enfin la présence de signes inflammatoires. L'interrogatoire permet de déterminer l'intervalle de temps séparant la date de découverte de la tumeur du moment où est entrepris le traitement. Le temps de doublement d'un adénocarcinome mammaire étant d'environ 3 mois, pour atteindre une taille cliniquement décelable (environ 1 cm), il faut environ 10^9 cellules soit environ 7 ans. [1]

Sur le plan biologique, hormis un bilan dit "standard" (numération - formule plaquettes, créatinine, coagulation) il est préférable de doser le marqueur tumoral CA 15.3 qui peut donner une indication, s'il est élevé, sur la présence d'éventuelles métastases infra cliniques ou radiologiques.

Les métastases des cancers du sein intéressent essentiellement le squelette, les poumons, la plèvre, le foie, les ganglions périphériques et le revêtement cutané. Ces différents sites doivent être recherchés par l'examen clinique, 8 à 12 % des cancers du sein traités en France actuellement sont déjà métastatiques au moment du diagnostic.

Le bilan de l'extension à distance doit comporter une radiographie pulmonaire, une échographie hépatique et une scintigraphie osseuse (pour les tumeurs de plus de 1 cm) avec, éventuellement, radiographies centrées sur les zones hyperfixantes. [1]

L'examen clinique et la recherche des métastases doit aboutir à classer la tumeur selon la classification TNM (tableau 6).

Tableau 6 - Classification TNM du cancer du sein

<i>TUMEUR T</i>	
<i>Tis</i>	<i>in situ ou maladie de Paget <= 2 cm</i>
<i>T1</i>	<i>> 2 et <= 5 cm</i>
<i>T2</i>	<i>>5 cm</i>
<i>T3</i>	<i>a : sans fixation au fascia pectoral sous-jacent et/ou au muscle b : avec fixation au fascia pectoral sous-jacent et/ou au muscle</i>
<i>T4</i>	<i>tumeur de toute taille avec extension directe à la paroi thoracique ou à la peau a : fixation à la paroi thoracique b : extension cutanée</i>
<i>ADÉNOPATHIE</i>	
<i>N</i>	
<i>N0</i>	<i>pas de ganglion axillaire</i>
<i>N1</i>	<i>ganglions axillaires homo latéraux mobiles</i>

N2	<i>a : considérés comme non envahis</i> <i>b : considérés comme envahis</i>
N3	<i>ganglions axillaires homo latéraux fixés entre eux ou à d'autres organes</i> <i>ganglions mammaires internes</i>
MÉTASTASE M	
M0	<i>absence de métastase</i>
M1	<i>présence de métastase</i>

IV. Méthodes thérapeutiques

Le traitement du cancer du sein a deux buts : d'une part contrôler la tumeur primitive et les territoires ganglionnaires de drainage, c'est le traitement locorégional, d'autre part traiter une éventuelle dissémination infra clinique à distance, c'est le traitement adjuvant.

A- CHIRURGIE

La chirurgie d'exérèse a longtemps représenté l'unique recours thérapeutique pour les femmes atteintes d'un cancer du sein. Une meilleure appréciation de ses limites, de ses objectifs a conduit à mieux codifier ses indications et ses modalités radicales.

L'intervention de Hals Ted qui comportait l'ablation totale du sein et des muscles pectoraux est actuellement abandonnée. Lorsque l'ablation totale du sein est nécessaire, l'intervention de Patey ou mastectomie radicale modifiée reste la référence. Il s'agit d'une ablation de la glande mammaire par une incision horizontale accompagnée d'un curage ganglionnaire axillaire. Le curage ganglionnaire axillaire doit comporter au moins 9 ganglions. Pour des tumeurs de moins de 3 cm, il est prouvé actuellement que l'on peut réaliser une mastectomie partielle qui enlève la tumeur mammaire associée à

un curage ganglionnaire axillaire qui évite une irradiation de l'aisselle. Toute chirurgie partielle de sein doit être suivie d'une radiothérapie de la glande. [2]

B- RADIOTHÉRAPIE

La radiothérapie tient une place importante dans le traitement locorégional des cancers infiltrant du sein, soit associée à la chirurgie à titre pré ou post-opératoire, soit seule. Elle est indispensable après traitement chirurgical conservateur. L'irradiation de base englobe la totalité du sein. Les doses sont de 55 Gy étalées sur 5 à 6 semaines à raison de 5 séances par semaine. Cette dose permet de contrôler la maladie infra clinique dans 90 % des cas. [2]

Après tumorectomie, un surdosage de 10 Gy est en général délivré sur le lit tumoral. S'il existe un envahissement ganglionnaire, il sera réalisé également une irradiation du creux sus claviculaire. [2]

En cas de tumeur primitive interne, la chaîne ganglionnaire mammaire interne sera irradiée. [2]

C- CHIMIOTHÉRAPIE

Le cancer du sein fut très vite reconnu comme une des localisations néoplasiques les plus sensibles aux produits cytotoxiques. D'abord utilisée dans les formes localement avancées et métastatiques, la chimiothérapie a vu par la suite ses indications s'élargir au traitement adjuvant. Récemment, la méta analyse de Peto a montré qu'elle améliorerait la survie globale chez certains groupes de patientes. Les produits utilisés les plus fréquemment sont les anthracyclines (Doxorubicine, Epirubicine), le cyclophosphamide, le 5 fluoro-uracile et, plus récemment, la vinorelbine et les Taxanes (tableau 7).

Tableau 7 - Protocoles de chimiothérapie couramment utilisés dans le cancer du sein

<i>Schéma</i>	<i>Composition</i>	<i>Doses</i>	<i>Voie</i>	<i>Administration</i>
<i>FAC</i>	<i>Adriamycine</i> <i>Cyclophosphamide</i> <i>Fluoro-uracile</i>	<i>50 mg/m²</i> <i>500</i> <i>mg/m²</i> <i>500</i> <i>mg/m²</i>	<i>IV</i> <i>IV</i> <i>IV</i>	<i>J1/3 semaines</i> <i>J1/3 semaines</i> <i>J1/3 semaines</i>
<i>CMF</i>	<i>Cyclophosphamide</i> <i>Méthotrexate</i> <i>Fluorouracile</i>	<i>600</i> <i>mg/m²</i> <i>40 mg/m²</i> <i>600</i> <i>mg/m²</i>	<i>IV</i> <i>IV</i> <i>IV</i>	<i>J1 et J8/4</i> <i>semaines</i> <i>J1 et J8/4</i> <i>semaines</i> <i>J1 et J8/4</i> <i>semaines</i>
<i>AC</i>	<i>Adriamycine</i> <i>Cyclophosphamide</i>	<i>50 mg/m²</i> <i>500</i> <i>mg/m²</i>	<i>IV</i> <i>IV</i>	<i>J1/3 semaines</i> <i>J1/3 semaines</i>
<i>VMM</i>	<i>Téniposide</i> <i>(VM26)</i> <i>Méthotrexate</i> <i>Mitomycine C</i>	<i>100</i> <i>mg/m²</i> <i>25</i> <i>mg/m²/j</i> <i>10 mg/m²</i>	<i>IV</i> <i>IV</i> <i>IV</i>	<i>J1/3 semaines</i> <i>J2 et J3/3</i> <i>semaines</i> <i>J2/3 semaines</i>

Ces produits sont utilisés en association de 2 ou plus souvent 3 drogues. Les cycles sont répétés tous les 21 jours. En adjuvant, 6 cycles au total sont délivrés. En phase métastatique, la durée de la chimiothérapie varie bien évidemment en fonction de la

réponse. Néanmoins, il est communément admis qu'il n'est pas nécessaire de poursuivre une chimiothérapie au-delà de 8 cycles.

D- HORMONOTHÉRAPIE

La première hormonothérapie est apparue en 1896, date à laquelle, Beatson a proposé de réaliser une ovariectomie à des patientes présentant un cancer du sein métastatique. Grâce à la découverte des récepteurs hormonaux en 1960, les moyens et les indications de l'hormonothérapie se sont affinés. Plusieurs types de traitement sont actuellement disponibles pour le clinicien. [1]

a) Hormonothérapie soustractive

* La castration :

Longtemps représentée par la castration chirurgicale ou radiothérapie, elle est actuellement le plus souvent réalisée grâce aux analogues de la LH-RH (Zoladex*, Décapeptyl*, Enantone*). Les résultats de ces différentes méthodes sont semblables : 35 à 40 % de réponses objectives en phase métastatique chez les femmes pré ménopausées. [1]

b) Hormonothérapie additive

* Le tamoxifène

Le tamoxifène (Nolvadex*, Tamofène*, Oncotam*, Kessar*) est la molécule le plus couramment utilisé en matière de cancer du sein que ce soit en phase adjuvante ou métastatique. Il s'agit d'un anti-oestrogène qui agit en se fixant aux récepteurs cytosoliques des oestrogènes de la cellule tumorale. Il est actuellement utilisé à la posologie de 20 mg/j en une ou deux prises. Le tamoxifène est utilisé généralement chez la femme ménopausée. [1]

Le tamoxifène est un produit remarquablement bien toléré. Le pourcentage de patientes interrompant le traitement pour effets secondaires (bouffées vasomotrices, métrorragies, nausées) est de l'ordre de 3 %.[1]

** Les progestatifs :*

Ils sont représentés essentiellement par l'acétate de médroxyprogestérone (Farlutal, Prodason*) et l'acétate de mégestrol (Megace*). Le taux de réponse objective varie entre 30 et 50 %. Les effets secondaires observés sont généralement une rétention hydrosodée avec prise de poids. Le risque de thrombophlébite doit être considéré chez la femme à risque. [2]*

** Les antiaromatases :*

Une seule molécule était autrefois disponible : l'aminogluthétimide (Orimétène). Il est préférable de l'associer systématiquement à l'hydrocortisone compte-tenu de l'effet anticorticosurrénalien. Il n'est actif que chez la femme ménopausée. Il agit en inhibant l'aromatisation périphérique des androgènes en oestrogènes. Les effets secondaires surviennent souvent en début de traitement. Il sont, le plus souvent, d'ordre neurologique (sommolence, vertiges) ou à type de rashs cutanés et de nausées. Actuellement est disponible également, le Formestane (Lentaron*), antiaromatase de 2^e génération. Chez la femme ménopausée, le Formestane par voie IM aux doses de 250 mg tous les 15 jours réduit les oestrogènes circulants de 60 %. L'adjonction de corticoïdes n'est plus nécessaire. Un certain nombre de nouveaux antiaromatases sont à l'étude. [2]*

IV. Qualité de vie et cancer du sein

L'utilisation de thérapeutiques locales mais agressives, l'étude des séquelles tardives et de la réinsertion des malades témoignent d'une réflexion sur la qualité de vie. [4]

Après chirurgie du sein, les deux principales séquelles sont la limitation du jeu articulaire de l'épaule et le lymphoedème. Les séquelles articulaires peuvent bénéficier d'une rééducation de l'épaule. Le lymphoedème est prévenu par des conseils locaux (limiter les injections, les mouvements de force et les traumatismes locaux). La limitation des indications d'irradiation du creux axillaire a entraîné une diminution de l'incidence des "gros bras". Lorsqu'il apparaît, le lymphoedème peut être combattu par des traitements à base de coumarine (Lysedem) ou par des moyens locaux mécaniques (massages, drainages lymphatiques, pressothérapie). Le traitement est d'autant plus efficace qu'il est entrepris tôt avant l'apparition d'une fibrose locale. Les séquelles*

radiothérapies neurologiques et articulaires sont devenues exceptionnelles. Les séquelles pulmonaires peuvent survenir lorsqu'une grande partie du poumon était incluse dans les champs d'irradiation. [4]

Les complications immédiates de la chimiothérapie sont maintenant mieux contrôlées notamment les vomissements grâce aux antagonistes des récepteurs à la sérotonine. L'alopécie reste un problème bien que certaines drogues actuellement utilisées soient très peu aloplasties (Mitoxantrone, Vinorelbine). Sur le plan de l'hormonothérapie, les analogues de la LH-RH est beaucoup mieux acceptés que la castration chirurgicale ou radiothérapie. [4]

A côté des complications propres des traitements, il est évident que la chirurgie mammaire altère l'image corporelle de la femme. Actuellement, les indications de tumorectomie ou de mastectomie partielle représentent la majeure partie des interventions mammaires pour cancer. La reconstruction mammaire apporte une réponse aux patientes perturbées par la perte du sein. Les différentes techniques de reconstruction font appel soit à la mise en place de prothèses internes, soit à des reconstructions par lambeaux (grand dorsal, TRAM).

Cependant, les études de qualité de vie montrent que 20 à 30 % des femmes ont une qualité de vie altérée à long terme. Les patientes atteintes de cancer du sein présentent cependant relativement peu de problèmes sexuels (12 % de plus que les témoins). D'une façon générale, les études montrent une amélioration des troubles psychologiques avec le temps. [4]

La qualité de vie passe également par la possibilité d'une réinsertion professionnelle. Dans une enquête récente, sur 160 patientes, la durée moyenne d'arrêt de travail était de 9 mois. 83 % des femmes ont pu reprendre leur travail dont 1/4 avec une réduction d'horaires. Les problèmes de fertilité et de grossesse possible après cancer du sein sont fréquemment posés chez les femmes jeunes. Des enquêtes épidémiologiques montrent qu'une grossesse n'altère pas l'évolution du cancer du sein stabilisé. Sur le plan du traitement substitutif de la ménopause, il n'est actuellement pas recommandé après traitement d'un cancer du sein. [4]

Les progrès des connaissances réalisés ces dernières années dans le cancer du sein ont permis d'étendre les indications de traitement conservateur, de diminuer les indications

d'irradiation, de développer une hormonothérapie avec peu de séquelles et de limiter les effets secondaires de la chimiothérapie. Dans l'attente de traitements plus efficaces à la phase métastatique de ce cancer, il est indispensable de développer le dépistage et les traitements adjuvants afin de faire reculer la mortalité engendrée par cette pathologie.

[4]

Chapitre II

Principe de La Classification

Abstrait

L'avancement de la technologie informatique a encouragé les chercheurs à développer des logiciels pour aider les médecins dans la prise de décision sans consulter directement les spécialistes. Le développement de logiciels exploite le potentiel de l'intelligence humaine telle que le raisonnement, la prise de décision, l'apprentissage (par l'expérience) et bien d'autres. L'intelligence artificielle n'est pas un concept nouveau, mais il a été accepté comme une nouvelle technologie en informatique. Elle a été appliquée dans de nombreux domaines tels que l'éducation, des affaires, médical et de la fabrication.

I. system d'aide au diagnostic médical

Dans la plupart des pays en développement l'insuffisance des médecins spécialistes a augmenté le taux de mortalité ce qui oblige toute fois Les établissements d'enseignement supérieur de prendre une action immédiate afin de former autant des médecins que possible. Toutefois, en attendant les élèves à devenir des médecins et des médecins à devenir des spécialistes, de nombreux patients peut-être déjà meurent.

La pratique actuelle pour le traitement médical requis aux patients de consulter le spécialiste pour un diagnostic. Autre médecin ne peut pas avoir suffisamment d'expertise ou d'expérience pour faire face à certaines maladies à haut risque.

Cependant, le temps d'attente pour les traitements prend normalement quelques jours, des semaines voire des mois, En ce moment la maladie peut être avoir déjà étalé.

Comme la plupart de la maladie à haut risque ne pouvait être guérie que dans un stade précoce.

La technologie informatique peut être utilisé pour réduire le nombre de mortalité et de réduire le temps d'attente pour voir le spécialiste. Un Programme informatique ou un logiciel développé par l'émulation l'intelligence humaine pourrait être utilisé pour aider les médecins dans la prise de décision sans consulter directement les spécialistes. Le logiciel n'a pas été conçu pour remplacer le spécialiste ou un médecin, mais il a été élaboré pour aider les praticiens généralistes et spécialistes dans le diagnostic et la prévision de l'état du patient à partir de certaines règles ou "l'expérience". Un Patient avec facteurs de risque élevé ou des symptômes ou prévus pour être très affecté à

certaines maladies ou de maladie, pourrait être présélectionné pour voir le spécialiste pour un traitement ultérieur.

Employer la technologie en particulier Intelligence Artificielle (IA) dans les applications médicales pourrait réduire le coût, le temps, l'expertise humaine et l'erreur médicale.

Un Programme informatique connu sous le nom médical aide à la décision du système a été conçu pour aider les professionnels de la santé prendre la décision clinique.

Le système traite les données médicales et domaine de connaissance dans le diagnostic de l'état des patients, ainsi que recommander des traitements appropriés pour les patients en particulier.

La technologie informatique aide aussi à réduire le coût et le temps pendant le processus d'inscription. Fréquentation des hôpitaux pourrait simplement la clé dans l'identité du patient et mise à jour du dossier du patient. le médecin peut se référer au dossier du patient pour voir sans historique médicale.

Les avantages des documents électroniques seraient nombreux, à savoir améliorer les fichiers traditionnels, le stockage et la récupération rapide, de et encourager la recherche dans les applications médicales.

II. Intelligence Artificielle en Médecine

Intelligence Artificielle (AI) est une étude visant à émuler l'intelligence humaine dans la technologie informatique.

*Les applications techniques de l'Intelligence Artificielle (AI) en médecine sans:
L'organisation, Produit de nouveaux outils pour soutenir la prise de décision médicale, de formation et de recherche. Intègre les activités dans le domaine médical, informatique, sciences cognitives et d'autres.*

La représentation et le catalogage des connaissances médicales Fournit un laboratoire pour l'examen et Offre une discipline riche en contenu pour l'avenir de spécialité médicale scientifique.

Beaucoup système intelligent ont été développés dans le but de renforcer la santé et de fournir une meilleure installation de soins de santé, réduire les coûts et etc.

L'Intelligent artificiel a été développé dans le domaine médicale pour aider les utilisateurs (en particulier les médecins et les patients) et de fournir un diagnostic précoce et de prévision afin de prévenir une maladie grave. Même si le système est équipé d'«humain» des connaissances, le système ne remplacera jamais l'expertise humaine en tant qu'êtres humains sont tenus de surveiller fréquemment et mettre à jour les connaissances du système. Par conséquent, le rôle de médecin spécialiste et les médecins (ou praticienne médicale) est important pour assurer la validité du système. Les premières études dans le système médical intelligente telles que MYCIN, CASNET, PIP et interniste-je.

MYCIN a été développé dans les années 1970 pour diagnostiquer certaines infections antimicrobiennes et recommande un traitement médicamenteux. Il dispose de plusieurs installations telles que les installations explication, les installations d'acquisition de connaissances, les installations d'enseignement et du système de construction des installations.

CASNET (causalité réseaux associatifs) a été développé dans les années 1960 au début est un outil général pour la construction de système expert pour le diagnostic et le traitement des maladies.

PIP une abréviation pour le programme de la maladie actuelle a été développée dans les années 1970 il à simulé le comportement d'un néphrologue expert en prenant l'histoire de la maladie actuelle d'un patient avec néphropathie sous-jacente.

Le travail sur interniste-I en début dans les années 1982a été concentrée sur l'étude des méthodes heuristiques pour imposer différentielles structures de travail de diagnostic sur la prise de décision clinique. Il a été appliqué dans les diagnostics de la médecine interne.

Dans les années 1990, les études en système intelligent a été amélioré pour utiliser le système en fonction des besoins actuels.

L'exploration de données est une technique AI pour la découverte de connaissances dans les bases de données de grande taille, pourrait être utilisé pour recueillir des informations cachées à des fins médicales (SitiNurulHuda et Miswan, 1999; Siti Fatma

et Rogayah, 1999;..Neves et al, 1999). Il pourrait également être combinés avec un réseau de neurones pour la classification du motif floue du VIH et du sida grâce à l'apprentissage non supervisé (SitiNurulHuda et Miswan, 1999).

La logique floue est une autre branche de techniques d'intelligence artificielle. Il traite de l'incertitude dans la connaissance qui simule le raisonnement humain dans les données incomplètes ou imprécises. Meng (1996) appliquée inférence floue relationnelle dans le diagnostic médical. Il a été utilisé au sein de la médecine fondée sur la connaissance du système, qui est dénommé Clinaid. Il traite de recommandations sur l'activité de diagnostic de traitement et de l'administration du patient.

Neural Network (NN) est l'un des techniques d'intelligence artificielle puissant qui a la capacité d'apprendre un ensemble de données et de matrices de poids des constructions pour représenter les modes d'apprentissage. NN est un réseau de plusieurs processeurs simples ou des unités (Sarle, 1999). Il simule la fonction du cerveau humain pour effectue les tâches.

le NN a été employé dans diverses applications médicales telles que l'artère coronaire (Lippmann, 1995), l'infarctus du myocarde (Heden et al, 1996.), Le cancer (Street et al, 1996;..Karkanis et al, 1999), la pneumonie (Caruana et al., 1996) et les troubles du cerveau (Pranckeviciene, 1999). En Karkanis et al (1999) NN a été exécuté comme un hybride avec la méthode description textuelle pour détecter des anomalies dans les mêmes images avec une grande précision.

La prévisibilité de l'indicateur de l'avenir des valeurs fondées sur des données passées et de la reconnaissance de tendance Automatisé en temps réel l'analyse et le diagnostic Permet une identification rapide et la classification des données d'entrée.

Élimine l'erreur associée à la fatigue humaine et de l'accoutumance.

III. les deux types de Classification

La classification automatique consiste à regrouper divers objets (les individus) en sous-ensembles d'objets (les classes). Elle peut être :

Supervisée : les classes sont connues à priori, elles ont en général une sémantique associée

Non supervisée : les classes sont fondées sur la structure des objets, la sémantique associée aux classes est plus difficile à déterminer

Dans les deux cas, on a besoin de définir la notion de distance entre deux classes : le critère d'agrégation.

III.1. Classification supervisée

Soit $D = \{d_1, d_2, \dots, d_i, \dots, d_m\}$ un ensemble de documents représentés chacun

par une description $\vec{d}_1, \vec{d}_2, \dots, \vec{d}_m$, et $C = \{C_1, C_2, \dots, C_k, \dots, C_c\}$ un ensemble de classes, la classification supervisée suppose connues deux fonctions. Le premier fait

correspondre à tout individu d_i une classe C_k . Elle est définie au moyen de couples

(d_i, C_k) donnés comme exemples au système. La deuxième fait correspondre à tout

individu d_i sa description \vec{d}_i . La classification supervisée consiste alors à déterminer une procédure de classification :

$$C^f : \vec{d}_i \rightarrow C_k \quad (1)$$

Qui à partir de la description de l'élément détermine sa classe avec le plus faible taux d'erreurs. La performance de la classification dépend notamment de l'efficacité de la description. De plus, si l'on veut obtenir un système d'apprentissage, la procédure de classification doit permettre de classer efficacement tout nouvel exemple (pouvoir prédictif).

III.2. Classification non supervisée

La classification non supervisée est utilisée lorsque que l'on possède des documents qui ne sont pas classés et on ne connaît pas leur classification. A la fin du processus de classification non supervisée, les documents doivent appartenir à l'une des classes générées par la classification. On distingue deux catégories de classifications non supervisées : hiérarchiques et non hiérarchiques.

Dans la **classification hiérarchique (CH)**, les sous-ensembles créés sont emboîtés de manière hiérarchique les uns dans les autres. On distingue la CH descendante (division) ou l'ensemble de tous les individus est fractionné en un certain nombre de sous-ensembles, chaque sous-ensemble étant alors fractionné en un certain nombre de sous-ensembles, et ainsi de suite. Et la CH ascendante (ou agglomération) qui part des individus seuls que l'on regroupe en sous-ensembles, puis regroupe en sous-ensembles qui sont à leur tour, et ainsi de suite. Pour déterminer quelles classes on va fusionner, on utilise le critère d'agrégation.

Dans la **classification non hiérarchique**, les individus ne sont pas structurés de manière hiérarchique. Si chaque individu ne fait partie que d'un sous-ensemble, on parle de partition. Si chaque individu peut appartenir à plusieurs groupes, avec la probabilité P_i d'appartenir au groupe i , alors on parle de recouvrement.

VI. Critère d'agrégation

Le critère d'agrégation permet de comparer les classes deux à deux pour sélectionner les classes les plus similaires suivant un certain critère. Les critères les plus classiques sont le plus proche voisin, le diamètre maximum, la distance moyenne et la distance entre les centres de gravités.

VI. 1. Plus proche voisin

La distance entre la classe C_p et la classe C_q est la plus petite distance entre un élément de C_p et un élément de C_q .

$$D(C_p, C_q) = \min\{dist(i, j); i \in C_p, j \in C_q\} \quad (2)$$

VI.2 Diamètre maximum

La distance entre la classe C_p et la classe C_q est la plus grande distance entre un élément de C_p et un élément de C_q .

$$D(C_p, C_q) = \max\{dist(i, j); i \in C_p, j \in C_q\} \quad (3)$$

VI.3 Distance moyenne

La distance entre la classe C_p et la classe C_q est la moyenne des distances entre les éléments de C_p et les éléments de C_q .

$$D(C_p, C_q) = \frac{\sum_{i,j} \{dist(i, j); i \in C_p, j \in C_q\}}{Card(C_p) \times Card(C_q)} \quad (4)$$

VI.4 Distance entre les centres de gravité

Si G_p est le centre de gravité de la classe C_p et si G_q est le centre de gravité de la classe C_q alors la distance entre la classe C_p et la classe C_q est la distance entre leurs centres de gravités.

$$D(C_p, C_q) = dist(G_p, G_q)$$

Ce critère n'a de sens que si le calcul du centre de gravité possède lui-même un sens sur les données de l'étude.

V. critère d'évaluation d'un système de classification

Nous présentons ici une méthode permettant d'évaluer une classification supervisée, et des techniques classiques pour mesurer et comparer des systèmes de classifications non supervisées.

Corpus de test (cas supervisé)

Pour tester la qualité d'une procédure de classification supervisée, on sépare aléatoirement les éléments classés entre une base de référence (R) et une base de test (T). Ensuite, on détermine la procédure de classification C^f à partir des exemples de la base de référence. Puis, on utilise C^f pour retrouver la classe des éléments de la base de test. Enfin, on estime l'erreur de la procédure de classification.

Pour estimer le taux d'erreur TE d'une procédure de classification C^f , une méthode simple est de calculer le nombre d'éléments mal classés sur le nombre d'éléments à classer :

$$TE(C^f) = \frac{1}{\text{card}(T)} \sum_{t=1}^{\text{card}(T)} (C^f(\vec{d}_t) \neq C_{d_t}) \quad (5)$$

où C_{d_t} est la classe d'origine de d_t .

Dans les cas de classifications simples, on peut être amené à calculer l'erreur résultant d'une classification purement aléatoire C^a pour la comparer avec l'erreur faite par notre procédure C^f afin de vérifier la performance de notre système.

Soit P_k la fréquence (ou probabilité à priori) de la classe k dans la base de test, on appelle erreur TE_a du système aléatoire :

$$TE_a = 1 - \sum_{k=1}^c (P_k)^2 = 1 - \sum_{k=1}^c \left(\frac{\text{card}(C_k|T)}{\text{card}(T)} \right)^2 \quad (6)$$

Où c est le nombre de classes et $\text{card}(C_k|T)$ est le nombre d'éléments de T qui sont dans la classe C_k .

L'erreur apparente $TE(C^f)$ est dépendante de l'échantillon considéré. Cependant, plus le nombre d'éléments de l'échantillon est grand, plus l'erreur mesurée tend vers l'erreur réelle de C^f .

Cas non supervisé

Dans le cas non supervisé, on peut évaluer la classification par rapport à certaines de ces caractéristiques. On distingue d'une part, les caractéristiques numériques : le nombre de classes obtenues, le nombre d'éléments par classe, le nombre moyen d'éléments par classe, l'écart type des classes obtenues, et d'autre part, les caractéristiques sémantiques. Par exemple, si à un document est associé un ensemble de mots clés, la sémantique associée à une classe pourra se composer des mots les plus fréquents dans la classe.

Pour évaluer l'homogénéité du nombre d'images par classe, on peut utiliser la variance :

$$V = \sigma^2 = \frac{1}{c} \sum_{k=1}^c (\text{card}(C_k) - \text{moy})^2 \quad (7)$$

Où $\text{moy} = \frac{1}{c} \sum_{k=1}^c \text{card}(C_k)$ est le nombre moyen d'éléments par classe et c est le

nombre de classes obtenues. L'écart type $\sigma = \sqrt{V}$ permet d'exprimer la dispersion dans la même unité que la moyenne.

Les résultats d'une classification donnent

		MALADIE	
		Malade	Non Malade
TEST	Positif	VP	FP
	Négatif	FN	VN
		Malades $m = VP + FN$	Non malades $n = FP + VN$

VP : vrai positif

FP : faux positif

FN : faux négatif

VN : vrai négatif

m : nombre de malades

n : nombre de non-malades

Sensibilité du test

C'est la proportion de tests positifs (VP) chez les sujets malades

$$Se = VP / (VP + FN)$$

Moins les faux négatifs sont nombreux, plus le test est sensible

Spécificité du test

C'est la proportion de tests négatifs chez les sujets non malades

$$Sp = VN / (FP + VN)$$

Moins les faux positifs sont nombreux, plus le test est spécifique

IV Méthode Fuzzy C-Means

Fuzzy C-Means(FCM) est un algorithme de classification non supervisée floue. Issu de l'algorithme des C-moyennes (*C-means*), il introduit la notion d'ensemble flou dans la définition des classes : chaque point dans l'ensemble des données appartient à chaque cluster avec un certain degré, et tous les clusters sont caractérisés par leur centre de gravité. [6]

Comme les autres algorithmes de classification non supervisée, il utilise un critère de minimisation des distances intra-classe et de maximisation des distances inter-classe, mais en donnant un certain degré d'appartenance à chaque classe pour chaque vecteur . Cet algorithme nécessite la connaissance préalable du nombre de clusters et génère les classes par un processus itératif en minimisant une fonction objective. Ainsi, il permet d'obtenir une partition floue des données en donnant à chaque vecteur un degré d'appartenance (compris entre 0 et 1) à une classe donnée. Le cluster auquel est associé un vecteur est celui dont le degré d'appartenance sera le plus élevé. [5]

Les principales étapes de l'algorithme Fuzzy C-means sont :

1. La fixation arbitraire d'une matrice d'appartenance.
2. Le calcul des centroïdes des classes.
3. Le réajustement de la matrice d'appartenance suivant la position des centroïdes.
4. Calcul du critère de minimisation et retour à l'étape 2 s'il y a non convergence de critère.

L'algorithme *k-means* est l'algorithme de clustering le plus connu et le plus utilisé, du fait de sa simplicité de mise en oeuvre. Il partitionne les données d'une image en K clusters.

Contrairement à d'autres méthodes dites hiérarchiques, qui créent une structure en « arbre de clusters » pour décrire les groupements, *k-means* crée qu'un seul niveau de clusters. L'algorithme renvoie une partition des données, dans laquelle les objets à l'intérieur de chaque cluster sont aussi proches que possible les uns des autres et aussi loin que possible des objets des autres clusters. Chaque cluster de la partition est défini par ses objets et son centroïde. [5]

Le *k-means* est un algorithme itératif qui minimise la somme des distances entre chaque objet et le centroïde de son cluster. La position initiale des centroïdes conditionne le résultat final, de sorte que les centroïdes doivent être initialement placés le plus loin possible les uns des autres de façon à optimiser l'algorithme. *K-means* change les objets de cluster jusqu'à ce que la somme ne puisse plus diminuer. Le résultat est un ensemble de clusters compacts et clairement séparés, sous réserve qu'on ait choisi la bonne valeur K du nombre de clusters. [7]

Les principales étapes de l'algorithme *k-means* sont :

1. Choix aléatoire de la position initiale des K clusters.
2. (Ré- Affecter les objets à un cluster suivant un critère de minimisation des distances (généralement selon une mesure de distance euclidienne).
3. Une fois tous les objets placés, recalculer les K centroïdes.
4. Répéter les étapes 2 et 3 jusqu'à ce que plus aucune réaffectation ne soit faite.

La méthode FCM consiste donc à trouver l'appartenance exacte d'un vecteur de données à une classe. Au départ, à chaque vecteur est assignée une valeur qui représente son degré d'appartenance dans chaque classe.[6] Ce degré varie entre 0 et 1: c'est la fuzzyfication.

On applique la règle floue choisie, cette règle gère la défuzzyfication du système en affectant chaque pixel à une seule classe à savoir celle dont il a le plus grand degré d'appartenance. Le principe consiste à trouver une partition de l'image caractérisée par une matrice de partition U_{ij} qui caractérise le degré d'appartenance d'un vecteur x_j à une classe C_i

Soit la matrice U_{ij} donnée par:

$$U_{ij} = \begin{bmatrix} P_{C_1}(x_1) & \dots & P_{C_1}(x_n) \\ P_{C_2}(x_1) & \dots & P_{C_2}(x_n) \\ \dots & \dots & \dots \\ P_{C_n}(x_1) & \dots & P_{C_n}(x_n) \end{bmatrix}$$

Où: $P_{C_i}(X_j)$: degré d'appartenance du vecteur (x_j) à la classe C_i .

Avec la condition: $0 \leq P_{C_i}(x_j) \leq 1$ et $\sum P_{C_i}(x_j) = 1$

Cette matrice interprète la fuzzyfication du système et fournit les données à l'étape suivante.[6]

Cette étape constitue l'établissement des règles qui serviront par la suite à la défuzzyfication de la matrice.[6] Celle-ci consiste à trouver une partition stable des données en affectant chaque vecteur dans une seule classe. Le nombre de classes est choisi par l'opérateur.

Pour partitionner les données, il faut minimiser le critère de la somme des distances intra-classes généralisée au cas flou et donné par la formule suivante:

$$J = \sum_{i=1}^C \sum_{j=1}^N (U_{ij})^m d^2(x_j, C_i)$$

sous la contrainte suivante:

$$\sum_{i=1}^c U_{ij} = 1 \quad \forall j \in [0, N]$$

Où:

- $m \in]1, +\infty]$ est un paramètre qui caractérise le degré du flou.

- C: représente le nombre de classes,

- N: représente le nombre de pixels à classifier,

- C_i : est le vecteur caractéristique du centre de gravité de la classe d'indice i.

- $d^2(x_j; C_i)$ caractérise la distance entre un vecteur (x_j) et le centre de gravité de la classe (C_i)

Algorithme de partitionnement (F C.M)

Avant d'exécuter l'algorithme, on fixe le nombre de classes C et le degré de flou m. Pour minimiser [J] on procède par deux étapes:

Dans un premier temps, nous minimisons J par rapport aux centres (C_i)

Une fois les centres étant fixés, nous minimisons J par rapport à U_{ij}

L'algorithme suit les instructions suivantes:

1/ Initialisation des centres de gravités des classes prédéfinis Dans notre cas, pour tirer les centres, on fait un maillage qui consiste à partitionner Les données selon le nombre de classes choisies. [7]

2/ Fuzzyfication [8]

C'est le calcul de la matrice de partition U_{ij} . La matrice finale s'obtient après un nombre d'itérations qui détermine sa stabilité. C'est à dire que la configuration ne change plus.

U_{ij} est donné par les équations suivantes:

$$\begin{aligned} & \text{Si } I_j = \emptyset \\ & U_{ij} = \left[\sum_{k=1}^C \left(d^2(x_j, c_i) / d^2(x_j, c_k) \right)^{1/m-1} \right]^{-1} \\ & \text{Si } I_j \neq \emptyset \\ & U_{ij} = 0 \quad \text{pour } i \notin I_j \\ & U_{ij} = \frac{1}{|I_j|} \quad \text{pour } i \in I_j \\ & \text{Avec } I_j = \{i / 1 \leq i \leq C, d^2(x_j, c_i) = 0\} \end{aligned}$$

3/ Calcul des nouveaux centres multidimensionnels qui serviront à modéliser la matrice de partition dans une nouvelle itération. L'équation de mise à jour des centres C_i est donnée par:

$$C_i = \frac{\sum_{j=1}^N (U_{ij})^m * X_j}{\sum_{j=1}^N (U_{ij})^m}$$

$1 \leq i \leq C$

Où:

X_j : la valeur de chaque vecteur.

m : le degré de flou choisi.

N : le nombre de vecteur dans chaque classe i .

4/ Règle de décision

Dans cette étape, on peut juger de la stabilité de la matrice de partition en utilisant les deux

Cas suivants:

$$\begin{cases} \|U_{(n)} - U_{(n-1)}\| > \xi \\ \|U_{(n)} - U_{(n-1)}\| \leq \xi \end{cases}$$

Tel que: $U_{(n-1)}$ représente la matrice précédant et ξ le seuil de convergence.

Dans le cas où:

Si $\|U_{(n)} - U_{(n-1)}\| > \xi$. Retour à l'étape de Fuzzification pour la mise à jour de la matrice de

Partition par une nouvelle itération.

La comparaison se fait entre tous les vecteurs. Il suffit qu'il y ait un seul vecteur qui ne satisfait pas la comparaison pour décider du non convergence et donc l'instabilité de la partition.

Si $\|U_{(n)} - U_{(n-1)}\| \leq \xi$. La matrice converge, aller à l'étape suivante (dedéfuzzification).

5/ Défuzzyfication [8]

Après la satisfaction de la condition de convergence, la défuzzyfication consiste à appliquer

La règle floue choisie. Il s'agit de la technique du maximum d'appartenance. On affecte

Chaque vecteur à la classe pour laquelle il a le plus grand degré d'appartenance. [9]

IIV les réseaux de neurones

Dans le texte qui suit, des symboles et des variables sont utilisées pour représenter les composantes des réseaux de neurones. Voici les définitions de ce symbolisme.

S	Système que représente le réseau de neurone, de l'extérieur
X	Vecteur d'intrant du réseau, ou d'une manière locale, vecteur d'intrant d'une couche du réseau.
x_i	i-ème élément du vecteur d'intrant X
Y	Vecteur de sortie du réseau, ou d'une manière locale, vecteur de sortie produit par une couche du réseau.
y_i	i-ème élément du vecteur de sortie Y
w_i	Poids du i-ème intrant du neurone
$w_{i,j}$	Poids sur le lien entre le neurone i et le neurone j
W_i	Vecteur de poids associé au neurone i
W	Matrice de poids, associée à une couche.
f()	Fonction d'activation d'un neurone
θ	Seuil d'activation d'un neurone

LISTE DES ABBRÉVIATIONS

PMC Perceptron Multicouche

ACP Analyse des Composantes Principales

Les réseaux de neurones sont une des composantes importantes du domaine de l'intelligence artificielle. Les réseaux de neurones « artificiels », pour les différencier des réseaux de neurones biologiques, sont composés de neurones artificiels simples, petites fonctions mathématiques, qui permettent, montés en réseau de former des fonctions complexes très utiles. Par analogie aux neurones biologiques, les neurones artificiels ont pour but de reproduire des raisonnements « intelligents » d'une manière artificielle. Ces neurones peuvent adopter de certaines qualités habituellement propres au biologique, c'est-à-dire, la généralisation, l'évolutivité, et une certaine forme de déduction.

La structure d'un neurone artificiel est en fait copiée sur la structure des neurones biologiques. Les principales structures biologiques des neurones ont toutes leur équivalent artificiel (voir plus bas), ceci ayant pour but de reproduire leur fonctionnement de la meilleure façon possible (d'une manière logique, simple et facilement représentable sur informatique). Un neurone est une minuscule structure qui traite les influx nerveux qui arrivent (*inputs*), chacun selon son importance relative, et qui émet un (et un seul) signal de sortie (*output*). Les neurones artificiels reproduisent le même procédé, recevant chaque signal d'entrée (*input*) tempéré par un poids (*weight*). Ces poids sont aussi appelés poids synaptiques, par analogie. Les intrants pondérés, habituellement (mais

pastoujours) sommés, sont ensuite comparés à un seuil d'activation θ^1 et passés dans la fonction du neurone (habituellement une fonction très simple), qui produit l'extrait

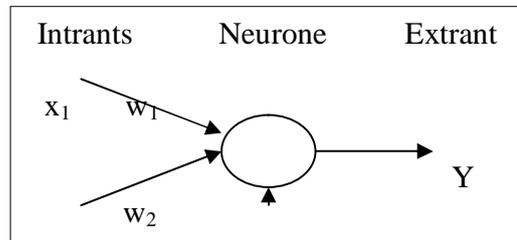


Figure 1 : Schéma d'un neurone artificiel (deux signaux d'intrants seulement)

IIV.1. Naissance

En 1943, Warren McCulloch et Walter Pitts ont proposé un des premiers modèles de neurone artificiel, qui deviendra la base des réseaux de neurones artificiels. Ce modèle de neurone (également appelé neurone de McCulloch-Pitts) utilise la sommation des intrants pesés (*weighted inputs*) comme intrant « net » X (intrant direct de la fonction du neurone) et comme fonction d'activation, la fonction de signe (comparé avec le seuil θ) .

$$X = \sum_{i=1}^n x_i w_i$$

$$Y = \begin{cases} +1 & \text{si } X \geq \theta \\ -1 & \text{si } X < \theta \end{cases}$$

² Par sa simplicité et son « intuitivité » c'est cette forme d'intrant net qui est utilisée par défaut..

Tableau 2 : Autres fonctions d'activation³ (comparées au seuil θ)

Étage	Signe	Sigmoïde ⁴	Linéaire
$Y^{étage} = \begin{cases} 1 & \text{si } X \geq \theta \\ 0 & \text{si } X < \theta \end{cases}$	$Y^{signe} = \begin{cases} +1 & \text{si } X \geq \theta \\ -1 & \text{si } X < \theta \end{cases}$	$Y^{sigmoïde} = \frac{1}{1 + e^{-\lambda(X-\theta)}}$	$Y^{linéaire} = X - \theta$

La fonction d'activation par signe est très utile, mais d'autres fonctions peuvent être utilisées. Par exemple, la fonction par étage ($Y = \{0,1\}$) est meilleure pour les sorties par bits et pour les tables de vérités. La fonction sigmoïde présente davantage de souplesse et de nuance lors du traitement des erreurs.

Malgré la grande répartition de ces neurones, il en existe au moins un autre type. Les neurones, on en effet deux fonctions d'activation, la première étant la production de l'intrant net (ci-haut, la somme des intrants pesés), le seconde étant la fonction d'activation. Dans le cas du neurone de McCulloch-Pitts, ainsi que dans la plupart des cas, les neurones sont des neurones « produit scalaire », car l'intrant net est produit de la même manière que le produit scalaire entre le vecteur d'intrant et le vecteur de poids associés à ce neurone. Dans d'autres cas, on peut aussi utiliser des neurones « distance ». Comme le nom l'indique, la production de l'intrant net dans ce cas est produit par le calcul d'une distance entre le vecteur d'intrant et le vecteur de poids. Habituellement, une autre manière de calculer l'intrant net amène à faire des choix différents par rapport à la fonction d'activation utilisée. Dans le cas des neurones « distances ». Les neurones

³ Ce tableau n'est pas exhaustif, mais regroupe les fonctions d'activations les plus utilisées

⁴ Dans cette formule, le λ permet de modifier l'importance de la variation. Plus λ est grand (>1), plus la fonction sigmoïde tend vers la fonction par étage. Plus λ est petit (< 1), plus la fonction tend vers une droite. Par défaut, $\lambda=1$.

« distance » sont utilisés d'une manière plus spécifique, particulièrement dans le cas des réseaux de Kohonen. Il est important de noter que les poids d'un réseaux doivent être interprétés de manière différente selon les type de neurone utilisé.

Le seuil d'activation d'un neurone peut également être considéré comme un intrant fixe de valeur -1 et de seuil θ . Cette manière de procéder permet d'inclure la modification du seuil dans un algorithme d'entraînement, augmentant ainsi l'apprentissage du réseau.

Cependant, cette modification n'est correcte que lorsque la fonction d'activation s'effectue sur l'intrant net auquel on soustrait le seuil ($f(X - \theta)$). Par exemple, dans le cas d'un neurone « distance », cette modification serait tout a fait incorrecte. Certaines personnes utilisent la notion de biais pour les neurones. Il s'agit d'un intrant fixe de 1 et de poids déterminé selon le réseau. Généralement, l'utilisation du biais sur un neurone ne se fait pas en même temps que l'utilisation de seuils d'activation, mais rien ne l'empêche, sinon les quelques complications que cela occasionne. De fait, dans plusieurs cas, la double utilisation de biais et de seuils d'activation est superflue. On peut facilement considérer que le biais sur un neurone est équivalent à un seuil d'activation, mais si ceci se révèle vrai dans plusieurs cas, il serait incorrect de généraliser. Il est important de faire la différence entre le biais d'un neurone et le biais dans le cas d'une méthode statistique. Dans la suite du texte, le biais fera référence à celui des méthodes statistiques, sauf indication contraire.

IIV.2. Architecture Des RNS

Un réseau de neurones est, comme son nom l'indique, un ensemble de neurones en réseau, de sorte que les signaux sortant (*outputs*) des neurones deviennent des signaux entrant (*inputs*) dans d'autres neurones. L'architecture générale des réseaux de neurones

consiste en la représentation des neurones en couches (*layers*) successives, la première représentant la couche d'intrants (*input layer*), la dernière étant la couche de sortie (*output layer*), les couches intermédiaires étant les couches cachées (*hiddenlayers*) du réseau. Ces couches sont dites cachées car de l'extérieur du réseau, on ne peut analyser clairement leur fonctionnement. On ne connaît vraiment que les signaux intrants et extrants du réseau. Les neurones de la couche d'intrants ne sont pas vraiment des neurones traitants (*computingneurons*), mais ont pour seule utilité de normaliser l'entrée des signaux ainsi que la distribution des signaux intrants. Dans cette architecture normalisée, les couches de neurones sont totalement interconnectées, c'est-à-dire que les neurones d'une couche sont tous reliés à tous les neurones des couches adjacentes. Cette architecture normalisée peut sembler rigide, mais elle permet une représentation correcte de la plupart des réseaux de neurones, tout en permettant l'utilisation d'algorithmes d'entraînement plus généraux.

D'une manière générale, un réseau de neurones se comporte, d'un point de vue extérieur, comme une fonction S qui traite des données (*inputs*) et produit une réponse correspondante (*output*). Les données entrées peuvent être de n'importe quel type représentable de manière binaire ou numérique. Ces données peuvent également être vues comme des vecteurs, et le réseau de neurones une application vectorielle.

Les réseaux de neurones peuvent être soit statiques, soit dynamiques⁵. Les réseaux dynamiques ont la capacité de s'actualiser (modifier les poids des liaisons) en tout temps. Ces réseaux possèdent habituellement des algorithmes d'entraînements non-supervisés,

⁵ Les réseaux sont des structures statiques, c'est-à-dire que, une fois construit, leur architecture ne change pas. Ici, on parle plutôt du caractère statique ou dynamique des poids du réseau lors de son utilisation, c'est-à-dire, si les valeurs des poids des réseaux peuvent ou non se modifier lors de l'utilisation (à différencier de l'entraînement) du réseau.

ce qui permet au réseau en question d' « évoluer » à chaque utilisation. Les réseaux statiques se séparent en deux catégories. La première concerne les réseaux entraînés, c'est-à-dire les réseaux qui ont vu leurs poids modifiés par un algorithme d'entraînement, sous la supervision d'un instructeur jusqu'à ce que ces réseaux donnent des résultats satisfaisants aux données d'entraînement. La seconde catégorie concerne les réseaux non-évolutifs, comme les réseaux de mémoire. Les poids de ces réseaux sont choisis (et non pas modifiés par un entraînement), pour donner l'effet escompté.

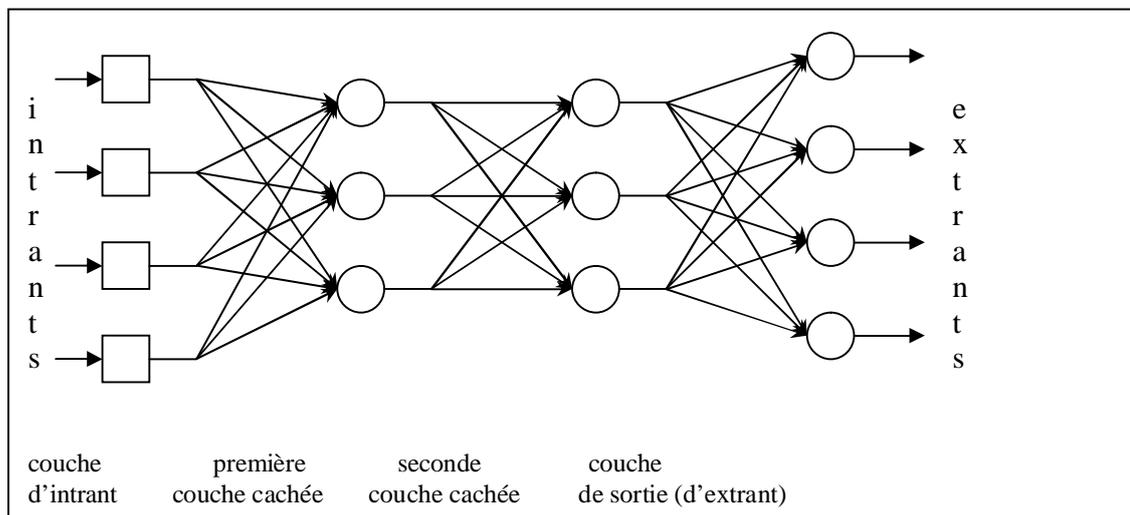


Figure 2 : Architecture d'un réseau de neurones (avec deux couches cachées)

IV.3. Le Perceptron

En 1958, Frank Rosenblatt proposa le perceptron, accompagné d'un algorithme d'entraînement. Un perceptron est un réseau de neurones avec des neurones « produit scalaire » fonctionnant avec une fonction par étage ou sigmoïde (la fonction est un cas particulier de la fonction sigmoïde). On présente ici un algorithme d'entraînement pour un perceptron à un seul neurone (le cas le plus trivial)

Algorithme de traitement

Initialisation

Mettre les poids initiaux w_1, w_2, \dots, w_n ainsi que le seuil θ à des valeurs aléatoires de l'intervalle $[-0,5, 0,5]$. Mettre le taux d'apprentissage α à une petite valeur positive.

2- Activation

Activer le perceptron en appliquant les intrants $x_1(p), x_2(p), \dots, x_n(p)$ et l'extrant désiré $Y_d(p)$. Calculer l'extrant actuel à l'itération $p = 1$.

$$Y(p) = \text{étage} \left[\sum_{i=1}^n x_i(p)w_i(p) - \theta \right]$$

Où n est le nombre de signaux intrants, et étage est la fonction d'activation par étage.

3- Entraînement des poids

Mettre à jour les poids du perceptron

$$w_i(p+1) = w_i + \Delta w_i$$

où $\Delta w_i(p)$ est la correction de poids à l'itération p .

La correction de poids est calculée par la loi delta :

$$\Delta w_i = \alpha x_i(p) \times e(p)$$

où α est le taux d'apprentissage et $e(p)$ l'erreur à l'itération p (la différence entre l'extrant désiré et l'extrant actuel du perceptron)

4- Itération

Augmenter p de 1, retourner à l'étape 2 et répéter le procédé jusqu'à convergence.

Exemple d'entraînement d'un perceptron (un neurone, deux intrants), $\theta = 0,35$, $\alpha = 0,2$:

Période	Intrants		Sortie désirée	Poids Initiaux		Sortie actuelle	Erreur	Poids Finaux	
	x_1	x_2	Y_d	w_1	w_2	Y	e	w_1	w_2
0,4	0,12	0	0,30	-0,20	0,4	0	0	0,30	-0,20
0,99	0,5	0	0,30	-0,20	0,99	0	0	0,30	-0,20
0,82	0,87	1	0,30	-0,20	0,82	0	1	0,46	-0,03
0,34	0,6	1	0,46	-0,03	0,34	0	1	0,53	0,09
0,4	0,12	0	0,53	0,09	0,4	0	0	0,53	0,09
0,99	0,5	0	0,53	0,09	0,99	1	-1	0,33	-0,01
0,82	0,87	1	0,33	-0,01	0,82	0	1	0,50	0,17
0,34	0,6	1	0,50	0,17	0,34	0	1	0,57	0,29
0,4	0,12	0	0,57	0,29	0,4	0	0	0,57	0,29
0,99	0,5	0	0,57	0,29	0,99	1	-1	0,37	0,19
0,82	0,87	1	0,37	0,19	0,82	1	0	0,37	0,19
0,34	0,6	1	0,37	0,19	0,34	0	1	0,44	0,31
0,4	0,12	0	0,44	0,31	0,4	0	0	0,44	0,31
0,99	0,5	0	0,44	0,31	0,99	1	-1	0,24	0,21

0,82	0,87	1	0,24	0,21	0,82	1	0	0,24	0,21
0,34	0,6	1	0,24	0,21	0,34	0	1	0,31	0,33

L'entraînement de ce perceptron sera terminé après 8 périodes (c'est-à-dire, 32 tests).

Les poids finaux seront $w_1 = 0,03$ et $w_2 = 0,64$.

Cependant, les perceptrons à un neurone ne peuvent apprendre que des fonctions séparables linéairement. Dans le cas présenté ci-haut, la fonction apprise est séparable par la droite $x_1w_1 + x_2w_2 = \theta$, soit $0,03x_1 + 0,64x_2 = 0,35$ ou encore $x_1 = -64x_2/3 + 0,35$. De plus, comme Mansky et Papert l'ont montré dans leur livre *Perceptrons* en 1969, les perceptrons ne peuvent généraliser sur des exemples appris localement. Il est cependant possible de surpasser ces limitations grâce à un algorithme différent, comme il sera vu plus tard.

IIV.4.Réseaux Multicouche (Multilayer Neural Networks)

Les réseaux de neurones multicouches sont habituellement bâtis selon le modèle « normalisé » et comprennent 3 ou 4 couches en tout (donc 1 ou 2 couches cachées). S'il est théoriquement possible de construire des réseaux avec un très grand nombre de couches cachées, les réseaux comprenant plus de couches cachées sont très rares, étant donné que chaque nouvelle couche augmente la quantité de calculs d'une manière exponentielle. La plupart des réseaux de neurones multicouches sont, dans la pratique, des perceptrons multicouches (PMC).

IIV.4.1. Rétro-Propagation Du Gradient (Back-Propagation)

La méthode de modification des poids est très simple avec l'algorithme de Rosenblatt, mais il implique quelques limitations d'apprentissage. Dans le cas de perceptrons multicouches, comme on ne sait pas les extrants (outputs) désirés des couches cachées, mais seulement de la dernière couche, il faut propager la responsabilité des erreurs de la dernière couche à la première, dans le sens contraire de l'exécution de réseau, d'où le nom rétro-propagation. De plus, les perceptrons multicouches utilisent des neurones munis de la fonction d'activation sigmoïde, celle-ci permettant les nuances nécessaires à une bonne utilisation de la rétro-propagation.

les principales lignes de la méthode de rétro-propagation sont les suivants:

Calcul de l'erreur :

$e_k(p) = y_{d,k} - y_k$: l'erreur du neurone k est la différence entre la valeur de l'extrait désiré et la valeur actuelle de l'extrait du neurone k)

Correction de poids (par la loi delta):

$\Delta w_{jk}(p) = \alpha \times y_j(p) \times \delta_k(p)$: La correction du poids du lien du neurone j au neurone k , où $\delta_k(p)$ est le gradient d'erreur du neurone k à l'itération p .

Gradient d'erreur (pour les neurones de la couche de sortie) :

$$\delta_k(p) = F'[X_k(p)] \times e_k(p) = y_k(p) \times [1 - y_k(p)] \times e_k(p)$$

La deuxième partie est obtenue par la dérivée de la fonction sigmoïde appliquée à l'intrant net du neurone k à l'itération p

Gradient d'erreur (pour les neurones d'une couche cachée) :

$$\delta_j(p) = y_j(p) \times [1 - y_j(p)] \times \sum_{k=1}^l \delta_k(p) w_{j,k}(p)$$

où l est le nombre de neurones sur la couche suivante.

Algorithme de traitement

1- Initialisation

Mettre tous les poids et les seuils d'activation du réseau à des valeurs aléatoires uniformément distribuées dans un petit intervalle :

$$\left(-\frac{2,4}{F_i}, \frac{2,4}{F_i} \right) \text{ (Selon Haykin, 1994), où } F_i \text{ est le nombre total de signaux intrants}$$

du neurone i dans le réseau. Cette initialisation est faite un neurone à la fois.

Mettre la valeur du taux d'apprentissage à une petite valeur positive.

2- Activation

Activer le réseau de neurones en appliquant les signaux intrants et les extrants désirés

Calculer les signaux sortant des neurones des couches successives, de la première couche cachée à la couche de sortie.

3- Entraînement des poids

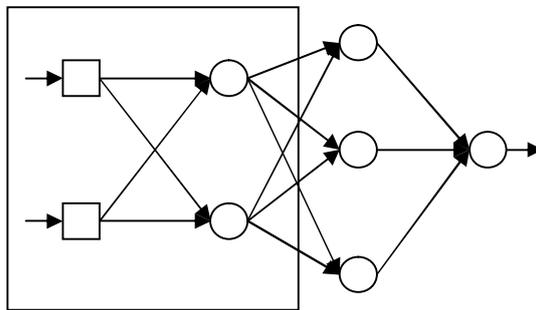
Mettre à jour les poids du réseau en propageant les erreurs dans le sens inverse :

Calculer les gradients d'erreurs pour les neurones des couches successives (ordre inverse), de la couche de sortie à la première couche cachée.

Calculer les corrections de poids pour chaque lien et mettre à jour les poids.

4- Itération

Incrémenter p , retourner à l'étape 2 et répéter le procédé jusqu'à ce que le critère



d'erreur soit atteint.

Figure 3 : Simple réseau de neurones multicouche

La méthode de rétro-propagation peut mettre énormément de temps avant que la sommation des erreurs carrées ne soit réduite à zéro (Ceci peut être impossible selon les cas). C'est pourquoi on utilise plutôt un critère d'erreur, soit une certaine valeur critique, qui doit être atteinte par la sommation des erreurs carrées (habituellement entre 0,001 et 0,1). La loi delta, utilisée par la méthode de rétro-propagation, a pour objectif de réduire la sommation des erreurs carrées produites par le réseau. Si on représente la courbe de la fonction de la somme des erreurs carrées par rapport à un poids du réseau, on peut

observer que la variation idéale dépend de la pente de la courbe de la fonction, d'où l'utilisation de la dérivée dans le calcul du gradient d'erreur (et par le fait même, de la variation de poids).

Comme cette recherche, telle la méthode du *hillclimbing*, recherche les extrema par une vérification de la pente, elle peut être limitée par l'atteinte de minima locaux.

L'entraînement d'un perceptron multicouche avec l'algorithme de rétro-propagation du gradient se fait d'une manière similaire à l'entraînement d'un perceptron à neurone unique, mais il faut procéder neurone par neurone et poids par poids, en commençant par la dernière couche. Il faut calculer le gradient d'erreur pour chaque neurone, puis calculer la variation de poids pour les liens rattachés à ce neurone. Ensuite on passe à la couche précédente. Cette méthode devient très rapidement lourde en calculs.

La différence entre l'utilisation de la fonction d'activation par étage et sigmoïde (où encore la différence entre un petit λ et un λ élevé tient au degré de généralisation que l'on veut atteindre. Avec la fonction par étage, la courbe produite par le système S du réseau de modifiera jusqu'à ce que les réponses soit correctes et cessera d'évoluer sitôt après.

Dans le cas de la fonction sigmoïde, il n'existe pas vraiment de réponse correcte et incorrecte de manière claire, la réponse étant une valeur entre 0 et 1. Cette façon de faire permet d'avoir de plus grandes nuances quant aux erreurs produites par le système.

Même avec des réponses « correctes », il reste une petite marge d'erreur, et la courbe du système tendra à démarquer le plus possible les valeurs d'apprentissage. De cette manière, une réponse correcte de 0,60 (arrondie à 1 dans un modèle binaire) ne sera pas considérée satisfaisante, et l'algorithme permet encore de modifier les poids pour que la réponse soit de plus en plus claire (de tendre vers 1). Bref, l'utilisation de la fonction

sigmoïde par rapport tient à l'évolutivité même si les réponses données par le réseau sont bonnes. Un λ trop grand (tend vers la fonction par étage) fera que le réseau se modifiera moins a posteriori que pour $\lambda = 1$, alors qu'un λ trop aura de la difficulté à s'arrêter. Il n'est pas toujours possible de « clarifier » complètement, et les cas limites devront se contenter de réponses moins claires. Dans ces cas, la limite entre les classes déterminées sera légèrement mouvante. De toute façon, l'atteinte d'un taux d'erreur très très petit est extrêmement longue à atteindre, et n'est pas toujours souhaitée (voir plus bas).

La grande majorité des PMC sont des réseaux avec apprentissage supervisé, mais il en existe quelques-uns qui peuvent apprendre sans professeur pour les guider. Une classe de ces réseaux non-supervisés est celle des réseaux auto-associatifs. Par défaut, un PMC est un réseau hétéro-associatif, ce qui veut dire que la sortie est différente de l'entrée. Dans le cas particulier où l'entrée doit être la même que la sortie, ces réseaux étant appelés auto-associatifs, on n'a pas besoin des valeurs de sortie que devraient produire le PMC, celles-ci étant les mêmes que l'entrée. Ces PMC auto-associatifs ont ceci d'intéressants que la première partie du réseau constitue un codage des données, alors que la seconde partie est le décodage, pour retrouver l'information entrée.

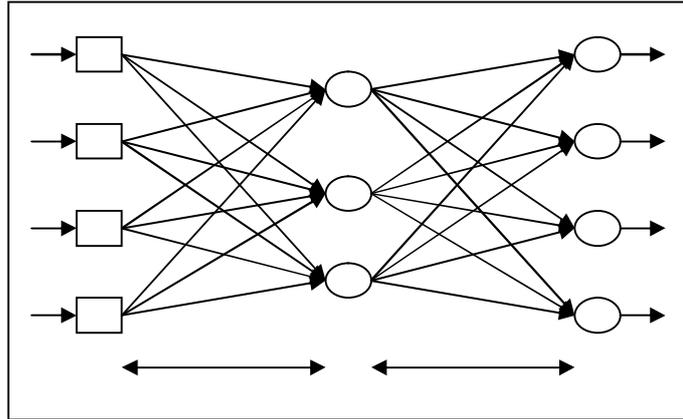


Figure 4 : PMC auto-associatif

Quelques améliorations ont été proposées pour augmenter la vitesse d'apprentissage des PMC :

L'utilisation de la fonction tangente hyperbolique $Y^{\tan h} = \frac{2a}{1 + e^{-bX}} - a$, où a et b sont

constants. Les valeurs de $a = 1,716$ et $b = 0,667$ sont efficaces (Guyon, 1991).

L'avantage de cette fonction par rapport à la fonction sigmoïde de base a été testé empiriquement.

L'utilisation d'un terme de momentum (*momentumterm*) dans l'équation de la loi delta :

$\Delta w_{jk}(p) = \beta \times \Delta w_{jk}(p-1) + \alpha \times y_j(p) \times \delta_k(p)$, avec $0 \leq \beta < 1$ (une bonne valeur de β est 0,95). Cette nouvelle équation porte le nom de loi delta généralisée. L'ajout du terme de momentum permet de stabiliser l'apprentissage, tout en augmentant la descente de la courbe de la somme des erreurs carrées selon les itérations.

L'utilisation d'un taux d'apprentissage variable plutôt que constant. Si le changement dans la somme des erreurs carrées porte le même signe algébrique, le taux d'apprentissage devrait être augmenté. Si le signe algébrique du changement de la somme des erreurs carrées alterne d'une itération à l'autre, le taux d'apprentissage devrait être décrétement.

IV.4.2. Problèmes Pour La Généralisation

Les réseaux de neurones devant généraliser sur les exemples de l'ensemble d'apprentissage, ils ne sont qu'une approximation des fonctions que l'on recherche vraiment. Quelques problèmes de l'approximation de réseaux de neurones sont identifiés ici et des solutions y sont proposées.

Compromis Biais-Variance

Le paramètre à minimiser dans le cas de l'entraînement d'un réseau de neurone est l'erreur sur les résultats donnés par le réseau de neurone. Cependant, celle-ci ne tient compte que des valeurs de l'ensemble d'entraînement, alors qu'on devrait tenir compte de l'erreur sur toutes les données à traiter. Ceci étant impossible, on base plutôt l'optimisation sur un risque moyen. Le risque quadratique moyen est un bon paramètre à optimiser. Ce risque se décompose en trois termes : l'erreur bayésienne, le biais et la variance. L'erreur bayésienne tient à l'apprentissage, mais est indépendante de la procédure d'apprentissage. La difficulté de cette optimisation est de contrôler à la fois le biais et la variance. Alors que la variance est monotone croissante, amis converge, le biais a un comportement non-monotone. Il commence par décroître, puis croît lentement. Il faut donc trouver un compromis où la somme du biais et de la variance est minimale. Cela revient à accepter un certain biais pour maintenir la variance relativement faible. Dans le cas des réseaux de neurones, c'est le contrôle de la complexité du réseau de neurone qui permet de trouver l'estimateur réalisant le bon compromis biais-variance.

Choix de l'architecture

Le choix de l'architecture d'un réseau détermine la classe des fonctions calculables par celui-ci, ou encore sa complexité potentielle. C'est évidemment le premier paramètre sur lequel les utilisateurs de réseaux ont joué pour contrôler les performances d'un système. La démarche la plus évidente pour choisir la meilleure architecture est bien entendu de tester plusieurs modèles différents, changeants les types de neurones, le nombre de couches, le nombre de neurones cachés. Cependant, l'évaluation comparative des réseaux ainsi créé pose problème, de nombreuses méthodes existant mais étant beaucoup trop lourdes en calculs. Pour cette raison, la communauté de réseaux de neurones a adopté des procédures sous-optimales. La plus courante consiste en l'utilisation d'un ensemble de validation, le réseau offrant les erreurs les moindres sur cet ensemble étant considéré le meilleur. Cette méthode est également coûteuse en temps de calculs et soumise à de nombreux aléas.

Stopper l'apprentissage avant convergence

Trop apprendre à partir d'un ensemble limité de données d'apprentissage nuit aux performances de généralisation du réseau. C'est le phénomène du sur-apprentissage : après avoir induit des règles statistiques représentatives, le réseau capture les idiosyncrasies de l'ensemble d'apprentissage. L'erreur calculée sur l'ensemble d'apprentissage décroît d'une manière continue et se stabilise ensuite, mais l'erreur de test passe par un minimum avant de croître. Il faut donc stopper l'apprentissage là où l'apprentissage produira les meilleures performances de généralisation. La sélection du meilleur moment pour stopper l'apprentissage pose quelques problèmes. L'utilisation d'un ensemble de validation reste une des méthodes reste une des techniques les plus

utilisées. Il suffit de stopper l'apprentissage avant que l'erreur sur l'ensemble de validation augmente.

Régularisation

Les techniques d'apprentissage des réseaux de neurones sont principalement basées sur l'optimisation, à savoir, la recherche de l'erreur, du risque ou du coût minimal, selon les cas. Les méthodes de régularisation visent à modifier les fonction de coûts naturelles en ajoutant des termes supplémentaires aux critères des fonctions de coûts. La nouvelle fonction de coût Q , est représentée comme suit : $Q = Q_1 + \lambda Q_2$, où Q_1 est la fonction de coût naturelle modifiée, λ le paramètre de pondération de Q_2 et Q_2 le terme

supplémentaire. Ici, on montre le cas d'un seul terme ajouté à la fonction de coût de base, mais plusieurs autres termes peuvent être ajouté. Par exemple, le choix de

$Q_2 = \sum_i w_i^2$ aurait pour effet de contraindre les poids à tendre vers 0. L'utilisation de ce

terme correspond à la technique connue sous le nom de *weightdecay*, est très facile à implanter à l'aide de l'algorithme de rétro-propagation du gradient. Cette technique vise à mettre à zéro les poids inutiles au réseau. Une variation du terme précédent a été proposé par Weigend (1991), qui permet de faire tendre vers zéro les poids faibles, alors que les

autres ne seront pas soumis à cette contrainte : $Q_2 = \sum_i \frac{w_i^2/K}{1 + w_i^2/K}$, où K est une

constante, permettant de définir quels poids sont faibles et quels poids sont forts.

Ces deux termes sont plutôt rustiques, étant donné que les poids d'un réseau peuvent avoir des significations différentes, ce nécessitant des contraintes différentes.

Élagage

L'élagage est une technique simple qui vise à éliminer les poids ayant une utilité faible. La plus populaire des techniques, appelée *Optimal Brain Damage* ou OBD, consiste en deux phases, entraîner le réseau jusqu'à l'atteinte d'un minimum local, puis d'éliminer les poids d'utilité faible. L'utilité d'un poids w_i , dans l'OBD est définie comme étant la variation du coût résultant de sa suppression qui correspond à poser $w_i = 0$. Par quelques approximations, on obtient que l'utilité d'un poids w_i est donnée par :

$$\Delta Q(w_i) = \frac{1}{2} w_i^2 \frac{\partial^2 Q}{\partial w_i^2}. \text{ Comparée à un seuil, on décide si un certain poids est utile ou non.}$$

Cette procédure appliquée plusieurs fois peut conduire à une réduction importante du nombre de poids du système sans baisse de performances.

Bruitage

Le bruitage des données est une autre technique empirique qui permet d'augmenter la qualité de la généralisation. Elle consiste à ajouter un léger bruit à chaque vecteur d'entrée pendant l'apprentissage, la sortie désirée demeurant inchangée. De cette façon, on associe une petite zone de l'espace d'entrée centrée autour d'un vecteur d'entrée à une même sortie désirée. Cela permet d'assurer un lissage de la fonction apprise par le réseau autour des points d'apprentissage. Cette technique favorise les solutions qui sont peu sensibles à des variations de la taille du bruit appris, sur les entrées.

IIV.5 Réseaux Non Supervisés

Il existe également des réseaux de neurones qui s'entraînent sans besoin de supervision, c'est-à-dire, sans que l'on ait besoin de signifier au réseau comment il doit (devrait) se comporter.

Ces réseaux, habituellement, sont toujours en train de s'entraîner. Il en existe plusieurs types, mais voici deux exemples plus marquants.

IIV.5.1. Réseau De Kohonen (Carte De Kohonen)

Ce type de réseau de neurones consiste en une situation où plusieurs neurones entrent en compétition pour être activé. Dans cette situation, un seul neurone est activé (*winner-takes-all neuron*) suite à la compétition. Ces réseaux ont été étudiés par Grossberg, von der Malsburg, Fukushima, mais c'est TuevoKohonen qui a vraiment lancé ce domaine en introduisant une nouvelle classe de réseau de neurones, soit les « cartes topologiques

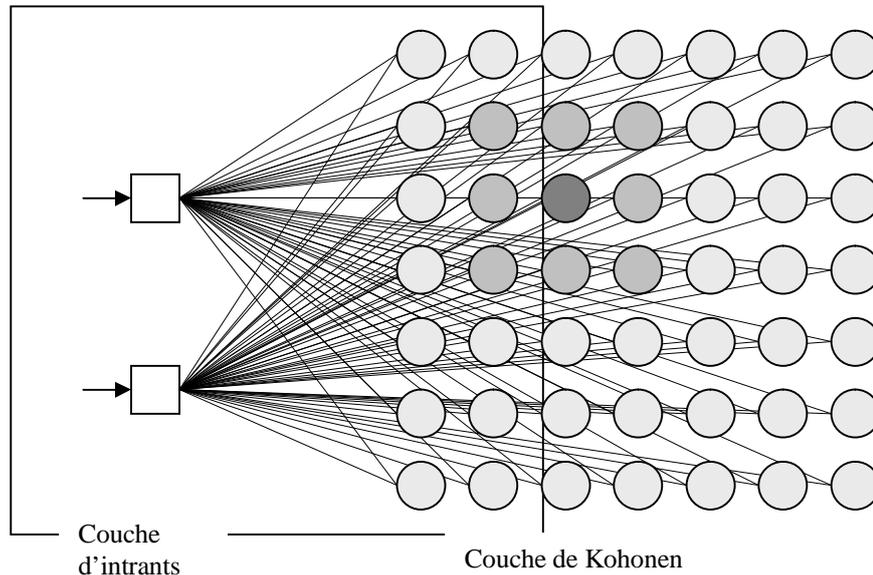


Figure 5 : Couche de Kohonen avec deux intrants

d'auto-organisation » (*self-organizingfeaturemap*).[10] Il est à noter que les termes réseau de Kohonen, carte de Kohonen et couche de Kohonen, bien qu'ils réfèrent à des principes légèrement différents, sont souvent utilisés sans grande distinction dans la littérature. [10]

Le neurone gagnant est le plus foncé et les neurones de son voisinage sont en gris.

Dans ces réseaux, seul le neurone gagnant, selon une certaine fonction, ainsi que, dans certains cas, un certain voisinage de ce neurone, verra son poids modifié. Seul le neurone gagnant et son voisinage profiteront de l'apprentissage. [10] On calcule le neurone gagnant en choisissant celui dont le vecteur de poids est le plus près du vecteur d'intrants selon une distance euclidienne. L'apprentissage consiste à rapprocher les poids des neurones gagnants des valeurs d'intrants. De cette façon, une application successive de plusieurs signaux intrants différents « organisera » les poids du réseau. Les neurones d'une couche de Kohonen possèdent également des liaisons entre eux, dans le but de « compétition » qui est la base du réseau de Kohonen. La taille et la manière de calculer le voisinage d'un neurone de la couche de Kohonen peut varier d'un réseau à l'autre. Ce voisinage peut aussi bien être nul, c'est-à-dire que seul le neurone activé est entraîné à chaque présentation d'un élément.

Algorithme de traitement

1- Initialisation

Mettre les poids synaptiques à de petites valeurs aléatoires et donner un petite valeur positive au taux d'apprentissage $\tilde{\eta}$

2- Activation

Activer le réseau de Kohonen en appliquant le vecteur d'intrants X , et trouver le neurone gagnant $j_X(p)$ par la distance euclidienne minimale.

$$j_x(p) = \min_j \|X - W_j(p)\| \left\{ \sum_{i=1}^n [x_i - w_{ij}(p)]^2 \right\}^{1/2} ; j = 1, 2, \dots, m$$

où n est le nombre de neurones de la couche d'entrée et m le nombre de neurones de la couche de Kohonen.

3- Apprentissage

Mettre à jour les poids synaptiques

$$\Delta w_{ij}(p) = \begin{cases} \alpha [x_i - w_{ij}(p)], & j \in \Lambda_j(p) \\ 0, & j \notin \Lambda_j(p) \end{cases}$$

où $\Lambda_j(p)$ est le voisinage du neurone gagnant

4- Itération

Incrémenter p , retourner à l'étape 2 jusqu'à ce que le critère de distance euclidienne minimale soit atteint ou jusqu'à ce que les poids convergent.

Les cartes de Kohonen se trouvent utiles dans le cas d'une classification automatique. En effet, en utilisant les neurones de la couche de Kohonen comme étant les indicateurs de l'appartenance à une classe. Pour un certain neurone, le vecteur W de poids lui étant associé représente le représentant de la classe que ce neurone a la tâche de représenter. Ce neurone sera le neurone activé si l'élément présenté au réseau est plus près du représentant de cette classe que du représentant de n'importe quelle autre classe, c'est-à-dire, que l'élément présenté devrait appartenir à la classe que représente ce neurone. Bien sûr, le problème de l'utilisation de réseaux de Kohonen pour effectuer ce type de classification automatique est dans le nombre de classes, et en même temps, dans le nombre de neurones de la couche de Kohonen. Un réseau de Kohonen trouvera des classes même s'il n'en existe pas. Un autre léger problème est l'entraînement de poids. La modification des poids d'un neurone, c'est-à-

dire la mise à jour du représentant d'une classe, se fait selon un critère d'apprentissage α .

Mais celui-ci, par défaut, ne tient pas compte du nombre d'éléments dans la classe ou d'autres contraintes.

IIV.5.2. Apprentissage Hebbien

Un autre type d'apprentissage non supervisé est l'apprentissage Hebbien (*Hebbian Learning*), tiré de la loi de Hebb, tel qu'énoncé par le neuropsychologue Donald Hebb en 1949. Celle-ci stipule que si deux neurones de chaque côté d'une liaison sont activés synchroniquement, le poids de cette liaison augmente. Si, d'un autre côté, deux neurones de chaque côté d'une liaison sont activés asynchroniquement, le poids de cette liaison décroît.

La modification des poids suit la règle suivante :

$$\Delta w_{ij}(p) = \alpha y_i(p)x_j(p) - \phi y_i(p)w_{ij}(p)$$

Où ϕ représente le facteur d'oubli et α le taux d'apprentissage.

Cette règle peut également être écrite comme la règle générale du produit d'activité (*generalised activity product rule*) :

$$\Delta w_{ij}(p) = \phi y_i(p) [\lambda x_j(p) - w_{ij}(p)]$$

où $\lambda = \alpha / \phi$

Algorithme de traitement

1- Initialisation

Mettre les poids synaptiques et les seuils à de petites valeurs aléatoires. Donner de petites valeurs positives au taux d'apprentissage α et au facteur d'oubli ϕ .

2- Activation

Activer les neurones du réseau selon leur fonction d'activation.

3- Apprentissage

Mettre à jour les poids du réseau selon la règle générale du produit d'activité

4- Itération

Incrémenter p , retourner à l'étape 2 jusqu'à ce que les poids convergent.

Il existe également une méthode d'apprentissage Hebbien supervisé. Le seul changement à apporter est qu'on remplace la valeur de $y_i(p)$ par $y_{i,d}(p)$, la valeur désirée de l'extrait. Cependant, cette méthode d'apprentissage supervisée, contrairement à la rétro-propagation ne tient pas compte des réponses déjà données par le réseau et met l'accent plutôt sur ce qui doit être appris que sur la correction des erreurs. La méthode d'apprentissage de Hebb applique aux réseaux de neurones les théories behavioristes (expérience du chien de Pavlov), comme quoi, en présentant un stimulus effectif et un stimulus neutre, on peut, avec entraînement, associer la réponse du stimulus effectif avec la présence seul du stimulus neutre

La carte de Kohonen

La carte de Kohonen effectue une compression sur les données reçues en entrée en représentant des vecteurs de dimension N dans un espace de dimension moindre. La propriété la plus intéressante de cette carte est la faculté de préserver les relations topologiques des vecteurs d'entrées. Il est intéressant de noter que des fonctions similaires sont présentes dans les cerveaux humain et animal ce qui est un premier point en faveur de la préservation de la topologie comme un mécanisme essentiel dans le traitement du signal.

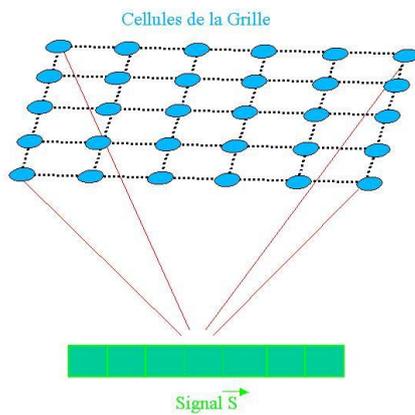


Figure.1. Carte de Kohonen.

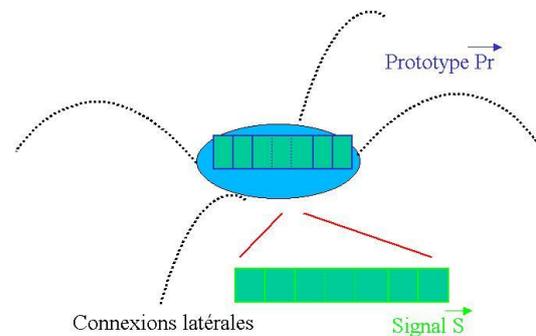


Figure.2. Cellule de Kohonen

La carte se présente sous la forme d'une grille de cellules (cf. 0). Un vecteur d'entrée $\vec{S}^{(t)}$ est présenté à toutes les cellules de la grille. Chaque cellule réagit préférentiellement à un vecteur prototype \vec{Pr} spécifique (cf. 0). Des interactions de voisinage entre cellules de la grille sont matérialisées par des connexions latérales. Ces interactions locales assurent un mécanisme de compétition entre les cellules pour favoriser la cellule la plus excitée par le vecteur d'entrée $\vec{S}^{(t)}$.

Ainsi, pour calculer les excitations consécutives à la présentation aux cellules de la grille du vecteur d'entrée $\vec{S}^{(t)}$, on calcule la distance δ_{ij} séparant ce vecteur et aux vecteurs prototypes $\vec{Pr}_{ij}^{(t)}$ des cellules c_{ij} conformément à l'équation 0 :

$$\delta_{ij}(\vec{S}^{(t)}) = \left\| \vec{S}^{(t)} - \vec{Pr}_{ij}^{(t)} \right\|$$

A partir de cette distance, deux méthodes s'offrent à nous pour mettre en œuvre le mécanisme de compétition qui va favoriser les cellules les plus excitées :

Par un calcul local, utilisant les influences exercées par les neurones les uns sur les autres (via les liens latéraux), qui fait émerger une activité en chapeau mexicain autour des neurones les plus actifs,

par application a posteriori sur le voisinage du neurone le plus actif d'une activité en chapeau mexicain (algorithme du '*winnertakes all*') (cf. 0).

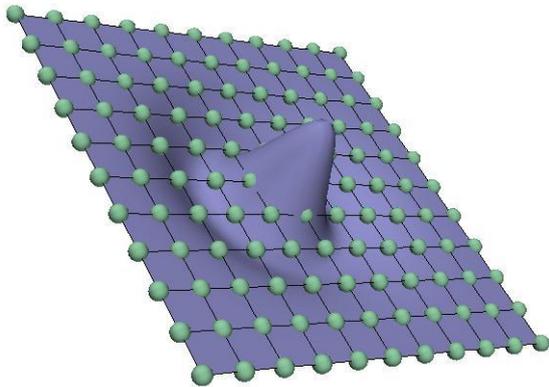


Figure.3. activité en chapeau mexicain autour du neurone le plus excité (ici, le neurone central de la grille).

On représente en vert les cellules de la grille, en violet la nappe d'activité générée par un neurone fortement excité. Lorsqu'on utilise la première méthode, on définit pour calculer les influences des neurones de la grille sur un neurone N des poids dont les valeurs se répartissent autour de N en suivant un chapeau mexicain

Désignons alors par c_* la cellule dont le prototype $\vec{\text{Pr}}_*$ est le plus semblable au vecteur d'entrée.

$$c_* = \left\{ c_{i^*j^*} / \delta_{i^*j^*} = \min_{ij} \left(\delta_{ij}(\vec{S}^{(t)}) \right) \right\}$$

L'apprentissage non supervisé s'opère par une modification des prototypes des cellules appartenant au voisinage N_* de c_* afin d'augmenter la ressemblance entre le vecteur d'entrée et les prototypes les plus similaires. Le processus d'auto organisation s'opère selon l'équation 0.

$$\vec{\text{Pr}}_{ij}^{(t+1)} - \vec{\text{Pr}}_{ij}^{(t)} = a(t) \cdot \rho_R(L_1(c_*, c_{ij})) \cdot \left[\vec{S}^{(t)} - \vec{\text{Pr}}_{ij}^{(t)} \right] \quad \forall c_{ij} \in N_*$$

avec $a(t)$ le coefficient d'apprentissage, $\rho_R(r)$, une fonction en forme de 'chapeau mexicain' de rayon R ,

$$L_1(c_{i1j1}, c_{i0j0}) = \sqrt{(i0 - i1)^2 + (j0 - j1)^2} .$$

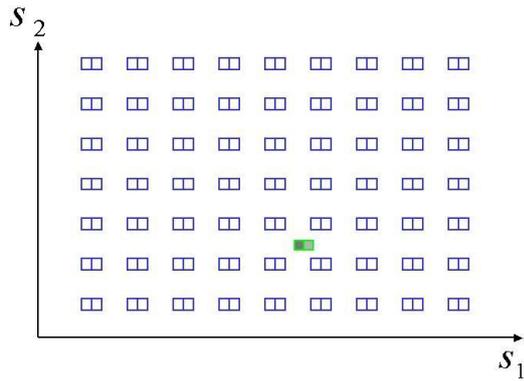


Figure.4. Processus d'auto organisation, avant modification des prototypes.

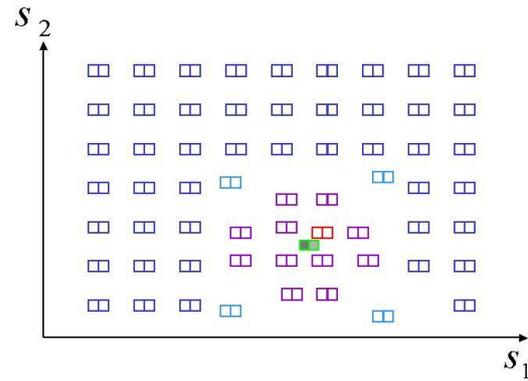


Figure.5. Processus d'auto organisation, après modification des prototypes.

Le processus d'auto organisation consiste en une modification des prototypes. Les figures ci-dessus représentent un espace d'entrée (S_1, S_2) bidimensionnel pour permettre la visualisation du phénomène. Pour clarifier l'exemple, les prototypes sont répartis de manière homogène dans l'espace d'entrée alors que cette répartition est aléatoire expérimentalement. Le vecteur d'entrée est représenté en vert, un neurone est placé à l'endroit de son prototype en bleu. L'apprentissage a pour effet de 'regrouper' les prototypes du voisinage autour du vecteur d'entrée : le neurone gagnant (en rouge) ainsi qu'une partie de son voisinage se rapprochent du vecteur d'entrée (en mauve, région où le chapeau mexicain est positif) alors que les cellules inhibées par les interactions locales (en bleu clair, région où le chapeau mexicain est négatif) s'en éloignent. [11]

Au cours de l'apprentissage, on réitère le processus pour l'ensemble des vecteurs $\vec{S}^{(t)}$ perçus et les vecteurs prototypes de la grille se stabilisent vers un ensemble de prototypes représentatifs de la distribution des vecteurs de l'espace d'entrée. [11]

L'une des forces de la carte de Kohonen (et des cartes auto-organisatrices en général) est de conserver les relations topologiques entre les différents vecteurs de l'espace d'entrée. Ainsi,

des prototypes proches sur la grille sont proches dans l'espace d'entrée ce qui autorise une certaine résistance au bruit pour les traitements qui s'appuient sur ces prototypes. Toutefois, outre le nombre d'itérations à effectuer pour obtenir un apprentissage efficace, le réglage du rayon et du voisinage en chapeau mexicain dépend fortement des données à traiter. De plus, la topologie de l'espace d'arrivée (généralement bidimensionnelle) peut être très différente de la topologie de l'espace d'entrée du fait de la définition statique de la configuration de la grille. Remarquons que des améliorations ont été proposées afin de combler certaines limitations des grilles de Kohonen, en particulier celles inhérentes à la topologie de la carte. [11]

IIV.6.1. Les grilles dynamiques ⁶

Les grilles dynamiques de Fritzke sont une extension des cartes de Kohonen. Le fonctionnement de base reste identique à celui de la carte de Kohonen. La grille de départ est une carte rectangulaire M dont chaque nœud est une cellule c_{ij} . A chaque cellule c_{ij} de la grille est associé un accumulateur noté τ_{ij} .

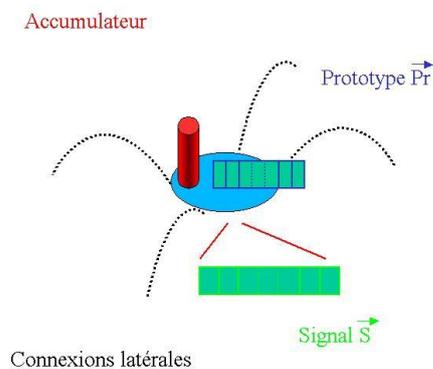


Figure.1. Cellule de Fritzke.

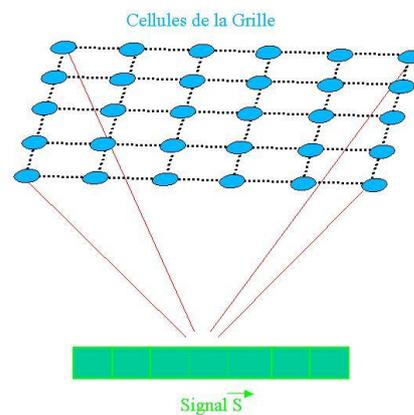


Figure.2. Carte de Fritzke

⁶ « growinggrid ».

Cet accumulateur est destiné à fournir une information statistique pour décider de l'endroit d'insertion d'une nouvelle ligne (ou d'une nouvelle colonne) de cellules. A chaque vecteur d'entrée $\vec{S}^{(t)}$ présenté, on choisit la cellule c_* la plus proche du vecteur d'entrée. Les vecteurs prototypes de la grille sont alors modifiés selon l'équation suivante :

$$\vec{Pr}_{ij}^{(t+1)} - \vec{Pr}_{ij}^{(t)} = \varepsilon \exp\left(-\frac{L_1(c_{ij}, c_*)}{2\sigma^2}\right) \left[\vec{S}^{(t)} - \vec{Pr}_{ij}^{(t)}\right]$$

$$\text{Avec } L_1(c_{i_1j_1}, c_{i_0j_0}) = |i_0 - i_1| + |j_0 - j_1|,$$

ε : Un pas d'apprentissage constant σ : l'étendue du voisinage. [12]

De plus, on incrémente l'accumulateur τ_* d'une certaine valeur (1 ou $\left\|\vec{S}^{(t)} - \vec{Pr}_{ij}^{(t)}\right\|$).

L'insertion d'une nouvelle colonne ou d'une nouvelle ligne est décidée selon des cycles dépendant de la taille de la grille. Cette insertion (cf. 0, 0) est réalisée autour de la cellule N_+ possédant l'accumulateur le plus élevé. Cet accumulateur, lorsque la cellule associée est gagnante, correspond alors au prototype qui est le plus souvent utilisé (pour un incrément de

1) ou à celui pour lequel l'erreur est la plus importante (pour un incrément de $\left\|\vec{S}^{(t)} - \vec{Pr}_{ij}^{(t)}\right\|$

).[12]

L'ajout de colonne est réalisé entre la cellule N_+ et sa voisine la plus éloignée (au sens des prototypes). Le vecteur prototype d'une cellule c appartenant à une nouvelle colonne (resp. une nouvelle ligne) est donné par l'équation 0 (resp. 0). [12]

$$\vec{\text{Pr}}_{ij'}^{(t+1)} = .5 \left(\vec{\text{Pr}}_{ij}^{(t)} + \vec{\text{Pr}}_{ij+1}^{(t)} \right)$$

$$\vec{\text{Pr}}_{i'j}^{(t+1)} = .5 \left(\vec{\text{Pr}}_{ij}^{(t)} + \vec{\text{Pr}}_{i+1j}^{(t)} \right)$$

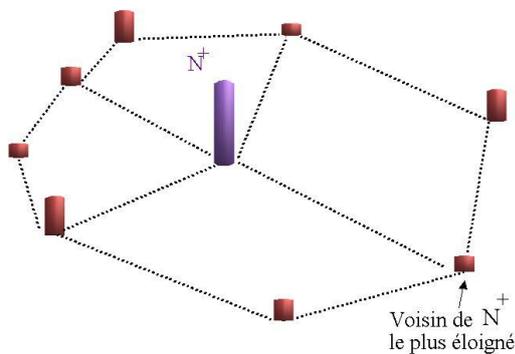


Figure.1. Choix du lieu d'insertion.

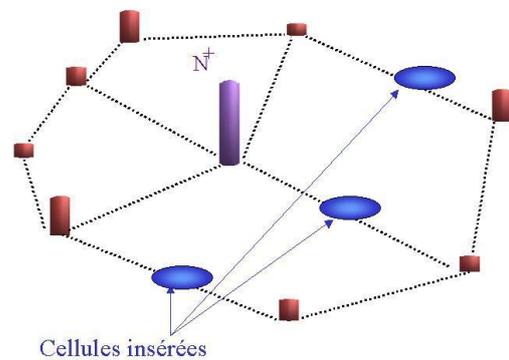


Figure.2. Insertion des cellules.

L'arrêt de la croissance de la grille est défini arbitrairement. Lorsque la croissance est terminée, un affinement des prototypes peut être réalisé en modifiant le taux d'apprentissage. Les grilles dynamiques sont donc un réseau dont la taille s'accroît au fur et à mesure de l'apprentissage. Remarquons qu'un réseau dynamique avait déjà été proposé dans lequel la croissance de la grille était réalisée dans toutes les directions. Les deux objectifs principaux de ces grilles sont de s'affranchir de la définition du coefficient et du voisinage d'adaptation variable dans le temps et de s'adapter à la topologie de l'espace d'entrée pour un surcoût de calcul relativement faible. Fritzke a proposé d'autres systèmes incrémentaux (« growingcell structures » ; «growing neural gas ». Dans ces nouveaux modèles, on insère et/ou supprime à chaque cycle une cellule unique. [12]

L'intérêt de ces structures incrémentales réside dans l'approximation plus fine de la densité autour d'un prototype. Cela est particulièrement intéressant lorsque la grille incrémentale est isolée de toutes autres structures. Toutefois, dans une architecture complexe constituée de plusieurs grilles connectées entre elles, l'insertion et la déletion de cellules d'une grille risquent de perturber l'apprentissage des cartes auto organisatrices voisines. [12]

IIV.6.2. LES MÉCANISMES D'APPRENTISSAGE

III.2.1. La règle de Hebb (1949)

Parmi les différents paradigmes de l'apprentissage, l'une des hypothèses utilisées a été formulée en 1949 par Hebb. La règle d'apprentissage de Hebb (biologiste) se formule de la manière suivante :

« quand une cellule A excite de manière répétée une cellule B,

l'efficacité de A à exciter B est améliorée par des changements métaboliques »

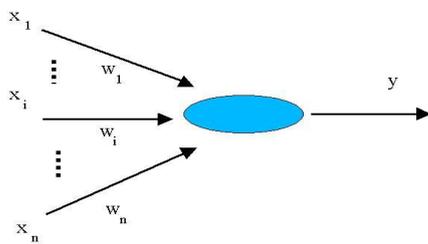


Figure.1. Le neurone formel de Mc Culloch et Pitts (**Erreur ! Source du renvoi introuvable.**, 1943)

Considérons un neurone formel N possédant les entrées pré-synaptiques $x_i, i = 1..n$; les poids de connexion $\omega_i, i = 1..n$; et une sortie post-synaptique y calculée selon l'équation :

$$y = f\left(\sum_{i=1}^n \omega_i . x_i\right)$$

avec f une fonction sigmoïde.

La règle de Hebb peut alors se formuler mathématiquement de la façon suivante :

$$\omega_i^{(t+1)} - \omega_i^{(t)} = a \cdot y \cdot x_i$$

avec a une constante définissant le pas d'apprentissage.

On constate donc que le mécanisme d'apprentissage s'attache à mettre en relief les corrélations entre le signal de sortie et les signaux d'entrée. En effet, plus le signal de sortie y et le signal d'entrée x_i sont similaires, plus le poids ω_i sera augmenté ce qui renforcera l'influence de l'entrée x_i dans la réponse y . Toutefois la règle de Hebb n'est pas suffisante pour rendre compte des expériences de conditionnement classique. [13]

IIV.6.2. La règle de Widrow-Hoff (1972)

La règle de Widrow-Hoff stipule que l'apprentissage ne se réalise que lorsqu'un événement viole les prédictions d'un organisme. La formulation mathématique de Rescola-Wagner intègre le rôle d'un signal de sortie désirée z afin de rendre compte du mécanisme d'apprentissage. [14]

$$y = f\left(\sum_{i=1}^n \omega_i \cdot x_i\right)$$

Avec f une fonction sigmoïde.

$$\omega_i^{(t+1)} - \omega_i^{(t)} = a \cdot [z - y] \cdot x_i$$

Afin de comprendre le rôle de cette sortie désirée, on peut faire référence à l'expérience de conditionnement de Pavlov. Présentons à un animal un stimulus inconditionnel *ICS* (un morceau de viande) produisant naturellement la réponse inconditionnelle *ICR* (la salivation). Le mécanisme de conditionnement consiste à associer un stimulus conditionnel *CS* (le tintement d'une cloche) au stimulus inconditionnel *ICS*. Pour ce faire, le stimulus *CS* est

présenté suivi du stimulus ICS . Lorsque le conditionnement aura opéré, la réponse CR issue de la présentation du stimulus CS sera relativement proche de la réponse ICR . [14]

L'équation 0 permet de rendre compte du phénomène d'association d'un ensemble de stimuli conditionnels CS_i à la réponse inconditionnelle ICR . Assimilons z au stimulus inconditionnel ICS , x_i à un stimulus conditionnel CS_i et considérons la force d'association ω_i qui lie ICS à CS_i . En l'absence de tout stimulus conditionnel, la force d'association n'est pas modifiée et la réponse y est invariante. Durant la phase de conditionnement, la présence conjointe des stimuli conditionnels et du stimulus inconditionnel provoque une modification des poids, modification d'autant plus importante que la réponse est différente du stimulus inconditionnel. Par conséquent, lorsque le conditionnement est réalisé, la réponse y générée par les stimuli conditionnels est proche de celle générée par le stimulus inconditionnel. Remarquons que cette règle d'apprentissage est compétitive. Ainsi lorsqu'un stimulus conditionnel a été associé au stimulus inconditionnel, le terme $[z-y]$ est nul ce qui bloque toute nouvelle association. [14]

Chapitre III

Résultats et discussions

I. La base de données

Pour cette base de données le diagnostic se fait avec précision en fonction des masses mammaires uniquement par une aspiration utilisant une aiguille fine pour obtenir une échantillon de la lésion appelé (fna)[22] le procédé d'extraction de caractéristiques est effectué comme suit:

L'échantillon (fna) est tiré de la masse de sein est posé sur une lame de microscope coloré pour mettre en évidence les noyaux cellulaires. À partir de la lame est à l'aide d'un appareil photonumérique et une carte de capteur d'images sont identifiées neuf caractéristiques pertinentes évaluées visuellement sur un échantillon (fna) [22] le jeu de données résulte et bien connu par données de Wisconsin du cancer du sein.

(Chaque malade est caractérisé par 16 attributs dont 7 attributs souvent manquants, la base est composée de 458 cas bénignes et 241 cas malignes)

Pour la sensibilité varier d'une technique à une autre :

-mammographie : la sensibilité varie de 68% à 79% .

-fna : la sensibilité varie de 65% à 98%.

- biopsie chirurgicale la sensibilité varie de 98% à 100%.

L'ensemble de données Wisconsin cancer du sein a été initialement créé pour mener des expériences qui ont montré l'intrait médical de l'automatisation du diagnostic cytologique aspiration à l'aiguille bien chaque cas comprend 11 attributs:

Le tableau 1 présente l'ensemble des attributs caractérisant le noyau cellulaire

Tableau 1. Les attributs d'ensemble des données

<i>ID d'échantillon Code d'identification</i>	<i>Nombre entier</i>
<i>A1 Épaisseur (ep)</i>	<i>1 à 10</i>
<i>A2 L'uniformité de la taille de la cellule (utc)</i>	<i>1 à 10</i>
<i>A3 Uniformité de la forme e cellulaire (ufc)</i>	<i>1 à 10</i>
<i>A4 adhérence marginal (am)</i>	<i>1 à 10</i>
<i>A5 à l'unité Taille des cellules épithéliales (utce)</i>	<i>1 à 10</i>
<i>A6 noyaux nus (nnus)</i>	<i>1 à 10</i>
<i>A7 Bland chromatine (bc)</i>	<i>1 à 10</i>
<i>A8 normale nucléoles (nnuc)</i>	<i>1 à 10</i>
<i>A9 mitoses (m)</i>	<i>1 à 10</i>
<i>A10</i>	<i>valeur 2 pour bénigne, valeur 4 pour maligne</i>

II. classification (fcm)

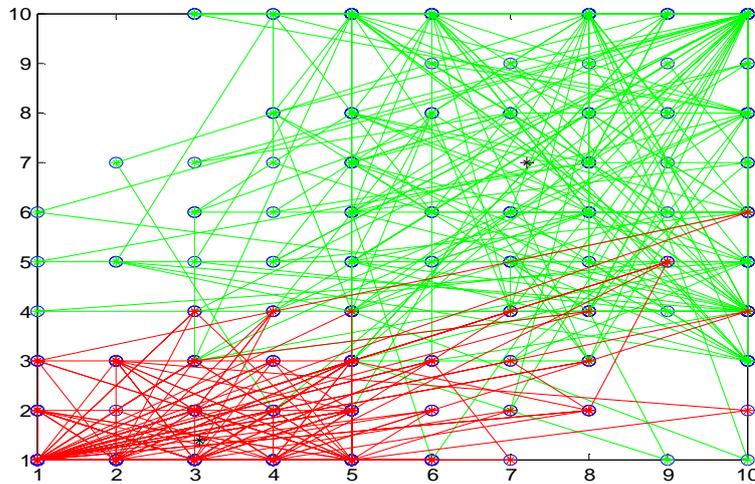


Figure 2 classifications par la méthode (fcm)

La figure (2) montre les deux classes (maligne en rouge et bénigne en vert)

Nous avons obtenu les résultats suivants

La sensibilité :($S_e = 0.9247$), la spécificité :($S_p = 0.9797$)

Le nombre de vrai positif :($V_p=221$), Le nombre du vrai négatif :($V_n=435$), Le nombre du faux positif :($F_p=9$), Le nombre du faux négatif :($F_n=18$).

CONCLUSION

Nous remarquons clairement dans la figure (2) les classes bénignes en vert et les classes malignes en rouge avec deux centres distincts C1 et C2.

Donc cette tâche du classificateur (fcm) est aisée pour séparer deux classes différentes.

Concernant les performances de la classification (fcm) à savoir la sensibilité S_e et la spécificité S_p , sont similaires avec une légère amélioration pour la spécificité S_p .

Les résultats obtenus sont expliqués par le fait que le nombre des cas non pathologiques est plus grand par rapport aux cas pathologiques.

Nous prenons des noyaux cellulaires de la base de données citées ci-dessus avec chaque noyau cellulaire caractérisé de 9 attributs (A1- A2) plus le code d'identification (Id) et sa classe (classe :2 bénigne, classe :4 maligne), ces caractéristiques sont déjà définies dans le tableau 1.

<i>Code Id</i>	<i>A1 Ep</i>	<i>A2 Utc</i>	<i>A3 Ufc</i>	<i>A4 Am</i>	<i>A5 Utce</i>	<i>A6 Nn us</i>	<i>A7 Bc</i>	<i>A8 Nn uc</i>	<i>A9 M</i>	<i>CLASS E</i>
109126 2	2	5	3	3	6	7	7	5	1	4
109680 0	6	6	6	9	6	5	7	8	1	2
110052 4	5	6	5	6	10	1	3	1	1	4
110257 3	5	6	5	6	10	1	3	1	1	4
106572 6	5	2	3	4	2	7	3	6	1	4
106637 3	3	2	1	1	1	1	2	1	1	2
106697 9	5	1	1	1	2	1	2	1	1	2
105955 2	1	1	1	1	2	1	3	1	1	2
116743 9	2	3	4	4	2	5	2	5	1	4
116747 1	4	1	2	1	2	1	3	1	1	2
117041 9	10	10	10	8	2	10	4	1	1	4

Le patient (Id : 1096800) et Le patient (Id : 1059552) appartient à des classes différentes mais leur position sont rapprochées.

Ces deux cas causent difficultés de reconnaissance pour la classification (FCM) ce qui justifier énormément la présence des faux positifs (fp) et des faux négatifs (fn).

III- apprentissage par la carte kohonen SOM

- En premier cas en fait une classification de la base de donnée dans une carte kohonen à bas de deux neurones

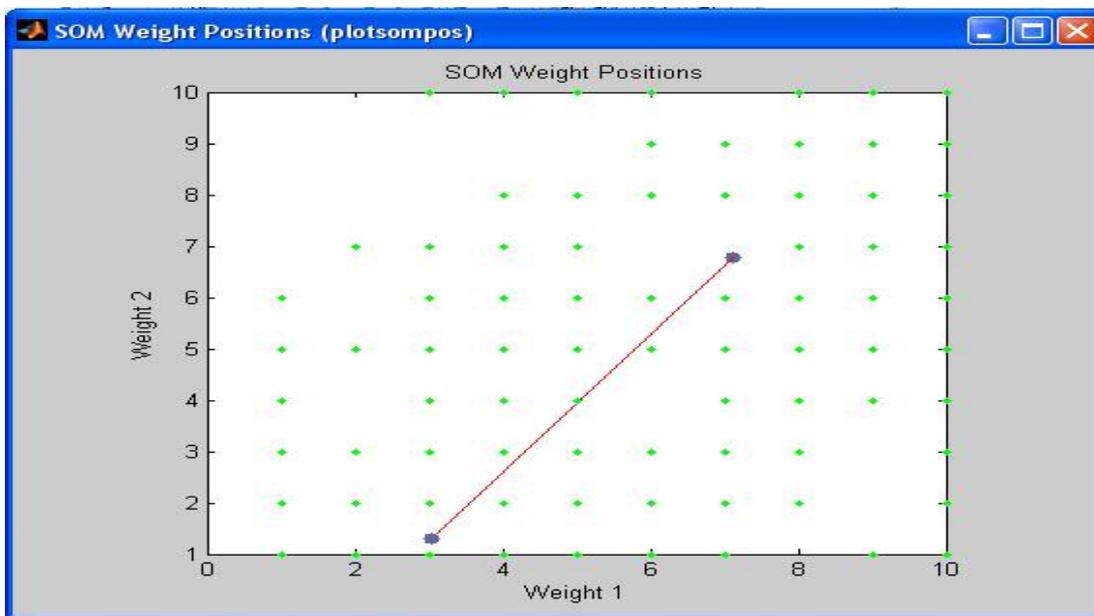


Figure 3(position des poids)

Les poids du premier neurone ayent valeurs de (3,1.3) les deux poids sont diffirant danc ce neurone et sensible au donnée ayant une grande diffirace mais appartient à la même classe c' est le cas d'une classe maligne.

Les poids de duxieme nerone ayent valeur de (7,7) les deux poids sans identique danc ce neurone et sensible au donnée ayant une léger diffirant et appartient à la même calasse c' est le casd'une classe benigne.

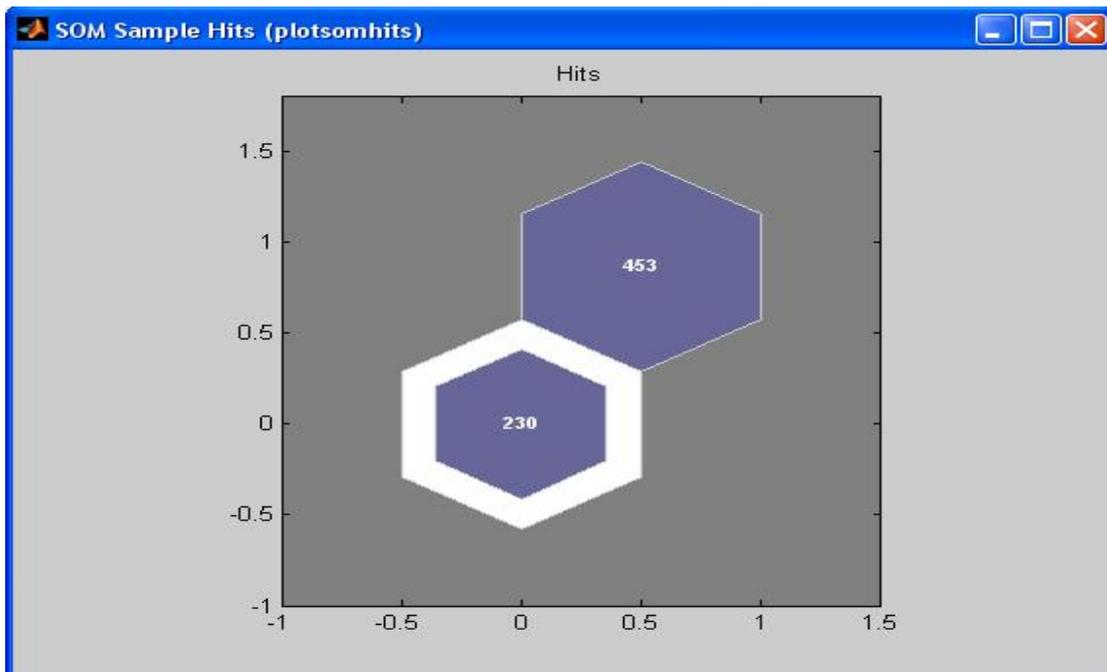


Figure 4(représentation des valeurs de chaque classe)

La figure 4 montre les valeurs de chaque classe ou la première classe contient 230 cas (classe maligne), et la deuxième classe contient 453 cas (classe bénigne).

La sensibilité ($Se = 0.9247$), la spécificité ($Sp = 0.9797$), nombre des vrais positifs ($Vp=221$), nombre des vrais négatifs ($Vn=435$), nombre des faux positifs ($Fp=9$), nombre des faux négatifs ($Fn=18$).

-En deuxième cas en fait une classification de la base de données dans une carte Kohonen au bas de quatre neurones

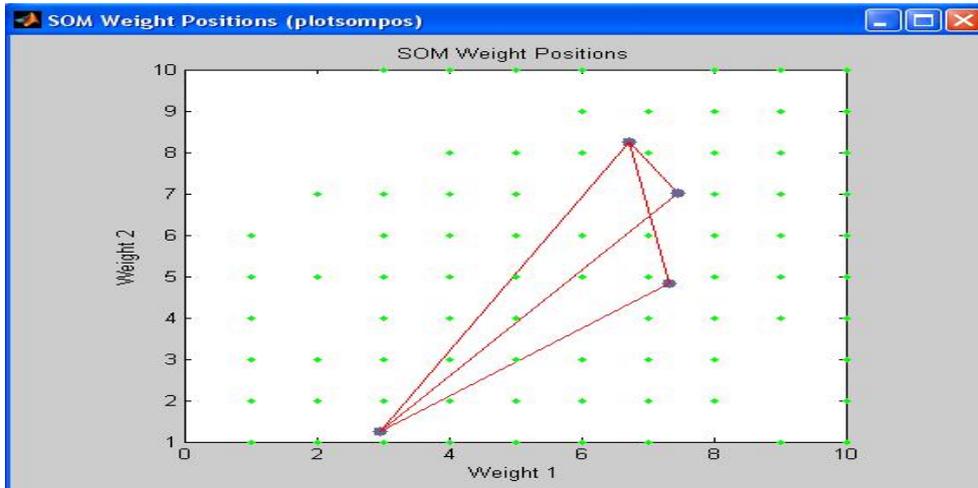


Figure 5 (représentation de position des pois)

La figure (5) montre les positions des poids pour les quatre neurones

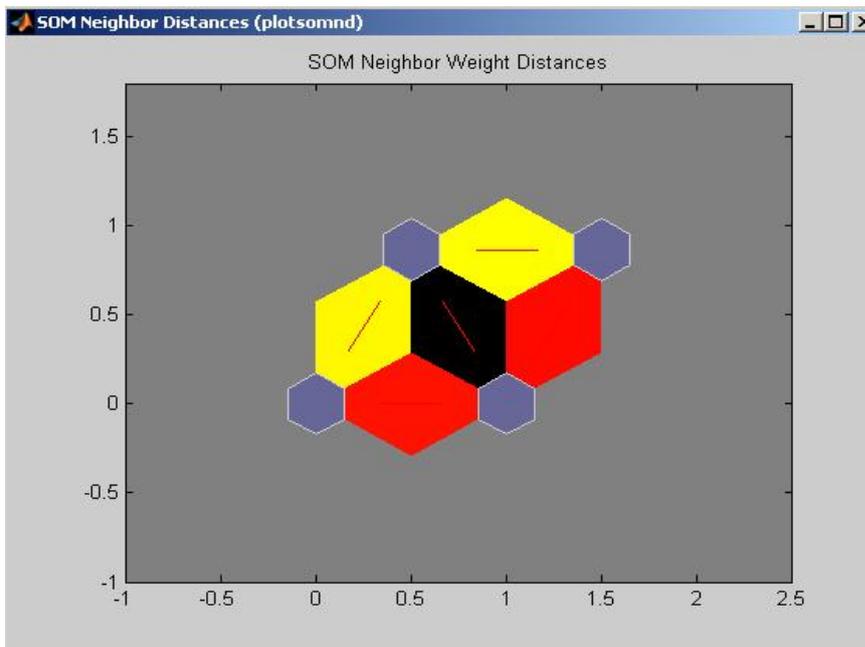


Figure 6 (distances entre les poids des quatre neurones)

La figure (6) montre les distances entre les poids des quatre neurones sous forme de la couleur Tel que la couleur j'aune signiphie que les deux neurone sont plus proche et la

couleur noire signifie que les deux neurones sont trop loin et la couleur rouge c'est une distance intermédiaire

Dans en remarquant que trois neurones appartiennent à la même classe et un autre neurone à une autre classe

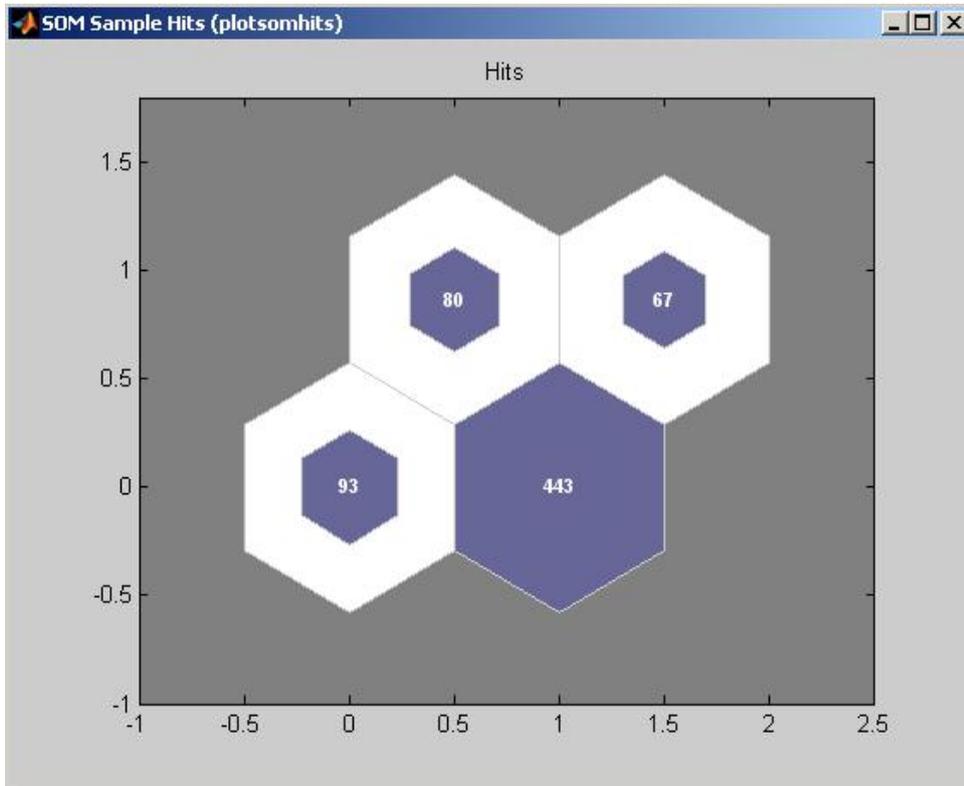


Figure 7(représentation des classes)

Figure (7) montre le nombre des cas dans chaque neurone puisque trois neurones forment une classe cette classe de 229 cas maligne et l'autre neurone forme une classe de 443 cas bénigne.

Nombre des vrais négatifs ($Vn = 433$)

Nombre des vrais positifs ($Vp = 229$)

Nombre des faux positifs ($Fp = 10$)

Nombre des faux négatifs ($Fn = 11$)

-Maintenant en prend d'échantillons pour l'entraînement de la carte kohonen et d'échantillons pour le test

Tableau 8 Résultat d'apprentissage par la carte kohonen SOM

Nombre des chantions		Résultat d'apprentissage						
Nombre d'échantillon d'entraînement	Nombre d'échantillon du test	Erreur d'entraînement	Erreur de test	spécificité	Sensibilité	Faux positif	Faux négatif	Vrai négatif
359	324	0,01	0,04	93,34	98,66	12	2	170
369	314	0,02	0,06	90,04	98,23	20	2	181
379	304	0,01	0,04	93,58	98,29	12	2	175
389	294	0,02	0,03	95,5	98,3	6	2	170
399	284	0,01	0,03	95,65	99	8	1	176
409	274	0,02	0,04	93,14	98,98	12	1	163
419	264	0,01	0,03	95,22	98,97	8	1	158
429	254	0,01	0,05	91,55	98	13	2	141
439	244	0,02	0,01	98,61	99	2	1	142
449	234	0,02	0,03	95,8	98,51	6	1	137
459	224	0,02	0,01	98,38	99	2	1	122
469	214	0,01	0,03	95,93	97,93	5	2	118
479	204	0,01	0,04	93,06	99	10	1	94
489	194	0,02	0,04	94,62	98,01	5	2	88
499	184	0,01	0,01	93,47	98,9	6	1	86
509	174	0,02	0,01	99	98,93	1	1	99
519	164	0,02	0,02	97,84	98,61	2	1	91
529	154	0,02	0,05	92,42	97,72	5	2	86
539	144	0,01	0,02	98,36	97,59	1	2	81
549	134	0,01	0	100	100	0	0	74

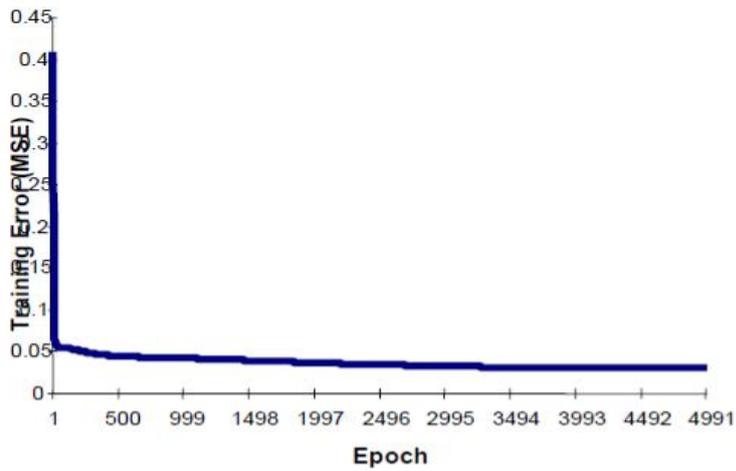


Figure 9: la valeur d'erreur par rapport au nombre d'epoch

Nos remarquons dans la figure (9) que l'erreur diminuée quand en augmente le nombre d'epoch

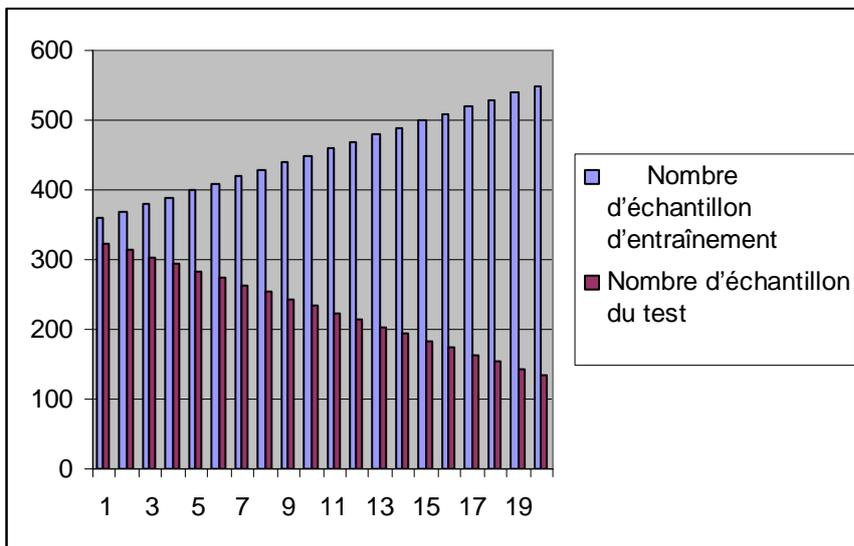


Figure 10 : histogramme (nombre d'échantillon d'entraînement, nombre d'échantillon test)

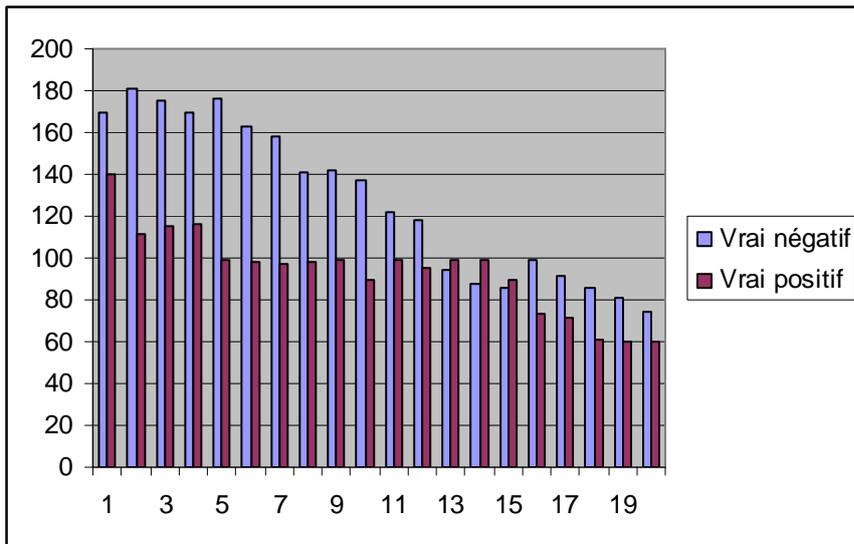


Figure 11 Histogramme (vrai négatif, vrai positif)

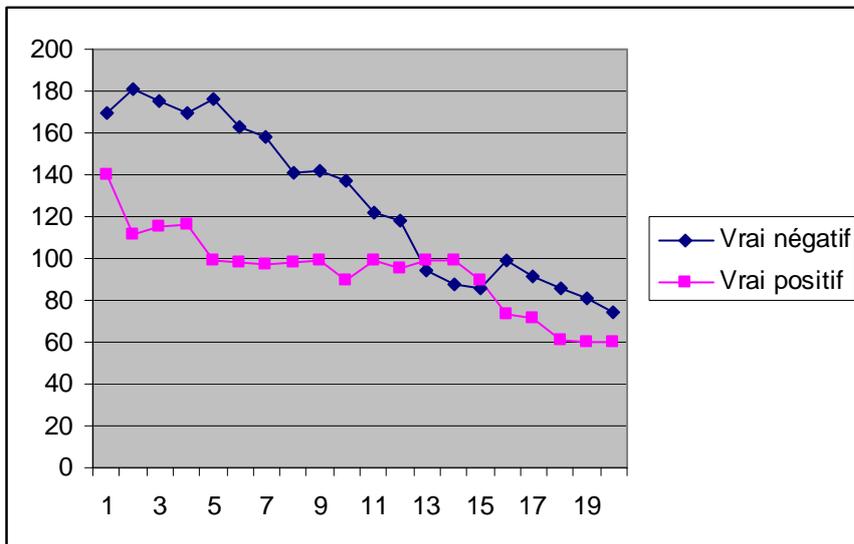


Figure 12 courbes (vrai négatif, vrai positif)

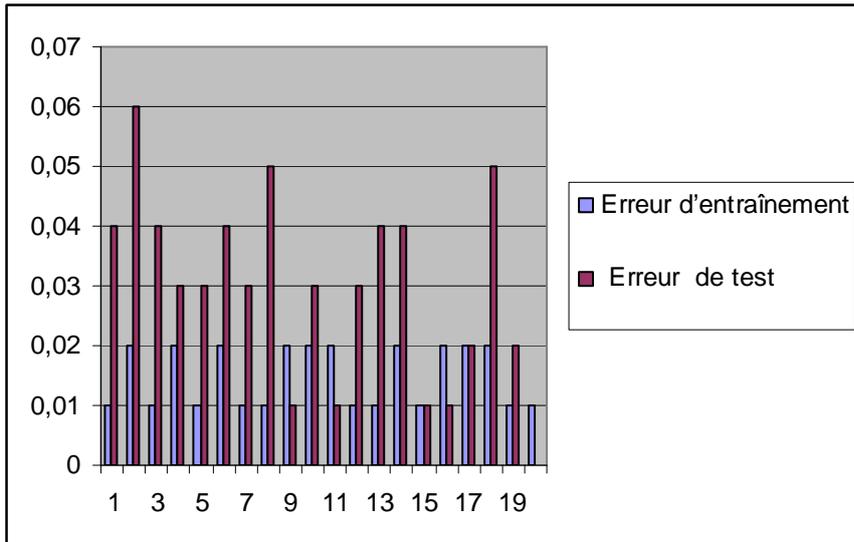


Figure 13 Histogramme (erreur d'entraînement, erreur de test)

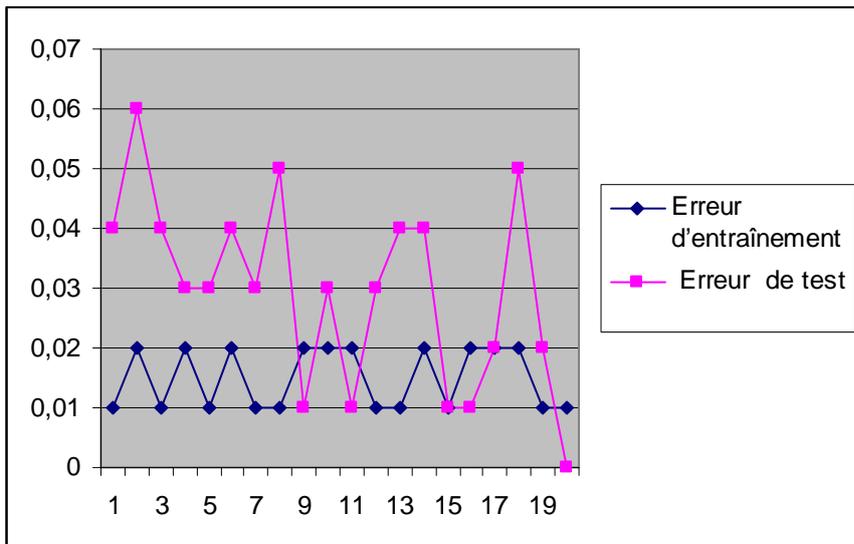


Figure 14 courbe (erreur d'entraînement, erreur de test)

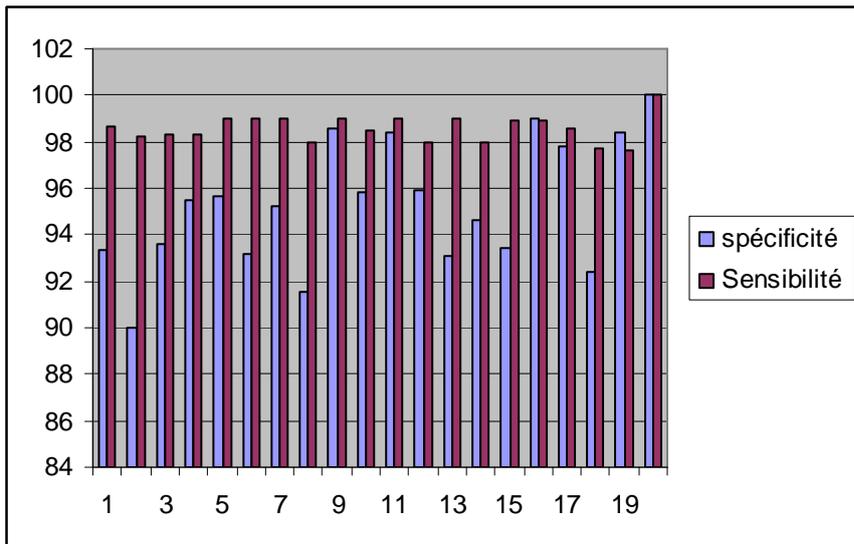


Figure 15 Histogramme (spécificité, sensibilité)

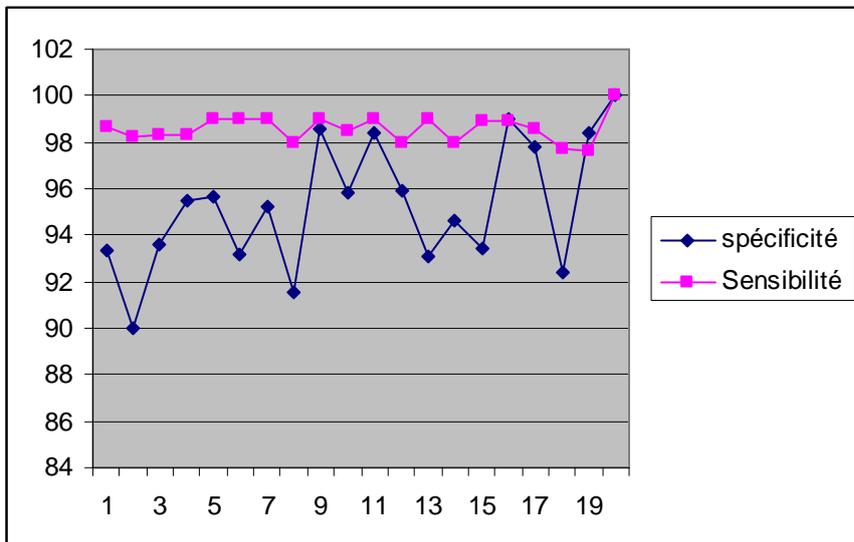


Figure 16 courbes (spécificité, sensibilité)

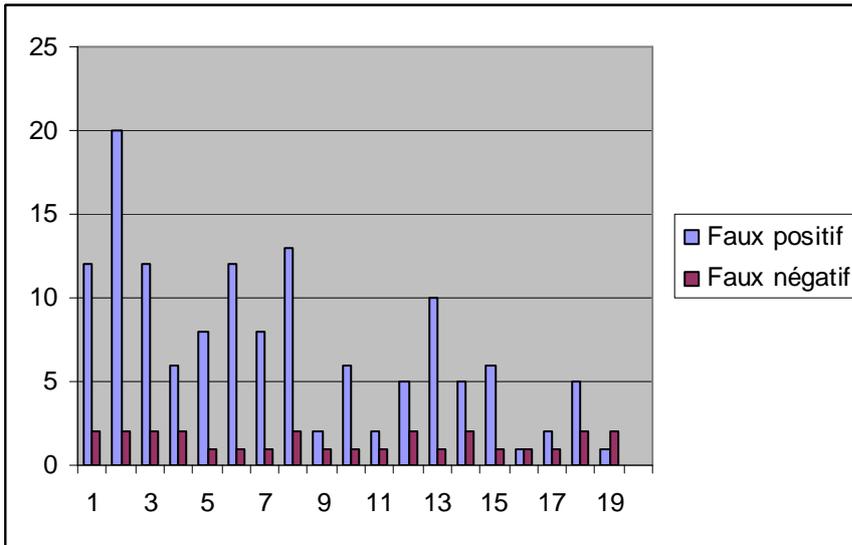


Figure 17 histogrammes (faux positif, faux négatif)

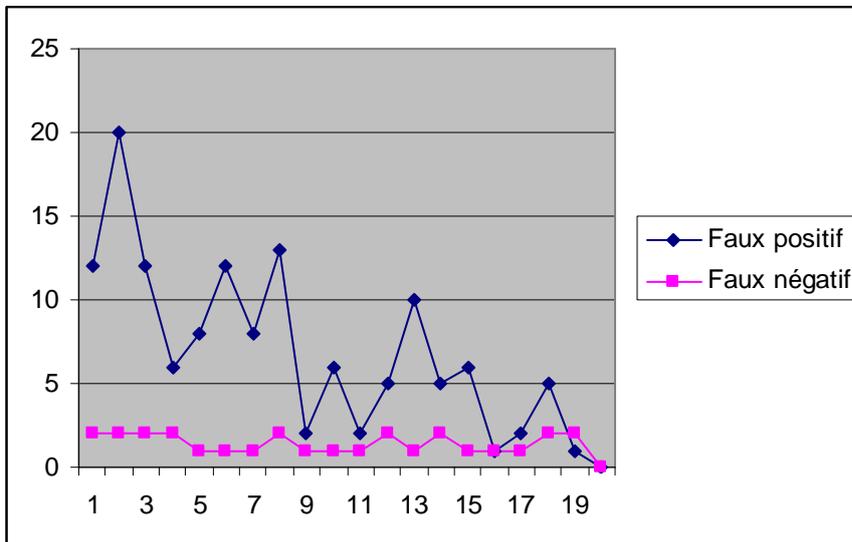


Figure 18 courbes (faux positif, faux négatif)

CONCLUSION

Nos remarquons que pendant les différentes itérations, l'erreur d'apprentissage (bleu) est largement inférieure à l'erreur du test nous lisons à partir de la figure 13, que l'erreur d'apprentissage e_{app} est limitée entre 10^{-2} et $2 \cdot 10^{-2}$ par contre l'erreur du test e_{test} est limitée entre 0 et $6 \cdot 10^{-2}$ et les fp plus grande que les fn figure (17,18).

Cette contribution est très logique pour que le classifieur (carte kohonen) est déjà entraîné sur les exemples d'apprentissages.

Concernant les performances du classifieur (carte kohonen) à savoir la sensibilité S_e et la spécificité S_p sont similaires avec une légère amélioration pour la sensibilité (voir figure 14,15).

Les résultats obtenus sont expliqués par le fait que le nombre des cas non pathologiques est plus grand par rapport aux cas pathologiques. Autre chose nous remarquons que la sensibilité est légèrement stable par rapport à la spécificité.

Conclusion générale et perspective

Dans ce mémoire, on s'est intéressé à l'étude des systèmes d'aide au diagnostic dédiés au dépistage du cancer du sein, notamment la classification des noyaux cellulaires. En effet, les statistiques confirment que le cancer du sein représente une menace prépondérante pour la vie de la femme.

Cependant, une telle menace n'est gérée que par la prise en charge rapide de la maladie pour maximiser les chances de survie. D'où, le recours au moyen informatique automatisant la procédure de diagnostic appelée: le Diagnostic Assisté par Ordinateur (DAO).

Ce travail de recherche a nécessité tout d'abord la maîtrise de certaines connaissances dans le domaine du cancer du sein. Nous avons, alors, étudié en détail la base de données nécessaire pour réaliser nos classifieurs.

Afin d'améliorer le résultat de classification des noyaux cellulaires (malignes ou bénignes), un état de l'art détaillé concernant les étapes d'un diagnostic assisté par ordinateur a été réalisé afin de réussir la classification des noyaux cellulaires. Dans ce contexte, une étude concernant les performances du classifieur à base de (fcm) à savoir la sensibilité S_e et la spécificité S_p , nous avons remarqué avec une légère amélioration pour la spécificité S_p .

Et Concernant les performances du classifieur (carte de kohonen) nous avons remarqué que S_e et S_p sont similaires avec une légère amélioration pour la sensibilité.

Les résultats obtenus sont expliqués par le fait que le nombre des cas non pathologiques est plus grand par rapport aux cas pathologiques.

La difficulté rencontrée concerne l'existence des noyaux cellulaires de la base de données pour appartenant à des classes différentes mais leurs positions sont tous rapprochées.

Ce problème cause des difficultés de reconnaissance pour la classification ce qui justifie énormément la présence des faux positifs (fp) et des faux négatifs (fn).

Il apparaît qu'une fusion d'entre l'approche FCM et l'approche basé sur Kohonen permet de résoudre et de diminuer le nombre des FPS et des FNS. Cette future vision peut être concrétisée dans ce travail.

Bibliographie

- [1] :http://www.doctissimo.fr/html/dossiers/cancer_sein/cancer_sein.htm
- [2] :<http://www.inserm.fr/thematiques/cancer/dossiers/cancer-du-sein>
- [3]<http://www.cancerdusein.org/>
- [4]<http://sante-az.aufeminin.com/w/sante/s201/maladies/cancer-sein.html>
- [5]home.deib.polimi.it/.../Clustering/.../cmeans.htm
- [6]www.sciencedirect.com/.../0098300484900207
- [7]www.mathworks.com/.../fuzzy/.../fuzzy-c-means
- [8]rss.acs.unt.edu/Rdoc/library/e1071/.../cmeans.html
- [9]ce.sharif.edu/~m.../y_fcmc_presentation_v0.8.ppt
- [10]networkcameras.free.fr/AutoOrganisation/.../kohonen.pdf
- [11]tutoriels-data-mining.blogspot.com/2008/.../les-cartes-de-kohonen.ht...
- [12]-PR Gérard Dreyfus (Apprentissage Statistique : Réseaux de neurones • Cartes topologiques Machines à vecteurs supports)
- [13]-PR Claude TOUZET (LES RESEAUX DE NEURONES ARTIFICIELS)
- [14]: M.Boughanem, Thèse de doctorat de l'université P.Sabatier, France 1992.
- Un modèle connexionniste pour la recherche d'information.
- [15]: E. Davalo, P. Naïm, Des Réseaux de neurones, Edition Eyrolles, 1993.
- [16]: J. A. Freeman, D. Skapma, Neural Networks, Algorithms, and programming technique, Addison-Wesley publishing company 1992

- [17]: J Herault, réseaux neurones et traitement de signal, Édition HERMES, 1994.
- [18]: T. Kohonen, Self-Organized Formation of topologically correct features maps, Biological Cybernetics, 1988.
- [19]: T. Kohonen, Self-Organising Maps, Edition Springer, second edition 1997.
- [20]: Xia Lin, D. Soergel, A self-organizing semantic map for information retrieval, Int AMC/SIGIR Conf on R&D information retrieval, pp 262-269, 1991
- [21]: M. Cord Marilyn, Nelson and W. T. Illengworth, A practical guide to neural nets, Addison-Wesley Publishing company 1991.
- [22] Breastcancer databases University of Wisconsin**
Hospitals, Madison from Dr. William H. Wolberg