

République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche scientifique
Université Abou BEKR BELKAID
Faculté de Technologie
Département de Génie Biomédical
LABORATOIRE DE GENIE BIOMEDICAL

THÈSE DE DOCTORAT

Présentée et soutenue publiquement par

BENCHAIB YASMINE

Spécialité : ELECTRONIQUE BIOMEDICALE

Sujet de Thèse :

Techniques Auto Améliorantes de l'Apprentissage des
Classifieurs : Application aux Données Médicales

Soutenue en Octobre 2015, devant le jury composé de :

Pr.BESSAID Abdelhafid	Université de Tlemcen	Président
Pr.ATMANI Baghdad	Université d'Oran	Examineur
Pr.RAHMOUN Abdellatif	E S I Sidi Bel Abbes	Examineur
Pr.CHIKH Mohamed Amine	Université de Tlemcen	Directeur de thèse

Année universitaire 2015-2016

Dédicaces

Je dédie cette thèse à ma très très chère maman que dieu lui préserve bonne santé et longue vie

A ma fille chérie Nawel

A mon mari

A mes deux chères sœurs ilhem et nawel et à leurs maris

A mon regretté père jamais oublié ,en témoignage de profonde affection et de vive reconnaissance

A mon cousin préféré Mourad, sa femme farida, ses enfants titae, malek et Elhadi

A feu mon oncle Mohamed

A mon oncle Bachir

A tous mes cousins et cousines

A toute ma famille et belle famille

A tous mes amis et collègues de l'université de Tlemcen.

Remerciements

Je remercie le dieu tout puissant de m'avoir donné le courage et la santé pour finir ce travail.

Je tiens à remercier Monsieur Chikh Mohamed Amine , Professeur à l'université de Tlemcen ,qui m'a encadré avec beaucoup d'efficacité et a été la source de précieux conseils et encouragements.

Mes remerciements s'adressent à Monsieur Bessaid Abdelhafid, Professeur à l'université de Tlemcen ,pour l'intérêt qu'il a accordé à ce travail en acceptant de le juger et de présider le jury.

Je tiens à remercier les membres du jury qui m'ont fait l'honneur de participer à l'examen de cette thèse :

Mr Atmani ,Professeur à l'université d'Oran.

Mr Rahmoun Professeur à l'université de Sidi BelAbbes.

Je tiens à remercier toute personne qui m'a soutenue et aidé de près ou de loin à la réalisation de ce travail.

Résumé

L'Algorithme de RétroPropagation ,ARP, est l'algorithme le plus connu et le plus utilisé pour l'apprentissage des Réseaux de Neurones Artificiels RNAs. L'ARP a su prouver son efficacité et sa réussite dans plusieurs domaines de la classification : en Médecine, bioinformatique, télécommunications, etc. Néanmoins, l'ARP présente quelques limitations qui empêchent d'atteindre un niveau d'efficacité optimal (problèmes de lenteur, de la convergence et de précision dans la classification). Les inconvénients de cet algorithme ont mené à un grand nombre de recherches pour essayer de surmonter ces problèmes. Cependant, malgré toutes les modifications et améliorations proposées pour l'ARP, il n'existe toujours pas de solution optimale, qui peut résoudre tous les problèmes.

Dans cette thèse de doctorat nous proposons un algorithme pour améliorer certaines de ces lacunes. L'algorithme proposé est un classifieur neuronal basé sur la propriété biologique de la Méta-plasticité.

La Méta-plasticité est un concept biologique largement utilisé dans les domaines de la biologie, de l'informatique médicale, des neurosciences, de physiologie, de neurologie et beaucoup d'autres. La Méta-plasticité est étroitement liée au processus de mémorisation et de l'apprentissage.

L'un des avantages majeur de l'algorithme de Méta-plasticité Artificielle est sa facilité d'implémentation dans les réseaux de neurones artificiels. Dans cette thèse l'algorithme est appliqué à un perceptron multi couches, MAPMC, (Méta-plasticité Artificielle appliquée sur un Perceptron Multi Couches) afin de modéliser la Méta-plasticité biologique.

De tous les modèles de méta-plasticité proposés jusqu'à maintenant, le plus efficace du point de vue de l'apprentissage et la performance est celui qui utilise la théorie d'information de Shannon, qui établit que les paramètres les moins fréquents contiennent plus d'information que les paramètres les plus fréquents. Le concept de méta-plasticité va être appliqué durant la phase d'apprentissage de telle sorte que les paramètres les moins fréquents auront des poids synaptiques plus forts que les paramètres les plus fréquents.

Cet algorithme est basé sur une fonction d'apprentissage qui modélise le fonctionnement biologique des synapses et de leur méta-plasticité. Cette fonction d'optimisation permet de modifier les poids synaptiques lors d'un processus d'apprentissage.

L'application du MAPMC vise à améliorer les résultats de l'apprentissage et par conséquent les performances de classification. Notre algorithme permet de pondérer les données les moins fréquentes (minoritaires) afin de mener un apprentissage convenable du classifieur.

L'algorithme proposé MAPMC est appliqué à différents problèmes de classification des données médicales. Le MAPMC a démontré une supériorité dans les résultats en terme de bonne classification, de vitesse de convergence, de fiabilité et une faible complexité dans le calcul, en comparaison avec d'autres algorithmes proposés dans la littérature et appliqués dans le même domaine.

Mots clés : Apprentissage, Amélioration, Algorithme de RétroPropagation, Méta-plasticité, Classification des données médicales.

Abstract

The backpropagation algorithm BPA is one of the most known and used algorithm to training the artificial neural networks, ANNs. The BPA has been success used in problems of pattern classification in areas such as: Medicine , Bioinformatics , telecommunications ,etc. however the BPA has some limitations that prevent to reach an optimal efficiency level (slowness problems, convergence and classification accuracy). These problems have provoked big number researches to improve the BPA. However, in general none of the modifications have been capable of delivering satisfactory performance for all problems.

In this doctoral thesis is proposed an alternative to improve some of the BPA deficiencies.the suggested algorithm , is a neuronal model based on the biological property of the Metaplasticity. The Metaplasticity is a biological concept widely known in the fields of biology , medical computer science , neuroscience, physiology , neurology and others. The Metaplasticity is related to the process of memory and of the learning.

The main advantage of the suggested Artificial Metaplasticity Algorithm,AMP ,is that , it is able implementing in any ANNs, in this thesis, algorithm was implemented in a multi layer perceptron MLP. The most efficient AMP model is the approach that connects metaplasticity and Shannon's information theory , which establishes that less frequent patterns carry more information than frequent patterns. This model defines artificial metaplasticity as a learning procedure that produces greater modifications in the synaptic weights with less frequent patterns , as a way of extracting more information from the former than from the latter.

The Artificial Metaplasticity on Multi Layer Perceptron AMMLP model was applied in the ANNs training phase. During the training phase, the AMMLP algorithm assigns higher values for updating the weights in the less frequent activations than in the more frequent ones. AMMLP achieves a more efficient training and improves MLP performance.

The suggested AMMLP algorithm was applied to different related problems to the medical classification. The AMMLP demonstrated the superiority in terms of classification accuracy, speed, reliability and low computational cost than others algorithms recently suggested by other researchers and applied in this area.

Keywords : Learning, Improvement, Back Propagation Algorithm, Metaplasticity, Medical data Classification.

الخوارزمية العكسية هي واحدة من الخوارزمية الأكثر المعروفة والمستخدمة لتدريب الشبكات العصبية الاصطناعية. أكدت الخوارزمية العكسية نجاحها في عدة مجالات التصنيف مثل: الطب، المعلوماتية الحيوية، والاتصالات، وغيرها. إلا أن بعض القيود تمنع الوصول إلى مستوى الكفاءة الأمثل (مشاكل بطء والتقارب ودقة التصنيف). وقد أثارت هذه المشاكل عدد كبير من البحوث لتحسين الخوارزمية العكسية. ورغم ذلك لم تصل إلى تقديم الأداء المرضي لجميع المشاكل. في هذه الأطروحة الدكتوراه نقترح بديلا لتحسين بعض النقايس. الخوارزمية المقترحة هي نموذج من الخلايا العصبية التي تستند على الطابع البيولوجي للميتابلاستيك.

الميتابلاستيك هو مفهوم بيولوجي معروف على نطاق واسع في مجالات البيولوجيا وعلوم الحاسوب الطبية، علم الأعصاب، علم وظائف الأعضاء وغيرها. ويرتبط الميتابلاستيك لعملية الذاكرة والتعلم في دماغ الإنسان.

الميزة الرئيسية للخوارزمية المقترحة الميتابلاستيك الاصطناعية، هي أنها سهلة التنفيذ في أي الشبكات العصبية الاصطناعية. في هذه الأطروحة، تم تنفيذ الخوارزمية في شبكة عصبية متعددة الطبقات. النموذج الأكثر كفاءة هو الذي يربط الميتابلاستيك ونظرية المعلومات شانون، التي تنص على أن أنماط أقل تواترا تحمل مزيد من المعلومات من أنماط متكررة. تم تطبيق الميتابلاستيك الاصطناعية على شبكة عصبية متعددة الطبقات في مرحلة تدريب الشبكات العصبية. خلال مرحلة التدريب، الخوارزمية تقوم بتعيين قيم أعلى لتحديث أوزان الأعصاب أقل عددا.

الخوارزمية المقترحة تحقق التدريب أكثر كفاءة وتحسن أداء الشبكة العصبية متعددة الطبقات.

تم تطبيق الخوارزمية المقترحة الميتابلاستيك الاصطناعية لمختلف المشاكل للتصنيف الطبي.

أظهرت الخوارزمية المقترحة تفوق من حيث دقة التصنيف والسرعة والتكلفة المنخفضة من خوارزميات حسابية مقترحة من قبل باحثين آخرين و مطبقين في نفس المجال.

كلمات البحث: التعلم، التحسين، الخوارزمية العكسية، الميتابلاستيك الاصطناعية، تصنيف المعطيات الطبية.

Table des matières

Dédicaces	i
Remerciements	ii
Résumé	iii
Abstract	iv
ملخص	v
Tables des Matières	vi
Tables des Figures	xii
Liste des Tableaux	xiv
Introduction Générale	1
1. Contexte général	2
2. Position du problème	2
3. Motivation et contribution de nos travaux	3
4. Organisation de la thèse.....	4
Chapitre 1 :Etat de l’art	5
1. Introduction	6
2. Etat de l’art	6
3. Conclusion	11
Chapitre 2 : Apprentissage Naturel et Métaplasticité	12
1. Introduction	13
2. Mémoire et Apprentissage	14
2.1. Principes de l’apprentissage naturel.....	14
2.2. Notions de mémoire	14
2.2.1. Mémoires à court terme	16
a) Mémoire sensorielle	16

b) Mémoire de travail	16
2.2.2 Mémoires à long terme	16
a) Mémoire déclarative	16
• Mémoire épisodique	16
• Mémoire sémantique	17
b) Mémoire non-déclarative	17
2.2.3 Passage de la mémoire à court terme à la mémoire à long terme	17
3. Réseaux de Neurones et Principes de la Plasticité	18
3.1. Plasticité des réseaux de neurones biologiques	18
3.2. La Transmission Neuronale	19
a) Principe	19
b) Intérêt des synapses	20
4. La Plasticité synaptique	21
4.1. Historique de la plasticité synaptique	21
4.2. Généralités sur la plasticité synaptique	21
4.3. La plasticité synaptique : substrat cellulaire de la mémoire	22
5. La Potentialisation à Long Terme (PLT)	23
5.1. Principe de la PLT	23
5.2. Processus de la PLT	24
6. La Depression à Long Terme (DLT)	27
6.1. Principe de la DLT	27
6.2. Mécanismes d'induction	28
7. Les types de la plasticité à long terme	30
7.1. PLT et DLT homosynaptiques	30
7.2. P LT et DLT associatives	31

8. Méta-plasticité	32
8.1. Concept de la Méta-plasticité	32
8.2. Méta-plasticité hétérosynaptique	33
8.3. Plasticité de la plasticité synaptique	34
9. Conclusion	35
Chapitre 3 : Apprentissage Artificiel et Méta-plasticité Artificielle	37
I. Apprentissage Artificiel des réseaux de neurones	38
1. Introduction	38
2. Le neurone artificiel	38
3. Apprentissage	39
4. l'Algorithme de RétroPropagation	42
II. Méta-plasticité Artificielle	44
1. Introduction	44
2. La Méta-plasticité et la théorie d'information de Shannon	44
3. Algorithme de rétropropagation et méta-plasticité artificielle	45
4. Implémentation du MAPMC dans la phase d'apprentissage du PMC	46
5. Description de l'algorithme MAPMC	49
6. Conclusion	50
Chapitre 4 Sélection de variables	51
1. Introduction	52
2. Extraction de caractéristiques	52
2.1 Méthodes linéaires	53
2.1.1 Analyse en Composantes Principales	54
2.1.2 AC Analyse des correspondances simples	54
2.1.3 ACM l'analyse des correspondances multiples	55

2.1.4 Analyse Linéaire Discriminante	55
2.1.5 Positionnement Multi-Dimensionnel	55
2.2. Méthodes non-linéaires	56
2.2.1 Isomap	56
2.2.2 Plongement localement linéaire	56
3 .Sélection de caractéristiques	56
3.1 Définition de la sélection	56
3.2 Processus général de la sélection de caractéristiques	57
3.2.1 La première étape	57
3.2.2 Procédure de recherche	57
3.2.3 Méthode d'évaluation	58
3.2.4 Critère d'arrêt	61
3.3 Présentation de quelques méthodes de sélection	61
3.3.1 Méthode Max-relevance, Min-Redundancy (mRMR)	61
3.3.2 Méthode FOCUS	62
3.3.3 Méthode Relief	62
3.3.4 Méthodes LVW et LVF	62
3.3.5 Méthodes SFS et SBS	63
3.3.6 Méthode Branch and Bound	64
3.3.7 Les algorithmes génétiques	64
3.3.4 Méthode SAC	65
4. Conclusion	65
Chapitre 5 :Résultats Expérimentaux et discussions	66
I. Introduction	67

II. 1 ^{ère} expérimentation : Classification des arythmies cardiaques	68
1. Introduction.....	68
2. Préparation de la base de données	69
2.1 Présentation du modèle de classification neuronale	69
2.2 Description de la base de données MIT-BIH	69
2.3 Sélection des descripteurs d'un cycle cardiaque	70
2.3.1 Introduction	70
2.3.2 Présentation des différents descripteurs	71
2.3.3 Sélection de la base d'exemples	74
3. Analyse de la base de données	75
3.1 Représentation boîte à moustaches ou diagramme en boîte	75
3.2. Représentation tridimensionnelle	77
4. Résultats	79
4.1 Dimensionnement du classifieur neuronal et choix des paramètres	79
4.2 Classification neuronale en utilisant deux fonctions d'apprentissage différentes	79
4.2.1 Apprentissage et test du classifieur PMC1	81
4.2.2 Apprentissage et test du classifieur PMC2.....	83
4.3 Sélection de l'architecture du classifieur MAPMC	84
4.4 Résultats du classifieur optimisé MAPMC	85
5. Discussion	87
6. Classification par MAPMC et Sélection de variables	88
6.1. Résultats de la classification avec sélection de variables	90
7. Conclusions	92
III. 2 ^{ème} expérimentation : Classification du cancer du sein	93
1. Introduction.....	93

2. Description de la base de données du cancer du sein(WBCD).....	94
3. Sélection de l'architecture du classifieur neuronal	95
4. Evaluation du classifieur optimisé MAPMC	97
4.1. Critères des performances	97
a) Matrice de confusion.....	97
b) Résultats de la courbe ROC	98
5. Comparaison avec l'état de l'art	100
6. Classification par MAPMC et Sélection de variables	101
7. Conclusion	102
Conclusion Générale	103
Conclusion et perspectives	104
Bibliographie	106
Annexe	121

Table des Figures

Figure 2.1 : Illustration d'un réseau de neurones dans le cerveau.....	15
Figure 2.2 : Similarité de forme entre l'hippocampe marin et l'hippocampe du cortex.....	17
Figure 2.3: Le neurone et la synapse.....	18
Figure 2.4: La plasticité synaptique.....	19
Figure 2.5: Connexion entre deux neurones.....	20
Figure 2.6: Synapse en fonctionnement normal.....	20
Figure 2.7: Principe de la Potentialisation à Long Terme PLT.....	24
Figure 2.8 :Etape1.....	25
Figure 2.9 :Etape2.....	25
Figure 2.10 :Etape3.....	25
Figure 2.11 :Etape4.....	26
Figure 2.12: Principe de la Dépression à Long Terme DLT.....	28
Figure 2.13: Mécanisme d'induction et d'expression de la PLT et de la DLT dans la région CA1 de l'hippocampe (Modèle de Lisman, 1989).....	29
Figure 2.14: Règles de plasticité synaptique.....	31
Figure 2.15 : Les changements dans la force synaptique dus à l'activité post-synaptique des neurones biologiques.....	33
Figure 2.16: Effet de la Méta-plasticité. Décalage vers la droite du seuil de la PLT en accord avec les poids initiaux des synapses.....	34
Figure2.17 : Le switch méta-plastique des récepteurs NMDA permet aux synapses fibres moussues hippocampiques d'exprimer la plasticité synaptique conventionnelle des synapses glutamatergiques.....	35
Figure 3.1 :Mise en correspondance : neurone biologique/neurone artificiel.....	38
Figure 3.2 : Schéma bloc du fonctionnement de l'algorithme MAPMC.la fonction de méta-plasticité est appliquée durant l'apprentissage.....	49
Figure4.1 : Processus de Sélection de variables.....	57
Figure 4.2 : La procédure du modèle « filter ».....	59
Figure 4.3 : La procédure du modèle "wrapper".....	60
Figure 5.1 : Protocole d'expérimentation.....	67
Figure 5.2 : Schéma global de la classification.....	69
Figure 5.3 : Interface de Mesure des Paramètres de l'ECG (IMPE).....	71

Figure 5.4 : Tracé ECG de l'enregistrement 119 de la base de données MIT-BIH montrant les caractéristiques d'une extrasystole ventriculaire ('V').....	72
Figure 5.5 : Mesure du temps d'apparition de la déflexion intrinsécoïde (DI).....	72
Figure 5.6 : Tracé ECG de l'enregistrement 212 de la base de données MIT-BIH montrant l'apparition de la déflexion intrinsécoïde (DI) dans le cas d'un bloc de branche droit.....	73
Figure 5.7 : signal ECG de l'enregistrement 214 de la base de données MIT-BIH montrant l'onde T négative et l'allongement de PR , QRS et QT dans le cas d'un bloc de branche gauche	73
Figure5.8(a) : Représentation de la durée P en fonction des classes.....	75
Figure 5.8(b) : Représentation de PR en fonction des classes.....	75
Figure 5.8 (c) : Représentation de la largeur du complexe QRS en fonction des classes.....	76
Figure 5.8 (d) : Représentation de ST en fonction des classes.....	76
Figure 5.8 (e) : Représentation de QT en fonction des classes.....	76
Figure 5.8 (f) : Représentation de RR précédent en fonction des classes.....	76
Figure 5.8 (g) : Représentation de RR suivant en fonction des classes.....	76
Figure 5.8 (h) : Représentation de RDI fonction des classes.....	76
Figure 5.8 (i) : Représentation de la durée du battement en fonction des classes.....	77
Figure 5.8 (j) : Représentation du rapport RRs/RRp en fonction des classes.....	77
Figure5.9(a) : Représentation 3D de la durée P et l'intervalle PR en fonction des classes.....	78
Figure 5.9(b) : Représentation 3D de la largeur du QRS et le RRs/RRp en fonction des classes.....	78
Figure 5.10 : exemple d'un faux négatif sur un enregistrement 212 de la base MIT-BIH.....	81
Figure 5.11: Les performances du PMC1 en utilisant les deux fonctions d'apprentissage 'trainbfg' et 'trainlm'.....	82
Figure 5.12: Les performances du PMC2 en utilisant les deux fonctions d'apprentissage 'trainbfg' et 'trainlm'.....	83
Figure5.13 : courbe ROC des deux classifieurs, nous pouvons remarquer clairement la supériorité des résultats du classifieur MAPMC par rapport à l'algorithme de rétropropagation classique.....	99

Liste des Tableaux

Tableau 4.1 : Différentes méthodes linéaires.....	54
Tableau 5.1 : Les différents descripteurs du vecteur d'entrée.....	74
Tableau 5.2 : Nombre de battements sélectionnés dans la base d'exemples.....	75
Tableau 5.3 : résultats des tests de performances du PMC1.....	82
Tableau 5.4 : résultats des tests de performances du PMC2.....	83
Tableau 5.5 : Résultats obtenus par le MAPMC pour différentes architectures proposées.....	85
Tableau 5.6 : Résultats obtenus par validation croisée pour l'algorithme de métaplasticité artificielle « MAPMC » et le perceptron multi couches classique « PMC »	85
Tableau 5.7 : Taux de classification correcte obtenus par notre algorithme et d'autres classifieurs de l'état de l'art.....	86
Tableau 5.8 : Classification des blocs de branche droit et gauche par le MAPMC et le PMC.....	87
Tableau 5.9 : meilleure performance des classifieurs PMC et MAPMC.....	87
Tableau 5.10 : performances en moyenne de 50 simulations des classifieurs PMC et MAPMC.....	87
Tableau 5.11 : Taux de classification des classifieurs PMC et MAPMC et temps d'exécution.....	90
Tableau 5.12 : Taux de classification correcte obtenus par notre algorithme et d'autres classifieurs de l'état de l'art.....	91
Tableau 5.13:Description de la base de données WBCD.....	95
Tableau 5.14: résultats obtenus par le MAPMC pour différentes architectures et différents paramètres de la métaplasticité.....	96
Tableau 5.15: Présentation des différents paramètres pour les deux classifieurs (MAPMC et PMC).....	96
Tableau 5.16 : matrice de confusion de la meilleure classification des deux classifieurs proposés.....	97
Tableau 5.17 : meilleur résultat pour chaque classifieur du cancer du sein.....	98
Tableau 5.18 : la moyenne des 50 simulations pour les deux classifieurs.....	98
Tableau 5.19 : Taux de classification correcte obtenus par notre algorithme et d'autres classifieurs de l'état de l'art.....	101
Tableau 5.20: comparaison de différentes méthodes de sélection de variables.....	101

Introduction Générale

Introduction Générale

Le thème de notre thèse de doctorat s'intitule Techniques Auto-améliorantes de l'Apprentissage des Classifieurs : Application aux Données Médicales.

1. Contexte général

Un diagnostic médical est le résultat du raisonnement d'un médecin avec une décision très souvent prise à partir de plusieurs informations.

Pour assurer un diagnostic exact, des systèmes automatisés sont proposés pour aider le médecin à réduire au minimum les erreurs possibles qui peuvent survenir pendant le diagnostic d'une maladie, et classer les patients suivant deux catégories malade et non malade, en analysant différentes valeurs des paramètres pris en considération pour le dépistage de cette maladie.

Des techniques telles que les réseaux de neurones sont le plus couramment utilisés dans les systèmes de classification et sont développés par le grand nombre d'équipes de recherche. Généralement la présence d'attributs redondants ou attributs fortement bruités dans les bases de données, diminue la performance des classifieurs.

Les techniques de sélection de variables permettent de choisir un ensemble optimal de caractéristiques les plus pertinentes pour faire un meilleur apprentissage et assurer une bonne performance du système de classification.

La mise en œuvre de tels systèmes d'aide au diagnostic est une tâche difficile, elle a été largement étudiée dans la littérature. Les difficultés se situent à différents niveaux comme par exemple la sélection des variables d'entrée, et le choix du classifieur le mieux adapté pour la classification des données médicales.

La variabilité des données médicales constitue un autre handicap pour les classifieurs, d'où le besoin d'avoir un bon algorithme d'apprentissage.

Le but d'une classification médicale de pathologies est de rassembler les cas qui ont des similitudes biologiques et qui sont susceptibles de partager certains facteurs étiopathologiques. L'identification des classes est importante car elle permet, d'une part de comprendre le processus de la maladie et d'autre part d'instaurer l'approche thérapeutique adéquate.

Les classifieurs doivent être en mesure de traiter des données hétérogènes caractérisant les différentes maladies. Nous sommes confrontés à une variabilité et une similarité de ces caractéristiques entre les différents patients.

La classification dépend fortement de l'application et il n'existe pas de solution générale mais un ensemble d'outils mathématiques et algorithmiques qu'il est possible de combiner pour résoudre des problèmes spécifiques.

2. Position du problème

Les modèles de classification sont des processus complexes composés d'une chaîne séquentielle de traitements devant aboutir à la reconnaissance d'un objet.

L'élaboration de tels modèles concerne plusieurs aspects comme : la caractérisation des données à l'entrée du classifieur, la sélection des variables les plus pertinentes pour mener à bien un apprentissage, et la dernière étape du processus le choix d'un bon modèle de classification.

Le classifieur doit être adaptable et robuste pour donner des résultats aussi stables que possible sur un échantillon de données tout en limitant le risque d'erreurs.

Dans le domaine médical, les médecins sont confrontés à des maladies qui n'apparaissent pas beaucoup mais qui sont très significatives et peuvent être fatales pour la vie du patient.

En général la plupart des bases de données médicales ne sont pas équilibrées. Le nombre des patients et les variables qui caractérisent les maladies changent. Nous sommes en présence d'un déséquilibre dans les données.

Dans le cadre de cette thèse nous proposons une solution pour aider les médecins à renforcer leur diagnostic, sachant que les maladies rares et atypiques sont généralement non identifiées par les classifieurs, ce qui constitue un danger réel pour la vie du patient, nous citons en particulier les arythmies rares qui sont à l'origine des morts subites.

Les réseaux de neurones artificiels RNAs sont très sollicités dans le domaine de la classification et en particulier la classification des données médicales. L'algorithme d'apprentissage le plus utilisé est la rétropropagation de gradient, il permet de modifier les paramètres du RNA (poids, biais,...) en minimisant une fonction de coût suivant une erreur quadratique.

Néanmoins cet algorithme présente quelques limitations qui empêchent d'atteindre un niveau optimal de performance. L'apprentissage prend souvent beaucoup de temps pour converger et peut tomber dans les minimums locaux sans atteindre le minimum global. Pour résoudre quelques inconvénients plusieurs recherches ont été développées comme modifications ou variantes de cet algorithme.

L'étape de sélection de variables est très importante pour assurer un apprentissage robuste et aussi pour mener à une classification fiable. Elle est utilisée généralement en phase de prétraitement. Les méthodes de sélection de caractéristiques sont généralement classées en deux groupes: les méthodes du type Filter et les méthodes du type Wrapper.

Une bonne sélection de variables pertinentes nous permet de gagner en complexité, temps d'exécution et simplicité d'interprétation des résultats.

3. Motivation et contribution de nos travaux

Notre objectif dans le cadre de cette thèse est l'amélioration du temps d'apprentissage et du taux de classification d'un classifieur neuronal.

Les travaux cités dans la littérature pour lever les limitations de l'algorithme de rétropropagation du gradient sont scindés en deux catégories.

La première catégorie comprend le développement de techniques heuristiques, basées sur l'étude des propriétés des performances de l'ARP en général. Ces techniques heuristiques concernent en particulier le taux d'apprentissage, le momentum (terme d'inertie ajouté pour sortir des minimums locaux), l'ajustement du gain de la fonction d'activation et l'optimisation de la topologie du réseau de neurones.

La deuxième catégorie concerne le développement d'algorithmes d'optimisation de plus en plus sophistiqués, basés sur des techniques d'optimisation numériques pour améliorer l'apprentissage de l'ARP.

Dans cette thèse nous proposons un algorithme pour améliorer quelques limitations de l'ARP. L'algorithme proposé est basé sur la Métaplasticité Artificielle et appliqué à un Perceptron Multi Couches « MAPMC ».

Comme nous travaillons avec un classifieur neuronal qui est un modèle inspiré des réseaux de neurones biologiques du cerveau humain, l'idée de rechercher une solution pour l'amélioration de l'apprentissage inspirée de la biologie est très intéressante et justifiable. Ce concept de métaplasticité biologique prend en considération les données minoritaires et accélère le temps d'apprentissage et de mémorisation.

Sachant que notre algorithme ne pourra pas trouver des solutions à toutes les limitations de l'algorithme de rétropropagation, néanmoins nous le proposons pour améliorer le temps d'apprentissage et de convergence, pour avoir une bonne généralisation et traiter les données d'apprentissage avec des poids proportionnels pour atteindre un taux de bonne classification très satisfaisant.

L'algorithme de métaplasticité permet de modifier la distribution des données lors de la phase d'apprentissage, il donne plus de poids aux paramètres les moins fréquents et moins de poids pour les plus fréquents donc il permet de pondérer les données les moins fréquentes (minoritaires) afin de mener un apprentissage convenable du classifieur.

Notre dernière contribution concerne la sélection de variables pertinentes pour aboutir à un bon apprentissage, et réduire le temps d'exécution et aussi pour améliorer le taux de classification.

4. Organisation de la thèse

Ce document est organisé en 5 chapitres .dans le chapitre 1, nous présentons un état de l'art des techniques d'améliorations de l'algorithme de rétropropagation .Dans le chapitre 2, nous expliquons la propriété biologique de la métaplasticité, les concepts et les mécanismes qui participent à sa génération. Nous démontrons aussi la relation entre plasticité et métaplasticité.

Dans ce chapitre, nous commençons par examiner quelques notions biologiques relatives au cerveau, et à ses constituants les neurones et leurs synapses. L'organisation en réseaux des neurones permet d'illustrer les notions d'apprentissage et de mémorisation (modification des connexions). Ces données nous sont nécessaires pour aborder le chapitre suivant qui montre le passage des modèles de réseaux neuronaux biologiques à des modèles mathématiques : les réseaux de neurones artificiels.

Le chapitre 3 est consacré à notre travail de modélisation de la métaplasticité et son implémentation durant la phase d'apprentissage.

Dans Le chapitre 4, nous soulignons l'importance de l'utilisation d'une méthode de sélection de variables dans le domaine de la classification des données médicales. Nous présentons les plus importantes méthodes citées dans la littérature et nous expliquons notre choix de la méthode de sélection séquentielle.

Le chapitre 5 est dédié aux deux expérimentations menées sur deux bases données médicales.

La première concerne la classification des arythmies cardiaques en utilisant la base de données MIT-BIH,et la deuxième la classification du cancer du sein en utilisant la base de données WBCD.

Dans cette partie nous démontrons nos contributions.ces deux expérimentation sont organisés de la même manière : une classification sans étape de sélection de variables et une classification avec sélection pour pouvoir démontrer l'amélioration des performances du classifieur.

Comme tout travail qui s'achève, nous concluons par un bilan général sur l'apport de nos contributions et nous indiquerons les possibles améliorations et perspectives pour la continuation de recherches sur la thématique abordée dans ce travail.

Chapitre 1

Etat de l'Art

Chapitre 1 : Etat de l'Art

1. Introduction

Dans cette thèse nous proposons un classifieur capable de fournir une aide aux spécialistes dans le domaine médical .Ce classifieur est basé sur les réseaux de neurones artificiels.

La méthode la plus utilisée pour l'apprentissage des réseaux de neurones est la rétropropagation ,qui repose simplement sur la technique de calcul du gradient, appliquée à toute fonction dérivable.

La principale motivation de notre thèse est l'amélioration du temps d'apprentissage et du taux de classification d'un classifieur neuronal en utilisant l'algorithme de rétropropagation du gradient (l'ARP). Les inconvénients de cet algorithme ont mené à un grand nombre de recherches pour essayer de surmonter ces problèmes. Cependant, malgré toutes les modifications et améliorations proposées pour l'ARP, il n'existe toujours pas de solution optimale , qui peut résoudre tous les problèmes.

Dans ce travail nous proposons un algorithme pour améliorer certaines de ces lacunes. L'algorithme proposé est basé sur la métaplasticité artificielle et appliqué à un perceptron multi couches MAPMC.

Nous allons présenter dans ce chapitre quelques méthodes proposées dans la littérature pour améliorer l'apprentissage des réseaux de neurones en utilisant l'algorithme de rétropropagation du gradient.

2. Etat de l'art

En 1990 ,Leonard et Kramer ont développé une méthode basée sur le gradient conjugué ,son grand intérêt pratique du point de vue du temps de calcul vient de ce qu'une initialisation *astucieuse* (dite « préconditionnement ») qui permet d'aboutir en seulement quelques itérations à une estimation très proche de la solution exacte.ils ont démontré que le gradient conjugué peut être utilisé pour le réglage dynamique du taux d'apprentissage et améliorer les performances de classification.[Leo et Kra'90].

En 1991, Lee et al ont proposé que pour améliorer l'apprentissage des RNA,il était nécessaire de faire appel aux systèmes experts .ils ont conclu que l'utilisation des systèmes experts pour superviser l'apprentissage des RNA est plus efficace qu'un apprentissage des RNA seul.(en soi)[Lee et Wei'91].Kim et Ra ont formulé un algorithme nommé Dynamic of decision boundaries, pour choisir les valeurs initiales des poids ,qui sont obtenues de la règle delta généralisée, de sorte que la stabilité et la vitesse du processus d'apprentissage soit amélioré en éliminant les effets de retard et de la convergence de l'ARP.[Kim et Ra'91].

En 1992 ,Scalero ont développé un algorithme pour améliorer l'ARP basé sur la minimisation de l'erreur quadratique moyenne (MSE,Mean Square Error) ;entre la sortie désirée et la sortie réelle par rapport à la somme des sorties . Cela s'oppose avec l'ARP standard qui minimise l'erreur quadratique moyenne par rapport aux poids[Sca et Tep'92]. Karayiannis propose un critère général pour l'apprentissage des RNA progressifs (FFNN,Feed forward Neural Network),mono couche et multi couches basé sur la règle delta.ces algorithmes sont nommés Efficient Learning Algorithms ELEANNE[Kar et Ven'92].

En 1993,Anand et al ont analysé le problème de la lenteur de convergence de l'ARP pour les problèmes de deux classes avec des ensembles d'apprentissage déséquilibrés.ils proposent un algorithme Descent Vector qui calcule un vecteur de descente de gradient , dans le but de minimiser les erreurs entre les deux classes. Par conséquent,

les erreurs de réseau pour les deux classes dominantes et subordonnées diminuent par le changement des poids dans la direction du vecteur descendant[Ana et al'93].

En 1994,Riedmiller a procédé à un examen général de plusieurs algorithmes ,toutes les approches décrites dans cet article font en quelque sorte l'utilisation de la dérivée première de chaque partie en poids par rapport à l'erreur globale du réseau[Rie'94].

Chen et jain proposent un algorithme nommé Robust BP algorithm.(RBPA) .cet algorithme est résistant aux effets de bruit et est capable de rejeter la majeure partie des erreurs pendant le processus d'apprentissage [Che et Jai'94].

Selon Riedmiller ,cet algorithme présente Trois avantages distincts par rapport l'ARP standard :

- a)le RBPA utilise la corrélation entre les paramètres en vue d'interpoler les échantillons d'apprentissage .
- b) est robuste face aux erreurs
- c) le taux de convergence est amélioré car il supprime l'influence des échantillons erronés.

En1995 Alpsan et al ont réalisé une étude comparative de différentes méthodes proposées pour améliorer l'ARP appliqué à des problèmes médicaux réels, basée sur des méthodes heuristiques et d'optimisation. Alpsan et al ont conclu que l'ARP standard peut être assez rapide et peut donner une bonne généralisation en fonction du type du problème. dans le cas particulier de problèmes médicaux ,il est très important d'avoir une grande généralisation avec une tolérance d'erreur minimale ;pour que le système soit aussi fiable que possible[Alp et al'95].

En 1996 Solomon et al proposent un nouvel algorithme génétique, basé sur l'auto adaptation dynamique (Dynamic self adaptation ,DS) pour améliorer et accélérer l'apprentissage de l'ARP.DS algorithme, prend la valeur du taux d'apprentissage(LR learning rate) de l'étape précédente, puis il le fait varier en l'augmentant et diminuant légèrement, et évalue la fonction de coût pour les nouvelles valeurs de la vitesse d'apprentissage, et la plus petite valeur de la fonction de coût est sélectionnée[Sol et Van'96].

Fu et al présentent une nouvelle méthode d'apprentissage incrémental pour la reconnaissance de formes.IBPLN (Incremental BackPropagation Learning Network) ;elle emploie une modification des poids et une adaptation structurelle des règles d'apprentissage, et applique les connaissances initiales nécessaires pour arrêter le processus d'apprentissage [Fu et al'96].

En 1997 ;Magoulas et al ont élaboré une méthode BPVS (Back Propagation Variable Stepsize)basée sur une modification déterministe de la descente la plus raide(deepest descent) permettant de varier le pas ,comme conséquence de la minimisation de la fonction objective et de l'observation de la trajectoire dans l'espace des poids.ils ont obtenus de bons résultats dans différents types de problèmes[Mag et al'97].

Yam et al,ont formulé une nouvelle approche basée sur l'erreur quadratique.(Least squares) pour calculer les poids initiaux optimaux des RNAs.une fois les poids déterminés , l'erreur initiale est sensiblement plus petite et par conséquent le nombre d'itérations nécessaires pour atteindre le critère d'erreur établi est réduit[Yam et al'97].

En1998 ; Sexton et al ont montré que les restrictions imposées par certains chercheurs dans l'espace de recherche ou la restructuration de l'architecture des RNA ; ne sont pas nécessaires si une architecture initiale suffisamment complexe et un algorithme de recherche global approprié est utilisé. Sexton et al ont utilisé les AG afin d'atteindre une meilleure généralisation de l'ARP [Sex et al'98(a)]. Durant la même année Sexton et al ont présenté un autre algorithme, et comme dans l'article précédent ils utilisent des algorithmes de recherche globale. Cette fois, ils appliquent une version plus généralisée de l'algorithme Tabu Search,TS, comme une alternative possible à la problématique de l'ARP[Sex et al'98(b)].

En 1999, Kamarthi et Pittne ; proposent une amélioration de l'algorithme de rétropropagation pour l'apprentissage des réseaux Feed Forward, cette méthode est connue sous le nom de BPWE , Back Propagation by Weight Extrapolation. Cet algorithme est basé sur le concept d'extrapolation des poids calculés par le réseau. En extrapolant les poids, il est possible de diminuer le nombre d'itérations requises pour l'apprentissage du réseau avant d'atteindre un vecteur de poids convenable [Kam et Pit'99]. Lehtokangas, présente une méthode nommée Constructive Backpropagation ,CBP,cet algorithme est inspiré de l'apprentissage de corrélation en cascade. L'algorithme CBP commence avec un petit réseau puis ils rajoutent des couches cachées jusqu'à arriver à une solution satisfaisante [Leh'99]. Cho et Chow ont formulé un algorithme d'apprentissage hybride, rapide et robuste , basé sur la méthode des moindres carrés et de la recherche par pénalisation (Least Squares and Penalty),LS-PEN.la méthode LS est utilisée pour déterminer les poids des connexions entre la couche de sortie et la couche cachée.la méthode d'optimisation de pénalisation est utilisée pour évaluer les poids entre la couche d'entrée et la couche cachée[Cho et Chow'99].Ampazisa et al ,proposent un modèle dynamique du système, qui permet d'accélérer l'apprentissage et réduire le temps d'apprentissage. Pour cela ils ont utilisé des méthodes d'optimisation avec des restrictions qui atteignent la minimisation simultanée de la fonction de coût et la maximisation des grandes valeurs de la matrice Jacobienne, de telle manière à éviter que le réseau puisse tomber dans un minimum local et donc le temps d'apprentissage sera réduit[Amp et al'99].

En 2000, Yam et Chow ,ont présenté un algorithme pour déterminer les poids initiaux optimaux des réseaux FeedForward basé sur l'inégalité de Cauchy et une méthode algébrique linéaire. En utilisant la factorisation QR ;ils obtiennent une solution considérée comme la meilleure dans les cas des moindres carrés.les auteurs ont démontré une supériorité du temps de convergence du réseau.[Yam et Cho'00].Chaudhuri et Bhattacharya , ont proposé une méthode pour accélérer la rapidité de convergence de l'algorithme de rétropropagation , basée sur une sélection intelligente des échantillons d'apprentissage. Cette méthode n'implique aucune modification dans l'algorithme de rétropropagation. Elle est appliquée dans des cas de classification assez complexe où les classes ne sont pas linéairement séparables. leurs résultats ont démontré que les performances du Perceptron Multi couches ont été considérablement améliorées [Cha et Bha'00].

En 2001 ; Lee et al, proposent une méthode alternative à la descente de gradient appelée Error Saturation Prevention, ESP, pour prévention de la saturation de l'erreur dans les nœuds de la couche de sortie, et aussi dans la couche cachée pour ajuster le taux d'apprentissage. L'ESP ,améliore l'apprentissage et maintient un sens sémantique de la fonction MSE utilisée pour l'évaluation du critère d'erreur[Lee et al'01].

En 2002 ;Mandische ,propose une méthode évolutive d'apprentissage en utilisant les stratégies évolutives (ES, Evolution Strategies) , qui remplace la technique du gradient utilisée pour l'apprentissage des réseaux de neurones. Un avantage de cet algorithme sans gradient peut être utilisé dans les réseaux avec des fonctions d'activation différenciables[Man'02].

Hoo et al ,utilisent l'information obtenue par F-PM (First-Principle Model) ;pour de donner un sens à la direction pour l'estimation du RNA. ceci est obtenu par la modification de la fonction objective afin d'inclure un terme additionnel qui est la différence entre la dérivée des résultats ,estimée par le réseau, et des sorties du modèle durant l'étape d'apprentissage[Hoo et'02].

En 2003, Eom et al ont proposé une méthode appelée Fuzzy General Backpropagation ,FGBP, pour améliorer les performances de l'algorithme de rétropropagation. Cette méthode utilise un système de logique floue qui ajuste

automatiquement le paramètre de gain de la fonction d'activation, par rapport à un ensemble de règles heuristiques du domaine du problème donné grâce à l'étude d'une simulation préliminaire[Eom et al'03].

Zweiri et al, ont introduit un nouveau paramètre nommé Proportional Factor PF, en plus des paramètres du taux d'apprentissage et le facteur de momentum utilisés habituellement. Donc ils utilisent trois paramètres dans l'algorithme de rétropropagation (Three-term BP algorithm). L'algorithme ci-dessus est plus robuste pour le choix des poids initiaux, en particulier lorsque des valeurs relativement élevées pour les paramètres d'apprentissage sont choisies, cet algorithme est également applicable pour n'importe quel réseau avec des fonctions d'activation différentes [Zwe et al'03].

En 2004, Abraham a présenté un algorithme basé sur les algorithmes évolutionnaires, nommé MLEANN « Meta Learning Evolutionary Artificial Neural Network », pour une optimisation adaptative des RNAs, où l'architecture, la fonction d'activation, les poids, l'algorithme d'apprentissage et ses paramètres sont adaptés en fonction du problème[Abr'04]. Wang et al, présentent un algorithme dans lequel chaque paramètre d'apprentissage a ses propres fonctions d'activation de neurones dans la couches cachées. Les fonctions d'activation sont ajustées pour l'adaptation des paramètres de gain durant le processus d'apprentissage. Ces ajustements sont fait pour éviter que le réseau ne tombe dans un minimum local [Wan et al'04]. Mohammad et Pavesié ont proposé une amélioration de l'ARP nommé BackPropagation with Selective Training, BST, et appliqué à l'apprentissage d'un réseau à fonctions à bases radiales, RBFN(Radial Basis Function Networks) ce qui a amélioré les performances du RBF, en terme de rapidité de convergence et en terme de taux d'erreur. Donc le BST a résolu trois problèmes de l'ARP :le surapprentissage, le temps de convergence et l'incapacité d'apprentissage avec un nombre minimal de paramètres. [Moh et Pav'04].

En 2005 Pernia-Espinoza et al ;proposent un algorithme d'amélioration, appelé TAO-Robust BackPropagation Learning Algorithm. Ce dernier utilise un estimateur échelle, qui est variable et il dépend d'une fonction Huber des erreurs obtenues dans chaque itération, avec cette approche ils obtiennent 2 propriétés importantes : robustesse face aux données aberrantes dans la base de données avec point de rupture élevée et un modèle normal de haute efficacité[Per et al'05].

En 2006, Steil présente un algorithme nommé BackPropagation Decorrelation, BPDC, pour tester et contrôler la stabilité pour les grands réseaux où seule la couche de sortie est adaptée.cet algorithme combine trois principes :une étape de rétropropagation des erreurs, l'utilisation de la mémoire tampon dynamique et adaptée en se basant sur la dé-corrélation des fonctions d'activation et dernier point concerne l'implémentation de neurones qui ne sont pas adaptatifs pour réduire la complexité du modèle[Ste'06].

Behera et al, proposent deux nouveaux algorithmes d'apprentissage LF I et LF II pour la modification des poids des réseaux FeedForward en se basant sur la fonction de Lyapunov.la contribution de cette recherche était de démontrer les similarités entre les algorithmes LF I, II et l'ARP. Ils ont démontré que ces algorithmes ont la même structure que l'ARP,seule différence que le taux d'apprentissage de l'ARP est fixe et dans les algorithmes LF, le taux d'apprentissage est adaptatif [Beh et al'06].

En 2007, Wang et al, proposent un modèle interactif pour améliorer l'apprentissage de l'ARP. le modèle combine avec succès une nouvelle technique pour ajuster le taux d'apprentissage nommé Individual Inference Adjusting Learning Rate technique(IIALR), et une nouvelle manière de mettre à jour la fréquence des poids, nommé The Batch mode of Weight Updating Frequency BOWUF[Wan et al'07].

En 2008, Kashman , présente une modification de l'algorithme d'apprentissage de rétropropagation ,BackPropagation Emocional ,EmBP. L'algorithme est basé sur deux émotions qui peuvent affecter l'apprentissage, qui sont l'inquiétude et la confiance. Lorsqu'une nouvelle tâche est apprise, le niveau d'inquiétude est élevé au début et le niveau de confiance est bas. après un temps d'apprentissage, le niveau d'inquiétude diminue et le niveau de confiance augmente. Pour cette raison l'algorithme EmBP utilise les poids d'émotions pour modifier les poids dans la phase d'apprentissage [Kas'08]. Yang et al , proposent que l'intervalle des valeurs initiales des poids quand un algorithme est entraîné avec la division complexe (Split Complex BackPropagation ,SCBP) devrait être meilleure .l'utilisation de ce critère peut réduire le déséquilibre des poids et des biais durant l'apprentissage, aussi éviter que les performances du SPBP soient dépendantes des poids initiaux[Yan et al'08].

Zhang et al , pour améliorer l'efficacité des réseaux Feed Forward ont proposé un algorithme nommé Product Unit Neural Networks with Exponential Weights ,PUNNs[Zha et al'08].Silva et al ont formulé une nouvelle fonction d'erreur E_{EXP} ,inspiré de la « Error Density at the Origin », Z-EDM, qui est capable de simuler le fonctionnement d'autres fonctions d'erreur pour l'ajustement d'un seul paramètre avec des valeurs réelles [Sil et al'08].Soliman et Mohamed , proposent une version modifiée de l'ARP sur la base de la multiplication de matrices pour le traitement parallèle. pour cela, ils ont mis en œuvre la Matrix Back Propagation ,MBP ,qui utilise un ensemble d'architectures de matrices scalaires et vectorielles (Matrix Instruction Set Architectures,ISAs) [Sol et Moh'08].

En 2009, Cheng et Park , ont développé un algorithme pour améliorer les performances de l'ARP,appelé Learning Phase Evaluation BackPropagation neural network, LPEBP. Cet algorithme divise la phase d'apprentissage en plusieurs étapes et évalue les effets de chaque phase.ils appliquent aussi une technique pour réduire la dimension et la construction sémantique de ses termes d'apprentissage, nommée Singular Value decomposition (SVD)[Che et Par'09].

Kathirvalavakumar et Jayaseeli , présentent un algorithme d'apprentissage pour un réseau à une couche cachée« Algorithm Neighborhood Modified BackPropagation, ANMBP ». Ce dernier est basé sur le voisinage de la structure du réseau ,pour remplacer les paramètres d'apprentissage fixes par des paramètres adaptatifs. L'ANMBP est efficace en terme d'erreur ,et de temps de convergence[Kat et Jay'09]. Bai et al ,ont formulé un algorithme appelé BP algorithm with varying slope of activation function,BPVSAF, basé sur la variation de la pente de la fonction d'activation (La pente d'une fonction d'activation indique une valeur signifiant la rapidité de la fonction d'activation pour aller du minimum au maximum)avec différents taux d'apprentissage.les résultats obtenus ont démontré que l'algorithme de rétropropagation classique peut avoir de très bonnes performances en ajustant seulement les deux paramètres du taux d'apprentissage et la pente de la fonction d'activation de la couche de sortie[Bai et al'09].

En 2010 , Sun a développé un algorithme nommé Local Coupled Feed-Forward Neural Network ,LCFNN ; où à chaque neurone caché est attribué une direction de l'espace d'entrée, et chaque entrée active seulement les neurones à proximité d'elle. Aussi la dimension de l'espace de recherche n'augmente pas avec l'augmentation de la taille du réseau , et le coût de calcul de chaque échantillon d'apprentissage n'augmente pas avec l'augmentation de la taille du réseau[Sun'10].

En 2011 ,kocer et canal ont démontré que l'utilisation de la méthode d'optimisation par les algorithmes génétiques pour sélectionner les entrées pour la classification des épilepsies donne de très bons résultats. En effet les signaux EEG sont représentés par l'analyse FFT(Fast Fourier Transformation) , ces différentes représentations seront optimisées par les Ag pour faire l'apprentissage par l'algorithme de rétropropagation du gradient[Koc et Can'11].

En 2012, Mirjalili et al ont proposé une méthode hybride composée du GSA gravitational search algorithm et de PSO particle swarm optimization .Cette hybridation a été utilisée comme une nouvelle méthode d'apprentissage des réseaux de neurones. Les GSA sont de nouvelles méthodes heuristiques basées sur la loi de gravité ,elles sont très performantes pour la recherche du minimum global ,leur inconvénient est qu'elles sont très lentes .pour cette raison l'utilisation d'une méthode d'optimisation PSO a été proposée pour résoudre ce problème. Le PSOGSA a donné de très bons résultats en terme de rapidité de convergence et éviter les minima locaux[Mir et al'12].

En 2013 , Janaki et Geetha , ont proposé une méthode appelée MLP ABC(Multi layer perceptron artificial bee colony).ils utilisent un algorithme d'optimisation pour la recherche des valeurs des poids optimales , qui seront utilisées dans l'apprentissage des réseaux de neurones.cette méthode a été utilisée pour la classification du cancer du sein par images IRM et elle a donné de bons taux de reconnaissance des tumeurs[Jan et Gee'13].

En 2014 , Ashraf Osman et al ,ont proposé un algorithme pour résoudre les problèmes de classification en terme de complexité et de rapidité de convergence. L'algorithme est une hybridation de NSGA-II et l'algorithme de rétropropagation, NSGA-IIBP(non-dominated sorting genetic algorithm-II Back Propagation).leurs résultats ont été comparés avec une méthode hybride MOGATBP(multi objective genetic algorithm based Training Back Propagation network).Les résultats obtenus ont prouvé l'amélioration de l'apprentissage et par conséquent les performances du classifieur [Ash et al'14].

3. Conclusion

Nous avons présenté dans ce chapitre différentes approches et variantes pour l'amélioration de l'algorithme de rétropropagation du gradient ,proposées dans la littérature de 1990 jusqu'à 2014.

Dans le chapitre suivant nous montrerons l'intérêt essentiel de la métaplasticité dans l'apprentissage naturel et dans la mémorisation.

Chapitre 2

Apprentissage Naturel et Métaplasticité

Chapitre 2 : Apprentissage Naturel et Méta-plasticité

1. Introduction

Ce chapitre présente le principe et concepts des processus biologiques qui permettent la génération de la méta-plasticité, et également il explique la relation entre l'apprentissage et les changements qui se produisent dans les structures neuronales.

La méta-plasticité est un concept biologique largement connu et utilisé dans différents domaines tels que la biologie, l'informatique médicale, la psychologie, la neurologie et la neurophysiologie.

La méta-plasticité est un phénomène très lié au processus d'apprentissage au niveau de l'hippocampe, Certains chercheurs comme Abraham ont affirmé que « la méta-plasticité est l'élément clé dans la mémorisation et l'apprentissage » [Abr'99].

Le scientifique espagnol Santiago Ramón y Cajal en 1894 ; fut le premier à décrire les différents types de neurones [Caj'94]. En même temps il stipule que le système nerveux est constitué de neurones individuels, ces neurones communiquent entre eux par des contacts fonctionnels appelés synapses. L'hypothèse de Cajal s'oppose à d'autres scientifiques de son temps qui ont décrit le système nerveux comme un réseau de fibres nerveuses reliées entre elles, formant ainsi un réseau continu. (par analogie avec les vaisseaux sanguins et la circulation du sang). Ramon a indiqué que la présence de changements durables dans les synapses est liée à l'apprentissage.

A la même période, il a découvert que l'apprentissage peut produire des changements dans la communication entre les neurones, et que ces changements pourraient être les mécanismes essentiels de la mémoire. en 1948 Konorski suggère que l'activation simultanée de deux ou plusieurs neurones transformerait les connexions potentialisées en connexions excitatrices.

Il affirme également que le processus comme les changements de la plasticité des neurones pourrait avoir une relation directe avec la formation et la multiplication de nouvelles liaisons synaptiques entre les terminaisons des axones d'un neurone et le soma d'un autre neurone. Konorski a parlé en quelques sortes des changements plastiques persistants en mémoire [Kon'48].

Hebb en 1949, a postulé, que dans la phase d'apprentissage, les connexions synaptiques se renforceraient, du fait de la corrélation de l'activité des neurones présynaptiques et des neurones post-synaptiques [Heb'49].

Pour une bonne compréhension du mécanisme de la méta-plasticité, nous présentons en premier lieu une définition détaillée de la plasticité synaptique et une description de tous les phénomènes biologiques associés.

2. Mémoire et Apprentissage

2.1. Principes de l'apprentissage naturel

L'apprentissage naturel est un processus permettant de conserver des informations acquises, des états affectifs et des impressions capables d'influencer le comportement. L'apprentissage est la principale activité du cerveau, c'est-à-dire modifier constamment sa structure pour mieux exploiter les expériences rencontrées.

L'apprentissage est une modification relativement permanente du comportement qui marque un gain de connaissance, de compréhension ou de compétence grâce aux souvenirs mémorisés. La mémoire est le fruit de cet apprentissage, la trace concrète qui en est conservé dans nos réseaux de neurones.

Chaque fois que nous apprenons quelque chose, des circuits nerveux sont modifiés dans notre cerveau. Ces circuits sont constitués d'un certain nombre de cellules nerveuses (ou neurones) qui communiquent entre elles par des jonctions particulières appelées synapses. Ce sont ces synapses qui augmentent leur efficacité suite à un apprentissage, facilitant ainsi le passage de l'influx nerveux . Par exemple, lorsqu'on entend un nouveau mot, de nouvelles connexions entre certains de nos neurones sont sollicitées : certains du cortex visuel pour en reconnaître l'orthographe, d'autres du cortex auditif pour en entendre la prononciation, d'autres encore dans les régions associatives du cortex pour le relier à d'autres connaissances. Or pour apprendre ce nouveau mot, on se le répète un certain nombre de fois, ce qui a pour effet de sélectionner et de renforcer les connexions entre ces différents circuits du cortex. Et c'est cette nouvelle association durable entre certains neurones qui formera le souvenir de ce mot.

2.2. Notions de mémoire

Notre mémoire est fondamentalement associative : on retient mieux lorsqu'on peut relier la nouvelle information à des connaissances déjà acquises et solidement ancrées dans notre mémoire. Et ce lien sera d'autant plus efficace qu'il a une signification pour nous. Et contrairement à l'image classique d'une vaste collection de données archivées, la plupart de nos souvenirs sont des reconstructions. En effet, les souvenirs n'étant pas stockées dans le cerveau comme des livres dans une bibliothèque, leur rappel exige à chaque fois une reconstruction à partir d'éléments épars dans différentes aires cérébrales.

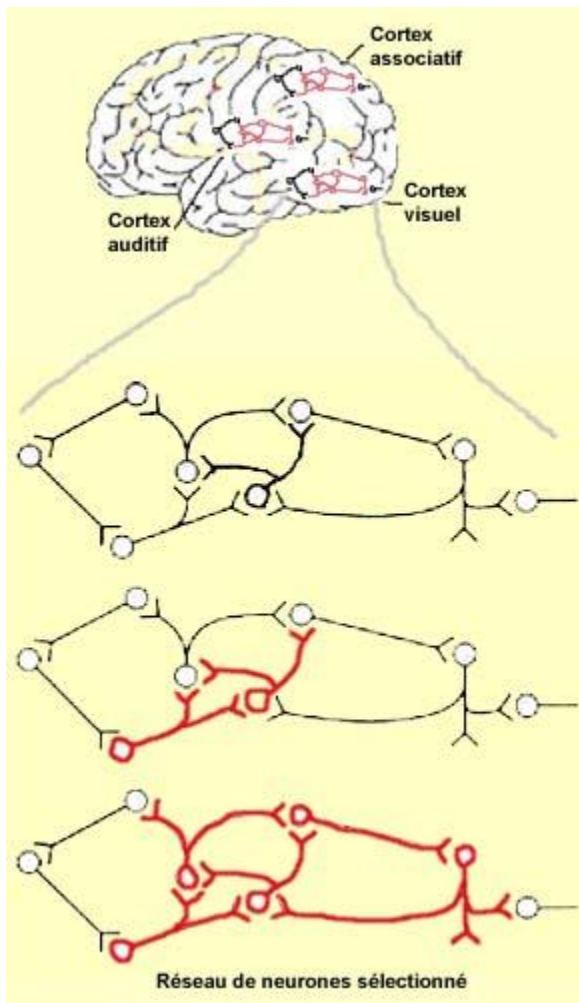


Figure 2.1 : Illustration d'un réseau de neurones dans le cerveau

Pour se remémorer une information apprise des jours ou des années plus tard, il faut réussir à réactiver ces circuits nerveux. En général il sera d'autant plus facile de le faire que le circuit aura été fortement façonné par un passage répété de l'influx nerveux dû à un long apprentissage. Au contraire, si l'on ne s'est répété l'information que quelques fois, les connexions entre les nouveaux neurones sont plus faibles, et le nouveau circuit plus difficile à réactiver.

Tous nos souvenirs (événements, mots, images, émotions, etc.) correspondent donc dans notre cerveau à l'activité particulière de certains réseaux de neurones ayant des connexions renforcées entre eux.

Notre cerveau retient une information quelques dixièmes de seconde ou toute une vie. C'est la durée d'une information, trace mnésique, qui permet de distinguer les mémoires à court terme des mémoires à long terme.

2.2.1. Mémoires à court terme

Il existe deux mémoires à court terme : la mémoire sensorielle et la mémoire de travail.

a) Mémoire sensorielle

Elle provient de nos sens, principalement de l'ouïe (mémoire échoïque) et de la vision (mémoire iconique). Par exemple, quand vous menez une expérimentation, vous traitez et reprenez en permanence des informations sur la procédure à suivre.

La plupart des informations provenant la mémoire iconique (visuelle) s'effacent après quelques millisecondes (500ms).

La mémoire auditive ou échoïque est un peu plus résistante : les informations sont stockées jusqu'à une dizaine de secondes.

b) Mémoire de travail

Elle joue un rôle central dans l'accomplissement de nos tâches quotidiennes : la mémoire de travail est la clé de l'attention et de la concentration.

Les exemples sont nombreux : retenir l'essentiel du cours d'un enseignant, faire un calcul mental, se souvenir d'une adresse en l'écrivant, composer une phrase dans un courrier, etc.

La mémoire de travail mobilise et retient les informations utiles à toutes nos activités cognitives, pendant une durée de quelques secondes à quelques dizaines de secondes.

Cette capacité de retenir temporairement une information en vue de mener à bien une tâche est spécifiquement humaine. Elle rend très active certaines régions de notre cerveau, en particulier le lobe préfrontal.

2.2.2. Mémoires à long terme

Notre cerveau est capable de stocker aussi des informations sur une très longue période, souvent la vie entière : ce sont les mémoires à long terme.

Nous distinguons deux systèmes de mémoire à long terme : déclarative (ou explicite) et non-déclarative (ou implicite).

a) Mémoire déclarative

Elle rassemble deux types de mémoire : épisodique et sémantique. Nous pouvons rappeler les souvenirs concernés, et ils deviennent alors explicites.

- **Mémoire épisodique**

La mémoire épisodique est la somme des événements que nous avons vécus. Le rappel des souvenirs de la mémoire épisodique est volontaire comme à titre d'exemple se rappeler d'une expérimentation réalisée dans notre projet de fin d'études de graduation ou involontaire par exemple un élève traite un exercice d'examen qui lui rappelle un exercice similaire appris pendant sa révision.

- **Mémoire sémantique**

La mémoire sémantique désigne l'ensemble des connaissances pratiques ou théoriques que l'on a acquises et conservées. Son domaine est très large : le sens des mots, la manière de rédiger un article, la signification de chaque onde de l'électrocardiogramme, etc.

La mémoire sémantique concerne des domaines génériques d'apprentissage, indispensables au bon déroulement de l'existence.

b) Mémoire non-déclarative

Elle concerne des informations acquises et durablement retenues, mais qui ne font pas l'objet d'un rappel conscient.

Un exemple simple concerne les formes procédurales de la mémoire : nous apprenons à nager, à faire du vélo, à conduire... et ces règles restent inscrites sans effort. Elles nous reviennent quand nous sommes en situation de les appliquer. Plus nous pratiquons, plus elles reviennent facilement.

La mémoire implicite se confond en partie avec l'inconscient. Beaucoup d'expériences sur les amnésiques ont montré que si leur mémoire épisodique est effacée, leur mémoire procédurale reste intacte. Elle ne renvoie pas au même système neuronal dans le cerveau.

2.2.3. Passage de la mémoire à court terme à la mémoire à long terme

Le passage de la mémoire à court terme (ou la mémoire de travail) à la mémoire à long terme s'effectue grâce à l'hippocampe, dont la forme incurvée rappelle la queue d'un hippocampe marin. L'hippocampe est une partie très ancienne du cortex situé dans le repli interne du lobe temporal. L'hippocampe ou corne d'Ammon (CA), est une structure cérébrale appartenant au système limbique, qui comporte trois régions : CA1, CA2 et CA3.

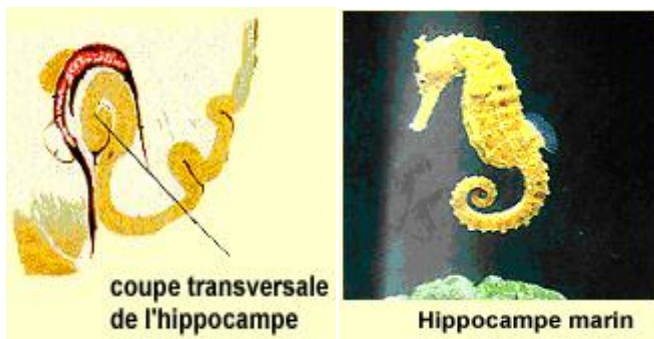


Figure 2.2 : Similarité de forme entre l'hippocampe marin et l'hippocampe du cortex.

Toutes les informations décodées dans les différentes aires sensorielles du cortex convergent vers l'hippocampe qui les retourne ensuite d'où elles viennent. C'est un peu comme un centre de tri qui comparerait ces sensations

nouvelles avec celles déjà enregistrées. L'hippocampe crée aussi des liens entre les différentes caractéristiques d'une chose.

La répétition ou les différents concepts qui nous permettent de retenir des faits nouveaux correspondent donc à de multiples passages dans l'hippocampe. Celle-ci va renforcer les liens entre ces nouveaux éléments, si bien qu'au bout d'un certain temps, son travail ne sera plus nécessaire : le cortex aura appris à lier lui-même ces différentes caractéristiques pour en faire ce qu'on appelle un souvenir.

La caractéristique fondamentale qui permet la mémoire et l'apprentissage est la plasticité de notre cerveau. Autrement dit, la capacité des neurones à modifier leurs connexions pour rendre certains circuits nerveux plus efficace.

3. Réseaux de Neurones et Principes de la Plasticité

3.1. Plasticité des réseaux de neurones biologiques

Les neurones sont des cellules spécialisées du système nerveux qui permettent la transmission et le traitement de l'information. Le cerveau humain est composé de plus de cent milliard de neurones qui communiquent entre eux constamment par l'intermédiaire de connexions appelées synapses (figure 2.3). Les neurones peuvent ainsi s'organiser en réseau, ce qui permet de traiter et de répondre de façon adaptée à des stimuli complexes provenant de l'environnement extérieur ou de l'organisme lui-même. Au fur et à mesure, des réseaux de neurones vont permettre au cerveau de mémoriser des situations afin de pouvoir y répondre de plus en plus efficacement.

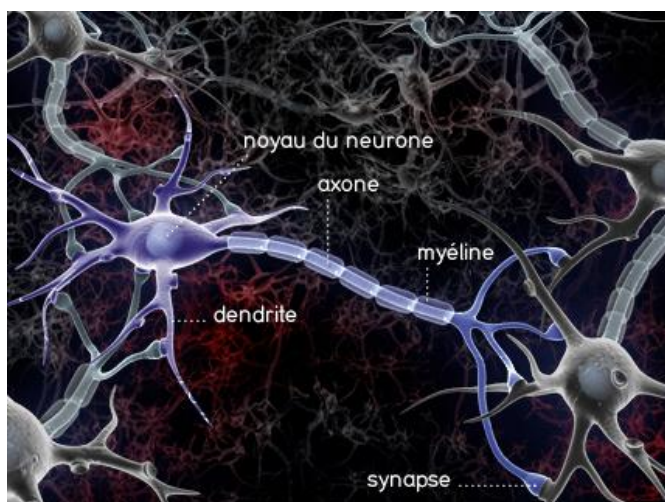


Figure 2.3: Le neurone et la synapse.

Un neurone est une cellule avec de nombreux prolongements (dendrites), qui reçoit les informations provenant d'autres neurones, d'un soma, qui peut aussi recevoir des contacts et qui assure un rôle central dans toutes les fonctions du neurone, et d'un axone, qui va transmettre l'information à d'autres neurones par l'intermédiaire des synapses. Les synapses excitatrices sont faciles à repérer car pour la majorité d'entre elles, la partie postsynaptique ressemble à une petite épine. Les contacts inhibiteurs, eux, se font directement sur les dendrites.

En général, nous considérons un réseau de neurones comme un groupe d'unités qui vont communiquer entre elles soit en se stimulant, soit en s'inhibant. Un neurone peut recevoir plusieurs centaines de milliers de connexions synaptiques de la part d'autres neurones. Au niveau de ces connexions synaptiques, l'influx électrique ne se propage pas directement (à quelques exceptions près): dans le neurone présynaptique, il provoque la libération de molécules appelées neurotransmetteurs dans la fente synaptique. Ces neurotransmetteurs vont alors se fixer sur des récepteurs spécifiques situés sur le neurone postsynaptique.

Une fois activés, ces récepteurs sont capables de recréer le signal électrique dans le neurone postsynaptique, qui va éventuellement le transmettre à d'autres neurones. Mais les synapses ne font pas que transmettre un influx

nerveux. Grâce à des enregistrements électrophysiologiques nous savons depuis plusieurs décennies qu'elles ont la capacité de le moduler. C'est ce que l'on appelle la plasticité synaptique (figure 2.4).

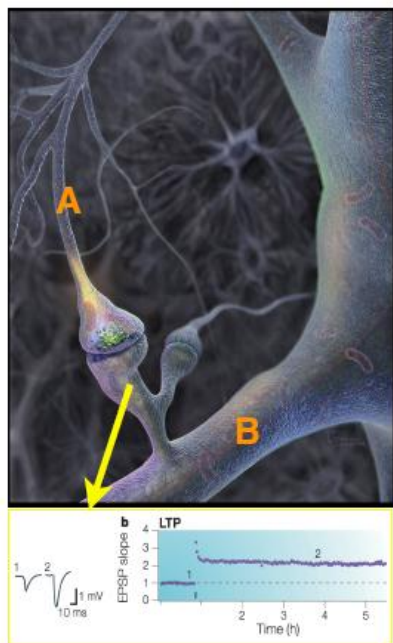


Figure 2.4: La plasticité synaptique.

Des expériences électrophysiologiques ont confirmé l'existence de la plasticité synaptique, c'est un type de «mémoire cellulaire»: Lorsque l'on stimule le neurone A et que l'on enregistre le neurone B qui était au repos, on obtient la réponse (1). Mais si lorsque que l'on stimule le neurone A alors que neurone B est rendu actif (par une stimulation à haute fréquence du neurone A par exemple), on obtient la réponse (2).

Cette dernière est plus élevée, alors que l'intensité de la stimulation est restée la même. Le neurone B répond désormais «plus intensément», et cet effet peut être observé pendant plusieurs heures, voire plusieurs jours.

L'électrophysiologie, c'est à dire l'enregistrement de l'activité électrique d'un ou de plusieurs neurones vivants et connectés entre eux a permis de comprendre et de dévoiler de nouvelles fonctions du système nerveux. Le développement de la microscopie confocale laser permet d'observer des neurones vivants sur le long terme. Actuellement nous pouvons savoir si un réseau de neurones, établi pendant le développement, reste immuable au cours de la vie, ou bien peut non seulement moduler la transmission synaptique pour permettre un apprentissage, mais aussi supprimer ou créer de nouveaux contacts. Il s'agit alors de plasticité structurelle d'un réseau de neurones.

3.2. La Transmission Neuronale

a) Principe

Dans le cerveau, les informations circulent sous forme de messages électriques, appelés influx nerveux de neurones en neurones. Depuis les dendrites, elles vont vers le corps cellulaire où elles sont traitées et arrivent jusqu'aux synapses via l'axone.

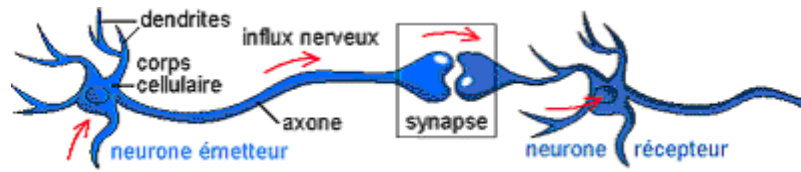


Figure 2.5: Connexion entre deux neurones

En 1897, Le neurophysiologiste anglais Charles Sherrington a constaté que la vitesse de l'influx nerveux depuis le cortex jusqu'aux membres était inférieure à celle de la propagation au niveau d'une seule fibre nerveuse [Pea'04]. Il conclut alors qu'il devait y avoir une sorte d'interruptions qui ralentissent la transmission de l'influx. C'est ainsi qu'il apporta le concept de Synapse (du grec, syn qui signifie ensemble, haptain veut dire toucher ou saisir; c'est-à-dire connexion).

b) Intérêt des synapses

Les synapses évoluent avec le temps, certaines disparaissent, d'autres se créent, mais toutes se modifient et tendent, soit à renforcer, soit à affaiblir, la communication entre deux neurones. La plasticité synaptique serait ainsi à la base des processus d'apprentissage et de mémorisation. En effet, les neurones communiquent les uns avec les autres grâce aux synapses. Sur un seul neurone on peut trouver de 1000 à 10000 synapses (environ 300 000 au niveau des cellules de Purkinje cérébelleuses. Multipliez ce nombre par 100 milliards de neurones pour avoir une idée sur le nombre des communications au sein du système nerveux !).

Les synapses constituent les zones d'échanges d'informations entre les neurones. Les neurones sont unis entre eux par des milliers de connexions synaptiques. A ce niveau, l'information est échangée sous forme de messages chimiques. Des substances chimiques appelées neurotransmetteurs sont sécrétées et se lient à des récepteurs spécifiques.

Dans le cerveau, plus d'une centaine de neurotransmetteurs ont été identifiés, mais concernant la mémoire, le glutamate, le GABA et l'acétylcholine sont les plus impliqués. Ils font tous cruellement défaut dans la maladie d'Alzheimer.

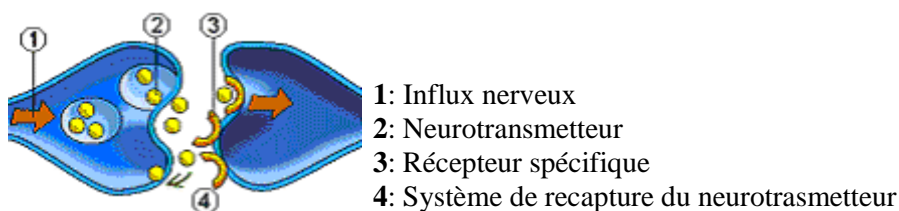


Figure 2.6: Synapse en fonctionnement normal

4. La plasticité synaptique

Le système nerveux central possède une capacité remarquable d'adaptation aux modifications de son environnement. Cette caractéristique implique des phénomènes de remodelage appelés plasticité.

Nous pouvons ainsi donner deux exemples dans lesquels la plasticité s'illustre particulièrement, l'un physiologique et l'autre pathologique. Durant le développement, elle permet de mettre en place l'ensemble des connexions neuronales, et, en cas de lésions cérébrales, elle contribue aux importants remaniements auxquels le cerveau est confronté dans cette situation. Mais au-delà de ces deux cas, de nombreuses fonctions cérébrales résultent de la capacité du cerveau à se restructurer en permanence, la principale étant celle de la mémoire qui implique le stockage et ensuite le rappel d'expériences passées.

4.1. Historique de la plasticité synaptique

L'observation que le nombre de neurones du cerveau n'augmente pas dans un cerveau adulte fût faite dès la fin du 19^{ème} siècle. Ceci impliquait que la mémoire ne pouvait pas être expliquée par la production de nouveaux neurones qui apparaîtraient au fur et à mesure des besoins de stockage des informations.

La question qui peut être posée concerne la mémorisation d'informations sans l'aide de nouveaux neurones ? Un premier élément de réponse fût apporté par le neuroanatomiste Santiago Ramón y Cajal [Caj'94] qui proposa que les souvenirs étaient stockés dans les connexions entre les neurones et qu'ils étaient codés grâce au renforcement des connexions nerveuses. Cajal proposa que ce renforcement permettrait aux neurones d'augmenter leur efficacité de communication et de pouvoir ainsi mémoriser des informations. Cette théorie, espagnole, fut ensuite précisée par un postulat, écossais. En effet, en 1949, Donald Hebb [Heb'49] a confirmé que les neurones devaient augmenter la force de leurs connexions pour améliorer leur capacité à communiquer. Le postulat de Hebb est que si un neurone A (pré-synaptique) participe de façon répétée et persistante à l'excitation d'un neurone B (post-synaptique), alors la transmission synaptique de A vers B est facilitée.

Le postulat de Hebb reçut une validation expérimentale grâce à la mise en évidence de la potentialisation à long-terme (PLT).

Terje Lomo, observa que l'amplitude des potentiels post-synaptique excitateur (PPSE) pouvait être augmentée pendant une longue période suite à l'application d'un train de stimulation à haute fréquence à cette synapse. Ce phénomène fut appelé « long-lasting potentiation » et sa première caractérisation, dans l'hippocampe, fut publiée en 1973 [Bliss et Lom'73]. Une dizaine d'années passa avant que ne fut mise en évidence dans le cervelet une forme de plasticité opposée à la PLT : une diminution persistante de l'efficacité de la transmission synaptique appelée la dépression à long-terme (DLT) [Ito et al'82].

4.2. Généralités sur la plasticité synaptique

Il existe plusieurs formes de plasticité synaptique qui se distinguent par leur effet, leur durée d'expression et par leurs mécanismes d'induction et de maintien. Une potentialisation ou une dépression de l'efficacité synaptique peuvent se produire pendant des temps courts de l'ordre de quelques millisecondes à quelques minutes, on parlera alors de plasticité à court-terme, ou sur des temps beaucoup plus longs, allant de quelques dizaines de minutes à plusieurs jours, voire des mois, on parle alors de plasticité à long-terme (PLT).

La plasticité à court-terme s'exprime par un accroissement (facilitation) ou une diminution (dépression) transitoire de l'amplitude de la réponse post-synaptique.

Beaucoup de travaux ont montré que la plasticité à court-terme est due à des modifications au niveau pré-synaptique et dépend généralement des niveaux de calcium dans la terminaison pré-synaptique [Zuc et Reg'02].

La plasticité à long-terme, quant à elle, correspond à un renforcement ou une diminution durable de la force synaptique. Selon le postulat initial de Hebb, la mise en place de changements d'efficacité synaptique à long-terme nécessite la coactivation des éléments pré- et post-synaptiques. La PLT est classiquement induite par des stimulations à haute fréquence (stimulations téaniques à 100 Hz, HFS).

D'une manière générale, un accroissement important et transitoire de la concentration de calcium dans le neurone post-synaptique est à la base de la PLT.

Cette entrée de calcium se fait principalement via les récepteurs 'NMDA' (acide N-méthyl-D-aspartique) qui sont activés par une dépolarisation et perméables aux ions calcium. L'augmentation de la concentration de calcium active des protéines kinases, telle que la CaMKII (protéine kinase de type II dépendante du calcium) par exemple, qui modulent l'expression et l'activité de protéines nécessaires pour la transmission, comme les récepteurs AMPA [Bli et Col'93] ; [Mal et Nic'99]. Néanmoins, il faut préciser que, au-delà des récepteurs NMDA, la PLT peut également résulter d'une augmentation de la libération de glutamate et donc être sous-tendue par des mécanismes pré-synaptiques [Lyn'04]. La DLT est classiquement induite par des stimulations à basse fréquence (le plus souvent à 1 Hz, LFS). Elle dépend également de la concentration de calcium post-synaptique mais se développe si l'augmentation de calcium est faible et lente [Lin et Con'95].

La majorité des synapses du système nerveux central sont capables de développer à la fois de la PLT et de la DLT et leur occurrence dépend des protocoles de conditionnement cellulaire utilisés. La possibilité pour une synapse de développer à la fois de la PLT et de la DLT permet l'existence d'une homéostasie de la transmission. Ce qui protège les synapses du blocage ou de la saturation.

Une problématique posée est de comprendre précisément les paramètres physiologiques qui orientent les synapses vers une PLT ou une DLT. Un modèle, dit BCM, établi par Bienenstock, Cooper et Munro [Bie et al'82], propose que toute synapse active sera potentialisée si l'activité post-synaptique correspondante est supérieure à une valeur critique et sera déprimée si l'activation post-synaptique est inférieure à ce même seuil. Ce modèle a été validé expérimentalement dans l'hippocampe [Dud et Bear'92] et dans le cortex [Sjo et al'01] où il a été observé qu'au-delà d'une certaine fréquence d'activation, seule une PLT pouvait être induite. Dernièrement, plusieurs travaux ont montré que la séquence temporelle entre les activations pré- et post-synaptiques pouvait également être déterminante dans l'orientation de la plasticité synaptique vers une potentialisation ou une dépression : cette forme de plasticité est appelée « spike-timing dependent plasticity » (STDP) [Dan et Poo'06].

4.3. La plasticité synaptique : substrat cellulaire de la mémoire

Il est admis aujourd'hui que les changements d'efficacité synaptique à long-terme seraient un des substrats cellulaires des phénomènes d'apprentissage et de mémorisation. Par contre, ce postulat n'a pas encore été totalement justifié car il est difficile expérimentalement de faire le lien entre des événements observés au niveau cellulaire et des tâches comportementales et mnésiques. De nombreux travaux ont tout de même fait nettement avancer nos connaissances des propriétés plastiques des circuits neuronaux et leur relation avec la mémoire.

Le lien entre la plasticité synaptique et les phénomènes d'apprentissage et de mémorisation a été étudié dans différentes structures cérébrales telles que le cortex, l'hippocampe, le cervelet, le striatum ou encore l'amygdale. Néanmoins, la majorité des études se sont concentrées sur l'hippocampe et le cortex. En effet, l'hippocampe est reconnu pour avoir un rôle prépondérant dans les phénomènes de mémorisation et a été décrit comme la structure cérébrale siège de la mémoire. Plus récemment, des études au niveau du néocortex se sont développées car son interaction directe avec les entrées sensorielles permet de tester plus facilement la plasticité induite par des entrées synaptiques et pas seulement des stimulations électriques. Les expériences menées dans l'hippocampe et

le cortex ont permis de faire un parallèle entre les changements d'efficacité synaptique à long terme et la mémoire. Il est considéré que la PLT serait le substrat de l'apprentissage tandis que la DLT celui de « l'oubli ». Cette vision est quelque peu caricaturale puisqu'une diminution de l'efficacité synaptique pourrait être en soi un apprentissage tout aussi bien qu'une augmentation de la transmission. Néanmoins, la majorité des études sur la mémoire ont été menées sur la PLT.

5. La Potentialisation à Long Terme (PLT)

5.1. Principe de la PLT

La plasticité synaptique correspond aux modifications morphologiques, chimiques et fonctionnelles qui interviennent au cours du temps au niveau de la synapse. La majorité des travaux sur la plasticité synaptique sont basés sur le postulat de Hebb qui a proposé que, lorsqu'un neurone prend part de façon répétée à l'activation d'un autre neurone, l'efficacité des connexions entre ces neurones est augmentée. Depuis, les manifestations de la plasticité qui font l'objet d'études intensives sont la PLT ou LTP (Long Term Potentiation) et la DLT ou LTD (Long Term Depression) qui représentent, respectivement, une augmentation et une diminution de l'efficacité de la transmission synaptique à long terme.

L'hippocampe est une structure cérébrale grandement impliquée dans la mémoire à long terme. Cependant la transmission synaptique au sein de la formation hippocampique est principalement assurée par le glutamate et/ou l'aspartate en ce qui concerne l'excitation et par le GABA (acide γ -aminobutyrique) pour l'inhibition. Ces neurotransmetteurs, synthétisés et libérés par les neurones de la formation hippocampique, modulent l'activité cellulaire en activant différents types de récepteurs.

Les neurones de l'hippocampe ont une plasticité remarquable. Certains de ceux-ci, lorsqu'ils reçoivent plusieurs influx nerveux sur une courte période, renforcent la connexion synaptique qui a reçu ces influx. Cette augmentation de l'efficacité de cette synapse peut durer des heures, des jours, voire des semaines. On utilise l'expression potentialisation à long terme pour décrire ces modifications moléculaires probablement à la base de bien des apprentissages à long terme.

La caractéristique la plus intéressante de ce mécanisme (PLT) découvert en 1973 est qu'il permet le renforcement durable des synapses entre deux neurones qui sont activés simultanément. Exactement, donc, le genre de mécanisme d'association qu'avait imaginé Hebb un quart de siècle auparavant.

La Potentialisation à Long Terme correspond à une augmentation d'amplitude de la réponse post-synaptique à la suite d'une intense activation pré-synaptique. Celle-ci est généralement courte (< 1 seconde) mais de fréquence élevée (> 100 Hz). Ce type de stimulation provoque une dépolarisation post-synaptique suffisante pour évacuer les ions magnésium qui bloquent le récepteur NMDA et permettre ainsi l'entrée massive de calcium dans le dendrite.

Or, l'ion calcium est un messenger intracellulaire extrêmement important qui active un grand nombre d'enzymes en modifiant leur conformation. C'est le cas de la calmoduline qui devient active lorsque quatre ions calcium s'y fixent. Elle devient alors la Ca^{2+} -calmoduline, second messenger principal de la PLT capable d'activer à son tour d'autres enzymes qui jouent un rôle clé dans ce processus comme l'adenylate cyclase et la protéine kinase II calmoduline-dépendante (CaM kinase II).

Ces enzymes vont à leur tour modifier la conformation spatiale d'autres molécules, le plus souvent en leur ajoutant un ion phosphate, un processus catalytique fréquent appelé phosphorylation.

Ainsi, l'adénylate cyclase activée fabrique de l'adénosine mono-phosphate cyclique (ou AMPc) qui catalyse à son tour l'activité d'une autre protéine, la kinase A (ou PKA). On est donc en présence d'une cascade de réactions biochimiques dont les effets peuvent être multiples.

On sait par exemple que les récepteurs AMPA vont être phosphorylés par la PKA, leur permettant de rester ouverts plus longtemps suite à la fixation de glutamate. Ceci entraîne par conséquent une plus grande dépolarisation post-synaptique et contribue ainsi à l'établissement de la PLT.

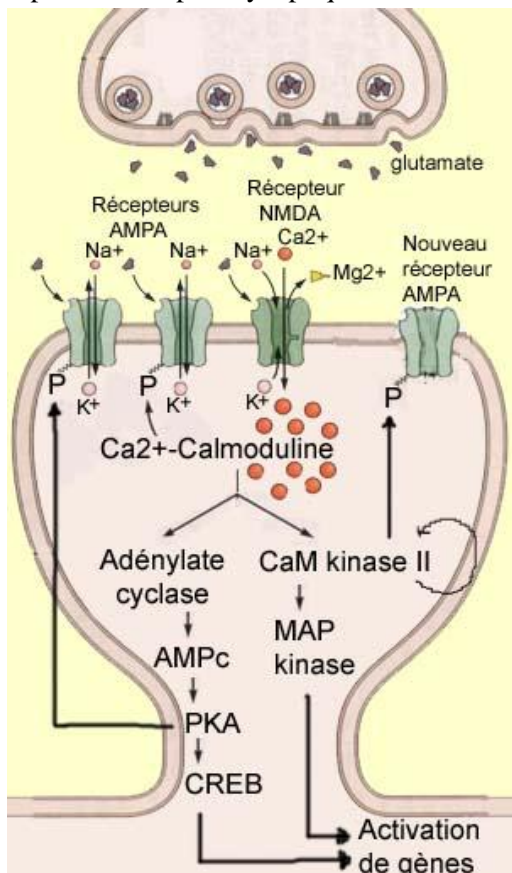


Figure 2.7: Principe de la Potentialisation à Long Terme PLT.

5.2. Processus de la PLT

Le mécanisme qui mène à la potentialisation à long terme est brièvement schématisé dans les 4 étapes ci-dessous:

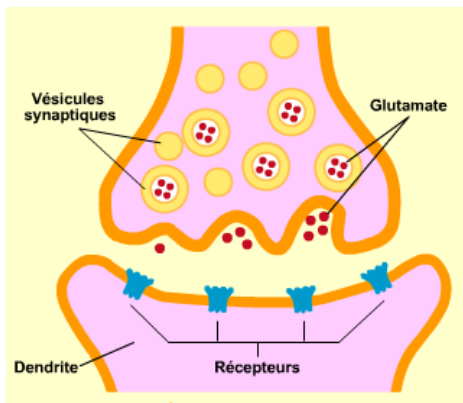


Figure 2.8 :Etape1: les neurones de l'hippocampe relâchent un messager chimique qui se nomme glutamate.

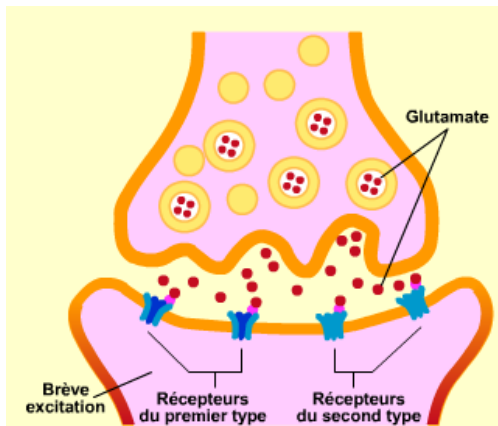


Figure 2.9 :Etape 2 : le glutamate va se fixer sur des récepteurs qui lui sont propres sur le neurone suivant .ce neurone subit une brève excitation qui, si elle est suffisante, pourra générer l'influx nerveux dans l'autre neurone.

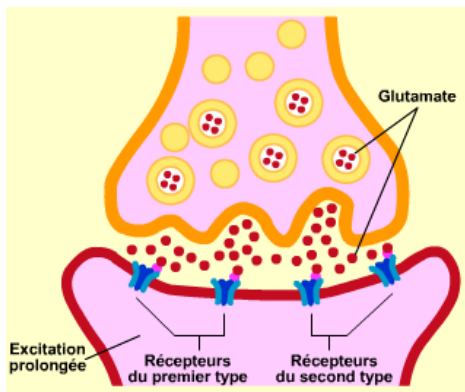


Figure 2.10 :Etape3 : Les conditions qui produisent de la PLT sont un peu différentes de ce cas classique de neurotransmission. En effet les influx qui la provoquent sont de très rapprochés dans le temps (on dit aussi qu'ils ont une fréquence élevée).Ceci a pour effet de maintenir le second neurone dans un état d'excitation prolongé qui va permettre l'activation d'un second type de récepteurs au glutamate.

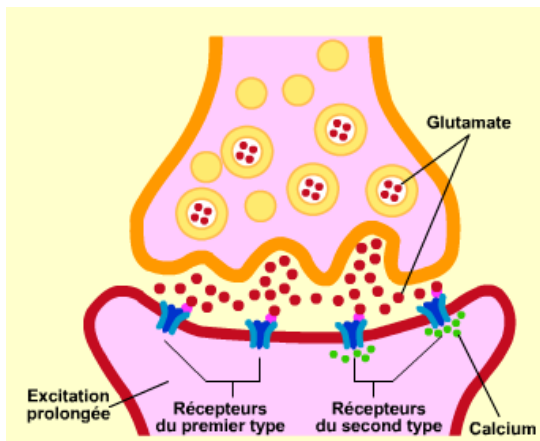


Figure 2.11 : Etape 4 : En effet, ce second type de récepteur ne devient actif que lorsqu'un neurone est dans un état d'activation prolongé justement. Il permet alors à des petites molécules de calcium d'entrer à l'intérieur du neurone.

Les deux types de récepteurs sont donc nécessaires pour assurer l'entrée de calcium qui va aboutir rapidement au renforcement de cette synapse après une série de réactions chimiques.

- ✓ Le glutamate, qui est le neurotransmetteur relâché dans ces synapses, se fixe sur plusieurs sous-types de récepteurs différents dont deux sont particulièrement important pour la PLT : les récepteurs AMPA et NMDA.
- ✓ Le récepteur AMPA est couplé à un canal ionique qui provoque l'entrée de sodium dans le neurone post-synaptique lorsque le glutamate s'y fixe. Cette entrée de sodium amène la dépolarisation locale du dendrite et, si cette dépolarisation atteint le seuil de déclenchement du potentiel d'action, la transmission de l'influx nerveux dans le neurone suivant.
- ✓ Le récepteur NMDA est également un récepteur couplé à un canal ionique, mais c'est le calcium qu'il laisse entrer de façon privilégiée dans la cellule. Au potentiel de repos de celle-ci, ce canal calcique est cependant bloqué par des ions magnésium (Mg^{2+}) qui, même si du glutamate s'y fixe, empêche l'entrée de calcium dans le neurone. Pour que ceux-ci se retirent du canal, le potentiel membranaire de la dendrite doit être dépolarisé. Et c'est justement ce qui va se produire lors de la stimulation à haute fréquence à l'origine de la PLT : une dépolarisation du neurone post-synaptique suite à l'activation soutenue des récepteurs AMPA. Le magnésium va alors se retirer des récepteurs NMDA et permettre l'entrée massive de calcium dans la cellule.

Or cette concentration accrue de calcium dans la dendrite mettra en branle plusieurs réactions biochimiques qui conduiront à l'augmentation prolongée de l'efficacité de cette synapse.

6. La Dépression à Long T (DLT)

6.1. Principe de la DLT

On peut considérer la dépression à long terme (DLT) comme un mécanisme complémentaire à la potentialisation à long terme (PLT). Dans l'hippocampe, on pense que la DLT permettrait de ramener les synapses potentialisées par la PLT à un niveau normal pour les rendre disponible au stockage de nouvelles informations. Mais ailleurs dans le cerveau, elle pourrait être activement responsable du stockage de nouvelles informations, comme dans le cervelet.

La DLT se développe lorsqu'un neurone pré-synaptique est actif à basse fréquence (1-5 Hz) sans que le neurone post-synaptique ne subisse une forte dépolarisation comme c'est le cas avec la PLT. Cette absence d'association entre les deux neurones a pour conséquence d'élever la concentration de calcium dans le neurone post-synaptique, mais beaucoup moins que dans la PLT.

Par conséquent, ce ne sont pas des protéines comme la CaM kinase II ou la kinase A qui seront activées mais bien des phosphatases. Ces enzymes vont enlever certains groupements phosphate aux récepteurs AMPA, en d'autres termes les déphosphoryler.

Dans l'hippocampe, cette déphosphorylation du récepteur AMPA aurait pour effet de ramener l'amplitude du potentiel post-synaptique au niveau normal, celui d'avant la PLT.

Parallèlement, on pense que le nombre de récepteurs AMPA diminuerait lors de la DLT. Ces récepteurs seraient enlevés de la membrane post-synaptique et mis en réserve. L'opération contraire, en somme, de celle observée durant la PLT où de nouveaux récepteurs sont insérés dans la membrane.

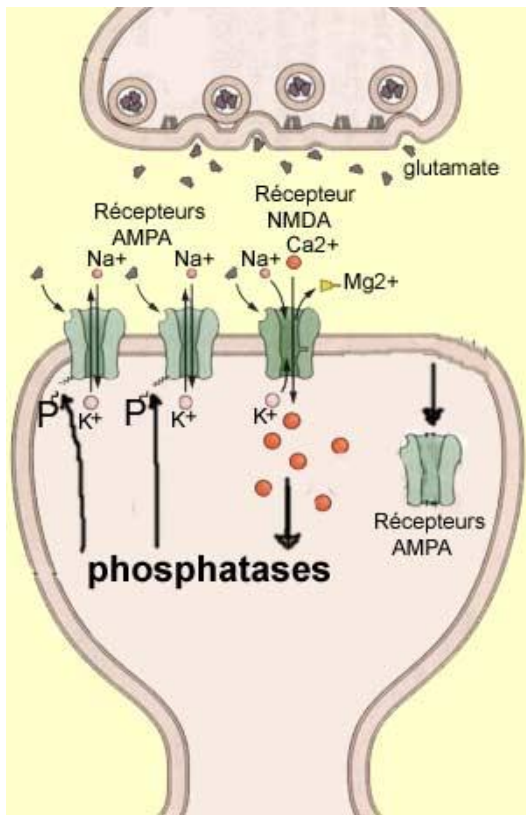


Figure 2.12 :Principe de la Dépression à Long Terme DLT.

6.2. Mécanismes d'induction

Dans l'hippocampe, l'induction de la PLT et de la DLT au niveau des synapses glutamatergiques par des protocoles de tétanisation (HFS/LFS) ou de STDP, repose majoritairement sur des changements de la concentration intracellulaire en calcium $[Ca^{2+}]$ et dépend uniquement de l'activation des récepteurs NMDA [Gus et al'87];[Bli et Col'93];[Deb et al'94]).

Comment le même signal peut-il être utilisé pour l'induction de la PLT ou de la DLT ? La quantité et la cinétique de l'influx calcique détermineraient le sens de la plasticité [Yan et Sur'96]. Ainsi pour une augmentation rapide et importante de calcium une PLT serait induite tandis qu'une augmentation plus modérée mais plus étalée dans le temps entraînerait une DLT.

En effet, une forte augmentation de $[Ca^{2+}]$ conduit à une activation des kinases sensibles au calcium alors que d'après le modèle proposé par Lisman une élévation plus modeste conduit à une activation des phosphatases [Lis et Spr'05]. Etant donné le fait que PLT et DLT partagent le même message d'induction (entrée de calcium via les récepteurs NMDA) activant sélectivement soit la voie des kinases, soit la voie des phosphatases, on peut envisager que des modifications opposées d'un même mécanisme puissent être responsables de l'expression de PLT et de la DLT. Ainsi, la DLT peut s'expliquer par une diminution des conductances AMPA et/ou NMDA due à une déphosphorylation ou une endocytose des récepteurs.

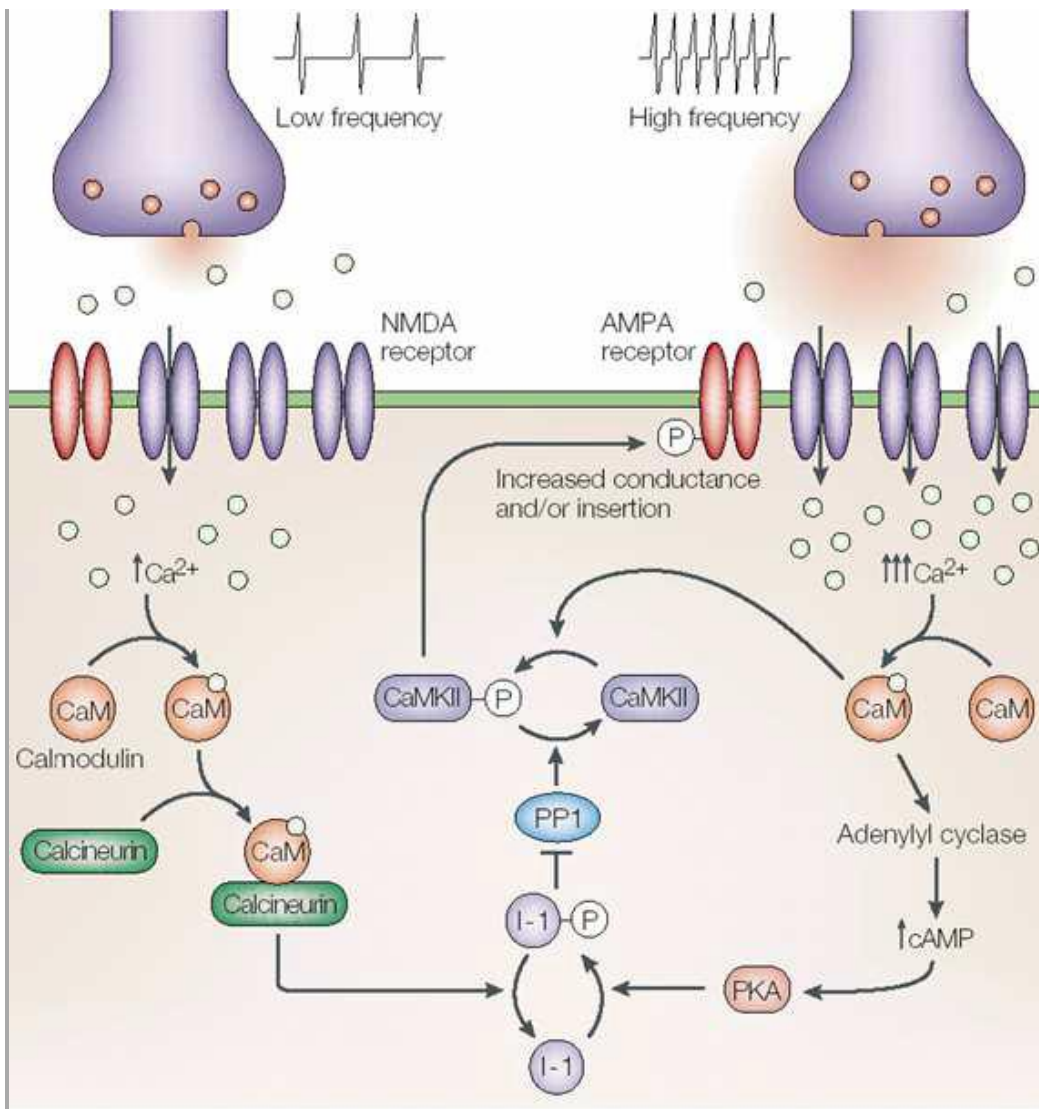


Figure 2.13: Mécanisme d'induction et d'expression de la PLT et de la DLT dans la région CA1 de l'hippocampe (Modèle de Lisman, 1989).

L'activation synaptique des récepteurs NMDA provoque une entrée d'ions calcium dans la cellule postsynaptique. D'après le modèle de Lisman (1989), la $[Ca^{2+}]$ va déterminer le sens de la plasticité synaptique, en activant spécifiquement les protéines kinases (CaMKII, PKA, ...) ou phosphatases (PP1, Calcineurine, ...). Dans l'hippocampe, la DLT est induite par une stimulation à basse fréquence des collatérales de Schaffer. L'activation synaptique à basse fréquence (*Low frequency*) qui provoque une entrée modérée d'ions calcium par les récepteurs NMDA et l'activation de la calcineurine, qui va désactiver l'inhibiteur-1 (I-1) de la PP1. La PP1 va à son tour déphosphoryler la CaMKII et la rendre inactive. La téτανisation à haute fréquence (100 Hz) des fibres collatérales de Schaffer induit la PLT. Une activation synaptique à haute fréquence induit une forte augmentation de calcium, propice à l'activation de protéines kinases, comme la PKA. La PKA active l'I1, qui empêche la PP1 de déphosphoryler la CaMKII. Cette dernière, alors active, peut aller phosphoryler les récepteurs AMPA et augmenter leur conductance ou induire l'insertion de nouveaux récepteurs à la membrane [Wind et Swe'01].

7. Les types de la plasticité à long terme

La plasticité à long terme correspond à un renforcement (PLT) ou une diminution (DLT) durable de l'efficacité synaptique. Deux formes de plasticité à long terme peuvent être distinguées en fonction de leur protocole d'induction : la plasticité homosynaptique et la plasticité associative.

La plasticité de la transmission synaptique a été étudiée dans de nombreuses synapses de l'hippocampe : les synapses glutamatergiques entre cellules pyramidales [Bli et Col'93];[Sou et Deb et al'98], cellules pyramidales sur interneurons [Oua et Lac'95];[Mac et McB'96];[McB et Mac'97] et les synapses GABAergiques entre interneurons et cellules pyramidales [Cai et al'99];[Woo et al'03];[Lam'10].

Cette plasticité se retrouve également dans d'autres structures du SNC, comme le cortex ou le cervelet [D'An et Ros'98].

7.1. PLT et DLT homosynaptiques

La PLT représente probablement l'une des formes de plasticité la plus étudiée en raison de son rôle présumé dans l'apprentissage et la mémoire. Ce phénomène a été pour la première fois mis en évidence par Bliss et collaborateurs dans l'hippocampe. Deux études réalisées chez des lapins anesthésiés [Bli et Lom'73] et éveillés [Bli et Gar'73] montrent qu'une stimulation tétanique à haute fréquence, appliquée sur des fibres afférentes, induit une augmentation de la réponse des cellules du gyrus denté. Cette potentialisation dure plus d'une dizaine d'heures chez les animaux anesthésiés et jusqu'à plusieurs semaines après plusieurs inductions consécutives chez les animaux éveillés [Abr et al'02]. Ces résultats ont été par la suite confirmés par des études in vivo et in vitro , puis dans diverses structures et types cellulaires chez différents animaux [Bli et Col'93]. Cette PLT induite par la seule activité de la synapse considérée, classiquement par des stimulations hautes fréquences (stimulations téaniques, 100Hz, HFS), est qualifiée de PLT homosynaptique.

Phénomène miroir de la PLT et mis en évidence plus tard, la DLT serait également impliquée dans les phénomènes d'apprentissage et de mémorisation. Sa découverte a constitué un apport fondamental à la compréhension du fonctionnement des réseaux neuronaux doués de plasticité et a donné un sens à l'existence même de la PLT. En effet, il a été suggéré très tôt qu'un réseau de neurones ne pouvant exprimer uniquement de la potentialisation est un système où les poids synaptiques ne peuvent que croître jusqu'à arriver à saturation et à terme deviendrait non plastique [Bie et al'82]. La DLT est donc essentielle pour la conservation d'un potentiel de plasticité et augmente les capacités de mémorisation. Les capacités d'apprentissage d'un réseau formel de neurones sont donc optimisées si des règles de dépression synaptique sont intégrées au modèle [Mor et al'86].

La DLT est induite par une stimulation répétée à basse fréquence (1 à 5 Hz) pendant plusieurs minutes [Dud et Bea'92]. Par analogie avec la PLT, cette dépression est dite homosynaptique et se traduit par une diminution de la réponse postsynaptique. Il existe une relation étroite entre la fréquence de la stimulation des fibres afférentes et le sens de la plasticité induite. En effet, les changements de l'efficacité synaptique sont directement fonction de la fréquence de stimulation utilisée lors du protocole conditionnant ; ainsi les stimulations basses fréquences (LFS) entraînent de la DLT et inversement les stimulations hautes fréquences (HFS) de la PLT [Deb et al'98].

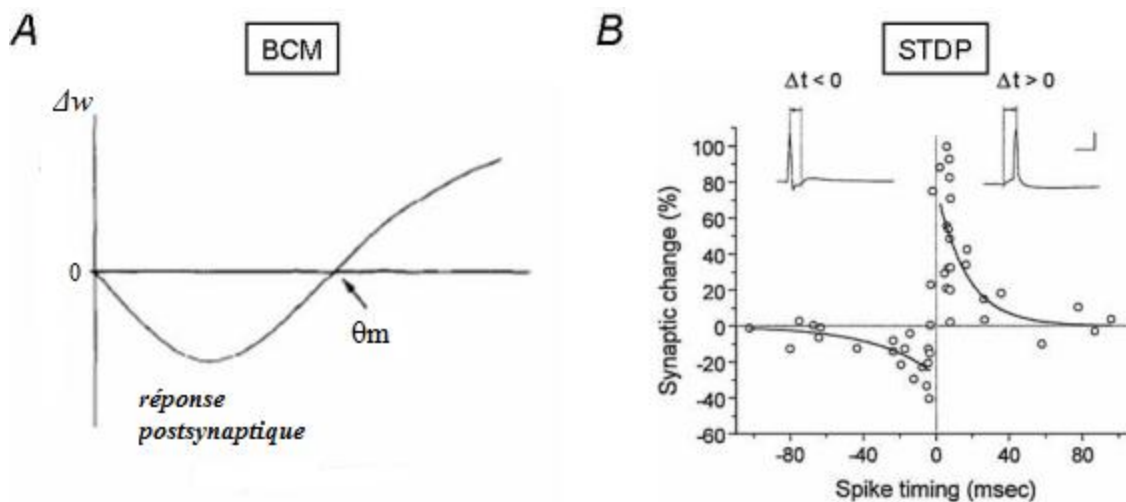


Figure 2.14: Règles de plasticité synaptique

A. Règle BCM (Bienenstock-Cooper-Munroe): [Bie et al.'82] proposent que les synapses actives sont potentialisées quand la réponse postsynaptique totale excède une valeur critique, appelée seuil de modification (θ_m), et que les synapses actives sont déprimées quand la réponse postsynaptique totale est comprise entre 0 et θ_m .

B. Règle STDP: induction de la PLT et de la DLT dépend de la relation temporelle entre l'activité pré et postsynaptique. Dans une échelle de quelques dizaines de ms, la LTP intervient si la cellule postsynaptique décharge juste après l'activation synaptique ($t > 0$). Inversement, si $t < 0$, les synapses sont déprimées [Bi et Poo'98].

7.2. PLT et DLT associatives

Une deuxième PLT qualifiée d'associative et reposant sur le postulat de Donald Hebb concernant la formation d'assemblées cellulaires a été décrite [Heb'49]. " *When an axon of cell A is near enough to excite cell B or repeatedly or consistently takes part in firing it, some growth or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased* ". Cette hypothèse stipule que l'efficacité de la connectivité entre deux cellules augmente lorsqu'il existe une relation causale entre les activités de ces cellules. Le mécanisme proposé par Hebb implique un renforcement des liens fonctionnels préexistants entre les neurones, ce qui aboutit à une stabilisation des schémas d'activation des chemins neuronaux empruntés et donc à un apprentissage.

La PLT associative peut en réalité être induite en synchronisant temporellement l'activité pré- et postsynaptique [Kel et al'86]; [Gus et al'87].

En 1973, Stent proposa un corrélat au postulat de Hebb. " *When the presynaptic axon of cell A repeatedly and persistently fails to excite cell B while cell B is firing some growth or metabolic change takes place in one or both cells such as A's efficiency, as one of the cell firing B, is decreased* ". Cette DLT associative est donc induite lors d'une désynchronisation des activités pré- et postsynaptiques. La plasticité associative dépend donc de la relation temporelle existant entre les activités pré- et postsynaptiques [Sten'73].

8. Méta-plasticité

8.1. Concept de la Méta-plasticité

La capacité des neurones à modifier leur structure est essentielle pour le développement normal, l'apprentissage et la réponse aux lésions cérébrales et aux maladies neurologiques. Au niveau synaptique, l'activité neuronale peut générer des formes persistantes de la plasticité synaptique comme la PLT et la DLT. Plusieurs études ont démontré l'efficacité de la PLT et la DLT dans le processus de mémorisation et d'apprentissage.

Des mesures doivent donc être mises en place pour éviter la saturation de PLT ou de la DLT, qui pourrait finalement nuire à la capacité des réseaux de discriminer les événements et de stocker des informations, et dans le cas des niveaux extrêmes de PLT, conduire à l'excitotoxicité.

Pour réguler la PLT et la DLT tout au long du temps, l'activité neuronale à un moment donné peut changer de cellules ou synapses telles que leur capacité à présenter PLT ou DLT après une période ultérieure de l'activité est modifiée. Cette forme de régulation de plasticité est appelée « Méta-plasticité ».

Le concept de méta-plasticité a été introduit par Abraham en 1996 [Abr et Bea'96]. C'est un procédé biologique largement utilisé dans les domaines de la biologie, de l'informatique médicale, des neurosciences, de physiologie, de neurologie et beaucoup d'autres ([Abr'99], [And et Pha'07],[Kin et al'07],[Mar et al'11],[Jed'02],[Rop et Piq'07]).

Le préfixe 'méta' vient du grec et signifie 'au-delà de', En neurosciences et d'autres domaines la "méta-plasticité" indique un niveau plus élevé de plasticité, exprimée comme un changement ou une transformation dans la façon dont l'efficacité synaptique est modifiée.

La direction et le niveau des changements synaptiques dépendent fortement de la dépolarisation postsynaptique qui se passe pendant l'activation synaptique. la PLT est induite lorsque nous avons de hauts niveaux de dépolarisation postsynaptique, alors que la DLT se produit lors de faibles niveaux de dépolarisation.

La Méta-plasticité est définie comme l'induction de changements synaptiques, qui dépend de l'activité synaptique antérieure [Abr et Bea'96] ,[Abr'99].

La méta-plasticité consiste en un décalage vers la droite du seuil de la PLT en accord avec les poids initiaux des synapses (Voir figure 2.15).

Pour des valeurs plus élevées de poids synaptiques initiaux, la courbe est allongée de sorte que la valeur du seuil PLT correspond à des valeurs plus élevées de l'activité post-synaptique.

Ces changements synaptiques facilitent la dépression synaptique dans les synapses potentialisées et inhibent la potentialisation synaptique dans les synapses déprimées (et vice versa dans les synapses potentialisées).

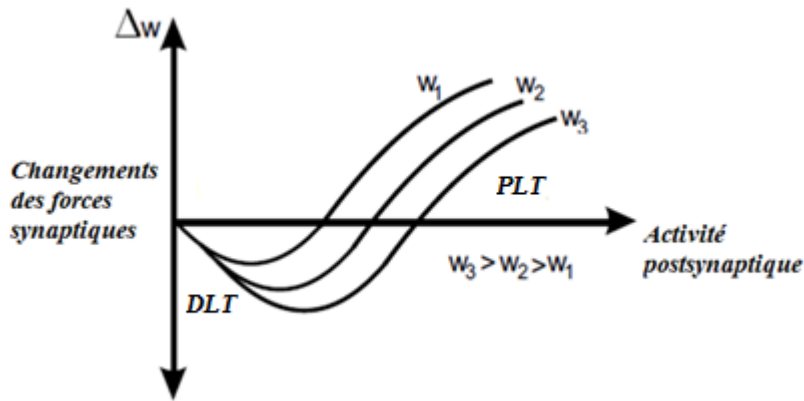


Figure 2.15 : Les changements dans la force synaptique dus à l'activité post-synaptique des neurones biologiques. Si l'activité est élevée, la courbe se déplace vers la droite, ce qui mène à renforcer la PLT. La figure démontre plusieurs courbes en fonction des variations des poids Δw . Pour les grandes valeurs w , la courbe est décalée vers la droite.

Essentiellement, la métaplasticité entraîne un changement dans l'état physiologique ou biochimique des neurones ou des synapses qui modifie leur capacité à générer de la plasticité synaptique. Une caractéristique clé de la métaplasticité est que ce changement dure longtemps et persiste au moins jusqu'à ce qu'un deuxième déclenchement d'activité se produit, ce qui induit une PLT ou une DLT [Abr'08].

8.2. Métaplasticité hétérosynaptique

Dans la nature, la métaplasticité est dans la plupart du temps homosynaptique, c.à.d les synapses qui sont activées lors d'une stimulation, sont également celles qui représentent un changement durable dans la plasticité. Toutefois, l'activité à un ensemble de synapses peut également affecter la plasticité ultérieure au niveau des synapses voisines.

Cette métaplasticité hétérosynaptique a été prédite par le modèle de Bienenstock, Cooper et Munro de plasticité synaptique (BCM) [Bie et al'82], dans laquelle des modifications dans le seuil de la PLT sont provoquées par l'état antérieur de la cellule post-synaptique.

Des Études dans l'hippocampe ont commencé à révéler une variété complexe d'interactions hétérosynaptique qui gouvernent l'induction et la persistance de la PLT.

Au niveau du CA1 in vitro, une stimulation forte d'une voie d'entrée a facilité l'induction de la DLT (et depotentialisation) et a inhibé l'induction de PLT dans un ensemble voisin de synapses.

Le modèle BCM de plasticité synaptique a été développé pour étudier l'activité dépendante de l'expérience dans la plasticité du cortex visuel d'un chaton. Par la suite il a été adapté pour une expérience sur la potentialisation à long terme (PLT) et métaplasticité chez le rat adulte au niveau du gyrus denté. Ils ont démontré que si les niveaux de fréquences sont maintenus à un niveau élevé, le seuil de modification se décale vers la droite, ce qui rend plus difficile à obtenir de la PLT et la DLT plus facile à obtenir.

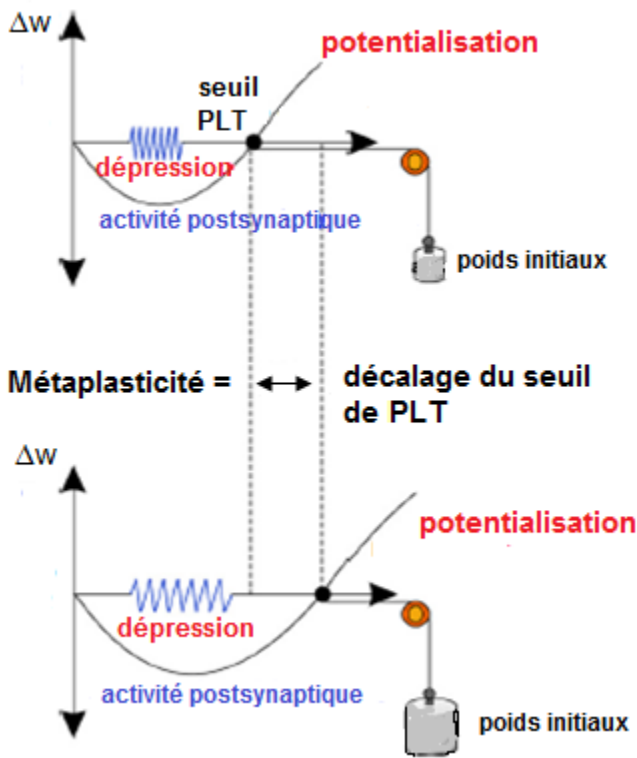


Figure 2.16: Effet de la Métaplasticité. Décalage vers la droite du seuil de la PLT en accord avec les poids initiaux des synapses.

Le concept de métaplasticité va être appliqué durant la phase d'apprentissage de telle sorte que les paramètres les moins fréquents auront des poids synaptiques plus forts que les paramètres les plus fréquents.(figure 2.16).

8.3. Plasticité de la plasticité synaptique

L'apprentissage entraîne des changements durables d'efficacité synaptique dans certains réseaux de neurones et l'inhibition de la plasticité synaptique pendant un apprentissage affecte la mémoire de cet évènement.

Certaines synapses de l'hippocampe sont reconnues pour leur inaptitude à la potentialisation à long terme des récepteurs synaptiques du glutamate, plasticité fondamentalement impliquée dans le processus de mémorisation. Des chercheurs en neurosciences de l'université de Bordeaux ont montré que ces synapses peuvent devenir compétentes pour une telle plasticité, mettant en évidence un phénomène nouveau de « plasticité de la plasticité synaptique »[Reb et al'11].

Les mécanismes moléculaires et cellulaires de la plasticité synaptique utilisent le glutamate comme neurotransmetteur. La libération de ce neurotransmetteur au niveau des synapses entraîne l'activation de récepteurs du glutamate de deux types, les récepteurs AMPA et les récepteurs NMDA. Les propriétés fonctionnelles qui distinguent ces récepteurs leur font jouer un rôle fondamentalement différent dans la plasticité synaptique (potentialisation à long terme) et donc la mémoire. Les récepteurs NMDA, du fait de leur forte perméabilité au calcium, sont essentiels à l'induction de la PLT, mais ne sont pas sujets à potentialisation. Les

récepteurs AMPA sont quant à eux, les cibles principales de cette plasticité : leur nombre s'accroît dans la zone post-synaptique, entraînant une augmentation du signal post-synaptique.

La grande majorité des synapses glutamatergiques suivent les mêmes mécanismes de PLT, certaines synapses n'apparaissent pas compétentes pour exprimer cette plasticité en conditions basales. C'est le cas des synapses entre les neurones du gyrus denté et les cellules pyramidales de CA3, qui constituent le premier niveau de traitement des informations corticales dans l'hippocampe.

Ces synapses sont appelées « fibres moussues hippocampiques », présentent une forme de plasticité qui se traduit par une augmentation durable des récepteurs NMDA. En revanche, ce protocole n'induit aucune PLT des récepteurs AMPA, malgré le renforcement de la réponse NMDA. Cependant, les chercheurs montrent aujourd'hui que si les synapses hippocampiques sont préalablement amorcées par une potentialisation des réponses NMDA, elles deviennent alors aptes à obéir aux mécanismes conventionnels de la PLT, comme les autres synapses glutamatergiques. On parle ici de métaplasticité, ce qui renvoie à la notion de « plasticité de la plasticité synaptique ».

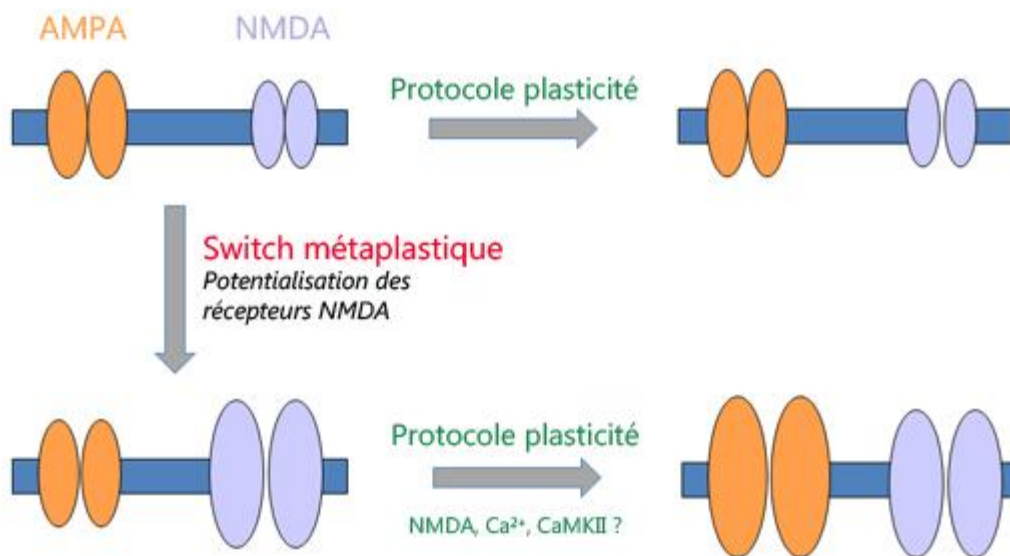


Figure 2.17 : Le switch métaplastique des récepteurs NMDA permet aux synapses fibres moussues hippocampiques d'exprimer la plasticité synaptique conventionnelle des synapses glutamatergiques.

La découverte de ce phénomène suggère que certaines synapses ont un fonctionnement relativement figé, mais que suite à un amorçage permis par la métaplasticité, elles deviennent compétentes, pendant un temps donné, pour exprimer la plasticité synaptique.

9. Conclusion :

La Métaplasticité synaptique est un concept biologique très lié à l'apprentissage des réseaux de neurones biologiques. Pour mieux comprendre ce concept nous avons commencé par une définition détaillée de la plasticité et une description de tous les phénomènes biologiques associés.

Pour résumer, la plasticité synaptique correspond aux modifications morphologiques, chimiques et fonctionnelles qui interviennent au cours du temps au niveau de la synapse.

Toutes les recherches citées dans la littérature nous mènent à affirmer que la métaplasticité entraîne un changement durable et persistant de l'activité synaptique.

Les renforcements de l'activité synaptique sont représentés par la potentialisation à long terme PLT et la dépression à long terme DLT, sachant que plusieurs études ont démontré l'efficacité de la PLT et la DLT dans le processus de mémorisation et d'apprentissage au niveau de l'hippocampe.

Nous allons présenter dans le chapitre 3 les relations existantes entre l'apprentissage naturel et l'apprentissage dans les réseaux de neurones artificiels, puis appliquer une modélisation de la métaplasticité sur un perceptron multi couches.

Chapitre 3

Apprentissage Artificiel et Métaplasticité Artificielle

Chapitre 3 : Apprentissage artificiel et Méta-plasticité artificielle

I. Apprentissage Artificiel des réseaux de neurones

1. Introduction

Nous avons déjà vu dans le chapitre précédent que l'apprentissage naturel implique des modifications physiques des connexions entre neurones.

Les réseaux de neurones biologiques réalisent facilement un certain nombre d'applications telles que la reconnaissance de formes, le traitement du signal, l'apprentissage par l'exemple, la mémorisation, la généralisation. Ces applications sont pourtant, malgré tous les efforts déployés en algorithmique et en intelligence artificielle, à la limite des possibilités actuelles. C'est à partir de l'hypothèse que le comportement intelligent émerge de la structure et du comportement des éléments de base du cerveau que les réseaux de neurones artificiels se sont développés. Les réseaux de neurones artificiels sont des modèles, à ce titre ils peuvent être décrits par leurs composants, leurs variables descriptives et les interactions des composants.

2. Le neurone artificiel

La figure 3.1 montre la structure d'un neurone artificiel. Chaque neurone artificiel est un processeur élémentaire. Il reçoit un nombre variable d'entrées en provenance de neurones amonts. A chacune de ces entrées est associée un poids w abréviation de weight (poids en anglais) représentatif de la force de la connexion. Chaque processeur élémentaire est doté d'une sortie unique, qui se ramifie ensuite pour alimenter un nombre variable de neurones avals. A chaque connexion est associé un poids.

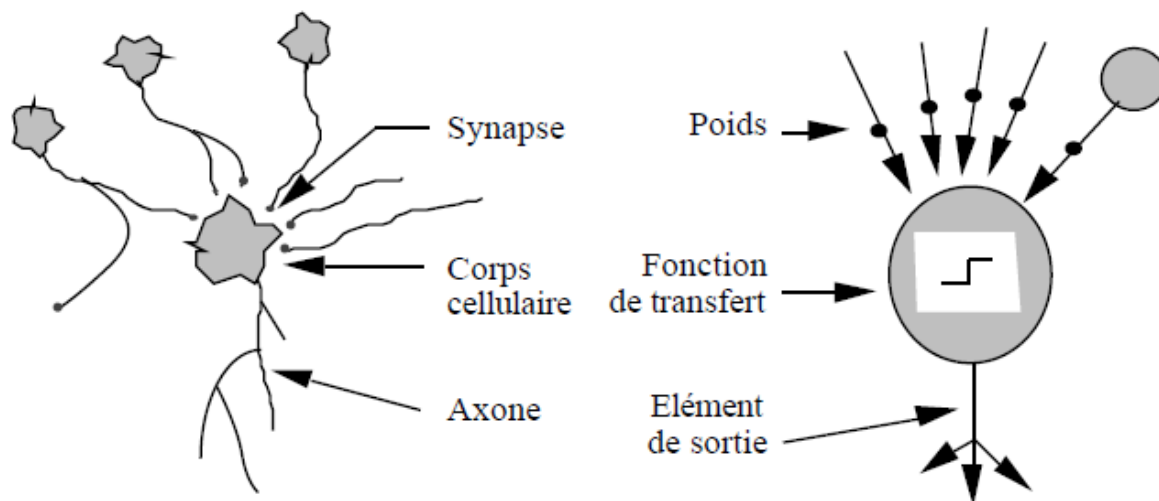


Figure 3.1 : Mise en correspondance : neurone biologique/neurone artificiel

3. Apprentissage

L'apprentissage est vraisemblablement la propriété la plus intéressante des réseaux neuronaux. Elle ne concerne cependant pas tous les modèles, mais les plus utilisés.

Définition :

L'apprentissage est une phase du développement d'un réseau de neurones durant laquelle le comportement du réseau est modifié jusqu'à l'obtention du comportement désiré. L'apprentissage neuronal fait appel à des exemples de comportement.

Apparus en 1985, les réseaux multicouches sont aujourd'hui les modèles les plus employés.

Plusieurs couches de traitement leur permettent de réaliser des associations non linéaires entre l'entrée et la sortie. Ils sont ainsi capables de résoudre le cas du "ou exclusif" [Ros'58].

On sait depuis les années soixante que les possibilités de traitement des réseaux multicouches sont supérieures à celle du Perceptron, cependant l'algorithme d'apprentissage manquait. Pour la couche de sortie, nous pouvons appliquer l'apprentissage du Perceptron, mais comment modifier les poids pour les connexions qui ne sont pas en relation avec un neurone de sortie ?

Le problème est ramené à l'obtention d'une estimation de la valeur désirée pour chaque neurone de la couche cachée. La rétropropagation de gradient est une solution à ce problème.

Cet algorithme a été proposé indépendamment par trois équipes en 1985, dont Y. le Cun[Cun'85]. Des recherches bibliographiques ont montré qu'il s'agit en fait d'une redécouverte.

Le principe utilisé par la rétropropagation ("backpropagation" en anglais) de gradient est la minimisation d'une fonction dépendante de l'erreur. Il s'agit d'une méthode générale, largement employée dans d'autres domaines tels que la physique. Une perception intuitive de cet algorithme consiste à considérer l'apprentissage comme la recherche sur la surface de coût de la position du coût minimal. A chaque configuration de poids correspond un coût. Le gradient est une estimation locale de la pente de la surface. La minimisation du gradient permet de parcourir cette surface orthogonalement aux courbes de niveau d'un pas fixé. Les problèmes rencontrés durant l'apprentissage résultent des zones très plates et des minima locaux.

L'algorithme de la rétropropagation de gradient, bien que très simple à implémenter, nécessite un certain savoir-faire pour une utilisation efficace.

En effet, la convergence de l'algorithme n'est pas prouvée et de multiples variables sont à ajuster précisément en fonction du problème traité. Parmi ces variables à fixer, citons par exemple : les paramètres apparaissant dans les différentes équations (gain de la procédure de gradient (μ), pente de la fonction sigmoïde), la sélection des exemples pour l'apprentissage ($X(t)$; $Y(t)$) et le test, l'ordre de présentation et les distributions relatives des exemples dans la base d'apprentissage, le choix du codage des informations en entrée et en sortie, la structure du réseau (présence éventuelle de connexions directes de la couche d'entrée sur la couche de sortie pour traiter à ce niveau la partie linéaire du problème, limitation pratique du nombre de couches, taille de la couche cachée), la configuration initiale des poids, le nombre d'itérations d'apprentissage,

Nous devons commencer par définir la relation entre les sorties du réseau d'une part, et les entrées et les poids d'autre part.

Dans un réseau à (l) couches ayant (n) entrées et (m) sorties, les relations entre les différents neurones sont données par les équations suivantes :

$$U_i^l(k) = f^l(P_i^l(k)) \quad (3.1)$$

$$P_i^l(k) = \sum_{j=1}^{N_{l-1}} W_{ij}^l U_j^{l-1}(k) \quad (3.2)$$

Où $i=1,2,\dots,N_l$ et $j=1,2,\dots,N_{l-1}$

N_l : nombre de neurones de la couche l

N_{l-1} : nombre de neurones de la couche $l-1$

L : nombre de couches

U : sortie du neurone i de la couche l .

W_{ij} : poids synaptique de la j ème entrée du neurone (i) de la couche (l).

$$U_i^0(k) = X_i(k) \quad , \quad i=1,2,\dots,n.$$

$$U_i^l(k) = Y_i(k) \quad , \quad i=1,2,\dots,m.$$

Où $X_i(k)$ et $Y_i(k)$ sont respectivement les entrées et les sorties du réseau.

Nous notons $\hat{Y}_j^{(l)}$ la sortie réelle du réseau, elle est calculée par l'équation suivante :

$$\hat{Y}_j^{(l)} = f_j^l(W_j^l \hat{Y}_j^{(l-1)}) \quad (3.4)$$

L'objectif de la méthode de la rétropropagation est d'adapter les paramètres des poids W de façon à minimiser une fonction de coût $\mathcal{E}(W)$.

Cette fonction est exprimée par l'erreur quadratique entre la sortie désirée et la sortie calculée par le réseau.

$$\mathcal{E}(W) = \frac{1}{2} \sum_{i=1}^m (y_i^l - \hat{y}_i^l)^2 \quad (3.5)$$

y_j^l : Le vecteur de sortie désirée.

\hat{y}_j^l : Le vecteur de sortie réelle du réseau.

L'approche la plus utilisée pour la minimisation de la fonction de coût est basée sur la méthode du gradient .nous commençons l'apprentissage par un choix aléatoire des poids initiaux.

Nous présentons le premier vecteur d'entrée, une fois nous obtenons la sortie du réseau, l'erreur correspondante et le gradient de l'erreur par rapport à tous les poids sont calculés .les poids sont alors ajustés. Nous exécutons la même procédure pour tous les exemples d'apprentissage. Ce processus est répété jusqu'à ce que les sorties du réseau soient suffisamment proches des sorties désirées.

- Adaptation des poids

Dans notre cas nous utilisons la méthode du gradient avec initialisation des poids aléatoirement et à chaque itération ces poids sont modifiés suivant cette équation :

$$w_{ij}^l(t+1) = w_{ij}^l(t) - \eta \nabla_{ij}^l(t) \quad (3.6)$$

$l=0,\dots,L$ indice des couches , $l=L$ couche de sortie.

t : correspond au compteur des itérations lors de l'apprentissage.

η : facteur d'apprentissage entre 0 et 1.

∇ : gradient de la fonction à minimiser $\mathcal{E}(W)$ défini par : $\nabla = \frac{\partial \mathcal{E}(W)}{\partial w_{ij}^l}$

La vitesse de convergence du réseau dépend du facteur η . sa valeur est généralement choisie expérimentalement.

L'expression du gradient peut être développée selon l'équation suivante :

$$\frac{\partial \varepsilon(W)}{\partial w_{ij}^l} = \frac{\partial \varepsilon(W)}{\partial \hat{y}_j^l} \cdot \frac{\partial \hat{y}_j^l}{\partial w_{ij}^l} \quad (3.7)$$

$$\frac{\partial \hat{y}_j^l}{\partial w_{ij}^l} = \frac{\partial f_j^l(W_j^l \hat{Y}_j^{(l-1)})}{\partial (W_j^l \hat{Y}_j^{(l-1)})} \cdot \frac{\partial (W_j^l \hat{Y}_j^{(l-1)})}{\partial w_{ij}^l} \quad (3.8)$$

$$\text{Notons que : } \frac{\partial f_j^l(W_j^l \hat{Y}_j^{(l-1)})}{\partial (W_j^l \hat{Y}_j^{(l-1)})} = f_i^{l'} \quad (3.9)$$

$$\text{Et } \frac{\partial (W_j^l \hat{Y}_j^{(l-1)})}{\partial w_{ij}^l} = \hat{y}_i^{l-1} \quad (3.10)$$

En prenant l'équation (3.9) et (3.10) nous obtenons :

$$\nabla_{ij}^l = \frac{\partial \varepsilon(W)}{\partial \hat{y}_i^l} \cdot \hat{y}_i^{l-1} \cdot f_i^{l'} \quad (3.11)$$

Où :

W_j^l : Matrice des poids du neurone j dans la couche l .

$\hat{Y}^{(l-1)}$: vecteur d'entree de la couche l

f_j^l : fonction d'activation du noeud j de la couche l

La fonction d'activation est la même pour tous les nœuds.

$$\text{Le terme d'erreur delta peut être défini par : } \delta_j^l = \frac{\partial \varepsilon(w)}{\partial \hat{y}_j^l} \cdot f_j^{l'} = \frac{\partial \varepsilon(w)}{\partial w_{ij}^l} \cdot \frac{1}{\hat{y}_i^{(l-1)}} \quad (3.12)$$

Ainsi nous pouvons réécrire le gradient de la fonction de cout par l'équation suivante :

$$\nabla_{ij}^l = -\delta_j^l \cdot \hat{y}_i^{l-1} \quad (3.13)$$

De cette façon, nous pouvons propager la correction des poids en fonction de l'erreur à la sortie du réseau.

$$\partial \hat{y}_j^l = \sum_{n=0}^{N_{l+1}-1} \frac{\partial \varepsilon(W)}{\partial \hat{y}_n^{(l+1)}} \cdot \frac{\partial \hat{y}_n^{(l+1)}}{\partial \hat{y}_j^l} \quad (3.14)$$

$$\frac{\partial \hat{y}_n^{(l+1)}}{\partial \hat{y}_j^l} = \frac{\partial f_n^{(l+1)}(W_n^{(l+1)} \hat{Y}^{(l)})}{\partial (W_n^{(l+1)} \hat{Y}^{(l)})} \cdot \frac{\partial (W_n^{(l+1)} \hat{Y}^{(l)})}{\partial \hat{y}_j^l} \quad (3.15)$$

Notons que :

$$\frac{\partial (W_n^{(l+1)} \hat{Y}^{(l)})}{\partial \hat{y}_j^l} = W_{jn}^{(l+1)} \frac{\partial f_n^{(l+1)}(W_n^{(l+1)} \hat{Y}^{(l)})}{\partial (W_n^{(l+1)} \hat{Y}^{(l)})} = f_n^{(l+1)'} \quad (3.16)$$

De l'équation (3.16) nous obtenons :

$$\frac{\partial \mathcal{E}(w)}{\partial \hat{y}_j^l} = \sum_{n=0}^{N_{l+1}-1} \frac{\partial \mathcal{E}(W)}{\partial \hat{y}^{(l+1)}} \cdot w_{jn}^{(l+1)} \cdot f_n^{(l+1)'} \quad (3.17)$$

Finalement nous obtenons une équation du gradient en fonction du terme d'erreur et des poids de la couche précédente.

$$\frac{\partial \mathcal{E}(w)}{\partial \hat{y}_j^l} = \sum_{n=0}^{N_{l+1}-1} \delta_n^{(l+1)} \cdot w_{jn}^{(l+1)} \quad (3.18)$$

Pour minimiser l'erreur totale sur l'ensemble d'entraînement, les poids du réseau doivent être ajustés après présentation de tous les exemples.

Nous pouvons définir la variation des poids comme suit :

$$w_{ij}^{(l)}(t+1) = w_{ij}^{(l)}(t) + \eta \cdot \hat{y}_i^{(l-1)}(t) \cdot \delta_j^{(l)}(t) \quad (3.19)$$

Et le terme d'erreur est calculé par :

$$\delta_j^l(t) = f_j^{l'} \sum_{n=0}^{N_{l+1}-1} \delta_n^{(l+1)}(t) \cdot w_{jn}^{(l+1)}(t) \quad (3.20)$$

Le perceptron multi couches est capable de résoudre des problèmes non linéairement séparables en utilisant des fonctions non linéaires, dans notre travail nous utilisons la fonction sigmoïde définie par :

$$f(x) = \frac{1}{1+e^{-x}} \quad (3.21)$$

$$f'(x) = f(x) (1-f(x))$$

4. l'Algorithme de RétroPropagation

- 1- Initialisation des poids $w_{ij}^{(l)}$ aléatoirement
- 2- présentation du vecteur $X(t)$ à l'entrée du réseau sachant que pour chaque $X(t)$ nous avons un $Y(t)$ qui est la sortie désirée.
- 3- Calcul de la sortie du réseau en utilisant la formule suivante :

$$\hat{y}_j^{(l)} = f\left(\sum_{i=0}^{N_{l+1}-1} w_{ij}^{(l)}(t) \hat{y}_i^{(l-1)}(t) - \theta_j^{(l)}\right) \quad (3.22)$$

Avec $0 \leq j \leq N_l - 1$ et $0 \leq l \leq L$.

- 4- Après la modification des poids des couches antérieures, le terme d'erreur delta est calculé par :

$$\delta_j^{(L)} = \frac{\partial \mathcal{E}(w)}{\partial \hat{y}_j^{(L)}} f_j^{(L)'} \quad (3.23)$$

Les poids actualisés de la couche de sortie sont calculés par l'équation suivante :

$$w_{ij}^{(L)}(t+1) = w_{ij}^{(L)}(t) + \eta \cdot \delta_j^{(L)}(t) \hat{y}_i^{(L-1)} \quad (3.24)$$

Avec ; $0 \leq i \leq N_{L-1}$; $0 \leq j \leq N_{L-1}$

Le terme d'erreur pour les autres couches est calculé par :

$$\delta_j^{(l)}(t) = f_j^{(l)'}(t) \sum_{n=0}^{N_{l+1}-1} \delta_n^{(l+1)}(t) \cdot w_{jn}^{(l+1)}(t) \quad (3.25)$$

Et la modification des poids est obtenue par :

$$w_{ij}^{(l)}(t+1) = w_{ij}^{(l)}(t) + \eta \cdot \hat{y}_j^{(l-1)} \cdot \delta_j^{(l)}(t) \quad (3.26)$$

5. Si la condition sur l'erreur ou le nombre d'itérations est atteinte l'algorithme s'arrête (étape 6) , sinon nous passons à l'itération t+1 et retourner à l'étape 2 de l'algorithme.

6 .fin.

Les exemples sont présentés d'une manière récursive, lorsque tous les exemples sont présentés, le test s'effectue sur l'erreur de sortie et les poids sont ajustés au fur et à mesure, jusqu'à ce que l'erreur de sortie se stabilise à une valeur acceptable.

Pour l'apprentissage nous devons tenir compte de la valeur du coefficient d'apprentissage ; si la valeur est proche de 0, le réseau va poursuivre un apprentissage très lent mais mener à une convergence. tandis que si la valeur est proche de 1, le réseau aura une phase d'apprentissage rapide mais peut ne pas garantir la stabilité des résultats.

Bien que l'algorithme de rétropropagation soit le plus utilisé pour l'apprentissage supervisé des PMC, son implémentation se heurte à plusieurs difficultés techniques. Il n'existe pas de méthodes permettant de :

- Trouver une architecture optimale appropriée (nombres de couches, nombre de neurones).
- Choisir une taille et une qualité adéquate d'exemples d'apprentissage.
- Choisir des valeurs initiales satisfaisantes pour les poids, et des valeurs convenables pour les paramètres d'apprentissage permettant d'accélérer la vitesse de convergence.
- Problème de la convergence vers un minimum local, qui empêche la convergence et cause l'oscillation de l'erreur.

Pour résoudre certains des inconvénients de l'algorithme de rétropropagation (ARP), beaucoup de modifications et variantes ont été développées. la plus grande partie concerne le problème de la lenteur de convergence , tout en se concentrant sur l'obtention d'une meilleure généralisation.

Dans la section suivante, nous présentons la meilleure modification de l'ARP qui est basée sur la propriété biologique de la métaplasticité.

II. Méta-plasticité Artificielle

1. Introduction

Dans cette section nous allons appliquer le concept de méta-plasticité sur un perceptron multicouches MAPMC et nous allons décrire la base mathématique derrière cette théorie et la relation existante entre la méta-plasticité et la théorie de l'information de Shannon.

L'algorithme de méta-plasticité sera implémenté durant la phase d'apprentissage du perceptron multi couches suivant la rétropropagation du gradient.

Rappelons que la méta-plasticité est définie comme l'induction de modifications synaptiques, qui dépend de l'activité synaptique précédente. Elle correspond aux variations des niveaux de dépolarisation postsynaptique, qui induisent les changements synaptiques. Ces variations facilitent la potentialisation synaptique et inhibent la dépression synaptique dans les synapses déprimées.

Le renforcement de l'efficacité synaptique est appelée LTP et la diminution est appelée LTD.

2. La Méta-plasticité et la théorie d'information de Shannon

Dans le domaine des RNAs, Hebb a postulé en 1949 que, durant la phase d'apprentissage, les connexions synaptiques des neurones biologiques sont renforcées en raison de la corrélation de l'activité présynaptique et postsynaptique des neurones. [Heb'49].

La plasticité synaptique des réseaux de neurones biologiques a été modélisée dans plusieurs réseaux de neurones comme un changement des valeurs des poids (forces synaptiques) des connexions des neurones artificiels ou des nœuds. Les poids sont les paramètres qui jouent le rôle le plus important dans l'apprentissage et la performance des RNAs.

Des recherches récentes en neurosciences ont montré que ce paradigme classique peut être amélioré. Par exemple, des propriétés biologiques ont été découvertes, elles ont une relation directe avec les poids, telle que la méta-plasticité.

Ces procédés biologiques sont cruciaux pour le développement de nouvelles architectures pour les réseaux de neurones.

Autres mécanismes biologiques particuliers ont été cités dans la littérature, comme la synchronisation entre les neurones, permettant l'identification du neurone qui a la plus forte activation, et le comportement dual (haute/basse fréquence) de certains neurones biologiques peut être utilisé pour améliorer la performance des réseaux de neurones artificiels ([Mar et al'00], [Nev et al'08]).

En 2007, Kinto et al [Kin et al'07] ont appliqué la plasticité synaptique et la plasticité intrinsèque sur un RNA pour identifier la direction du mouvement d'un objet.

La plasticité intrinsèque est caractérisée par deux fonctions particulières: la première est d'assurer l'homéostasie de l'excitabilité intrinsèque du neurone, ce qui veut dire que le neurone s'adapte aux moindres changements de l'excitation pour toujours produire la même réponse. Notamment lorsque la fréquence d'entrée augmente progressivement, l'excitabilité du neurone baisse de manière à garder la même fréquence de sortie. et la deuxième fonction est qu'elle permet sous certaines conditions de potentialiser l'excitabilité du neurone, par exemple après

une série de potentiels d'actions à haute fréquence sur une faible durée, le neurone augmente sa fréquence de sortie en augmentant l'excitabilité.

Monteiro et al [Mon et al'08] ont modélisé avec succès un réseau de neurones pour simuler la maladie d'Alzheimer, en se basant sur la plasticité intrinsèque (la plasticité intrinsèque est liée directement avec la métaplasticité). Dans son travail Monteiro a renforcé l'hypothèse selon laquelle la dérégulation du calcium est liée à la maladie d'Alzheimer.

Actuellement plusieurs modèles de métaplasticité ont été proposés, mais le plus efficace du point de vue de l'apprentissage et la performance est celui qui utilise la théorie d'information de Shannon [Sha'48]. Sachant que la théorie d'information de Shannon vise à quantifier et qualifier la notion de contenu en information présent dans un ensemble de données.

C'est une théorie probabiliste permettant de quantifier le contenu moyen en information d'un ensemble de données, dont le codage satisfait une distribution statistique précise. Ce domaine trouve son origine scientifique avec Claude Shannon qui en est le père fondateur avec son article « *A Mathematical Theory of Communications* » publié en 1948. [Sha'48]

D'après ces études et travaux de recherche nous pouvons confirmer que la modélisation d'une de ces propriétés biologiques des neurones, comme la métaplasticité, pourra participer à améliorer l'apprentissage et la performance des RNAs, sachant que cette propriété est étroitement liée à l'apprentissage dans les réseaux de neurones biologiques.

Suivant cette stratégie nous allons tester le concept de métaplasticité artificielle sur un perceptron multi couches (MAPMC). Ce modèle se base sur la théorie de Shannon et établit que les variables d'entrée les moins fréquentes contiennent plus d'information que les variables les plus fréquentes.

3. Algorithme de rétropropagation et métaplasticité artificielle :

la stratégie que nous abordons dans notre processus d'apprentissage du perceptron multicouches, est de minimiser l'erreur ; E_M , définie par l'expression suivante :

$$E_M = \varepsilon \{E(x)\} \quad (3.26)$$

L'idée est d'estimer E par $E(x) = \int f(x) dx$

Avec ε représente l'espérance mathématique (la valeur moyenne que peut prendre une variable aléatoire X est son espérance).

Par définition nous posons que si la variable aléatoire X admet une densité de probabilité f , son espérance est définie comme : $E(x) = \int xf(x) dx$

Où X est la variable aléatoire des vecteurs d'apprentissage de l'entrée du réseau : $x = \{x_1, x_2, \dots, x_n\}$, $x \in \mathbb{R}^n$ où \mathbb{R}^n est l'espace n dimensionnel et $E(x)$ est la fonction d'erreur, qui définit l'erreur entre la sortie et l'entrée du réseau. Les poids sont actualisés et modifiés à chaque itération suivant la stratégie de convergence vers l'erreur minimale. Pour introduire la métaplasticité, nous utilisons la fonction suivante :

$$E_M = \int_{R^n} E(x) f_X(x) dx = \int_{R^n} e(x) \quad (3.27)$$

$$E_M = \int_{R^n} \frac{e(x)}{f_X^*(x)} f_X^*(x) dx = \mathcal{E}^* \left\{ \frac{e(x)}{f_X^*(x)} \right\} \quad (3.28)$$

Nous définissons $e(x) = E(x) - f(x)$

Et calculer E avec l'estimateur suivant : \hat{E}_M

Si nous disposons de plusieurs observations x nous utilisons la loi des grands nombres définie par l'équation qui suit :

$$\hat{E}_M = \frac{1}{M} \sum_{k=1}^M \frac{e(x_k^*)}{f_X^*(x_k^*)} \quad (3.29)$$

Où x_k^* , $k=1,2,\dots,M$, sont les vecteurs des échantillons indépendants dont la fonction de densité de probabilité (pdf probability density function) est f_X^* qui peut être choisie arbitrairement, si $f_X^*(x) \neq 0$, partout où $e(x) \neq 0$, $\forall x \in R^n$.

Nous pouvons dire que l'équation est idéale pour :

$$(f_X^*(x))_{opt} = \frac{1}{E_M} e(x) \quad (3.30)$$

En pratique ceci est impossible de prédire $(f_X^*(x))_{opt}$, mais nous pouvons utiliser une fonction d'optimisation. L'objectif de notre travail est de trouver une fonction dont l'effet sur l'équation est l'amélioration de la convergence dans l'apprentissage. Plus près de l'optimum ; les résultats obtenus seront meilleurs.

4. Implémentation du MAPMC dans la phase d'apprentissage du PMC

Lorsque l'on cherche à étudier une suite de mesures provenant de la répétition d'une expérience, une méthode de modélisation consiste à supposer que ces mesures sont des réalisations de variables aléatoires indépendantes. Comprendre ces mesures et la façon dont elles sont distribuées revient à étudier la loi de probabilité de la variable aléatoire sous-jacente. Lorsque l'on n'a pas d'idée a priori sur la forme particulière que peut prendre la densité f , nous pouvons utiliser un estimateur de f calculé à partir de la moyenne et de la variance, comme c'est le cas pour des lois gaussiennes.

Dans le cas d'un Perceptron Multi Couches qui utilise l'algorithme de rétropropagation pour l'apprentissage, appliqué pour la classification de L classes H_l , $l=0,1,2,\dots,L-1$.

Les études ont montré que la sortie de chaque classe dans le PMC est l'estimation de la probabilité a posteriori de la classe, basé sur le théorème de Bayes.

Ce type de méthodes repose sur l'idée que les variables et les classes peuvent être traitées comme des variables aléatoires. Ainsi, l'appartenance à la classe d'une observation tirée au hasard dans l'espace des observations de l'échantillon est une réalisation d'une variable aléatoire dont la valeur est le numéro de la classe (variable discrète).

Dans les approches statistiques relatives aux problèmes de classification et plus généralement dans la théorie de la décision [Bis'06], la règle de décision de Bayes est fondamentale, car elle fournit la limite théorique du taux d'erreur d'un modèle. [Dre et al'02] proposent la définition de la règle suivante :

Définition 1 : Pour affecter une observation à une classe, nous minimisons le risque d'erreur en prenant la décision d'affecter l'observation à la classe dont la probabilité a posteriori est la plus grande.

Dans le cadre de la classification, cette limite est obtenue par la connaissance exacte de la distribution des observations pour chacune des classes. Ainsi, les fonctions de densités de probabilité conditionnelles des l classes $p(x/ H_l)$, dites aussi fonctions de vraisemblance, et les probabilités a priori $P(H_l)$ sont connues. Dès lors, en manipulant les règles de la théorie des probabilités, [Bis'06] montre que le théorème de Bayes permet d'obtenir l'affectation d'une nouvelle observation aux classes, par l'intermédiaire des probabilités a posteriori $P(H_l/ x)$:

$$P(H_l/x) = P(x/H_l) \cdot P(H_l) / P(x)$$

Ainsi, selon la règle de décision, l'observation x est affectée à la classe H_l , où la probabilité a posteriori $P(H_l/x)$ est maximum.

Ce qui nous donne dans notre cas:

$$\hat{y}_l = P(H_l/x) = \frac{f_X(x/H) P(H_l)}{f_X(x)} \quad (3.31)$$

Cette équation va permettre l'implémentation de la métaplasticité.

En prenant compte du modèle de métaplasticité proposé précédemment, nous pouvons supposer pour chaque classe l que : $f_X^*(x) = f_X(x)$

En prenant l'équation (3.29) nous obtenons :

$$\hat{E}_{M_l} = \frac{1}{M_l} \sum_{k=1}^{M_l} \frac{E(x_k) f_X(x_k/H_l)}{f_X(x_k)} \quad (3.32)$$

$$\cong \frac{1}{M_l} E(x_k) \frac{\hat{y}}{P(H_l)} \quad (3.33)$$

Où $k=1, 2, \dots, M_l$ représentent les vecteurs d'échantillons indépendants de la classe l dans la base d'apprentissage. De l'équation (3.29) et l'équation précédente nous pouvons en déduire :

$$\frac{\hat{y}_l}{P(H_l)} \cong \frac{1}{f^*(x)} \quad (3.34)$$

Pour introduire d'une manière analytique la métaplasticité dans l'étape d'apprentissage d'un PMC, tout ce qui doit être fait est d'introduire la fonction de pondération dans la fonction d'erreur entre la sortie réelle et la sortie désirée du réseau, en fonction de la matrice de poids W (t) à chaque itération de formation, t, qui est donnée par l'équation suivante :

$$E^*[W(t)] = \frac{E[W(t)]}{f_X^*(x)} \quad (3.35)$$

L'équation indique que l'estimation de l'erreur $E(x)$ peut être pondérée par une fonction adéquate sans affecter l'objectif final, qui est minimiser l'erreur de chaque classe.

Cette équation est implémentée en incluant la fonction optimale dans l'équation du calcul de l'erreur de l'ARP :

$$\frac{\partial \varepsilon(W)}{\partial w_i^{(s)}} = \frac{\partial}{\partial w_i^{(s)}} \left(\frac{1}{2} \frac{(y - \hat{y}^{(s)})^2}{f_X^*(x)} \right) = \frac{1}{f_X^*(x)} \frac{\partial \varepsilon(W)}{\partial w_i^{(s)}} \quad (3.36)$$

$$\delta_j^{(s)} = (y - \hat{y}_j^{(s)}) \cdot \frac{f_j'^{(s)}}{f_X^*(x)} \quad (3.37)$$

Où s est le compteur des couches, $s=1,2,\dots,S$, j et i représentent les nœuds et les entrées .

La métaplasticité est introduite dans la phase d'apprentissage en affectant aux poids à chaque itération la fonction suivante :

$$w^*(x) = \frac{1}{f^*(x)} = \frac{\hat{y}_l}{P(H_l)} = \frac{f_X(x/H_l)}{f_X(x)} \quad (3.38)$$

De l'équation (3.32) nous obtenons l'estimation de la probabilité a posteriori de chacune des classes d'entrée et de sa sortie du PMC , dans ce cas la distribution statistique des échantillons d'apprentissage est utilisée pour quantifier la fréquence d'un échantillon. Notez que si ce n'est pas le cas, comme cela se produit dans les premiers pas d'algorithmes d'apprentissage de l'ARP, l'apprentissage peut ne pas converger. Dans les premières itérations, les sorties du PMC ne fournissent pas encore d'estimation valable des probabilités a posteriori, mais plutôt des valeurs aléatoires correspondant à des estimations des poids du PMC, W . Il est alors préférable dans ces premières étapes de l'apprentissage, soit d'appliquer l'algorithme classique de l'ARP ou d'utiliser une autre fonction de pondération pour que l'algorithme de rétropropagation commence à minimiser le taux d'erreur.

Le théorème de Bayes fait intervenir les densités de probabilité conditionnelles et les probabilités *a priori*, or nous avons évoqué qu'il est rare qu'elles soient connues exactement. C'est pourquoi, il est nécessaire de les estimer à partir des observations.

la fonction qui va suivre est proposée pour la modification des poids, elle suit une distribution gaussienne ,nous allons la nommer fonction de pondération par métaplasticité .

$$w_X^*(x) = \frac{A}{\sqrt{(2\pi)^N} \cdot e^{B \sum_{i=1}^N x_i^2}} \quad (3.39)$$

N représente le nombre de neurones dans la couche d'entrée du perceptron, A et $B \in \mathbb{R}^+$ sont des paramètres d'optimisation de l'algorithme déterminés expérimentalement.

Et le renforcement des poids dans chaque itération est donné par :

$$w_{ij}^{(l)}(t+1) = w_{ij}^{(l)}(t) + \eta \cdot \delta_j^{(l)} \cdot \hat{y}_i^{(l-1)} \quad (3.40)$$

$$\delta^{(L)} = (y - \hat{y}^{(L)}) \cdot \hat{y}^{(L)} \cdot (1 - \hat{y}^{(L)}) / f_X^*(x) \quad (3.41)$$

Donc, comme la pdf de la fonction de poids proposée est une distribution des échantillons d'entrée qui ne dépendent pas des paramètres du réseau, le MAPPMC peut être considéré comme étant une opération de mise à jour des poids à chaque itération de l'apprentissage du PMC représentée par l'équation qui suit :

$$\Delta^*w = w^*(x)\Delta w \tag{3.42}$$

Avec $\Delta w = w(t+1) - w(t)$ est le changement obtenu par l'algorithme de rétropropagation classique et $w^*(x)$ représente la contribution de la fonction de poids pour chaque échantillon d'entrée x .

Durant la phase d'apprentissage, le MAPPMC peut être considéré comme une nouvelle méthode de probabilité de la loi présynaptique, qui stipule ;durant la phase d'apprentissage l'algorithme affectera de plus grandes valeurs de poids aux éléments les moins probables et des valeurs de poids minimales pour les éléments les plus probables.

5. Description de l'algorithme MAPPMC

L'idée proposée dans cette thèse définit le MAPPMC comme une procédure d'apprentissage pour améliorer l'algorithme de minimisation des erreurs utilisé lors de l'entraînement du réseau de neurones. L'idée est d'utiliser une fonction sur les poids synaptiques de telle façon à donner plus de pertinence aux variables du vecteur d'apprentissage à l'entrée les moins fréquentes et à diminuer la contribution aux variables les plus fréquentes. Ce modèle de métaplasticité artificielle est représenté dans la figure suivante.

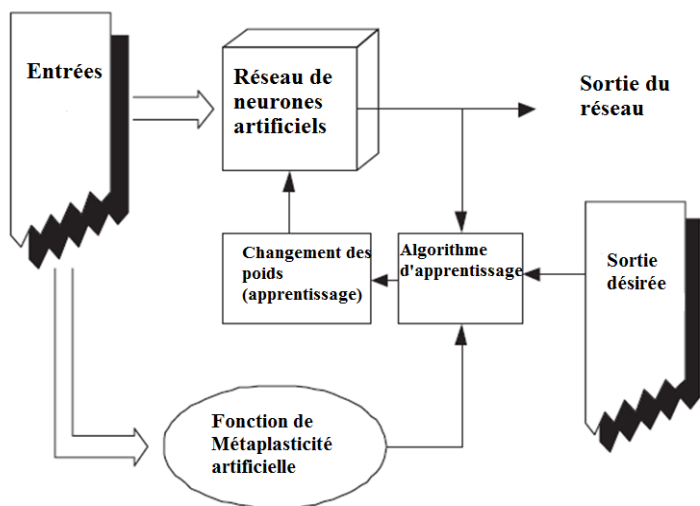


Figure 3.2 : Schéma bloc du fonctionnement de l'algorithme MAPPMC. la fonction de métaplasticité est appliquée durant la phase d'apprentissage.

6. Conclusion :

Nous avons appliqué dans ce chapitre la modification proposée dans cette thèse afin d'améliorer l'algorithme de rétropropagation en terme de rapidité de convergence et de fiabilité dans la classification. Le concept de métaplasticité est appliqué sur un perceptron multi couches durant la phase d'apprentissage. Des résultats très satisfaisants ont été obtenus, ils seront présentés dans le dernier chapitre.

Chapitre 4

Sélection de Variables

Chapitre 4

Sélection de variables

1 .Introduction

La sélection de variables est un problème difficile à résoudre. Il concerne le choix des variables pertinentes pour caractériser un objet ciblé.

Dans tout système de reconnaissance de formes et de classification, la qualité du classifieur dépend énormément de la bonne sélection des variables d'entrée.

Dans le monde réel nous faisons face à des systèmes complexes qui sont représentés par des vecteurs de caractéristiques (ou variables) d'une dimensionnalité trop importante, d'où la nécessité de faire appel à des méthodes de réduction et de sélection de variables. La résolution de phénomènes complexes avec des descripteurs de grande taille pourrait être gérée en utilisant peu de caractéristiques extraites des données initiales, il suffit qu'elles représentent convenablement le problème à résoudre.

Une méthode de réduction de la dimensionnalité est souvent définie comme un processus de pré-traitement de données qui permet de supprimer les informations redondantes et bruitées.

Dans la littérature les méthodes de réduction de la dimensionnalité sont généralement scindées en deux catégories:

- Une réduction basée sur une transformation des données appelée aussi une extraction de caractéristiques, elle consiste à remplacer l'ensemble initial des données par un nouvel ensemble réduit, construit à partir de l'ensemble initial de caractéristiques.

- Une réduction basée sur une sélection de caractéristiques qui consiste à sélectionner les caractéristiques les plus pertinentes à partir de l'ensemble de données des variables décrivant le phénomène étudié.

Dans cette thèse nous nous intéressons à la sélection de caractéristiques, néanmoins nous allons présenter quelques méthodes d'extraction de caractéristiques, en détaillant leurs algorithmes et leurs applications citées dans la littérature.

Nous citons en premier lieu, quelques méthodes d'extraction de variables ou de caractéristiques, bien connues et largement sollicitées dans la littérature.

2 .Extraction de caractéristiques

Nous soulignons la nécessité de réduire la dimension d'un problème, notamment pour améliorer les performances de classification [Lan et al'96]. Par ailleurs, la création d'un outil de classification revient à créer un modèle, et il paraît légitime que ce modèle soit le plus simple possible. Ainsi, tout en renforçant la classification, la simplicité du modèle améliore la vitesse d'exécution, le pouvoir de généralisation, ainsi que la non corrélation des entrées du modèle (les variables ou les caractéristiques nécessaires à la modélisation [Dre et al'02]).

Pour atteindre ce but, il existe des méthodes de réduction de dimension par des méthodes de projection linéaires et non linéaires. comme ACP(analyse en composantes principales), ACC (analyse en composantes curvilignes),ALD (analyse linéaire discriminante) et LLE(locally linear embedding ;plongement localement linéaire),AFD (analyse factorielle discriminante),MDS (Multi Dimensionnal Scaling,Positionnement Multi-Dimensionnel) , isomap, que nous détaillerons dans les sections suivantes. Ces dernières permettent d'obtenir une nouvelle représentation, par la transformation des variables originales en caractéristiques.

La réduction de la dimensionnalité par une transformation de données (appelée aussi extraction de caractéristiques) se fait par une construction de nouvelles caractéristiques obtenues en combinant les caractéristiques initiales. Une transformation de données risque de faire perdre la sémantique de l'ensemble initial de caractéristiques et donc l'utilisation de cette famille de méthodes n'est applicable que dans le cas où la sémantique n'intervient plus dans les étapes qui suivent la réduction.par exemple dans le cas d'un classifieur des arythmies cardiaques , les caractéristiques d'entrée ont toutes une signification précise pour la détection de la maladie. Lors de la transformation de ces caractéristiques par une méthode comme l'ACP ,nous allons obtenir des composantes principales à partir de ces caractéristiques, donc nous allons perdre la sémantique ,et l'interprétabilité des résultats obtenus sera difficile. Dans le cas de la classification des données médicales il faut prendre en considération le point d'interprétabilité , car ces résultats vont être présentés à un médecin pour l'aider à poser son diagnostic.

Cependant, malgré leur efficacité, deux inconvénients subsistent. Le premier concerne la difficulté d'obtenir des informations précises sur la constitution des nouvelles caractéristiques, diminuant par conséquent la compréhension du modèle. Le second, et non le moindre, concerne un aspect plus pratique relatif au processus global des systèmes de reconnaissance de formes. En effet, avec les méthodes d'extraction des variables, la nouvelle représentation des données nécessite l'ensemble des variables originales lors de la projection. En effet, les nouvelles caractéristiques sont des combinaisons linéaires ou non linéaires des variables. Par conséquent, les entrées du modèle et donc la dimension de l'espace d'apprentissage est bien réduite, mais le nombre de variables à recueillir reste inchangé.

Il existe aussi les problèmes causés par des variables redondantes, notamment d'un point de vue algorithmique. De plus, ces variables, repérables par des corrélations élevées, peuvent être considérées comme superflues, dans le sens qu'aucune information supplémentaire n'apparaît en les ajoutant.

Les sections suivantes décrivent brièvement plusieurs techniques de réduction connues. Elles sont généralement groupées en deux catégories : les méthodes linéaires et les méthodes non linéaires.

2.1 Méthodes linéaires

Il y a deux types de méthodes en statistique multidimensionnelle : les méthodes factorielles, qui consistent à projeter le nuage de points sur un sous-espace, en perdant le moins d'information possible ; et les méthodes de classification, qui tentent de regrouper les points.

Les méthodes factorielles regroupent trois techniques fondamentales : l'analyse en composantes principales « ACP » (plusieurs variables quantitatives), l'analyse des correspondances « AC » (deux variables qualitatives, représentées par un tableau de contingences) et l'analyse des correspondances multiples « ACM » (plus de deux variables qualitatives) .

Le tableau suivant regroupe ces différentes méthodes :

Méthode	Variables quantitatives	Variables qualitatives
ACP	plusieurs	aucune
AC	aucune	deux
ACM	aucune	plusieurs
ALD	plusieurs	aucune
MDS	plusieurs	aucune

Tableau 4.1 : Différentes méthodes linéaires.

2.1.1 Analyse en Composantes Principales

L'Analyse en Composantes principales (ACP) fait partie du groupe des méthodes descriptives multidimensionnelles appelées méthodes factorielles.

L'ACP est une méthode d'analyse de données très utilisée et très connue en statistiques et dans les sciences expérimentales, développée par Harold Hotelling [Hot'33].

Nous supposons disposer d'un ensemble X de N données, chacune décrites par P attributs. Si l'on considère les données comme des points dans un espace euclidien à P dimensions, l'objectif de l'ACP est de construire l'espace euclidien à P dimensions le plus caractéristique et le plus économique pour représenter ces points. L'objectif est donc de passer de l'espace des données à un espace de caractéristiques (feature space) ou espace factoriel.

Donc L'ACP est définie comme une technique qui permet de trouver des espaces de dimensions plus petites dans lesquels il est possible d'observer au mieux les individus. Sa démarche essentielle consiste à transformer les variables quantitatives initiales, plus ou moins corrélées entre elles, en des variables quantitatives, non corrélées, combinaisons linéaires des variables initiales et appelées composantes principales. Les composantes principales sont donc de nouvelles variables indépendantes, combinaisons linéaires des variables initiales, possédant une variance maximale.

Globalement l'ACP consiste à rechercher la direction suivant laquelle le nuage de points des observations s'étire au maximum. A cette direction correspond la première composante principale. La seconde composante principale est déterminée de telle sorte qu'elle soit la plus indépendante possible de la première ; elle est donc perpendiculaire à celle-ci. Ces deux composantes forment le premier plan principal. Cette opération est répétée de manière à trouver toutes les composantes principales expliquant le maximum de variance.

L'ACP étant une méthode de réduction de dimension, il est important de savoir qu'elle ne peut pas retenir la totalité de l'information contenue dans le nuage de points initial. Enfin, l'ACP prend uniquement en compte les dépendances linéaires entre les variables et ne peut donc pas fournir une projection pertinente pour une distribution non-linéaire de points.

2.1.2 AC Analyse des correspondances simples

Cette méthode est basée sur l'analyse en composantes principales : nous partons d'un tableau de nombres (les valeurs des variables ou une table de contingence), ses colonnes sont considérées comme des points d'un espace de dimension n (idem pour les lignes), et nous recherchons un sous-espace de dimension deux sur lequel nous voyons le plus de choses (faire une projection orthogonale sur ce sous-espace, au sens du produit scalaire canonique de \mathbb{R}^n ou d'un autre produit scalaire).

L'analyse des correspondances s'intéresse aux tableaux de contingence, i.e., aux variables qualitatives. D'un point de vue technique, l'analyse des correspondances ressemble beaucoup à l'analyse des composantes principales. Il s'agit de représenter graphiquement les points correspondant aux différentes lignes du tableau (resp, les colonnes), de manière à avoir un nuage de points les plus espacés possibles (i.e., en maximisant la variance). Nous procédons donc exactement comme pour l'ACP, à ceci près que nous ne mesurons pas les distances avec la métrique canonique, mais à l'aide de la « distance du khi2 ».

2.1.3 ACM l'analyse des correspondances multiples

L'analyse des correspondances simples s'intéressait à deux variables statistiques qualitatives ; l'analyse des correspondances multiples s'intéresse à plusieurs variables qualitatives. Les données ne sont généralement pas représentées par une matrice de contingence (on devrait plutôt dire une hypermatrice de contingence), car cette matrice aurait un nombre d'éléments vraiment énorme, et la plupart de ces éléments seraient nuls, mais par un tableau du même genre que pour les variables quantitatives.

L'ACM décrit les relations deux à deux entre p variables qualitatives à travers une représentation des groupes d'individus correspondant aux diverses modalités.

Une des variantes de cette méthode est le Modèle log-linéaire (régression de Poisson) : Le modèle remplit à peu près les mêmes rôles que l'analyse des correspondances multiples mais il a l'avantage, comme toute régression, de permettre de faire des tests.

2.1.4 Analyse Linéaire Discriminante

L'analyse linéaire discriminante, appelée aussi analyse discriminante linéaire de Fisher, est une méthode de réduction du nombre de dimensions proposée par Fisher en 1936 [Fis'36]. Cette méthode s'applique lorsque les classes des individus sont connues.

C'est une méthode descriptive et explicative, apparentée à l'analyse en composantes principales (ACP), s'appliquant à des données quantitatives sur lesquelles est déjà définie une typologie ou partition.

Par exemple nous cherchons à prédire une variable qualitative ,qui peut être « la patiente a-t-elle une tumeur bénigne ou maligne », à l'aide de variables quantitatives qui représentent différentes caractéristiques de ces tumeurs (la forme de la tumeur, l'épaisseur de cette tumeur, le nombre ,etc). nous essayons de trouver quels sont les caractéristiques les plus importantes, i.e, quelles variables déterminent le mieux la variable qualitative.

L'idée de Fisher a été de créer une méthode pour choisir entre les combinaisons linéaires des variables celles qui maximisent l'homogénéité de chaque classe. En d'autres termes, cette méthode consiste à chercher un espace vectoriel de faible dimension qui maximise la variance interclasse.

2.1.5 Positionnement Multi-Dimensionnel

Les techniques de positionnement multidimensionnel trouvent leur origine dans les études psychométriques [Ric'38] visant à comprendre comment les individus tissent des associations entre objets pour effectuer des regroupements, des classifications. Cependant, des études antérieures en biologie comparée, classant les espèces sur la base des réactions sériques interspécifiques , utilisent sans la nommer une construction géométrique s'apparentant au positionnement multidimensionnel[Des'05]. Depuis lors, le positionnement multidimensionnel est devenu une technique de représentation géométrique largement utilisée dans des domaines aussi divers que le marketing, la sociologie électorale, ou plus récemment l'analyse sensorielle.

L'idée fondamentale du positionnement multidimensionnel [Mes et Abe'56], MDS (Multi Dimensional Scaling) est de représenter chaque objet ou stimulus dans un espace euclidien, habituellement bi ou tridimensionnel, de telle sorte que deux objets semblables soient représentés par deux points proches l'un de l'autre, et un couple dissemblable par des points éloignés.

Son objectif est de construire, à partir d'une matrice de distances ou des mesures de similarité calculées sur chaque paire de points, une représentation euclidienne des individus dans un espace de dimension réduite qui préserve "au mieux" ces distances.

Comme l'ACP, l'algorithme de Positionnement Multi-Dimensionnel n'est valable que sur une distribution linéaire de données.

2.2 Méthodes non-linéaires

Les méthodes non linéaires ont pour objectif d'optimiser les représentations afin qu'elles reflètent au mieux la topologie initiale des données.

2.2.1 Isomap

Isomap [Ten et al'00] est une technique de réduction de dimensionnalité non-linéaire qui tient compte de l'ensemble des données.

Cette méthode est une extension de MDS (méthode de positionnement multidimensionnel), mais linéarité locale seulement elle part de la connaissance d'une matrice de dissimilarités entre les paires d'individus. Elle calcule la distance des voisins (K-plus proche ou radius) et calcule par la suite les plus courts chemins de tout les points à tout les points. Le but est cette fois de trouver une variété (non linéaire) contenant les données.

2.2.2 Plongement localement linéaire

Plongement localement linéaire, LLE (Locally Linear Embedding) en anglais, est une méthode de réduction non-linéaire proposée en 2000 [Row et Sau'00]. Elle utilise une approche différente d'Isomap mais partage la même philosophie : les points dans un espace de haute dimensionnalité étant voisins doivent se retrouver proches dans une projection de faible dimension. Donc elle s'attache à ce que la disposition des plus proches voisins soit préservée.

LLE n'a pas besoin de calculer les distances entre tous les points de la variété. Il agit sur le voisinage de chaque vecteur pour découvrir la structure globale de l'ensemble. L'idée principale consiste donc à déterminer pour chaque vecteur x_i , point de l'espace contenant les données à projeter, ses k voisins calculés sur la base de la distance euclidienne et de considérer que dans ce voisinage, les données sont linéaires. Une fois les voisins de x_i déterminés, nous calculons les poids de reconstruction w_{ij} représentant le voisinage de chaque vecteur x_i associé à chaque couple (x_i, x_j) , où x_j appartient au voisinage de x_i . Les w_{ij} permettront par la suite de reconstruire la topologie de voisinage de chaque vecteur x_i dans l'espace de projection.

3. Sélection de caractéristiques

3.1 Définition de la sélection

La sélection de caractéristiques est généralement définie comme un processus de recherche permettant de trouver un sous-ensemble "pertinent" de caractéristiques parmi celles de l'ensemble de départ. La notion de pertinence d'un sous-ensemble de caractéristiques dépend toujours des objectifs et des critères du système.

La sélection de caractéristiques pertinentes est une étape très importante dans un processus de classification ,en effet la performance de ce dernier est étroitement liée au bon choix des variables d'entrée. Des caractéristiques redondantes ou peu pertinentes peuvent réduire la performance du classifieur.

Dans la littérature, il existe plusieurs définitions de la pertinence d'une caractéristique, la plus connue est celle de [Joh et al'94], [Joh'97]. Selon cette définition, une caractéristique est classée comme étant très pertinente, peu pertinente et non pertinente.

Caractéristique très pertinente : Une caractéristique est dite très pertinente si son absence entraîne une détérioration significative de la performance du système de classification utilisé.

Caractéristique peu pertinente : Une caractéristique f_i est dite peu pertinente si elle n'est pas "très pertinente" et s'il existe un sous-ensemble V tel que la performance de $V \cup f_i$ soit significativement meilleure que la performance de V .

Caractéristique non pertinente : Les caractéristiques qui ne sont ni "peu pertinentes" ni "très pertinentes" représentent les caractéristiques non pertinentes. Ces caractéristiques seront en général supprimées de l'ensemble de caractéristiques de départ.

3.2 Processus général de la sélection de caractéristiques

Une sélection de caractéristiques présente plusieurs avantages liés à la réduction de la quantité de données (moins de caractéristiques). D'une part, cette réduction rend beaucoup plus facile de traiter les données et d'autre part, elle aide à mieux comprendre les résultats obtenus par un classifieur basé sur ces caractéristiques. Par exemple, pour un problème de classification, ce processus de sélection ne réduit pas seulement le temps d'apprentissage mais il aide aussi à mieux comprendre les résultats obtenus par le classifieur et à améliorer parfois la précision de la classification.

Une méthode de sélection passe généralement par quatre étapes [Liu et Yu '05] , nous les présentons dans la figure suivante:

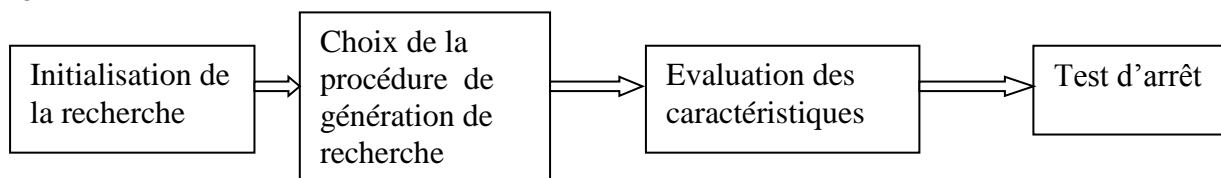


Figure4.1 : Processus de Sélection de variables

3.2.1 La première étape

Elle consiste à initialiser le point de départ à partir duquel la recherche va commencer : Par exemple, une recherche peut commencer par un ensemble vide de caractéristiques, et continuer par l'ajout successif, à chaque itération, d'une ou plusieurs caractéristiques. Inversement, la recherche peut commencer avec l'ensemble de toutes les caractéristiques et continuer le processus par la suppression séquentielle, à chaque itération, de la caractéristique la moins pertinente. Une autre façon est de commencer la recherche par un sous-ensemble aléatoire de caractéristiques.

3.2.2 Procédure de recherche

Une fois que le point de départ est bien choisi, nous passons à la deuxième étape qui consiste à choisir une procédure de recherche ,également connue sous le nom "organisation de la recherche", servant à générer des sous-ensembles de caractéristiques . En général, les stratégies de recherche peuvent être classées en trois catégories : exhaustive, heuristique et aléatoire.

a) Génération exhaustive

Dans cette approche, une recherche exhaustive sur tous les sous-ensembles de caractéristiques est effectuée afin de sélectionner le "meilleur" sous-ensemble de caractéristiques.

Cette stratégie de recherche garantit de trouver le sous-ensemble optimal. Le problème majeur de cette approche est que le nombre de combinaisons croît exponentiellement en fonction du nombre de caractéristiques. Pour un ensemble de N caractéristiques, et quand N devient grand, les 2^N combinaisons possibles rendent la recherche exhaustive impossible (problème NP-complet, [Blu et Riv '93]).

b) Génération heuristique

Dans cette catégorie, une approche heuristique est utilisée pour guider la recherche. Les algorithmes qui utilisent cette approche sont généralement des algorithmes itératifs dont chaque itération permet de sélectionner ou de rejeter une ou plusieurs caractéristiques.

Les avantages de ces algorithmes sont leur simplicité et leur rapidité. En revanche, ils ne permettent pas de parcourir totalement l'espace de recherche. Dans la littérature, les trois sous-catégories les plus connues de cette approche sont :

Forward : cette approche est également appelée ascendante, son principe est de commencer avec un ensemble de caractéristiques vide et à chaque itération une ou plusieurs caractéristiques seront ajoutées.

Backward : cette approche procède à l'inverse de "Forward". L'ensemble de départ représente l'ensemble total des caractéristiques et à chaque itération, une ou plusieurs caractéristiques seront supprimées. Cette approche est aussi qualifiée de descendante.

Stepwise : cette approche est un mélange des deux précédentes et consiste à ajouter ou supprimer des caractéristiques au sous-ensemble courant.

c) Génération aléatoire

La procédure de recherche aléatoire (appelée aussi stochastique ou non-déterministe) consiste à générer aléatoirement un nombre fini de sous ensembles de caractéristiques afin de sélectionner le meilleur. En outre, les stratégies de recherche aléatoires convergent en général rapidement vers une solution "semi-optimale", ce qui est préférable pour éviter le phénomène de sur-apprentissage.

3.2.3 Méthode d'évaluation

Une fois la stratégie de recherche définie, et les sous-ensembles générés, une méthode d'évaluation est définie dans la troisième étape.

Les méthodes utilisées pour évaluer un sous-ensemble de caractéristiques dans les algorithmes de sélection peuvent être classées en trois catégories principales : "filter", "wrapper" et "embedded".

Les méthodes « wrappers » utilisent les algorithmes de classification pour générer et ensuite évaluer la qualité des ensembles de variables candidats. Cette approche est généralement pertinente mais elle dépend de la représentativité de l'ensemble d'apprentissage.

Les méthodes « filters » sont basées sur une fonction critère pour mesurer la pertinence de l'information contenue dans les ensembles de variables candidats. Elles réduisent considérablement le temps de calcul alors que les méthodes « wrappers » permettent d'obtenir des meilleurs résultats.

Les méthodes « embedded » tentent de combiner les avantages de précédentes approches. Elles intègrent la phase de sélection des variables dans l'étape d'apprentissage. Néanmoins, le temps de calcul reste important. C'est probablement la raison principale pour laquelle les méthodes « filters » sont les plus populaires.

a) Méthode Filter

Cette méthode permet de rechercher des sous-ensembles de variables, sans utiliser un algorithme d'apprentissage : cette catégorie de méthodes est nommée filter [Koh et Joh'97]; [Yu et Liu'04]. En effet, en présence d'observations étiquetées, le choix d'un sous-ensemble de variables peut se faire en considérant l'habilité du sous-ensemble à discriminer les classes. Dans ce cas, la pertinence d'une variable pourrait être définie par une mesure de séparabilité des classes [Theo et Kou'03]. Cette pertinence s'obtiendrait indépendamment d'un algorithme d'apprentissage.

Les deux catégories, filter et wrapper, se distinguent donc en fonction de la participation de l'algorithme d'apprentissage dans la sélection du sous-ensemble de variables.

Le modèle "filter" a été le premier utilisé pour la sélection de caractéristiques. Dans celui-ci, le critère d'évaluation utilisé évalue la pertinence d'une caractéristique selon des mesures qui reposent sur les propriétés des données d'apprentissage. Cette méthode est considérée, davantage comme une étape de pré-traitement (filtrage) avant la phase d'apprentissage.

Les méthodes qui se basent sur ce modèle pour l'évaluation des caractéristiques, utilisent souvent une approche heuristique comme stratégie de recherche. La procédure du modèle "filter" est illustrée par la figure (4.2).

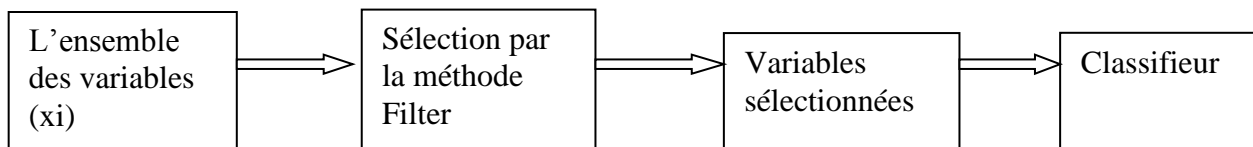


Figure 4.2 : la procédure du modèle « filter »

Le but d'une méthode d'évaluation "filter" est de calculer un score pour évaluer le degré de pertinence de chacune des caractéristiques (x_i). Ci-dessous, nous présentons quelques mesures utilisées dans la littérature comme score ou critère d'évaluation [Guy et Eli'03] :

Le critère de corrélation : ce score est utilisé dans le cas d'une classification binaire.

Le critère de Fisher : permet de mesurer le degré de séparabilité des classes à l'aide d'une caractéristique donnée [Dud et al'00], [Fur et al'00].

L'information mutuelle est une mesure de dépendance entre les distributions de deux populations [Fra et Swi'86]. SNR (Signal-to-Noise Ratio coefficient) est un score qui mesure le pouvoir de discrimination d'une caractéristique entre deux classes. D'une manière similaire au critère de Fisher, cette méthode classe les caractéristiques en calculant le rapport de la valeur absolue de la différence des moyennes des classes et de la moyenne des écart-types des classes.

Le principal avantage des méthodes de filtrage est leur efficacité calculatoire et leur robustesse face au sur-apprentissage. Malheureusement, ces méthodes ne tiennent pas compte des interactions entre caractéristiques et tendent à sélectionner des caractéristiques comportant de l'information redondante plutôt que complémentaire [Guy et Eli'03].

De plus, ces méthodes ne tiennent absolument pas compte de la performance des méthodes de classification qui suivent la sélection [Koh et Joh'97].

b) Méthode Wrapper

Un autre ensemble de méthodes permet de rechercher des sous-ensembles de variables, en utilisant un algorithme d'apprentissage : cette catégorie de méthodes est nommée Wrapper.

Le principal inconvénient des approches "filter" est le fait qu'elles ignorent l'influence des caractéristiques sélectionnées sur la performance du classifieur à utiliser par la suite.

Pour résoudre ce problème, Kohavi et John ont introduit le concept "wrapper" pour la sélection de caractéristiques [Koh et Joh'97]. Les méthodes "wrapper", appelées aussi méthodes enveloppantes, évaluent un sous-ensemble de caractéristiques par sa performance de classification en utilisant un algorithme d'apprentissage. La procédure du modèle "wrapper" est illustrée par la figure 4.3.

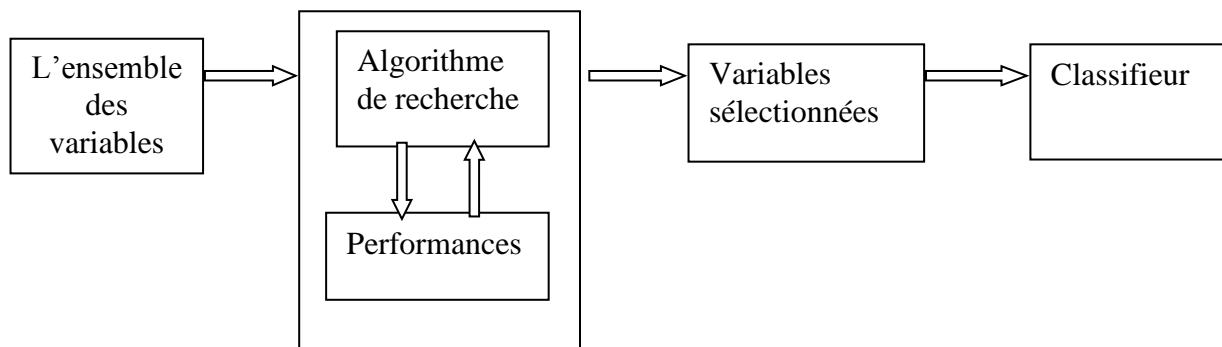


Figure 4.3 : La procédure du modèle "wrapper"

L'évaluation se fait à l'aide d'un classifieur qui estime la pertinence d'un sous-ensemble donné de caractéristiques. La complexité de l'algorithme d'apprentissage rend les méthodes "wrapper" très coûteuses en temps de calcul. En général, pour diminuer le temps de calcul et pour éviter les problèmes de sur-apprentissage, le mécanisme de validation croisée est fréquemment utilisé. Une meilleure performance des méthodes "wrapper" par rapport à certaines méthodes de filtrage a été démontré par [Koh et Joh'97].

Le problème de la complexité de cette technique rend impossible l'utilisation d'une stratégie de recherche exhaustive. Par conséquent, des méthodes de recherche heuristiques ou aléatoires peuvent être utilisées. La recherche devient néanmoins, de plus en plus irréalisable avec l'augmentation de la taille de l'ensemble initial de caractéristiques.

Les méthodes "wrapper" sont généralement considérées comme étant meilleures que celles de filtrage selon [Li et Guo'08], [Hua et al'08]. Elles sont capables de sélectionner des sous-ensembles de caractéristiques de petite taille qui sont performants pour le classifieur utilisé mais il existe deux inconvénients principaux qui limitent ces méthodes :

i) La complexité et le temps de calcul nécessaire pour la sélection.

Le principal inconvénient de l'approche "wrapper" est le temps nécessaire pour la sélection des caractéristiques, il est nettement plus long que celui des approches de filtrage et d'autres approches de sélection de caractéristiques. L'utilisation d'un classifieur pour évaluer les sous-ensembles ainsi que les techniques d'évaluation (validation croisée par exemple) rendent les méthodes "wrapper" très coûteuses en terme de temps de calcul.

ii) La dépendance des caractéristiques pertinentes sélectionnées par rapport au classifieur utilisé.

La deuxième limitation de l'approche "wrapper" est que l'évaluation des caractéristiques se fait par un seul classifieur lors de la sélection. Chaque classifieur a ses spécificités et ses hypothèses. Ainsi le sous-ensemble sélectionné dépend toujours du classifieur utilisé.

c) Méthode Embedded

A la différence des méthodes "wrapper" et "filter", les méthodes "embedded" (appelées aussi méthodes intégrées) incorporent la sélection de variables lors du processus d'apprentissage.

Un tel mécanisme intégré pour la sélection des caractéristiques peut être trouvé, par exemple, dans les algorithmes de type SVM, AdaBoost, ou dans les arbres de décisions. Dans les méthodes de sélection de type "wrapper", la base d'apprentissage est divisée en deux parties : une base d'apprentissage et une base de validation pour valider le sous-ensemble de caractéristiques sélectionné. En revanche, les méthodes intégrées peuvent se servir de tous les exemples d'apprentissage pour établir le système. Cela constitue un avantage qui peut améliorer les résultats. Un autre avantage de ces méthodes est leur plus grande rapidité par rapport aux approches "Wrapper" parce qu'elles évitent que le classifieur recommence de zéro pour chaque sous-ensemble de caractéristiques.

3.2.4 Critère d'arrêt

Les étapes deux et trois du processus de sélection (procédure de recherche et évaluation) se répètent jusqu'à un critère d'arrêt. Ce test d'arrêt représente la quatrième étape de la méthode.

Certains critères doivent être définis pour arrêter le processus de recherche sur les sous-ensembles de caractéristiques. Pour les méthodes de filtrage, le critère d'arrêt couramment utilisé est basé sur l'ordre des caractéristiques, rangées selon certains scores de pertinence (généralement des mesures statistiques). Une fois les caractéristiques ordonnées, celles qui ont les scores les plus élevés seront choisies et utilisées par un classifieur. Pour les méthodes de type "wrapper", le processus de recherche peut s'arrêter lorsqu'il n'y a plus d'amélioration de précision. En d'autres termes, lorsqu'il n'y a plus la possibilité de trouver un sous-ensemble meilleur que le sous-ensemble actuel. Un critère d'arrêt pour les méthodes enveloppantes est de continuer à rechercher jusqu'au moment où la précision dépasse un certain seuil défini par l'utilisateur.

3.3 Présentation de quelques méthodes de sélection

Dans cette section, nous présentons quelques méthodes de sélection de caractéristiques citées dans la littérature. Nous avons choisi de présenter de méthodes fondées sur les différentes techniques de recherche définies précédemment ainsi que différentes techniques d'évaluation.

3.3.1 Méthode Max-relevance, Min-Redundancy (mRMR)

"Max-relevance, Min-Redundancy" (mRMR) est une méthode de filtrage pour la sélection de caractéristiques proposée par Peng et al. en 2005 [Pen et al'05]. Cette méthode est basée sur des mesures statistiques classiques comme l'information mutuelle, la corrélation etc. L'idée de base est de profiter de ces mesures pour essayer de minimiser la redondance (mR) entre les caractéristiques et de maximiser la pertinence (MR). Les auteurs proposent deux variantes de leur méthode. Une pour des données discrètes et l'autre pour des données continues. Pour les données discrètes, les auteurs utilisent l'information mutuelle pour calculer les deux facteurs mR et MR. Pour les données continues, les auteurs ont remplacé l'information mutuelle par d'autres mesures. Pour la redondance ils ont utilisé la mesure de corrélation, par contre, la mesure F-statistique est utilisée pour calculer la pertinence.

3.3.2 Méthode FOCUS

Un algorithme de filtrage pour la sélection de caractéristiques, appelé FOCUS, a été proposé par Almuallim et Dietterich en 1991 [Alm et Die'91]. Cette méthode repose sur une recherche exhaustive sur l'ensemble initial de caractéristiques pour trouver le sous ensemble le plus performant de taille minimale. L'algorithme FOCUS commence par générer et évaluer tous les sous-ensembles de taille T (initialement un), puis tous les couples de caractéristiques, les triplets et ainsi de suite jusqu'à ce que le critère d'arrêt soit satisfait.

Les inconvénients de cette approche sont d'un côté la sensibilité de sa méthode d'évaluation au bruit et de l'autre côté son temps de calcul qui devient énorme avec l'augmentation de la taille de l'ensemble des caractéristiques et du nombre d'exemples de la base. Une année plus tard, les mêmes auteurs ont proposé FOCUS2 comme une amélioration de leur méthode initiale [Alm et Die'92]. FOCUS2 est beaucoup plus rapide que FOCUS, mais elle est toujours sensible au bruit.

3.3.3 Méthode Relief

Une des méthodes de filtrage les plus connues pour la sélection de caractéristiques est la méthode relief. Cette méthode fut proposée en 1992 par Kira et Rendell [Kir et Ren'92]. Son principe est de calculer une mesure globale de la pertinence des caractéristiques en accumulant la différence des distances entre des exemples d'apprentissage choisis aléatoirement et leurs plus proches voisins de la même classe et de l'autre classe. La simplicité, la facilité de la mise en œuvre ainsi que la précision même sur des données bruitées, représentent les avantages de cette méthode. En revanche, sa technique aléatoire ne peut pas garantir la cohérence des résultats lorsqu'on applique plusieurs fois la méthode sur les mêmes données. Par ailleurs, cette méthode ne prend pas en compte la corrélation éventuelle entre les caractéristiques.

Afin d'éviter le caractère aléatoire de l'algorithme, John et al. [John et al'94] ont proposé une version déterministe appelée ReliefD. D'autres variantes de cet algorithme, pour améliorer sa performance, sa vitesse ou les deux, ont été proposées dans [Kol et Sah'96], [Liu et al'02].

3.3.4 Méthodes LVW et LVF

LVW (Las Vegas Wrapper) est une méthode de sélection de caractéristiques proposée en 1996 par Liu et Setiono [Liu et Set'96]. Cette méthode consiste à générer aléatoirement et à chaque itération, un sous-ensemble de caractéristiques et à l'évaluer avec un classifieur.

Après avoir évalué, si sa performance est meilleure que la meilleure performance trouvée auparavant (au départ, l'ensemble de base est supposé comme le meilleur sous ensemble), ce sous-ensemble devient le meilleur sous-ensemble courant. Ce processus est répété jusqu'à ce que T essais consécutifs soient infructueux pour l'amélioration. Cette méthode présente l'inconvénient de ne pas garantir l'optimalité de la solution finale ainsi qu'un temps de calcul très élevé.

LVF (Las Vegas Filter) est une méthode de filtrage pour la sélection de caractéristiques, proposée deux ans plus tard par les mêmes auteurs [Liu et Set'98]. Elle est similaire à la méthode LVW mais l'évaluation des sous-ensembles se fait par le calcul d'une mesure appelée "taux d'incohérence" ou "taux d'inconsistance". L'inconsistance pour un sous-ensemble de caractéristiques est définie par le rapport entre le nombre d'exemples inconsistants de la base de données et le nombre total d'exemples. Un exemple est dit inconsistant s'il existe un

autre exemple qui a la même représentation dans l'espace des caractéristiques du sous-ensemble de caractéristiques étudié (appelé exemple équivalent), mais qui appartient à une autre classe.

Cette méthode présente les mêmes inconvénients que la méthode FOCUS. Elle est donc très sensible au bruit et comme toutes les méthodes de recherche exhaustive, elle est très coûteuse en temps de calcul.

3. 3.5 Méthodes SFS et SBS

Dans cette partie nous présentons deux méthodes, c'est-à-dire qu'il s'agit de méthodes nous donnant des résultats rapidement, mais qui ne sont pas forcément optimales. Il s'agit de méthodes approximatives. Nous trouvons "Forward Selection" qui correspond aux méthodes de recherche en avant, comme par exemple le SFS que nous expliquerons plus en détail dans la section suivante. Nous avons aussi les méthodes "Backward Selection" qui sont les méthodes de recherche en arrière comme le SBS, que nous expliquerons aussi par la suite.

SFS (Sequential Forward Selection) ou (sélection séquentielle croissante) est la première méthode proposée pour la sélection de caractéristiques. Cette méthode a été proposée en 1963 par Marill et Green [Mar et Gre'63]. Une approche heuristique de recherche est utilisée dans cette méthode, en commençant par un ensemble vide de caractéristiques.

A chaque itération, la meilleure caractéristique parmi celles qui restent sera sélectionnée, supprimée de l'ensemble de départ et ajoutée au sous-ensemble des caractéristiques sélectionnées. Le processus de sélection continue jusqu'à un critère d'arrêt.

En 1971, Whitney [Whi'71] a proposé une méthode similaire au SFS appelée SBS (Sequential Backward Selection) ou (sélection séquentielle arrière). A la différence de la méthode SFS, cette méthode commence par l'ensemble de toutes les caractéristiques et à chaque itération, la caractéristique la plus mauvaise sera supprimée.

Bien que les deux méthodes SFS et SBS semblent similaires, Ahan et Bankert [Aha et Ban'95] ont montré que la méthode SBS est plus performante parce qu'elle prend en considération l'interaction d'une caractéristique avec un ensemble de caractéristiques plus large, contrairement au SFS qui ne prend en considération que l'interaction de cette caractéristique avec le sous-ensemble déjà sélectionné. Par ailleurs, l'évaluation des sous ensembles de grande taille avec la méthode SBS pose un problème au niveau de temps de calcul.

En 1978, des généralisations des méthodes SBS et SFS appelées GSFS et GSBS, sont proposées par Kittler [Kit'78]. Dans ces méthodes, l'auteur propose, au lieu d'inclure (ou exclure) une caractéristique à chaque itération, d'inclure (ou exclure) un sous-ensemble de caractéristiques. Ces algorithmes ont montré une meilleure performance par rapport aux méthodes initiales, mais ils conservent toujours les mêmes problèmes que les méthodes de base.

Deux autres méthodes de la famille (FS, BS) qui limitent les inconvénients des méthodes décrites ci-dessous, appelées SFFS (Sequential Floating Forward Selection) et SFBS (Sequential Floating Backward Selection) sont proposées en 1994 par Pudil et al [Pudil et al'94].

Ces méthodes consistent à utiliser « l » fois l'algorithme SFS de manière à ajouter l variables, puis à utiliser « r » fois l'algorithme SBS afin d'en supprimer r. Ces étapes sont alors réitérées jusqu'à l'obtention du critère d'arrêt. La dimension du sous-ensemble à chaque étape sera alors dépendante des valeurs de l et r. Les valeurs optimales de ces paramètres ne pouvant pas être déterminées théoriquement, les auteurs proposent de les laisser flottantes au cours du processus de sélection afin de se rapprocher au maximum de la solution optimale.

3.3.6 Méthode Branch and Bound

La méthode Branch and Bound (procédure par évaluation et séparation progressive) est un algorithme qui construit l'arborescence des sous-ensembles de caractéristiques en évaluant a priori les chances de trouver la solution optimale dans une branche particulière.

Cette méthode consiste à énumérer les solutions en utilisant certaines propriétés du problème. Elle permet d'éliminer a priori des solutions partielles qui ne mènent pas à la solution que l'on recherche, ce qui évite de faire une recherche exhaustive et donc permet d'obtenir des résultats en un temps raisonnable.

Pour appliquer la méthode de B&B, nous devons avoir [Reb'05] :

- Un moyen de calcul d'une borne inférieure d'une solution partielle
- Une stratégie de subdiviser l'espace de recherche pour créer des espaces de recherche de plus en plus petits.
- Un moyen de calcul d'une borne supérieure pour au moins une solution.

A la racine de notre arborescence nous avons notre ensemble de caractéristiques.

Ensuite des procédures de calculs de bornes inférieures et supérieures sont appliquées à la racine. Lorsque les 2 bornes sont égales, cela veut dire que nous avons trouvé la meilleure solution. Sinon on divise l'ensemble des caractéristiques en plusieurs sous-ensembles. Lorsque l'on trouve une solution avec de très bons résultats, cela ne veut pas forcément dire que le résultat est optimal, car il est possible de trouver encore un meilleur résultat dans une autre branche. Par contre, ce résultat nous permettra de supprimer les branches ayant un moins bon résultat. Car si la borne inférieure d'un nœud dépasse la valeur d'une solution déjà connue, alors on peut affirmer que la solution optimale globale n'est pas présente dans ce nœud. Le système s'arrête lorsque tous les nœuds sont explorés ou éliminés.

De plus cet algorithme est efficace car il choisit le meilleur sous-ensemble sans recherche exhaustive.

L'inconvénient de cette méthode est son temps de calcul qui croît vite avec l'augmentation du nombre de caractéristiques et qui devient impraticable à partir d'un certain nombre (30 caractéristiques).

3.3.7 Les algorithmes génétiques

Les algorithmes génétiques (AG), ont été développés dans les années 1970, et utilisés dans le domaine de la sélection de caractéristiques afin d'accélérer la recherche et d'éviter les optima locaux.

Dans un AG [SkI'00], on a une population de base qui est souvent composée de chaînes de caractères correspondant chacune à un chromosome.

Souvent chaque chromosome est une chaîne binaire de taille n . La population initiale est générée aléatoirement.

Le mécanisme d'un algorithme génétique de base est assez simple [Gol'91].

Il est composé de 3 étapes, qui conduisent généralement à de bons résultats : La reproduction, le crossover et la mutation. La reproduction correspond à la copie de chaque chaîne en fonction des valeurs de la fonction à optimiser. Ce qui correspond à donner un poids d'importance à chaque chaîne. Après la reproduction, le crossover est appliqué. Il correspond à un croisement des différentes chaînes tirées lors de la reproduction. Pour finir la dernière étape est la mutation, elle permet de modifier aléatoirement un élément de chaîne.

De nombreuses études rapportées dans la littérature ont montré que les méthodes qui utilisent les AGs comme technique de recherche ont donné de meilleurs résultats que les résultats obtenus par les autres méthodes de sélection [Jai et Zong'97], [Kun et Jain'99], [Ish et Nak'00].

3.3.8 Méthode SAC

SAC (Sélection Adaptative de Caractéristiques) est une méthode de sélection de descripteurs proposée par Kachouri et al. en 2010 [Kac et al'10]. Cette méthode, développée dans le cadre d'un ensemble de descripteurs à plusieurs dimensions, peut être adaptée pour une sélection de caractéristiques. L'idée générale de la méthode est de construire un ensemble de classifieurs SVM appris sur chacun des descripteurs et de sélectionner les meilleurs par discrimination linéaire de Fisher (FLD). Ils proposent de considérer la performance d'apprentissage des modèles correspondant à ces descripteurs pour l'identification d'une meilleure discrimination de Fisher. La sélection adaptative permet de ne considérer que les caractéristiques les mieux adaptées.

Elle assure la reconnaissance et la classification des données en se servant hiérarchiquement des caractéristiques sélectionnées. Les résultats expérimentaux obtenus confirment l'efficacité et la robustesse de cette approche.

4. Conclusion

Dans ce chapitre, nous avons présenté quelques méthodes concernant la sélection de variables.

Dans un premier temps, nous avons détaillé, les techniques de réduction par une transformation de données, en présentant différentes approches linéaires et non linéaires. Ce sont des approches qui pourraient être incluses dans les méthodes de sélection de caractéristiques pour fournir à celles-ci des informations plus variées.

Dans la deuxième partie de ce chapitre, une revue du domaine de la sélection de caractéristiques a été présentée.

Après avoir exposé les différents composants nécessaires à un algorithme de sélection de caractéristiques, les alternatives possibles pour leurs mises en œuvre ont été présentées. Nous avons décrit un certain nombre d'algorithmes tout en démontrant leurs avantages et inconvénients.

L'approche par sélection permet de mieux appréhender la modélisation d'un problème et de limiter les mesures qui permettent la résolution du problème, par contre l'extraction de nouvelles caractéristiques conserve une vision globale des observations et ne permet pas d'économiser des mesures pour décrire le phénomène observé. Dans cette thèse, nous nous intéressons aux techniques de sélection de caractéristiques et plus précisément au type « wrapper ».

Les méthodes de sélection de caractéristiques sont classées généralement en deux groupes:

les méthodes "filter" et les méthodes "wrapper". La première approche (méthodes de filtrage) utilise des mesures statistiques calculées sur les caractéristiques afin de filtrer les caractéristiques peu informatives. Cette étape est généralement réalisée avant d'appliquer tout algorithme de classification. Et La seconde approche (méthodes enveloppantes ou "wrapper") est plus coûteuse en temps de calcul, mais en contre-partie, elle est souvent plus précise.

Les méthodes wrapper sont généralement considérées comme étant meilleures que celles de filtrage et de plus, elles sont capables de sélectionner des sous-ensembles de caractéristiques de plus petite taille, néanmoins aussi performants pour le classifieur utilisé.

Dans le chapitre suivant nous allons utiliser la méthode sélection de variables SBS « Sequential Backward Selection » pour sélectionner les caractéristiques les plus pertinentes pour la classification des données médicales.

Chapitre 5

Résultats Expérimentaux et Discussions

Chapitre 5

Résultats expérimentaux et discussions

I. Introduction

Ce chapitre est consacré à la présentation des résultats obtenus par notre classifieur optimisé MAPMC. Deux expérimentations ont été réalisées en utilisant deux bases de données médicales. Le schéma qui suit reflète les différentes étapes suivies dans ce chapitre tout en démontrant nos contributions.

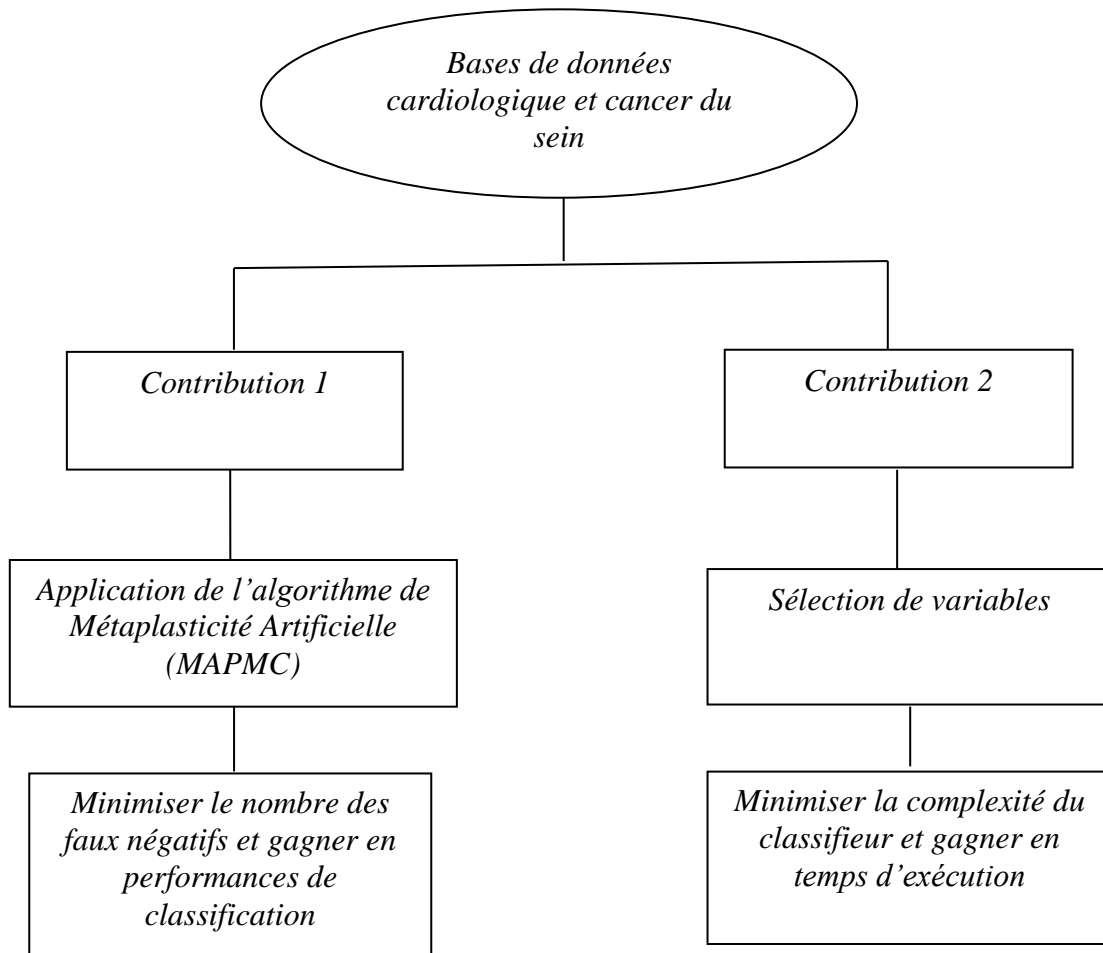


Figure 5.1 : Protocole d'expérimentation.

II. 1^{ère} expérimentation : Classification des arythmies cardiaques

1. Introduction :

D'après des statistiques de l'OMS (l'organisation mondiale de la santé), les maladies cardiovasculaires représentent la majeure cause de décès à travers le monde [WHO'11].

Les pathologies cardiovasculaires et les arythmies cardiaques provoquent chaque année 17 millions de décès à travers le monde [Mac et Men'04]. Elles représentent en Algérie la première cause de morbidité et de mortalité. Parmi ces troubles cardiovasculaires, les infarctus du myocarde représentent 10% des décès dans le monde et sont provoqués par une ischémie (défaut prolongé d'apport sanguin). En présence d'une ischémie, le tissu cardiaque tend à perdre ses propriétés de contractilité et le cœur ne peut plus assurer sa fonction de pompe ce qui provoque l'infarctus. L'infarctus du myocarde est accompagné de troubles graves du rythme cardiaque, appelés arythmies cardiaques, tels que la fibrillation ventriculaire.

Durant 2011 et 2012, 44.5% des décès en Algérie sont dues aux maladies cardiovasculaires en dépassant les chiffres de morts par accident de la route [Mou'12]. En 2030 environ 25 millions de personnes décèderont des maladies cardiovasculaires.

Les arythmies sont des troubles du rythme cardiaque dont certaines, telles que la fibrillation ventriculaire, peuvent entraîner des morts subites. Pour diagnostiquer les arythmies, la méthode la plus répandue est l'analyse de l'électrocardiogramme.

Le traitement de ces pathologies nécessite un personnel médical aguerri et du matériel spécialisé pour une intervention thérapeutique rapide.

Une détection précoce par des classifieurs intelligents de ces différentes cardiopathies permettrait d'assurer une prise en charge adéquate et d'améliorer la qualité de vie des patients.

Actuellement de nombreux travaux de recherche tentent de développer de nouvelles approches d'aide à la décision, le cas sur lequel nous avons travaillé relève du domaine de l'électrocardiologie et de l'interprétation automatique des ECG.

L'intérêt essentiel est de décharger les chercheurs de lecture de tracés de nombreuses heures et de réduire la subjectivité du jugement humain.

Plusieurs systèmes intelligents sont présents en littérature pour la détection des anomalies cardiaques.

Sikkandar et al en 2013 [Sik et al'13] ont utilisé la transformée de Hilbert–Huang (HHT) pour renforcer leur système de détection des arythmies cardiaques. En 2012, Kamath.C a présenté une étude pour estimer l'intérêt de l'analyse symbolique pour caractériser les propriétés non-linéaires des signaux ECG et faire une discrimination entre les signaux normaux et pathologiques. Il a proposé un réseau de neurones pour classer cinq différentes d'arythmies. Le système a donné un résultat de 93.5% de classification correcte moyenne [Kam'12].

Différents réseaux de neurones ont été proposés en littérature pour la détection des arythmies cardiaques.

Jadhav et al en 2012 ont proposé un perceptron multi couches feedforward avec un algorithme de rétropropagation statique pour différencier entre les cas normaux et les arythmies qui apparaissent sur plusieurs signaux ECG [Jad et al'12]. Les résultats expérimentaux ont donné un résultat de 86.67% de classification correcte.

En 2011 Gothwal et al. ont combiné la transformée de Fourier(FFT) et un réseau de neurones avec un algorithme de rétropropagation de Levenberg Marquardt ; pour analyser le signal Electrocardiogramme. Le meilleur résultat de classification correcte obtenu était de 98.48% [Got et al'11].

Yu et Chou en 2008 ont utilisé une analyse en composantes indépendantes (ICA) et un réseau de neurones pour classer différentes arythmies cardiaques, le résultat obtenu était de 98.71% [Yu et Cho'08].

Et en 2005 Gao et al. ont présenté un système de diagnostic des arythmies cardiaques, en utilisant les réseaux de neurones Bayésiens, Le taux de classification obtenu était de 90% [Gao et al'05].

Dans les domaines cliniques, on fait face au problème de développement de classifieurs qui sont capables de faire une discrimination non-linéaire entre les classes, de traiter des données d'entrée incomplètes ou ambiguës et de minimiser les fausses alarmes.

Il est donc nécessaire de développer de nouveaux systèmes de détection avec un niveau de précision élevé, de faibles statistiques de faux positifs et de faux négatifs, pour qu'ils soient utiles dans applications pratiques.

Dans le cadre des travaux de cette thèse, nous appliquons l'algorithme de métaplasticité artificielle sur un perceptron multi couches (Métaplasticité Artificielle appliquée sur Perceptron Multi Couches, MAPMC) pour classer différentes arythmies cardiaques. L'algorithme proposé est inspiré de la propriété biologique de métaplasticité des neurones et de la théorie d'information de Shannon. Le MAPMC a été testé sur la base de données des arythmies cardiaques MIT-BIH pour évaluer les performances de classification de cet algorithme. Nous avons utilisé les paramètres de performance : sensibilité, spécificité et taux de classification correcte. Les résultats démontrent la supériorité de la technique proposée pour la détection des anomalies cardiaques.

2. Préparation de la base de données :

2.1 Présentation du modèle de classification neuronale :

Lorsqu'une approche de reconnaissance des arythmies cardiaques est adoptée, les réseaux de neurones artificiels sont une bonne technique de classification. Pour notre étude, comme décrit auparavant, le réseau de neurones utilisé sera optimisé par l'algorithme de Métaplasticité artificielle. La stratégie que nous envisageons de mettre en œuvre est présentée dans la figure suivante.

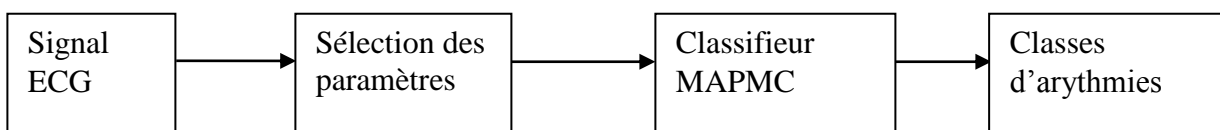


Figure 5.2 : schéma global de la classification

2.2 Description de la base de données MIT-BIH :

Les différents signaux ECG utilisés pour la validation de l'algorithme de métaplasticité artificielle, sont issus de la base de données MIT-BIH [Moo et Mar'01],[Har'80].

Depuis 1975, les laboratoires de l'hôpital de Beth Israël à Boston et MIT ont réalisé une base de données MIT-BIH, qui a été commencé à être distribuée en 1980. Cette base de données contient 48 enregistrements extraits d'une demi-heure des enregistrements ambulatoires à deux voies d'ECG, obtenus à partir de 47 sujets étudiés par

le laboratoire d'arythmie de BIH entre 1975 et 1979. Vingt-trois enregistrements ont été choisis au hasard d'un ensemble de 4000 enregistrements ambulatoires de 24 heures d'ECG rassemblées d'une population mélangée des patients hospitalisés (60 %) et des patients non hospitalisés (40 %) à l'hôpital de Beth Israël à Boston (les séries "100"), les 25 enregistrements restants ont été choisis parmi les mêmes enregistrements mais en considération des arythmies rarement observés qui ont une signification cliniques (les séries "200").

Les enregistrements ont été échantillonnés à une fréquence $f_e = 360 \text{ Hz}$ avec une résolution de 11 bits sur une gamme de 10 mV. Deux cardiologues ou plus ont indépendamment annoté chaque enregistrement, environ 110.000 annotations ont été incluses avec la base de données.

Les enregistrements comprenant suffisamment d'arythmies à classer ont été choisis pour construire la base d'exemples.

Les critères électrocardiographiques des différentes arythmies nous ont mené à retenir des descripteurs pertinents pour l'entrée des classifieurs neuronaux (voir Tableau 5.1 et annexe).

2.3. Sélection des descripteurs d'un cycle cardiaque

2.3.1 Introduction

Nous savons que lors de la mise en œuvre d'un classifieur la caractérisation des différentes formes des signaux ECG par des paramètres fiables et pertinents est une étape très importante.

Dans ce travail nous avons utilisé dix paramètres .notre choix est basé sur une approche intuitive qui tient compte de l'avis des spécialistes en cardiologie, nous avons gardé les descripteurs les plus importants.

La mesure de ces paramètres a été réalisée en utilisant un algorithme implémenté : **IMPE** « interface de mesure des paramètres de l'ECG » développée par Pr Chikh et Bechar en 2003 sous environnement Matlab.

Cette interface a été modifiée en 2006 par Benchaib Yasmine et Behadada Omar , en ajoutant trois options pour détecter le R_{pré} (le pic R du battement précédent), le R_{suiv} (le pic R du battement suivant) et le R_{di R'} (deuxième pic R correspondant au retard de la déflexion intrinsécoïde) afin de pouvoir obtenir toutes les mesures essentielles .Nous avons constitué une bonne base de données qui représente l'ensemble des arythmies présentes dans les différents enregistrements (figure 5.3).

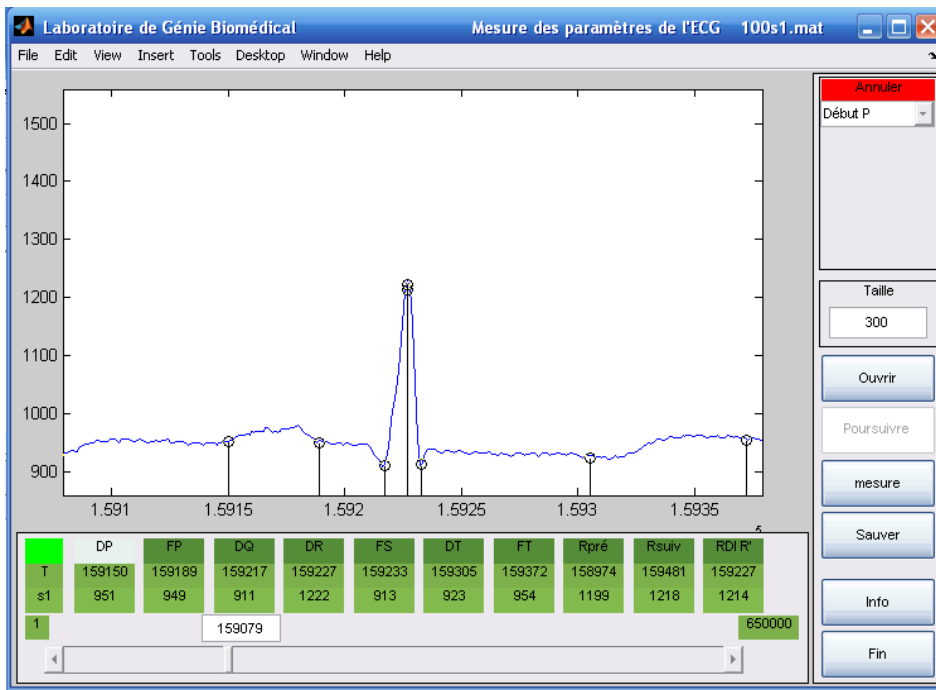


Figure 5.3 : Interface de Mesure des Paramètres de l'ECG (IMPE).

L'interface IMPE permet de localiser les différentes ondes d'une façon manuelle par un simple clic de souris ça nous permet de palier la difficulté de la détection automatique des ondes P, Q, R, S et T.

Ces points permettent de caractériser un battement cardiaque. Ils sont similaires aux paramètres sur lesquels est construite la connaissance médicale.

2.3.2 Présentation des différents descripteurs

A la suite de recherches sur les critères électrocardiographiques des différentes arythmies et de discussions avec des cardiologues, nous avons pu retenir les descripteurs les plus pertinents.

Notons en premier lieu que les Extrasystoles ventriculaires 'V' se caractérisent par :

- Un complexe QRS élargi, de durée égale ou plus souvent supérieure ou égal à 0.12s .Suivi immédiatement d'onde T sans segment RST visible.
- Une absence d'onde P avant le complexe QRS élargi.
- L'existence habituelle de repos compensateur.
- Un intervalle PR qui varie constamment d'un complexe à l'autre sans périodicité nette.

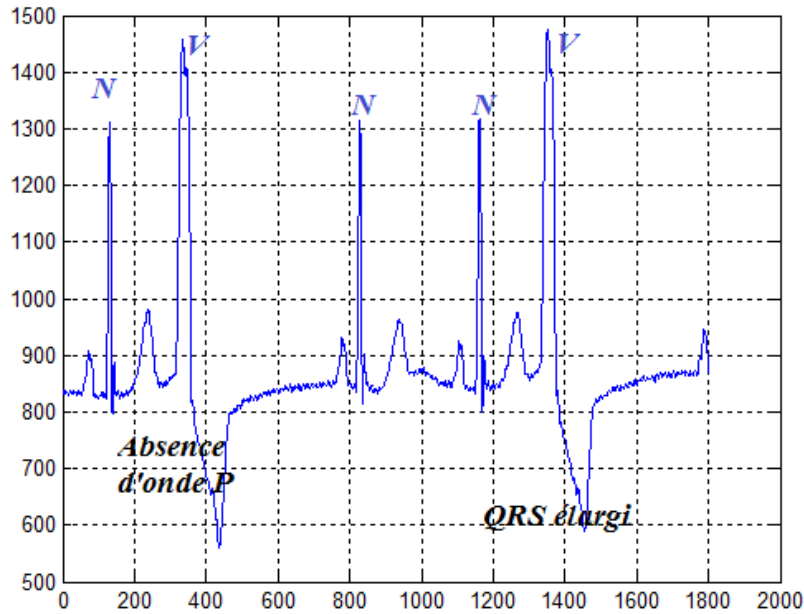


Figure 5.4 : Tracé ECG de l'enregistrement 119 de la base de données MIT-BIH montrant les caractéristiques d'une extrasystole ventriculaire ('V').

Notons en deuxième lieu que le diagnostic des blocs de branche gauche et blocs de branche droit repose sur les critères suivants :

- Un intervalle PR de durée supérieure ou égale à 0.12s.
- Un élargissement important du complexe QRS, toujours supérieur ou égal à 0.12s .
- Une Onde T négative avec un intervalle QT allongé. (voir figure)
- Un retard très important de la déflexion intrinsécoïde (RDI) supérieur ou égal à 0.08s.

DI : déflexion intrinsécoïde : La mesure de DI se fait du début du QRS jusqu'au sommet de l'onde R. si le complexe QRS comporte une onde R élargie et crochetée , la mesure se fait du début du QRS au sommet du dernier crochetage de R . (voir figure 2 et figure 3)

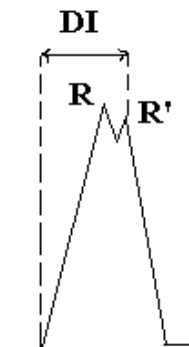


Figure 5.5 : Mesure du temps d'apparition de la déflexion intrinsécoïde (DI)

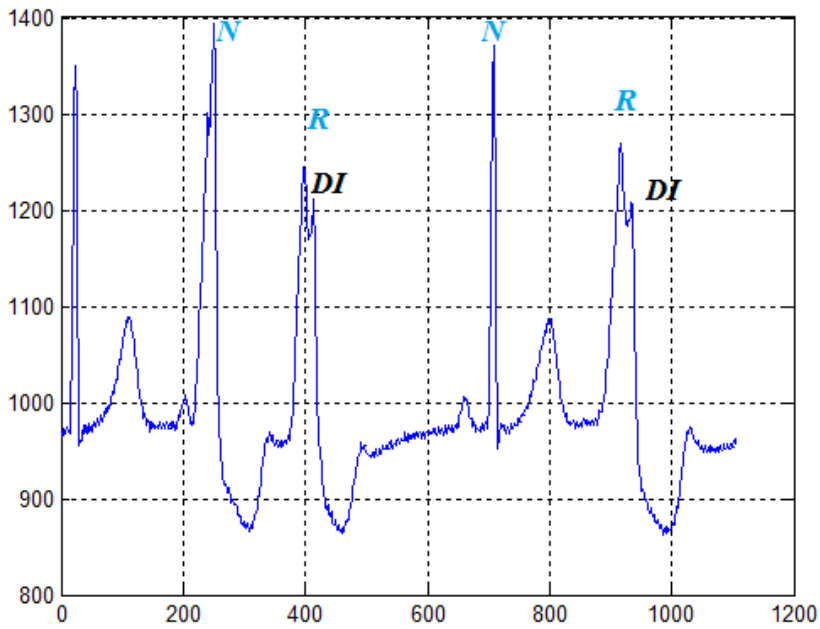


Figure 5.6 : Tracé ECG de l'enregistrement 212 de la base de données MIT-BIH montrant l'apparition de la déflexion intrinsécoïde (DI) dans le cas d'un bloc de branche droit.

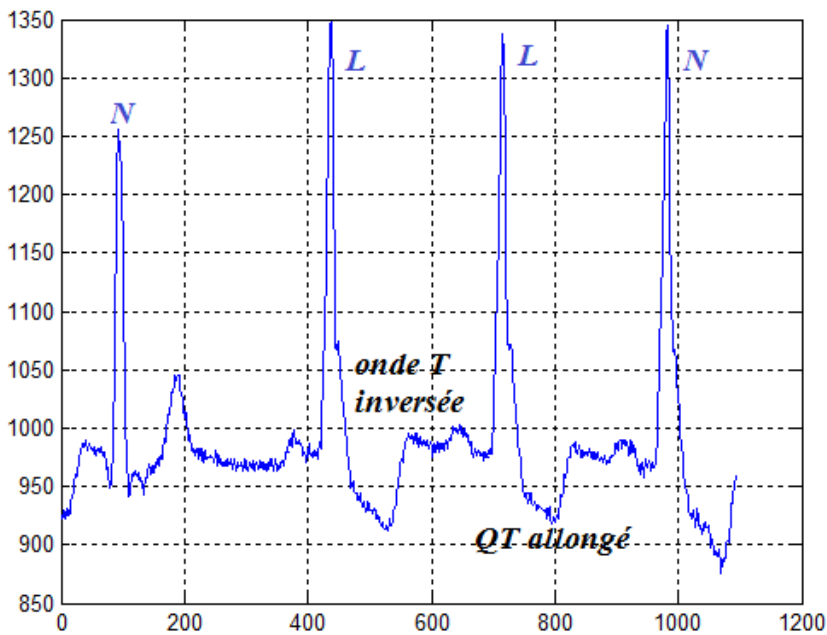


Figure 5.7 : signal ECG de l'enregistrement 214 de la base de données MIT-BIH montrant l'onde T négative et l'allongement de PR , QRS et QT dans le cas d'un bloc de branche gauche .

D'après ces informations les descripteurs qui ont été retenus pour l'entrée du classifieur neuronal sont représentés dans le tableau suivant : Durée P ; intervalle PR ; complexe QRS ; segment ST ; intervalle QT ; RR précédent ; RR suivant ; RDI (retard de la déflexion intrinsecoïde); durée battement ; rapport RR suivant \ RR précédent.

Ces différents descripteurs sont résumés dans le tableau suivant :

Attributs	Description
Durée P	Largeur de l'onde P
intervalle PR	du début de l'onde P jusqu'au début du QRS
complexe QRS	début de l'onde Q jusqu'à la fin de l'onde S
Durée T	Début de l'onde T jusqu'à la fin de l'onde de T
segment ST	de la fin de l'onde S ou R jusqu'au début de l'onde T
intervalle QT	du début du QRS jusqu'à la fin de l'onde T
RR précédent : RRp	la distance entre le pic R du présent battement et le pic R du battement précédent.
RR suivant : RRs	entre le pic R du présent battement et le pic R du battement suivant.
RDI (retard de la déflexion intrinsecoïde)	Du début du QRS jusqu'au sommet de la dernière positivité de l'onde R.
durée battement	début de l'onde P jusqu'à la fin de l'onde T.
RRs \ RRp	le rapport RRsuisant\RRp

Tableau 5.1 : Les différents descripteurs du vecteur d'entrée.

2.3.3 Sélection de la base d'exemples

Nous avons utilisé, dans le cadre de cette thèse , la base de données MIT-BIH. Cette base se compose de quarante huit enregistrements (deux voies ECG ambulatoire d'environ trente minutes), dont les instants d'apparition et la classe de tous les cycles sont annotés par deux cardiologues différents (voir annexe). Les ECG sélectionnés proviennent de l'électrode MLII (Modified Lead II) des enregistrements de la base d'arythmies MIT-BIH .

Les enregistrements comprenant suffisamment d'arythmies à classer ont été choisis pour construire la base d'exemple, voir tableau 5.2 et le tableau en Annexe.

Nous avons rassemblé les battements de chaque enregistrement en quatre groupes :

- _ le groupe N est composé des battements classés normaux.
- _ le groupe V est composé des battements classés extrasystoles ventriculaires.
- _ le groupe R est composé des battements classés blocs de branche droit.
- _ le groupe L est composé des battements classés blocs de branche gauche

La base d'apprentissage affecte la performance des classifieurs neuronaux. Le critère général à respecter pour l'élaborer est de sélectionner des exemples représentatifs de toutes les classes.

Dans notre cas nous avons sélectionné plusieurs cycles de type 'N' ,'V', 'R' et 'L' dans différents enregistrements présentés dans le tableau suivant.

<i>Nombre de battements N</i>	<i>Nombre de battements V</i>	<i>Nombre de battements R</i>	<i>Nombre de battements L</i>
345	301	121	233

Tableau 5.2 : Nombre de battements sélectionnés dans la base d'exemples.

3. Analyse de la base de données

Une exploration de la base de données permet de voir clairement la corrélation entre les différents descripteurs et la classe ciblée. Nous allons présenter dans la section suivante les représentations graphiques des 10 descripteurs de notre base de données en utilisant les diagrammes en boîte.

3.1 Représentation boîte à moustaches ou diagramme en boîte :

Nous avons testé séparément les descripteurs en fonction de leurs classes sachant qu'ils sont codés (0,1,2 et 3).

- 0 :c'est le cas normal (N).
- 1 :c'est le cas extrasystole ventriculaire (V).
- 2 : c'est les cas bloc de branche droit (R).
- 3 : c'est les cas bloc de branche gauche (L).

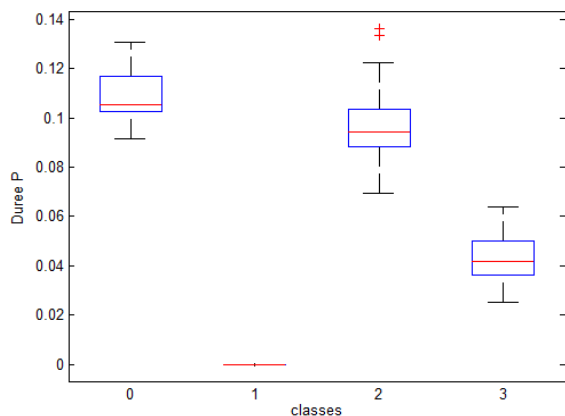


Figure 5.8(a) : représentation de la durée P en fonction des classes.

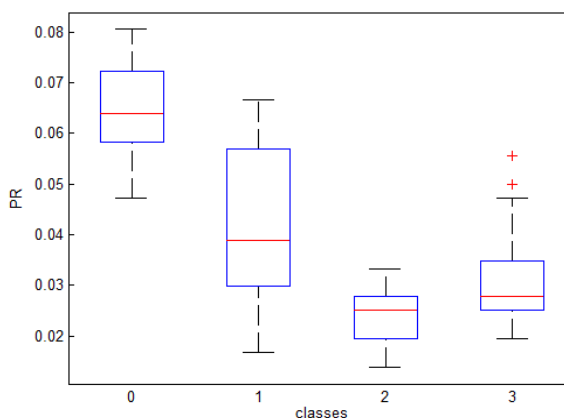


Figure 5.8(b) : représentation de PR en fonction des classes.

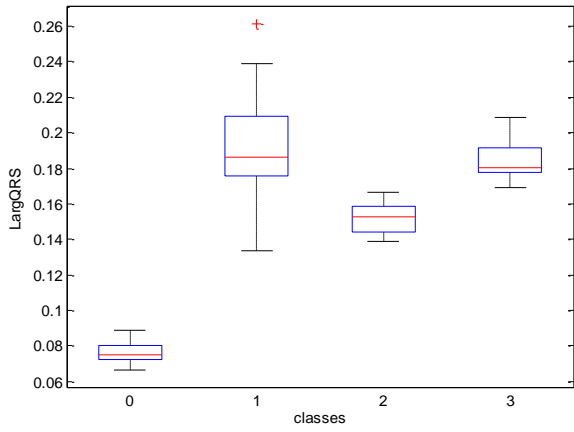


Figure 5.8 (c) : représentation de la largeur du complexe QRS en fonction des classes.

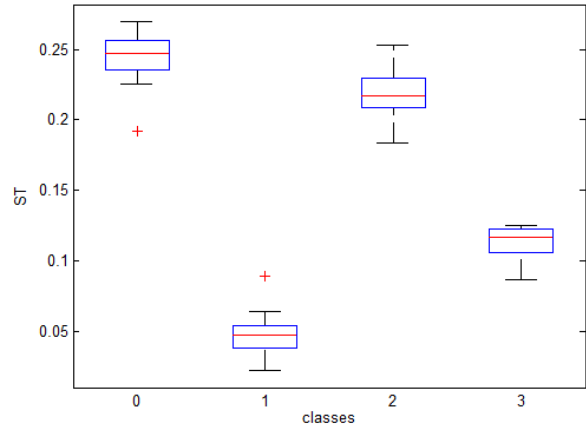


Figure 5.8 (d) : représentation de ST en fonction des classes.

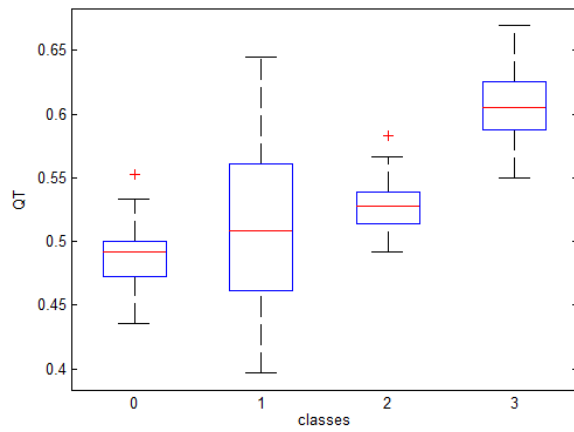


Figure 5.8 (e) : représentation de QT en fonction des classes.

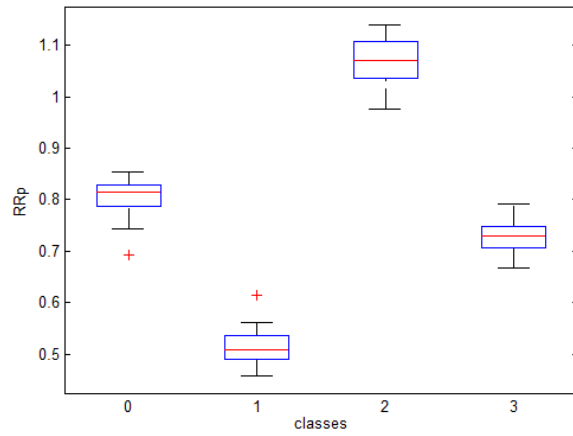


Figure 5.8 (f) : représentation de RR précédent en fonction des classes.

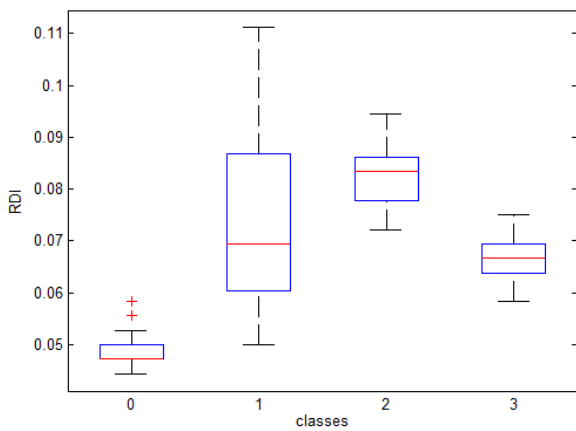
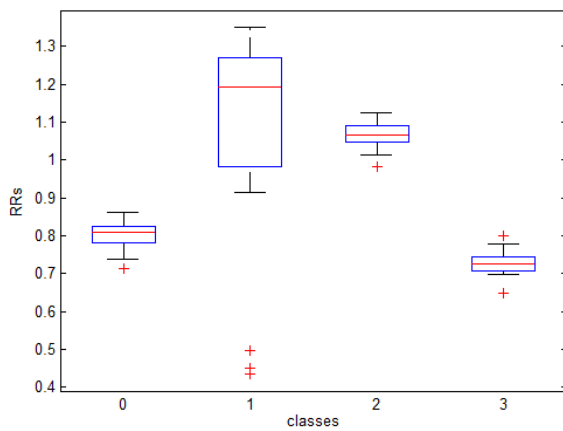


Figure 5.8 (g) : représentation de RR suivant en fonction des classes. Figure 5.8 (h) : représentation de RDI fonction des classes.

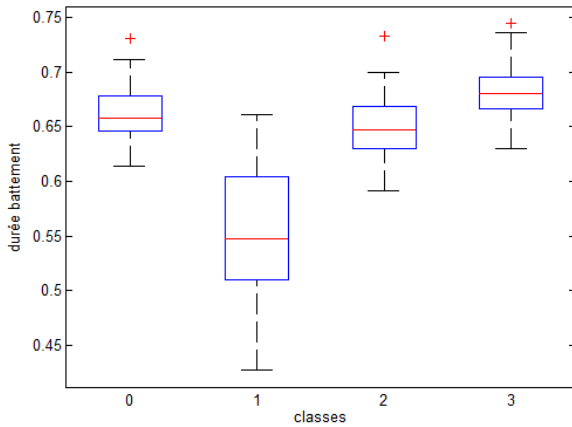


Figure 5.8 (i) : représentation de la durée du battement en fonction des classes.

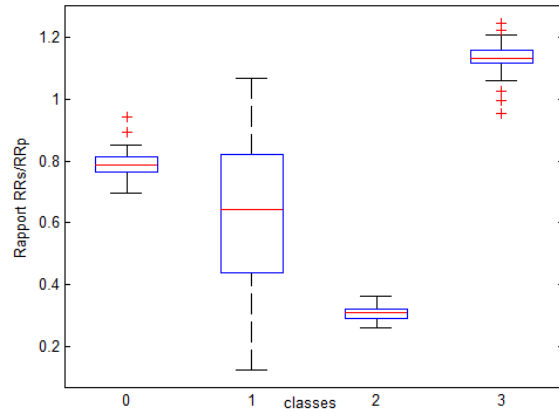


Figure 5.8 (j) : représentation du rapport RRs/RRp en fonction des classes.

- Interprétation des représentations graphiques des paramètres :

Les représentations graphiques de type boîte à moustaches sont un moyen rapide de figurer le profil essentiel d'une série statistique quantitative et servent à représenter schématiquement une distribution.

Tous les descripteurs varient différemment d'une classe à une autre, chacun occupe une plage de variations propre à lui. De ces représentations nous déduisons qu'un seul paramètre seul ne peut pas être l'unique discriminant pour les différentes classes.

Nous voyons clairement que la durée de P dans la (figure 5.8 (a)) a pratiquement la même plage de variation pour les 3 classes (N, R, et L) et une variation nulle pour la classe V ce qui est très logique du point de vue physiologique (dans le cas d'une extrasystole ventriculaire (V), l'onde P est absente).

Pour le cas de la figure 5.8 (b) ; nous remarquons que la variation de l'intervalle PR est distincte par rapport aux autres.

Quelques descripteurs se distinguent clairement par les différences de plages de variations pour chaque classe et montrent leur pertinence, c'est le cas des figures 5.8 (c-d-e-f-g-j).

Les descripteurs concernant les durées du pic R et R précédent et R suivant ont un impact très important et cela est justifié car nous sommes entrain de classer des arythmies cardiaques qui sont caractérisées par un élargissement ou par un complexe QRS prématuré donc un intervalle RR qui diminue.

Le complexe QRS varie différemment dans le cas normal(N) et dans les cas pathologiques (V, R, et L) (figure 5.8 (c)) donc son effet n'est pas négligeable à l'entrée du classifieur.

3.2. Représentation tridimensionnelle

Dans cette partie, nous avons testé l'effet de deux paramètres ensemble à la fois sur les 4 classes ciblées. La couleur rouge (+) représente le cas normal .La couleur verte (+) représente le cas extrasystole ventriculaire. La couleur bleue (+) représente le cas bloc de branche droit. La couleur violette (+) représente le cas bloc de branche gauche.

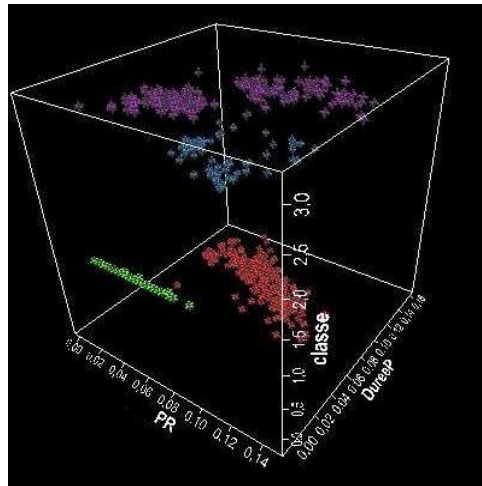


Figure 5.9(a) : représentation 3D de la durée P et l'intervalle PR en fonction des classes.

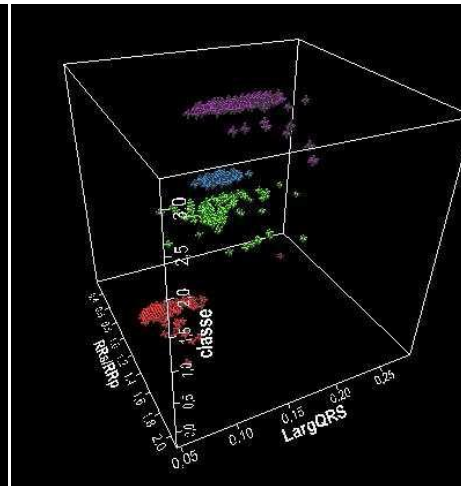


Figure 5.9(b) : Représentation 3D de la largeur du QRS et le RRs/RRp en fonction des classes.

- Interprétation tridimensionnelle des paramètres :

Nous avons déjà remarqué, la limite d'un seul paramètre à discriminer entre les différentes classes dans la 1ère expérimentation, ce qui nous a poussé à voir l'intérêt d'augmenter le nombre des attributs. Et nous remarquons clairement dans la (figure 5.9(a)) que seul le nuage de points pour le cas durée P nulle (couleur verte) se distingue des autres, ce qui correspond à la classe extrasystole ventriculaire, par contre pour le dernier couple (largeur QRS, RRs/RRp) (figure 5.9(b)) nous distinguons une bonne discrimination entre les pathologies, les nuages de points des différentes classes sont nettement séparés. Nous pouvons tirer de ces expérimentations menées sur les différents paramètres du vecteur d'entrée d'un classifieur des arythmies cardiaques les points suivants :

- 1- Les performances de tout modèle de classification, indépendamment de la technique utilisée, dépend énormément du vecteur d'entrée.
- 2- Une représentation géométrique des paramètres d'entrée peut être intéressante sur le plan visuel et sur le choix final de ces paramètres. Mais néanmoins une visualisation de plus de deux paramètres devient difficile à représenter.
- 3- Des paramètres comme : RRs, RRs/RRp, Durée P et larg QRS sont très pertinents pour la reconnaissance des pathologies comme V, R et L.

4. Résultats

4.1 Dimensionnement du classifieur neuronal et choix des paramètres :

Plusieurs paramètres sont à fixer au niveau de l'architecture des réseaux de neurones : le nombre de neurones dans la couche d'entrée, dans la couche de sortie et dans la couche cachée, les fonctions d'activation, et l'algorithme utilisé en apprentissage.

1. Couche d'entrée : elle est composée de 10 descripteurs décrits précédemment.

En premier lieu nous commençons par réaliser des expérimentations sur un perceptron multi couches PMC pour choisir la meilleure architecture et l'appliquer pour les expérimentations suivantes.

-en premier lieu nous appliquons l'algorithme d'apprentissage de la descente de gradient de Levenberg Marquardt (LM).

- en deuxième lieu nous appliquons l'algorithme d'apprentissage de la descente de gradient de Quasi-Newton (BFG)

2. Couche cachée : le choix de la taille de la couche cachée est réalisée souvent de façon heuristique et expérimentale, car aucune formulation théorique ne peut nous aider à l'optimiser. Nous avons remarqué qu'une seule couche cachée composée de huit neurones nous donne les meilleurs résultats.

3. Couche de sortie : La sortie est composée de cinq neurones qui indiquent les classes suivantes ('N', 'V', 'R' et 'L' et 'autre').

Après plusieurs essais nous avons fixé le nombre d'itérations à 500, ce qui a permis de réaliser un apprentissage satisfaisant.

Le nombre d'itérations est un paramètre important : s'il est trop grand, le réseau risque de réaliser un sur-apprentissage; au contraire, s'il est trop petit, l'apprentissage ne sera pas conduit à son terme.

Les poids synaptiques initiaux ont été choisis d'une façon aléatoire. Le seuil de l'erreur globale est fixé à 0.0001 pendant ces essais.

4.2 Classification neuronale en utilisant deux fonctions d'apprentissage différentes

Dans cette section, nous présentons les résultats de quelques classifieurs neuronaux choisis après plusieurs essais.

Nous comparons ces différents classifieurs en les testant sur une même base de test sachant que la base d'apprentissage concerne des patients différents de la base de test.

Nous commençons par présenter les différentes bases d'apprentissage et de test qui nous serviront à la conception de différents classifieurs neuronaux.

Nous adoptons les annotations suivantes pour les enregistrements utilisés :

A titre d'exemple ,les battements sélectionnés de classe normale N de l'enregistrement 106 de MIT :106n.

Les battements sélectionnés de classe normale bloc de branche gauche de l'enregistrement 207 de MIT :207l

Cette annotation est aussi utilisée pour les autres enregistrements, par exemple nous avons 119l et 214v.

Nous pouvons aussi constituer une base de test constituée de battements représentant toutes les classes, par exemple : 103n200v118r207l.

Nous allons présenter dans cette section les résultats obtenus par le classifieur neuronal avec l'algorithme de rétropropagation du gradient (PMC).

Pour mesurer la performance des classificateurs nous utilisons des lois statistiques qui sont : la sensibilité(Se), la spécificité(Sp) et le taux de classification correcte(Cc).

$$Se(i) = \frac{vp(i)}{vp(i) + fn(i)}$$

$$Sp(i) = \frac{vn(i)}{vn(i) + fp(i)}$$

$$Cc = \frac{vp(i) + vn(i)}{vp(i) + vn(i) + fp(i) + fn(i)}$$

Nous allons donner des exemples pour expliquer les différents termes utilisés dans ces calculs.

VP: l'entrée est classée comme arythmie et diagnostiquée arythmie par les médecins.

VN: l'entrée est classée normale et définie par les médecins comme personne saine.

FP: l'entrée est classée comme une arythmie, alors qu'elle est étiquetée comme une personne en bonne santé par des experts médicaux.

FN: l'entrée est classée comme normale mais en réalité elle est étiquetée comme arythmie par les médecins.

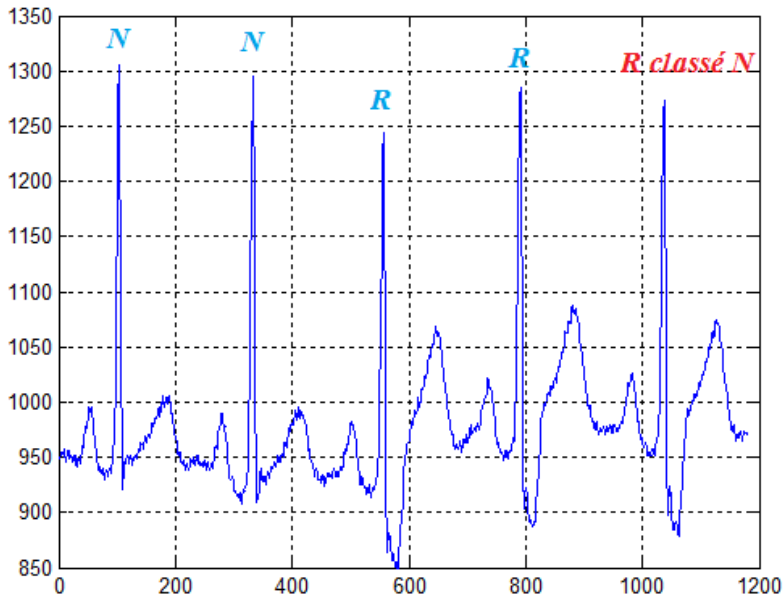


Figure 5.10 : exemple d'un faux négatif sur un enregistrement 212 de la base MIT-BIH. L'entrée est classée normale (N) alors que nous sommes en présence d'un bloc de branche droit.

4.2.1 Apprentissage et test du classifieur PMC1

Nous présentons dans le tableau 5.3 et la figure 5.11 les résultats des tests du classifieur neuronal PMC1 sur différentes bases. Sachant que nous avons effectué l'apprentissage par les deux méthodes celle de Quasi-Newton et celle de Levenberg-Marquardt.

	vp	vn	fp	fn	se	sp	cc	Fonction
PMC1 testé avec 106n106v	28	25	0	6	82.3529	100	89.8305	Trainbfg
	9	27	0	24	27.2727	100	60	Trainlm
PMC1 testé avec 214i214v	47	13	37	3	94	26	60	Trainbfg
	5	48	0	42	10.6383	100	55.7895	Trainlm
PMC1 testé avec 208v	151	0	0	1	99.3421	NaN	99.3421	Trainbfg
	0	0	0	150	0	NaN	0	Trainlm

PMC1 215n208v124r10911	testé avec	151	184	56	1	99.3421	77.0833	85.4592	Trainbfg
		0	237	3	150	0	98.75	60.7692	Trainlm
PMC1 testé avec 2071		40	0	0	0	100	NaN	100	Trainbfg
		0	0	0	40	40	NaN	0	Trainlm
PMC1 103n200v118r2071	testé avec	40	87	8	0	100	91.5789	94.0741	Trainbfg
		0	95	0	40	0	100	70.3704	Trainlm

Tableau 5.3 :résultats des tests de performances du PMC1

Nan : Absence de cas non ciblés.

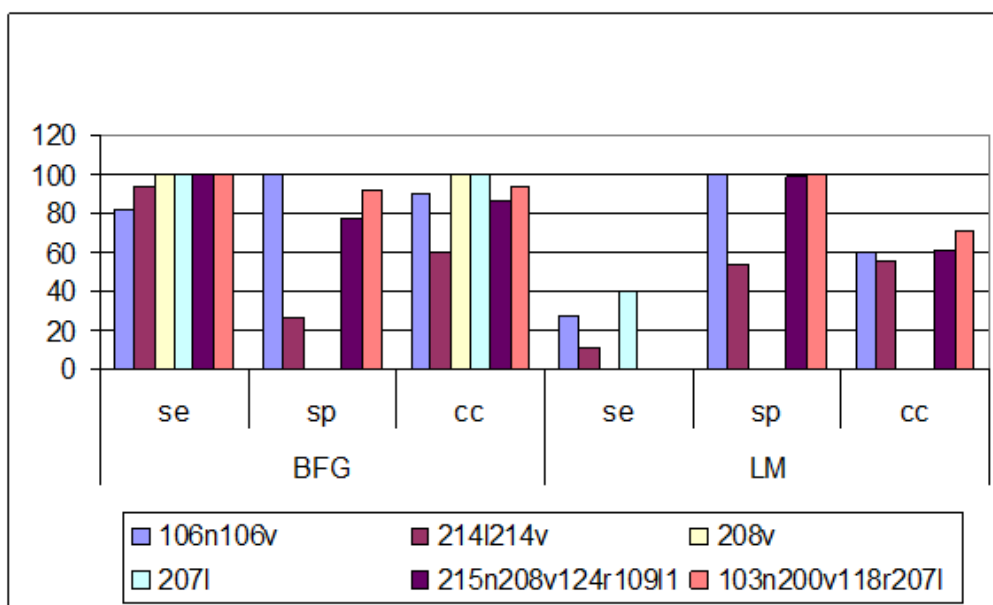


Figure 5.11: Les performances du PMC1 en utilisant les deux fonctions d'apprentissage 'trainbfg' et 'trainlm'.

L'analyse des résultats montre que les classifieurs dont leur fonction d'apprentissage est « trainbfg » donne de meilleures performances de classification que les réseaux qui ont comme fonction d'apprentissage « trainlm ».

4.2.2 Apprentissage et test du classifieur PMC2

Lors du test du PMC2 avec la base 119n119v (l'arythmie ciblée est l'ESV) les résultats obtenus ;présentés dans le tableau 5.4 et la figure 5.12 ; montrent clairement que la fonction d'apprentissage « trainbfg » nous a permis d'avoir les meilleures performances.

Nous avons ciblé l'anomalie BBD : 'R' lors du test du PMC2 avec la base 212n212r et l'anomalie BBG 'L' lors du test du PMC2 avec la base 214i214v.

	vp	vn	fp	fn	se	sp	cc	fonction
PMC2 testé avec 119n119v	34	50	0	0	100	100	100	Trainbfg
	4	27	23	30	11.7647	54	36.9048	Trainlm
PMC2 testé avec 212n212r	18	0	5	8	69.2308	0	58.0645	Trainbfg
	0	5	0	26	0	100	16.1290	Trainlm
PMC2 testé avec 214i214v	40	49	0	10	80	100	89.8990	Trainbfg
	3	48	2	47	6	96	51	Trainlm

Tableau 5.4 :résultats des tests de performances du PMC2

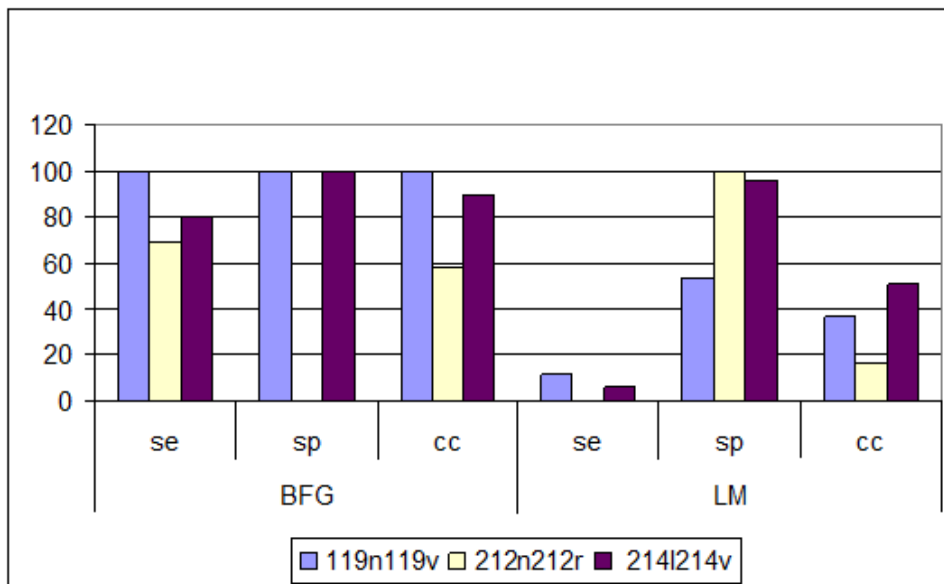


Figure 5.12: Les performances du PMC2 en utilisant les deux fonctions d'apprentissage 'trainbfg' et 'trainlm'

Le classifieur qui utilise la fonction d'apprentissage « trainlm » présente des résultats moins performants par rapport à celui qui utilise la fonction « trainbfg ».

Les performances des différents perceptrons multi couches sur multiples bases de test nous ont mené à choisir la fonction « trainbfg » pour l'apprentissage des différents classifieurs utilisés dans notre travail.

4.3 Sélection de l'architecture du classifieur MAPMC :

Pour choisir la meilleure architecture du classifieur et les paramètres de métaplasticité dans cette recherche, nous appliquons deux critères différents :

- 1) les paramètres de métaplasticité : un certain nombre de neurones a été fixé dans la couche cachée suffisamment élevé pour supposer que le réseau de neurones puisse réaliser la classification, tout en variant les paramètres de métaplasticité A et B, pour obtenir la valeur de $MSE=0.0001$ en un nombre minimum d'itérations.
- 2) Nombre de neurones dans la couche cachée : on fait varier dans ce cas le nombre de neurones dans la couche cachée jusqu'à atteindre le minimum d'erreur (les paramètres de métaplasticité sont maintenus fixes) avec un nombre minimum de neurones mais sans affecter les performances finales.

Toutes les expérimentations réalisées dans ce travail sont implémentées comme suit :

- Le classifieur MAPMC proposé a été implémenté sous l'environnement Matlab (Matlab 7.4, R2007a) sur un PC Pentium IV. cet algorithme a été appliqué sur deux bases de données très connues dans le thème de la classification des données médicales.
- La fonction d'activation utilisée est une sigmoïde et elle sera la même fonction pour tous les neurones. pour l'analyse comparative des performances des classifieurs, ils seront tous entraînés avec le même ensemble de données, dépendant de la base de données.
Dans toutes les expérimentations, la base de données est divisée en deux parties :
Une base d'apprentissage : elle contient 60% ou 70% des échantillons de la base de données.
Une base de test : elle est formée à partir des 30 ou 40% des échantillons qui ne sont pas présentés dans la base d'apprentissage.
Une base de validation : elle contient 10% de la base de données.
- Dans ce travail nous avons réalisé deux expérimentations dans chaque cas ; une expérimentation fut formée pour obtenir le meilleur résultat en terme de précision dans la classification pour chaque base de données proposée. pour le deuxième cas nous avons généré 50 MAPMC avec différents poids initiaux, dont leurs valeurs sont aléatoires avec une distribution normale entre (moyenne 0 et variance 1).
Dans chaque expérimentation nous avons entraînés 50 réseaux afin de trouver un résultat moyen qui ne dépend pas de la valeur initiale aléatoire des poids des réseaux de neurones.
- Nous avons appliqué deux critères pour arrêter l'apprentissage :
 - 1) l'apprentissage s'arrête quand le réseau converge vers une erreur de 0.0001.
 - 2) l'apprentissage s'arrête quand le nombre d'itérations est égal à 500.Dans cette première expérimentation nous allons rechercher la meilleure architecture possible et les paramètres de Métaplasticité pour le perceptron multi couches. En se basant sur les critères cités au début de cette section. les résultats obtenus seront présentés dans le tableau 5.5 .

Architecture du réseau			Paramètres de Méta-plasticité		Nombre d'itérations	Temps d'apprentissage	Taux de classification %
E	CC	S	A	B			
10	8	5	39	0.5	297	116.2312s	97.89
10	8	5	41	0.25	336	125.7929s	97.98
10	8	5	38	0.5	227	109.6756s	98.17
10	7	5	38	0.5	214	107.2312s	97.79
10	11	5	38	0.5	387	130.5231s	97.43

Tableau5.5 : Résultats obtenus par le MAPMC pour différentes architectures proposées.

E :Nombre de neurones dans la couche d'entrée.

CC :Nombre de neurones dans la couche cachée.

S :Nombre de neurones dans la couche de sortie.

Nous remarquons que les meilleurs résultats ont été obtenus avec 08 neurones dans la couche cachée. Donc tous les réseaux utilisés dans ce travail auront la même architecture, avec 10 neurones dans la couche d'entrée et 8 neurones dans une couche cachée et 5 neurones dans la couche de sortie. La fonction d'activation utilisée est une fonction sigmoïdale.

D'après les résultats du tableau les meilleures performances sont obtenues pour les paramètres A=38 et B=0.5 (choisis expérimentalement).

L'erreur a été fixée à 0.0001 et le nombre d'échantillons à l'entrée de chaque classifieur est de 500.

Dans ce cas, deux expérimentations sont menées : la première pour obtenir le meilleur taux de classification correcte en une seule simulation, et pour la deuxième obtenir la moyenne de 50 simulations.

L'intérêt de faire la moyenne de 50 simulations réside dans l'initialisation aléatoire des poids des différents classifieurs.

4.4 Résultats du classifieur optimisé MAPMC

Dans cette section nous allons présenter les résultats de la classification par MAPMC et les comparer avec les résultats de l'algorithme classique de rétropropagation par un perceptron multi couches PMC.

Nous avons utilisé la méthode de validation croisée pour évaluer les performances des classifieurs .Nous présentons dans le tableau 5.6 les résultats de la moyenne des 10 validation croisée.

Validation croisée	Perceptron Multi Couches			Metaplasticité Artificielle PMC		
	CC%	SE%	SP%	CC%	SE%	SP%
1	92.84	87.37	98.34	98.91	98.87	98.94
2	97.06	98.31	94.73	99.63	99.43	100
3	94.51	87.42	98.31	99.26	100	97.89
4	91.18	91.12	91.67	97.59	98.52	96.56
5	90.27	90.11	91.27	92.81	91.6	92.73
6	95.89	95.65	96.78	99.06	100	97.59
7	89.13	81.48	100	97.83	96.29	100
8	97.12	100	92.47	100	100	100
9	96.89	93.8	100	99.97	100	99.02
10	92.34	98.23	90.29	97.47	98.29	95.2
Moyenne	93.72	92.34	95.38	98.25	98.3	97.79

Tableau 5.6 :Résultats obtenus par validation croisée pour l'algorithme de méta-plasticité artificielle « MAPMC » et le perceptron multi couches classique « PMC ».les meilleurs résultats obtenus dans cette recherche sont représentés en gras.

Les résultats obtenus par l'algorithme de métaplasticité (MAPMC) sont supérieurs à ceux du perceptron multi couches(PMC).la validation croisée a démontré que le modèle de métaplasticité produit une meilleure classification. La moyenne des 10 tests du MAPMC était de 98.25% avec une sensibilité de 98.3% et une spécificité de 97.79%, alors que pour le PMC la meilleure moyenne de classification était de 93.72%, avec une sensibilité de 92.34% et une spécificité de 95.38%.

Nous proposons aussi un autre tableau 5.7 pour comparer nos résultats avec ceux d'autres auteurs qui ont utilisé la même base de données.

Nous avons utilisé aussi deux autres classifieurs très utilisés dans le domaine médical, pour enrichir nos résultats et effectuer une comparaison assez complète.

Ces deux classifieurs sont les SVM(Support Vector Machine) et les KNN (K Nearest Neighbors).

SVM est une méthode de classification par apprentissage supervisé, elle fut introduite par Vapnik en 1995. Elle est basée sur l'utilisation de fonctions dites noyaux (kernel) qui permettent une séparation optimale des données lors de l'apprentissage. Les SVMs ont une capacité de traiter des données de grande dimension, un faible nombre d'hyper paramètres à régler, et un grand pouvoir de généralisation.

L'algorithme KNN figure parmi les plus simples algorithmes d'apprentissage artificiel.

Dans un contexte de classification d'une nouvelle observation x , l'idée fondatrice est de faire voter les plus proches voisins de cette observation. La classe de x est déterminée en fonction de la classe majoritaire parmi les k plus proches voisins de l'observation x .

Auteurs (année)	Méthode	Taux de classification%
Osowski, S et al (2001)	Fuzzy Hybrid NN	96.06
Owis M.I et al (2002)	Blind Source Separation.	96.79
Prasad, G et al (2003)	ANN	85.04
Gao.D et al (2004)	NN Bayesian	90.00
Yu S.N et al (2008)	ICA-NN	98.36
Y.Benchaib et al (2009)	MLP BPA	95.12
Gothwal H et al (2011)	Fourier-NN	98.48
Jadhav et al (2012)	ANNs	86.67
Cette étude (2015)	PMC	93.72
Cette étude (2015)	KNN	93.30
Cette étude (2015)	SVM	97.67
Cette étude (2015)	MAPMC	98.25

Tableau 5.7 : Taux de classification correcte obtenus par notre algorithme et d'autres classifieurs de l'état de l'art.

Les résultats présentés dans le tableau ci-dessus montrent que notre algorithme MAPMC a obtenu des résultats de classification très intéressants.

Une des difficultés rencontrées dans la classification des arythmies cardiaques est le déséquilibre dans le nombre des données d'apprentissage caractérisant chaque classe. Nous allons mentionner dans le tableau 5.8 ,la

différence du nombre de battements des blocs de branche droit et gauche, par rapport aux battements Normaux et extrasystoles ventriculaires.

Naturellement le classifieur aura plus de facilité à reconnaître les battements qui sont en plus grand nombre dans la phase d'apprentissage. la caractéristique la plus importante de notre classifieur MAPMC est de donner plus de poids aux données les moins fréquentes, d'où l'amélioration de la classification des battements R et L. (qui sont moindres par rapport aux autres).

	<i>Battements R correctement classés</i>	<i>Battements L correctement classés</i>
<i>MAPMC</i>	106	202
<i>PMC</i>	97	145

Tableau 5.8 : Classification des blocs de branche droit et gauche par le MAPMC et le PMC.

Nous remarquons que quelques battements mal classés par le classifieur PMC ont été correctement reconnus par le classifieur MAPMC, donc un autre avantage de notre algorithme proposé vient d'être exposé. Le nombre des vrais positifs et la sensibilité du MAPMC ont augmenté. Le MAPMC a permis de diminuer le nombre de fausses alarmes, ce qui représente un point très important dans le diagnostic médical.

Dans les tableaux suivants nous présentons les meilleures performances, en terme de sensibilité, spécificité et taux de classification correcte obtenus par les deux classifieurs MAPMC et PMC en une seule simulation et aussi la moyenne de 50 simulations.

Classifieurs	Performances			Temps de convergence
	Sensibilité	Spécificité	Taux de classification correcte	
MAPMC	100	99.02	99.97	100.6756 s
PMC	93.8	100	96.89	122.3546 s

Tableau 5.9 : meilleure performance des classifieurs PMC et MAPMC

Classifieurs	Performances		
	Sensibilité	Spécificité	Taux de classification correcte
MAPMC	98.57±0.7	100±0.3	98.88±0.5
PMC	93.67±0.9	97.37±0.5	95.76±0.6

Tableau 5.10 : performances en moyenne de 50 simulations des classifieurs PMC et MAPMC.

Les tableaux démontrent une fois de plus la supériorité des performances de notre algorithme de métaplasticité artificielle dans le cas de la meilleure simulation obtenue et dans le cas de la moyenne de 50 simulations.

5. Discussion

Les résultats obtenus par l'algorithme de métaplasticité proposé dans cette thèse sont de loin les meilleurs en comparaison avec d'autres méthodes de l'état de l'art.

Nous savons que dans le domaine médical, le nombre des cas FP (faux positifs) et FN (faux négatifs) doit être minimisé, car dans ce domaine la vie des patients en dépend. Nous ne pouvons pas passer à côté d'une arythmie cardiaque (FN) qui peut mener à une fibrillation cardiaque puis à la mort. et dans le deuxième cas classer un

battement normal comme étant pathologique (FP) et induire le diagnostic en erreur et peut être mettre le patient dans une situation critique en lui prescrivant un traitement alors que le patient est sain.

Souvent en pratique un exemple FN est plus pénalisé qu'un exemple FP. en particulier lorsqu'il s'agit d'un diagnostic médical où nous devons encourager le rejet de FN lorsque la probabilité de leur appartenance à une classe est proche de 0.5.

Nous avons même comparé notre algorithme avec deux classifieurs performants dans le domaine médical qui sont le SVM et le KNN, ces derniers ont obtenus des résultats satisfaisants de 97.67% et 93.30% respectivement en terme de classification correcte mais ces résultats restent loin devant les performances réalisées par le MAPMC.

Nous remarquons que deux travaux ont donné des taux de classification plus élevés notamment celui Gothwal et al et Yu and Chou , qui ont obtenu des taux de 98.48% et 98.36%, respectivement (voir Tableau 5.7).

Dans ces deux méthodes ils ont inclus des étapes de prétraitement et d'extraction des données d'entrée.

De plus, dans le travail de Gothwal et al , ils n'ont utilisé que 5 neurones dans la couche d'entrée du réseau de neurones. Et dans l'application de Yu and Chou , les auteurs ont utilisé 23 composantes indépendantes dans l'entrée pour la reconnaissance des arythmies cardiaques.

Dans notre travail nous avons sélectionné 10 descripteurs dans la couche d'entrée du MAPMC mais sans application de méthodes d'analyse de données. ces descripteurs ont été choisis avec l'aide de cardiologues et d'après les caractéristiques des pathologies à classer.

Un des avantages de notre algorithme d'apprentissage est sa robustesse et sa rapidité de convergence.

Une des difficultés dans ce travail est la reconnaissance du bloc de branche droit (R) et bloc de branche gauche (L).

Les résultats obtenus par le MAPMC de 98.25% ont une relation directe avec les données d'apprentissage issues de la base de données MIT-BIH. Les enregistrements de la base de données sont de la dérivation DII alors que les blocs de branche droit et gauche apparaissent mieux dans les dérivations V1 et V6. et aussi le nombre de paramètres d'entrée représentant les blocs de branche sont moins nombreux que ceux représentant les cas normaux et d'extrasystole ventriculaires(V).

Mais cette difficulté de reconnaissance a été palliée en utilisant la fonction de distribution des poids de métaplasticité dans la phase d'apprentissage, ce qui a permis d'améliorer la reconnaissance des cas R et L et par conséquent le taux de classification correcte devient meilleur.

Le classifieur MAPMC a permis de diminuer le nombre de faux négatifs, ce qui représente un point très important dans le diagnostic médical.

Quelques battements non reconnus par le PMC ont été bien classés par notre algorithme, donc nous avons augmenté le nombre des vrais positifs et la sensibilité du MAPMC.

6. Classification par MAPMC et Sélection de variables

D'après les résultats de l'état de l'art, nous avons souligné la nécessité de réduire la dimension d'un problème, notamment pour améliorer les performances de classification [Lan'96]. Par ailleurs, la création d'un outil de classification revient à créer un modèle, et il paraît légitime que ce modèle soit le plus simple possible. Ainsi, tout en renforçant la classification, la simplicité du modèle améliore la vitesse d'exécution, le pouvoir de généralisation, ainsi que la compréhension des entrées du modèle (les variables ou les caractéristiques nécessaires à la modélisation [Dre'02]).

A la suite de tout ça, il nous semble très important d'inclure une étape de sélection de variables dans le système de classification. Nous avons choisi d'utiliser une méthode de la famille des « Wrapper » (méthodes enveloppantes) (voir chapitre 4). Elles sont généralement considérées comme étant meilleures que celles de filtrage selon [Li et Guo'08], [Hua et al'08]. Les méthodes Wrapper sont capables de sélectionner des sous-ensembles de caractéristiques de petite taille qui sont performants et pertinents pour le classifieur.

Parmi les méthodes utilisant une évaluation de type *wrapper*, les techniques heuristiques les plus connues sont fondées sur des sélections séquentielles ascendante et descendante, respectivement nommées *sequential forward selection* (SFS) et *sequential backward selection* (SBS). Ces méthodes identifient le meilleur sous-ensemble de variables en ajoutant ou en éliminant progressivement des variables. Ainsi, SFS commence avec un ensemble vide de variables, et choisit à chaque itération la variable améliorant le plus la valeur du critère. Dans cette procédure, le critère d'évaluation permet d'obtenir les performances des sous-ensembles de variables, il est généralement fondé sur les performances de classification. Dans notre exemple, nous chercherons à le maximiser. Inversement, SBS commence avec toutes les variables, et choisit d'éliminer à chaque itération la variable qui améliore le plus le critère.

Ahan et Bankert [Aha et Ban'95] ont montré que la méthode SBS est plus performante parce qu'elle prend en considération l'interaction d'une caractéristique avec un ensemble de caractéristiques plus large, contrairement au SFS qui ne prend en considération que l'interaction de cette caractéristique avec le sous-ensemble déjà sélectionné. Le processus de recherche peut s'arrêter lorsque il n'y a plus d'amélioration de précision. En d'autres termes, lorsqu'il n'y a plus la possibilité de trouver un sous-ensemble meilleur que le sous-ensemble actuel.

Les méthodes de sélection de type wrapper illustrent parfaitement ce fonctionnement, au sens où le processus de sélection des variables utilise les performances des algorithmes de classification pour optimiser la sélection.

L'évaluation se fait à l'aide d'un classifieur qui estime la pertinence d'un sous-ensemble donné de caractéristiques. Les sous-ensembles de caractéristiques sélectionnés par cette méthode sont bien adaptés à l'algorithme de classification utilisé, mais ils ne sont pas forcément valides si on change le classifieur. La complexité de l'algorithme d'apprentissage rend les méthodes "wrapper" très coûteuses en temps de calcul. En général, pour diminuer le temps de calcul et pour éviter les problèmes de sur-apprentissage, le mécanisme de validation croisée est utilisé.

SBS commence par l'ensemble de toutes les caractéristiques et à chaque itération, la caractéristique la plus mauvaise sera supprimée.

Dans cette thèse nous utiliserons la fonction « *sequentialfs* » dans Matlab de la Toolbox Statistics, elle effectue la sélection de fonction séquentielle. Des vecteurs d'entrée comprennent des variables d'entrée et les classes. Une fonction nous permet de spécifier la direction de la sélection ascendante ou descendante, SFS ou SBS. Et pour terminer la validation croisée est utilisée comme critère pour le choix des variables les plus pertinentes.

Dans notre cas nous avons choisi la direction descendante « SBS » tel que, à partir de toutes les variables, la fonction « *Sequentialfs* » crée des sous-ensembles en éliminant séquentiellement des caractéristiques. Pour chaque sous-ensemble candidat, on effectue la validation croisée par des appels répétés d'une fonction avec différents sous-ensembles d'apprentissage.

L'affichage itératif montre un changement dans la valeur de critère lorsque chaque nouvelle variable est supprimée du modèle. Le résultat final est un modèle réduit avec seulement cinq des dix caractéristiques d'origine: colonnes 1, 2,4, 8, et 10 .

A partir des 10 caractéristiques de la base de données des arythmies cardiaques , la méthode SBS a sélectionné les paramètres suivants :l'onde P ; PR ;ST ; RDI (retard de la déflexion intrinsécoïde) et le rapport RRp/RRs.

Nous rejoignons directement avec ces descripteurs retenus la nature physiologique des arythmies cardiaques à classer.Ces 5 descripteurs sont très pertinents et permettent de distinguer entre les différentes classes. Un exemple très clair peut être cité, revenons à la figure 5.8(d) où nous avons représenté le segment ST en fonction des différentes classes et où nous remarquons que chaque classe a une plage de variations différente ce qui montre l'importance de ce descripteur dans la classification.

Chaque descripteur retenu par la méthode de sélection a une signification logique et directe avec les différentes arythmies à classer. La durée de l'onde P est une variable discriminante pour la classification de l'extrasystole ventriculaire 'V', le RDI est un aspect propre aux blocs de branche droit et gauche, de plus le segment ST et l'intervalle PR permettent de différencier entre ces deux arythmies. Terminons par le rapport des rythmes précédent et suivant qui reflète l'élargissement du complexe QRS dans le cas des extrasystoles et des blocs de branche.

6.1. Résultats de la classification avec sélection de variables

Nous allons effectuer maintenant une classification en utilisant les variables sélectionnées. Les résultats obtenus sont présentés dans le tableau suivant :

Base de données	Variables sélectionnées	MSE	CC% PMC	CC% MAPMC	Temps d'apprentissage PMC	Temps d'apprentissage MAPMC
MIT-BIH	5	0.0001	98.22	99.14	99.2160 s	86.3478 s

Tableau 5.11 : Taux de classification des classifieurs PMC et MAPMC et temps d'exécution.

Nous remarquons que l'algorithme de métaplasticité (MAPMC) a obtenu un taux de classification très élevé de 99.14% et ce résultat est nettement supérieur à celui obtenu par l'algorithme de rétropropagation classique(PMC).

Nous allons comparer nos résultats avec des travaux présentés dans la littérature, qui ont utilisé d'autres méthodes de sélection de variables et les réseaux de neurones artificiels pour la classification des arythmies cardiaques.

Prasad et Sahambi en 2003 ont proposé une méthode de sélection de variables en utilisant la décomposition et transformée en ondelettes .Les différents attributs choisis représentent les entrées d'un classifieur neuronal des arythmies cardiaques ,avec l'algorithme d'apprentissage de rétropropagation du gradient. Ce classifieur proposé est capable de reconnaître 12 arythmies cardiaques, ils sont parvenus à atteindre un taux de classification correcte 96.77%[Pra et Sah'03].

En 2004 Acharya et al ,ont proposé trois méthodes de sélection de variables l'entropie spectrale, la méthode de Poincaré et les exposants de Lyapunov, tout en utilisant les réseaux de neurones pour la classification d'arythmies cardiaques ,les meilleurs résultats obtenus varient de 80 à 85% [Ach et al'04].

Polat et al 2007 Ont proposé la méthode d'ACP pour sélectionner 15 variables et ils ont utilisé le classifieur LS SVM (least square support vector machine pour la classification des arythmies cardiaques. Le meilleur résultat obtenu en terme de taux de classification est de 96.86% [Pol et al'07].

Récemment en 2014;Maliarsky et al,ont utilisé les algorithmes génétiques pour optimiser la sélection de variables pour la classification des arythmies cardiaques par réseaux de neurones (GANNs), ils ont obtenu de très bons résultats pour la reconnaissance de la tachycardie et la bradycardie ,le meilleur résultat obtenu pour la classification du bloc de branche droit est de 86.5% [Mal et al'14].Durant la même année Masood et al utilisent le CGP :Cartesian Genetic Programming pour la sélection de 5 variables,et la classification par réseaux de neurones des arythmies cardiaques. Le CGPANN a donné un taux de classification correcte de 96.84%[Mas et al'14].

Nous avons testé aussi une méthode de type Filter pour la sélection de variables qui est la mRMR (Minimum Redundancy and Maximum Relevance). C'est une méthode qui sélectionne les variables les plus pertinentes et élimine les redondances.

Néanmoins l'inconvénient des méthodes de filtrage est qu'elles ignorent l'impact des sous ensembles sélectionnés sur les performances de l'apprentissage.

Les résultats sont présentés dans le tableau 5.12.

Auteurs (année)	Méthode	Taux de classification%
Owis M.I et al (2002)	Blind Source Separation.	96.79
Prasad et al (2003)	DWT	96.77
Acharya et al (2004)	Spectral entropy	85
Polat et al (2007)	PCA	96.86
Yu S.N et al (2008)	ICA-NN	98.36
Gothwal H et al (2011)	Fourier-NN	98.48
Maliarsky et al (2014)	GA	86.5
Masood et al (2014)	CGPANN	96.84
Cette étude (2015)	mRMR +PMC	97.98
Cette étude (2015)	mRMR +MAPMC	98.74
Cette étude (2015)	SBS +PMC	98.22
Cette étude (2015)	SBS+MAPMC	99.14

Tableau 5.12 : Taux de classification correcte obtenus par notre algorithme et d'autres classifieurs de l'état de l'art.

D'après les résultats présentés dans le tableau 5.12, nous pouvons confirmer l'intérêt d'utiliser une méthode de sélection de variables .nous avons amélioré les performances du classifieur MAPMC avec un nombre minimes de variables (5). Nous avons gagné aussi en temps d'apprentissage ,le MAPMC a pu converger au bout de quelques itérations. La simplicité du modèle améliore la vitesse d'exécution.la méthode SBS « Sequential Backward Selection » a donné les meilleurs résultats en les comparant avec les travaux cités dans la littérature.

7. Conclusion

Dans cette étude, nous avons appliqué l'algorithme de métaplasticité artificielle avec un perceptron multi couches pour faire la classification des arythmies cardiaques.

L'algorithme MAPMC s'inspire de la propriété biologique de métaplasticité des neurones.

L'intérêt de cette recherche était d'améliorer notre algorithme MAPMC avec un perceptron classique avec rétropropagation, et aussi comparer les résultats obtenus avec d'autres méthodes proposées dans l'état de l'art.

La base données MIT-BIH a été utilisée pour évaluer notre algorithme. Nous avons remarqué que le MAPMC a donné de meilleurs résultats par rapport au perceptron multi couches classique.

Notre classifieur a donné de bons résultats par rapport à quelques récents travaux de l'état de l'art qui ont utilisé la même base de données.

Nous avons par la suite utiliser une méthode de sélection de variables « SBS » ,ce qui nous a permis de travailler avec juste 5 variables .les résultats obtenus sont très satisfaisants en terme de rapidité et de taux de classification des différentes arythmies cardiaques.

Les résultats obtenus par notre algorithme MAPMC ont démontré qu'il peut être utilisé comme un système d'aide à la détection des anomalies avec un taux de fausses alarmes minimales.

III. 2^{ème} expérimentation : Classification du cancer du sein

1. Introduction

Un des problèmes rencontrés de nos jours dans la société et dans le milieu médical est la détection précoce du cancer du sein. En connaissance de cette problématique l'algorithme MAPMC sera testé sur la base de données très connue WBCD (Wisconsin Breast Cancer Database).

Le cancer est une des principales causes de mortalité dans le monde, par conséquent la recherche dans le diagnostic et le traitement sont devenus des sujets d'importance vitale pour la communauté scientifique.

La prévention reste un défi, et le meilleur moyen d'augmenter la survie des patients est la détection précoce. Si des cellules cancéreuses sont détectées avant leur propagation à d'autres organes, le taux de survie est supérieur à 97 %.

Pour cette raison, l'utilisation et le développement des classifieurs automatisés d'aide au diagnostic médical ont augmenté considérablement ces derniers temps. Ces systèmes de classification tentent de minimiser les erreurs produites par des spécialistes, et peuvent effectuer un nombre de diagnostic important en un temps donné et avec un taux de réussite très élevé. La plupart de ces systèmes reposent sur des techniques d'intelligence artificielle, combinées avec les méthodes du traitement de signal, principalement : réseaux de neurones artificiels, wavelet analysis, analyse d'images à l'aide de modèles bayésiens, SVM, logique floue et fractale parmi d'autres techniques mathématiques puissantes.

Dans cette recherche nous nous intéressons aux réseaux de neurones combinés avec une nouvelle méthode de Métaplasticité artificielle, proposée par le professeur Diego Andina et appliquée aux données du cancer du sein.

Le cancer du sein est le cancer le plus fréquent chez les femmes dans le monde et représente 16% de l'ensemble des cancers féminins.

En Algérie selon les statistiques de l'Institut national de la santé publique, basées sur ce qu'on appelle les registres du cancer, il y a environ 7 000 nouveaux cas du cancer du sein par an, ce qui est important. L'âge moyen des femmes touchées par cette maladie est de 45 ans mais cela va de 19 à 97 ans.

Le cancer du sein touche la femme quel que soit son statut. Les études montrent que cette maladie affecte aussi bien les célibataires que les femmes mariées.

Le plus fréquent des symptômes est l'apparition d'un nodule sur le sein, d'une boule. Quand elle apparaît, elle fait déjà un centimètre. Pratiquement huit femmes sur dix ayant un cancer du sein vont consulter pour cette raison.

Les autres symptômes pouvant alerter une femme en dehors de la boule est l'écoulement par le mamelon de sang ou la présence de ganglions sous l'aisselle.

Grâce au dépistage précoce, la femme peut être traitée et guérie dans 100% des cas avec simplement de la chirurgie. Or, malheureusement, les femmes consultent très tardivement. Elles viennent avec des tumeurs qui dépassent trois centimètres, et dans ce cas-là c'est obligatoire de faire l'ablation de tout le sein, d'utiliser la chimiothérapie et la radiothérapie avec tous leurs inconvénients. D'où l'importance de sensibiliser les femmes sur

l'importance de se faire dépister. Il faut qu'elles sachent qu'en termes de prévention, la meilleure façon de guérir d'un cancer éventuel est de le diagnostiquer le plus tôt possible. C'est valable pour tous les autres cancers, plus il est détecté rapidement mieux il est facile à guérir.

2. Description de la base de données du cancer du sein(WBCD)

La base de données du cancer du sein dénommée « Wisconsin Breast Cancer Database » a été obtenue par l'Université du Wisconsin [Wol'94] à travers les travaux du Docteur William H. Wolberg .il a collecté les informations médicales de 699 cas cliniques relatifs au cancer du sein classés comme bénin ou malin. La base de données contient 16 données manquantes, nous nous sommes restreints à travailler sur 683 /699 patientes : 444 patientes (soit 65%) sont des cas bénins et 239 patientes (soit 35%) sont des cas malins.

Les attributs ont des valeurs comprises entre 1et 10, tels que les 1 sont plus proches des cas bénins et les 10 sont plus proches des cas malins.

Les patientes sont caractérisées par 11 attributs : le premier fait référence à l'identificateur de la patiente et le dernier représente la classe: le diagnostic est de 2 si le cas est bénin ,4 si le cas est malin quant aux 9 autres, ils représentent des cas cliniques suivants:

1-Clump Thickness: l'épaisseur de la membrane plasmique d'une cellule cancéreuse est plus importante que celle d'une cellule normale.

2. Uniformity of Cell Size : les cellules cancéreuses sont caractérisées par une anisocytose, à savoir une inégalité au niveau de la taille par comparaison avec les cellules saines.

3. Uniformity of Cell Shape : les cellules cancéreuses sont marquées par des contours irréguliers ainsi que des incisures.

4. Shape Marginal Adhesion: une surexpression de la protéine integrin beta3 au niveau de la surface de la cellule cancéreuse.

5 .Single Epithelial Cell Size: étant donné que les cellules épithéliales sont absentes à l'état naturel au niveau de la moelle osseuse et qu'elles ne sont pas détectées chez les individus sains, la moelle osseuse peut, de ce fait, être considérée comme un indicateur de maladie métastatique chez les patients atteints du cancer du sein au stade primaire.

6. Bare Nuclei: à l'état normal, les nucléoles se trouvent à l'intérieur du noyau. Dans le cas où ses derniers se trouvent confondus avec le cytoplasme cela indique que la cellule présente une anomalie et qu'elle est susceptible de devenir cancéreuse.

7. Bland Chromatin : H2az est une protéine qui induit l'expression du gène du récepteur d'oestrogènes.

La surproduction de cette protéine est un marqueur de présence de cellules cancéreuses au niveau du sein étant donné qu'elles sont hormono-dépendantes.

8. Normal Nucleoli : L'ADN est naturellement protégé par une membrane nucléaire. Une défaillance observée au niveau de cette membrane peut refléter une croissance tumorale.

9. Mitoses : La mitose est un processus de division cellulaire régulé permettant de reproduire des cellules filles génétiquement identique à la cellule parentale.

Les cellules malignes sont caractérisées par une division cellulaire anarchique et intense par comparaison avec une population cellulaire normale.

Numéro	Attributs	Valeur des attributs	Moyenne
1	Clump thickness	1-10	4.44
2	Uniformity of Cell Size	1-10	3.15
3	Uniformity of Cell Shape	1-10	3.22
4	Shape Marginal Adhesion	1-10	2.83
5	Single Epithelial Cell Size	1-10	2.23
6	Bare Nuclei	1-10	3.54
7	Bland chromatin	1-10	3.45
8	Normal nucleoli	1-10	2.87
9	Mitoses	1-10	1.60

Tableau 5.13:Description de la base de données WBCD.

3. Sélection de l'architecture du classifieur neuronal

Nous allons poursuivre notre expérimentation pour évaluer notre algorithme de Métaplasticité dans la classification du cancer du sein.

Tout d'abord nous allons choisir la meilleure architecture du réseau, qui donne les résultats les plus satisfaisants, et ceci en changeant les différents paramètres de la métaplasticité et le nombre de neurones dans la couche cachée.

Notons que nous allons suivre le même protocole de l'expérimentation 1 de la classification des arythmies cardiaques.

1. Algorithme de MAPMC

- 1) Structure du réseau utilisé dans les différentes expérimentations :
 - a) Le Nombre de neurones d'entrée est égal au nombre des attributs de la base de données.
 - b) Nombre de couches cachées :1.
 - c) Nombre de neurones dans la couche cachée : 8 (peut varier en fonction de l'expérimentation, de la base de données et des simulations précédentes)
 - d) Nombre de neurones de la sortie : 1 pour l'expérimentation du cancer du sein et 5 pour l'expérimentation des arythmies cardiaques.
 - e) facteur d'apprentissage $\eta=1$.
 - f) Fonction d'activation : sigmoïdale avec des valeurs entre [0,1].
- 2) Tous les poids de la matrice W sont initialisés aléatoirement, dans le rang [-0.6 , 0.6].(il a été démontré empiriquement que le travail est optimal avec une répartition uniforme dans cette gamme lors de la génération des valeurs initiales des poids et des seuils)
- 3) phase d'apprentissage :
 - a) l'algorithme de métaplasticité artificielle est modélisé en appliquant la fonction de poids de l'équation (5.1) pour le changement des poids de l'ARP pendant la phase d'apprentissage :

$$w_{ij}^{(l)}(t+1) = w_{ij}^{(l)}(t) + \eta \cdot \delta_j^{(l)} \cdot \hat{y}_i^{(l-1)} \quad (5.1)$$

$$\delta^{(L)} = (y - \hat{y}^{(L)}) \cdot \hat{y}^{(L)} \cdot (1 - \hat{y}^{(L)}) / f_x^*(x)$$

Où w_{ij} sont les poids des neurones j dans la couche l pendant l'itération t , et $\hat{y}_i^{(l-1)}$ les sorties des neurones i de la couche précédente, et $\delta(l)$ est le terme d'erreur habituel dans la rétropropagation du gradient, pour le cas de la fonction d'activation sigmoïdale et la couche de sortie L , seront actualisés suivant la fonction de pondération de métaplasticité, et y est la sortie désirée.

- b) Conditions d'arrêt de l'apprentissage :
- 1) Si le nombre d'itérations atteint 500
 - 2) Si l'erreur quadratique $MSE=0.0001$.

Le tableau 5.14 ci-dessous représente les résultats obtenus pour les différentes structures proposés et des paramètres de Métaplasticité.

Architecture du réseau			Paramètres de métaplasticité		MSE	N itérations	Temps d'apprentissage	Taux de classification apprentissage	Taux de classification test
E	CC	S	A	B					
9	8	5	39	0.5	0.0001	500	77.5132 s	99.75	99.63
9	8	5	41	0.25	0.0001	500	88.7929 s	98.89	98.71
9	7	5	39	0.25	0.0001	500	75.2212 s	99.11	98.71

Tableau 5.14: résultats obtenus par le MAPMC pour différentes architectures et différents paramètres de la métaplasticité.

Pour obtenir de bons résultats en termes de précision de classification et de rapidité d'apprentissage on fait varier les paramètres du réseau et de métaplasticité.

La fonction d'activation est une fonction sigmoïde d'intervalle $\{0,1\}$ et qui est le même pour tous les neurones. Pour évaluer et comparer les différents classifieurs, la base de données a été divisée en deux parties :60% pour l'apprentissage qui contient 410 malades dont 144 cas de tumeurs malignes et 266 de tumeurs bénignes .et 40% des données pour le test qui contient 273 malades dont 95 cas de tumeurs malignes et 178 de tumeurs bénignes.

Type de classifieur	Architecture du réseau			Paramètres de métaplasticité		MSE	N itérations	Nombre d'individus pour l'apprentissage	Nombre d'individus pour le test
	E	CC	S	A	B				
MAPMC	9	8	5	39	0.5	0.001	500	410	273
PMC	9	8	5	NA	NA	0.001	500	410	273

Tableau 5.15: Présentation des différents paramètres pour les deux classifieurs (MAPMC et PMC)

Dans ce travail deux expérimentations ont été menées : une pour obtenir le meilleur résultat en terme de précision dans la classification en une simulation et une autre pour obtenir la moyenne de 50 simulations. pour le second cas nous allons généré 50 MAPMC avec des initialisations de poids différents et dont les valeurs ont été distribués normalement aléatoirement (avec une moyenne 0 et une variance 1).

Dans chaque expérience 50 réseaux de neurones ont été formés pour obtenir une moyenne, qui est indépendante des valeurs aléatoires initiales des poids des réseaux. Deux critères d'arrêt de l'apprentissage ont été appliqués :

dans un cas l'apprentissage s'arrêtera quand l'erreur atteindra 0.0001 et dans le deuxième cas l'apprentissage s'arrêtera quand le nombre d'itérations sera 500.

4. Evaluation du classifieur optimisé MAPMC

Nous allons présenter dans cette section les résultats obtenus par le classifieur proposé MAPMC et aussi par le perceptron multi couches avec l'algorithme de rétropropagation du gradient (PMC).

Également une comparaison est effectuée entre les deux classifieurs.

Pour mesurer la performance des classificateurs nous avons mené deux évaluations : la première consiste à analyser la sensibilité, la spécificité et le taux de classification correcte ;et aussi la matrice de confusion, pour évaluer la précision de la classification des deux algorithmes. la seconde méthode représente de manière visuelle les résultats de la classification en utilisant la courbe « ROC » (Receiver Operating Characteristics) et le calcul de l'aire sous la courbe « AUC » (Area Under Curve).

Dans la section qui suit, Nous allons définir ces deux méthodes.

4.1. Critères des performances

a) Matrice de confusion :

La matrice de confusion contient des informations sur les classifications actuelles et celles prédites par les classifieurs. Le comportement d'un classifieur est généralement évalué en utilisant les données de ce tableau.

Dans le tableau suivant nous allons expliquer le fonctionnement d'une matrice de confusion pour un classifieur de deux classes.

La matrice prend la forme suivante :

observation	Positif	Négatif	Total
Positif	A	b	a+b
Négatif	C	d	c+d
Total	a+c	b+d	n=a+b+c+d

a :représente le nombre de prédictions correctes pour les entrées positives.

b :représente le nombre de prédictions incorrectes pour les entrées négatives.

c :représente le nombre de prédictions incorrectes pour les entrées positives.

d :représente le nombre de prédictions correctes pour les entrées négatives.

Classifieurs neuronaux	Résultats désirés	Résultats calculés	
		Tumeur bénigne	Tumeur maligne
MAPMC	Tumeur bénigne	178	1
	Tumeur maligne	0	95
PMC	Tumeur bénigne	175	4
	Tumeur maligne	5	90

Tableau 5.16 : matrice de confusion de la meilleure classification des deux classifieurs proposés.

Les résultats présentés par la matrice de confusion démontrent la supériorité de la classification par l'algorithme de Métaplasticité. L'algorithme a pu modifier les poids synaptiques et les ajuster pour atteindre un apprentissage optimal et donner une classification presque parfaite des tumeurs malignes.

La fonction de pondération utilisée va prendre en considération l'apport informationnel de chaque variable d'entrée et l'ajuster pour mener un apprentissage équilibré.

Le nombre de faux négatifs a sensiblement diminué en utilisant le classifieur MAPMC en comparaison avec la classification par PMC.

Le tableau 5.17 qui suit représente le résultat de la meilleure simulation pour chaque classifieur proposé en terme de sensibilité, spécificité et taux de classification correcte du cancer du sein.

Le tableau 5.18 contient la moyenne des 50 simulations pour les deux classifieurs.

Les classifieurs	Mesure des performances		
	Sensibilité	Spécificité	Taux de classification correcte
MAPMC	99.46	100	99.67
PMC	98.34	95.73	97.13

Tableau 5.17 : meilleur résultat pour chaque classifieur du cancer du sein.

Les classifieurs	Mesure des performances		
	Sensibilité	Spécificité	Taux de classification correcte
MAPMC	99.53±0.3	100±0.5	99.58±0.3
PMC	98.61±0.4	95.43±0.9	97.83±0.4

Tableau 5.18 : la moyenne des 50 simulations pour les deux classifieurs.

Nous avons mentionné dans la description de la base de données du cancer du sein, la différence du nombre de tumeurs malignes, par rapport aux tumeurs bénignes.

Naturellement le classifieur aura plus de facilité à reconnaître les tumeurs qui sont en plus grand nombre dans la phase d'apprentissage. la caractéristique la plus importante de notre classifieur MAPMC est de donner plus de poids aux paramètres les moins fréquents, d'où l'amélioration de la classification des tumeurs malignes. (qui sont moindres par rapport aux autres).

b) Résultats de la courbe ROC

L'analyse de sensibilité et de spécificité et les courbes ROC ont d'abord été développées pendant la seconde guerre mondiale pour la mise au point de moyens efficaces de détection des avions japonais. Elles ont ensuite été appliquées de manière plus générale en détection du signal, puis en médecine, où elles sont aujourd'hui très utilisées.

La problématique est la suivante : on étudie un phénomène, souvent de nature binaire (par exemple, la présence ou absence d'une maladie) et on souhaite mettre au point un test permettant de détecter efficacement la survenance d'un événement précis (par exemple, la présence de la maladie).

Nous proposons la courbe de ROC [Met'78], [Faw'06] comme dernière mesure pour évaluer des tests diagnostiques. Cette mesure est fondée sur la théorie statistique de la décision.

Elle a été développée initialement pour la détection de signaux associés aux radars durant la seconde guerre mondiale, où les opérateurs radars devaient distinguer les bateaux ennemis des bateaux alliés. Dans les années 1970, les courbes de ROC sont apparues en médecine afin d'améliorer la prise de décision en imagerie médicale [Han et McN'82]. Depuis quelques années, ces courbes sont présentes dans de nombreux domaines, notamment dans la recherche biomédicale où elles caractérisent les capacités (et les limitations) d'un test à prédire une maladie chez un patient.

Ces courbes analysent les variations de la sensibilité et de la spécificité, afin de visualiser la performance globale du test. Elles sont obtenues en traçant dans un plan la « sensibilité » en fonction de « 1- spécificité », en faisant varier le seuil de décision du modèle.

À partir de la courbe de ROC, l'indice permettant d'évaluer numériquement la courbe est l'**aire sous la courbe** de ROC (AUC, *Area Under the Curve*). Cet indice peut être analysé comme la probabilité du modèle à prédire correctement la pathologie chez un patient. Deux cas particuliers sont observés, ainsi :

- une aire AUC de 0.5 indique que le test est proche du hasard et il n'apporte rien au diagnostic: le signe informant de la présence de la maladie est alors indépendant de la pathologie.

En d'autres termes, le modèle ne différencie pas les deux classes, cela peut être observé par un recouvrement total des deux densités de probabilité, avec $V P = FP$;

- une aire égale à 1 correspond au meilleur résultat, avec une sensibilité et une spécificité de 100%, dans ce cas les deux classes sont totalement séparées.

Dans notre cas nous avons utilisé la courbe ROC pour montrer la supériorité du MAPMC sur le PMC. La figure suivante représente la courbe ROC à la fois pour le classifieur proposé MAPMC et pour le PMC.

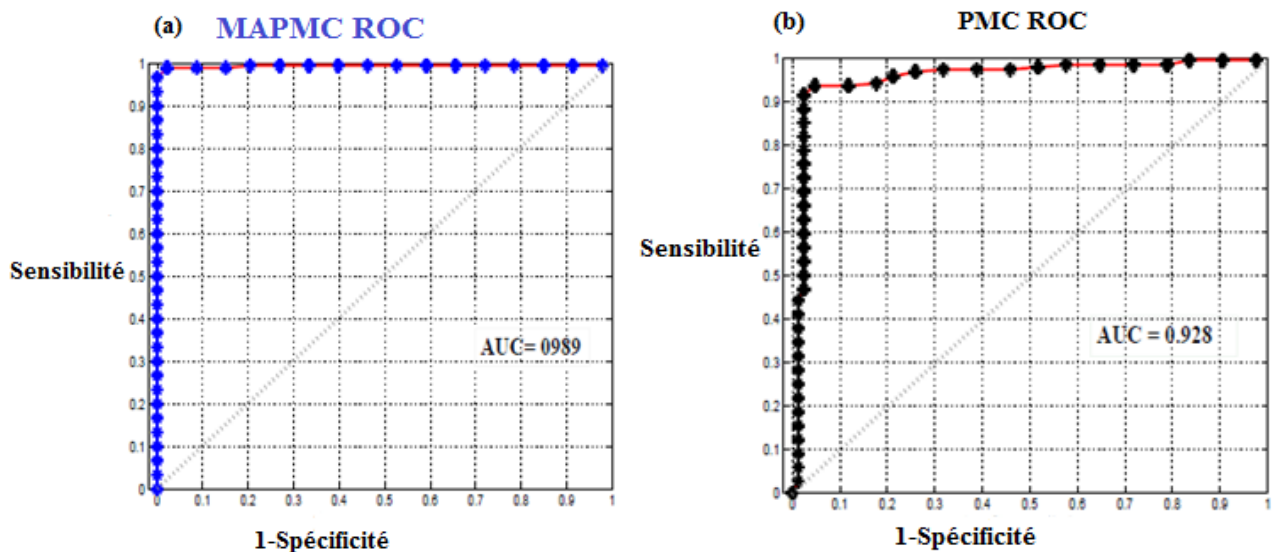


Figure 5.13 : courbe ROC des deux classifieurs, nous pouvons remarquer clairement la supériorité des résultats du classifieur MAPMC par rapport à l'algorithme de rétropropagation classique.

(a) Représente la courbe ROC et la AUC de 0.989 de MAPMC et (b) Représente la courbe ROC et la AUC de 0.928 du PMC.

Les résultats obtenus par calcul de l'AUC sont :0.928 et 0.989 pour le PMC et le MAPMC respectivement.

Les résultats indiquent une fois de plus la supériorité du MAPMC sur le PMC , dans ce cas en particulier.

5 .Comparaison avec l'état de l'art

Les résultats obtenus par le MAPMC ont été comparés aux résultats d'autres algorithmes de deux façons :

Premièrement en comparant nos résultats avec d'autres algorithmes proposés récemment appliqués sur la même base de données WBCD.et deuxièmement les résultats sont comparés aussi bien avec d'autres algorithmes développés pour diverses applications et qui ont choisis la même base de données.

Les résultats du MAPMC sont comparés avec les résultats d'autres algorithmes développés récemment et appliqués sur la même base de données WBCD, Conforti et Guido dans [Con et Gui'10] ont généré une fonction noyau optimale , en utilisant un algorithme de programmation semi défini(SDP Semi Defined Programming), Ils ont obtenu un taux de classification correcte de 96.79% . L'algorithme SDP est inefficace lorsque l'ensemble d'apprentissage est très grand, dans ces cas le SDP consomme beaucoup de temps dans la phase d'apprentissage et un grand cout computationnel. Peng et al ont présenté une technique de sélection de caractéristiques pour diminuer la grande dimension des données biomédicales .la technique est basée sur l'intégration de filtres et de méthodes de recherche séquentielle pour améliorer la classification des caractéristiques .la technique proposée fut implémenté en deux phases : premièrement ils ont ajouté une étape de présélection pour améliorer l'efficacité dans le recherche de sous ensemble de caractéristiques pour obtenir de meilleures performances de classification[Pen et al'09]. Et deuxièmement ils ont utilisé les courbes ROC pour représenter les performances de classification des caractéristiques individuelles et des sous ensembles de caractéristiques .Peng et al ont obtenu une valeur de 0.997 de l'aire sous la courbe ROC.

En 2009 , Akay présente un modèle basé sur les séparateurs à vaste marge (SVM Support Vector Machine) utilisant une recherche en grille(Grid Search) pour optimiser les paramètres du modèle et les caractéristiques d'entrée ont été sélectionnées en appliquant un calcul statistique nommé F-score (qui représente une mesure de la précision des tests de performances) .Akay a atteint une performance de 99.519% [Aka'09] . Ubeyli dans son experimentation a utilisé 5 classifieurs (SVM , Réseaux de neurones probabilistes ,réseaux de neurones récurrents, réseau combiné et un perceptron multicouches). Ubeyli a rapporté une performance de 99.54%[Ube'07].

Il est important de noter que les auteurs de ces études n'ont pas précisé si leurs résultats représentent la meilleure performance d'une simulation ou la moyenne de plusieurs simulations.

Notre méthode proposée nous a permis d'obtenir un résultat de 99.67% comme meilleure simulation et 99.58% comme moyenne de 50 simulations.

Les résultats obtenus ont été comparés avec d'autres résultats obtenus dans d'autres investigations menées sur la même base de données. Dans le tableau 5.19 nous représentons les différents auteurs et les résultats obtenus par les différentes méthodes proposées pour la classification.

Auteurs (année)	Méthode	Taux de classification%
Quinlan (1996)	C4.5	94.74
Hamiton et al (1996)	RAIC	95.00
Ster et Dobnikar (1996)	LDA	86.80
Nauk and Kruse (1999)	NEFCLASS	95.06
Pena Reyes and Sipper (1999)	Fuzzy-GA1	97.36
Setiono (2000)	Neuro rule 2a	98.10
Albrecht et al (2002)	LSA machine	98.80
Abonyi and Szeifert (2003)	SFC	95.57
Ubeyli (2007)	SVM	99.54
Polat and gunes(2007)	LS-SVM	98.53
Guijarro et al (2007)	LLS	96.00
Akay (2009)	SVM-CFS	99.51
Karabatak and Cevdet (2009)	AR+NN	97.40
Peng et al (2010)	CFW	0.997
Conforti and Guido (2010)	SVM-SDP	96.79
Cette étude (2014)	PMC	97.13
Cette étude (2014)	MAPMC	99.67

Tableau 5.19 : Taux de classification correcte obtenus par notre algorithme et d'autres classifieurs de l'état de l'art.

6. Classification par MAPMC et Sélection de variables

Comme présenté dans l'expérimentation de classification des arythmies cardiaques, nous allons montrer l'intérêt d'utiliser une méthode de sélection de variables. Nous remarquons que plusieurs auteurs utilisent des méthodes de sélection pour améliorer les performances de leurs modèles de classification.

Nous allons présenter les résultats obtenus après l'implémentation de la méthode SBS pour la sélection sur la base de données du cancer du sein WBCD.

Après la sélection des variables nous obtenons 3 paramètres parmi les 9 présents dans la base de données, qui sont : Clump thickness, single epithelial cell size et Uniformity of Cell Size.

Comparaison avec l'état de l'art

Auteurs	Méthode	Nbre de variables sélectionnées	Taux de classification%
Last et al 2001	IFN	3	32.2
Lee et al 2001	FEBFC	6	95.14
Chen et Shie 2005	FUZZY-ENTROPY	4	97.21
Muni et al 2006	GP	2,23	96.84
Bratu et al 2008	GBW,BFS	5	96.48
Notre étude 2015	mRMR,PMC	6	96.13
Notre étude 2015	mRMR,PMC	6	97.72
Notre étude 2015	SBS,PMC	3	98.23
Notre étude 2015	SBS,MAPMC	3	99.73

Tableau 5.20: comparaison de différentes méthodes de sélection de variables

Le tableau 5.20 résume quelques travaux d'auteurs qui ont utilisé la même base de données, nous présentons dans ce tableau la méthode utilisée pour la sélection de variables, le nombre de variables et le taux de classification correcte du cancer du sein.

Les résultats démontrent une fois de plus la supériorité de notre algorithme de métaplasticité et aussi l'efficacité de la méthode SBS par rapport à l'algorithme de rétropropagation classique et à d'autres classifieurs proposés récemment pour réduire la dimension des variables et pour la classification de la même base de données. De ces résultats nous pouvons conclure que la combinaison de la méthode SBS avec l'algorithme de métaplasticité MAPMC a réduit les variables d'entrée, a réduit le coût computationnel et a donné une haute précision de classification.

Nous avons suivi le même protocole de l'expérimentation 1, nous avons testé la méthode de filtrage mRMR pour la sélection de variables. Nous remarquons que la méthode de SBS donne de meilleurs résultats de classification avec un nombre minimum de variables. Dans ce cas nous sommes en présence d'un classifieur rapide, moins complexe et robuste.

7. Conclusions

Dans cette étude l'algorithme de métaplasticité MAPMC a été appliqué au problème critique et pertinent de la détection du cancer du sein. Pour évaluer les performances de l'algorithme proposé nous l'avons testé sur la base de données très reconnue du cancer du sein WBCD.

Le MAPMC a obtenu de meilleurs résultats par rapport à l'algorithme de rétropropagation du gradient classique appliqué sur le perceptron multi couches (PMC) et aussi par rapport à d'autres algorithmes récents de l'état de l'art appliqués sur la même base de données.

Les résultats démontrent que le MAPMC peut être utilisé pour la classification du cancer du sein car il améliore la performance actuelle des classifieurs et peut être utilisé pour comme un deuxième avis auprès des médecins pour prendre des décisions sur le diagnostic final.

La combinaison d'une méthode de sélection de variables « SBS » avec notre algorithme MAPMC permet d'améliorer le taux de reconnaissance et diminue considérablement le temps d'exécution.

Conclusion Générale

Conclusion et Perspectives

L'importance de traiter les données médicales à l'aide d'un système intelligent n'est plus un sujet à vérifier, en effet les méthodes d'intelligence artificielle ont pris une grande ampleur dans la classification et l'aide au diagnostic médical.

Les données extraites de différents signaux électro-physiologiques et différentes images médicales, sont prises en charge par un modèle intelligent qui va les exploiter durant la phase d'apprentissage. Cette phase joue un rôle majeur dans l'amélioration des performances des modèles intelligents.

Les réseaux de neurones s'avèrent des modèles intelligents très robustes et très utilisés dans le domaine médical.

L'algorithme le plus utilisé dans l'apprentissage des réseaux de neurones est l'algorithme de rétropropagation (ARP Algorithme de Rétropropagation du gradient). Ce dernier est appliqué en particulier au Perceptron Multi Couches (PMC).

Plusieurs difficultés sont rencontrées lors de l'apprentissage du PMC sur les données médicales.

Nous citons en particulier la dimension du vecteur d'entrée (le nombre de variables choisies), le nombre de classes en sortie, et les paramètres internes du PMC (poids, biais, ...etc). Dans le cadre des travaux de cette thèse, nous avons proposé un algorithme de Métaplasticité basé sur les propriétés biologiques de Métaplasticité des neurones pour améliorer l'apprentissage et les performances d'un classifieur neuronal.

Nous avons implémenté une fonction de métaplasticité artificielle qui permet de donner un poids plus grand aux données d'entrée minoritaires afin d'équilibrer l'apprentissage.

Notons que la métaplasticité biologique a une relation directe avec la plasticité et les mécanismes de mémorisation et d'apprentissage, et une relation aussi avec la théorie d'information de Shannon. L'algorithme de métaplasticité artificielle est appliqué sur le perceptron multi couches (MAPMC), en optimisant une fonction de modification des poids lors de la phase d'apprentissage du classifieur neuronal.

Nous avons testé le PMC optimisé sur deux bases de données médicales (cardiologique et cancer du sein). L'algorithme MAPMC nous a permis d'obtenir d'excellentes performances en terme de taux de classification, il a reconnu des cas d'arythmies minoritaires grâce à l'application de la fonction de pondération de métaplasticité artificielle lors de la phase d'apprentissage.

Le nombre de faux négatifs (FN) a sensiblement diminué lors de la classification par MAPMC. Sachant que les faux négatifs représentent un risque majeur dans le diagnostic médical. Nous avons remarqué aussi que la durée d'apprentissage a diminué en comparaison avec l'apprentissage par rétropropagation classique.

Nous avons utilisé aussi une méthode de sélection de variables pour diminuer la dimension du vecteur d'entrée du MAPMC, ce qui nous a permis d'avoir une architecture plus simple et de gagner en temps d'exécution.

Nous avons mené deux expérimentations distinctes : la première utilise une base de données cardiologique « MIT-BIH » et la deuxième une base de données du cancer du sein « WBCD ».

Les résultats obtenus par notre classifieur optimisé MAPMC ont été comparés avec les résultats du perceptron classique avec rétropropagation, et aussi à d'autres méthodes proposées dans l'état de l'art.

La base données MIT-BIH a été utilisée pour évaluer notre algorithme. Nous avons remarqué que le MAPMC a donné de meilleurs résultats par rapport au perceptron multi couches classique.

Notre classifieur a donné de bons résultats par rapport à quelques récents travaux de l'état de l'art qui ont utilisé la même base de données.

Les réseaux de neurones sont des modèles très robustes et peuvent être appliqués à la résolution de n'importe quel problème, le seul inconvénient c'est qu'ils sont considérés comme des boîtes noires.

Le MAPMC a été validé sur la base de données largement utilisée dans la littérature WBCD pour la classification du cancer du sein. Les résultats présentés sont très satisfaisants et dépassent les résultats obtenus par d'autres méthodes qui ont utilisé la même base de données. Nous avons utilisé la sélection séquentielle pour garder les variables les plus pertinentes pour améliorer le taux de reconnaissance des tumeurs.

Le MAPMC s'est avéré un algorithme qui améliore significativement les performances de classification des cas minoritaires dans une base de données, et améliore le temps d'exécution et de convergence du classifieur neuronal.

Nous pouvons conclure que l'algorithme de métaplasticité artificielle est facile à implémenter et à appliquer dans plusieurs domaines. Il réduit le coût computationnel et le nombre d'itérations durant la phase d'apprentissage, ce qui nous permet d'avoir une bonne généralisation et un apprentissage plus efficace pour obtenir de meilleurs taux de bonne classification.

Même si le MAPMC a obtenu d'excellents résultats, nous restons réalistes que notre algorithme n'est pas une solution définitive aux limitations de l'algorithme de rétropropagation du gradient.

Le MAPMC a démontré une grande capacité de généralisation dans la classification des arythmies cardiaques et du cancer du sein. Les résultats obtenus sont nettement supérieurs aux résultats de l'algorithme de rétropropagation du gradient et à d'autres classifieurs proposés dans la littérature.

Comme perspectives nous allons travailler sur la combinaison des réseaux de neurones avec des modèles qui permettent de donner une interprétabilité des résultats car ce point est très important dans le domaine médical.

Une autre approche est à prévoir, dans le cas de la classification des arythmies cardiaques, c'est d'utiliser d'autres signaux physiologiques avec l'ECG comme la pression artérielle et d'autres sources pour enrichir les données d'entrée ce qui nous permettra aussi de classer plusieurs autres arythmies cardiaques.

Et pour le cas de classification du cancer du sein, utiliser des méthodes de segmentation des images, pour ajouter d'autres variables pertinentes pour améliorer la reconnaissance des tumeurs malignes.

Les résultats obtenus par notre algorithme MAPMC ont démontré qu'il peut être utilisé comme un système fiable d'aide à la détection des données médicales.

Bibliographie

Bibliographie

- [Abo et Sze'03] Abonyi, J., & Szeifert, F. "Supervised fuzzy clustering for the identification of fuzzy classifiers". *Pattern Recognition Letters*, 14(24), 2195–2207. doi:10.1016/S0167-8655(03)00047-3.(2003).
- [Abr'04] Abraham Ajith " Meta-learning evolutionary artificial neural networks ". *Neurocomputing*, Vol. 56 , pp. 1-38. 2004.doi: 10.1016/S0925-2312(03)00369-2.
- [Abr et Bea'96] Abraham W.C et Bear M.F . "Metaplasticity :The plasticity of synaptic plasticity".*Trends in Neurosciences*.Vol 14(9),pp 126-130.1996.doi:10.1016/S0166-2236(96)80018-X.
- [Abr'99] Abraham W.C ."Metaplasticity: Key element in memory and learning?".*News in Physiological Sciences*,Vol 14(2).pp.85.1999.
- [Abr'08] Abraham, W.C." Metaplasticity: tuning synapses and networks for plasticity" *Nature Reviews Neuroscience* 9, 387 (May 2008) | doi:10.1038/nrn2356.
- [Abr et al'02] Wickliffe C. Abraham, Barbara Logan, Jeffrey M. Greenwood, and Michael Dragunow "Induction and Experience-Dependent Consolidation of Stable Long-Term Potentiation Lasting Months in the Hippocampus" *The Journal of Neuroscience*, November 1, 2002, 22(21):9626–9634.
- [Ach et al'04] Acharya R., Kumar A., Bhat P. S., Lim C. M., Iyengar S. S., Kannathal N., et al. (2004). Classification of cardiac abnormalities using heart rate signals. *Med. Biol. Eng. Comp.* 42, 288–293 10.1007/BF02344702 .
- [Aha et Ban'95] Aha, D. W. et Bankert, R. L. "A comparative evaluation of sequential feature selection algorithms". In *5th International Workshop on Artificial Intelligence and Statistics*, pages 1-7. Ft. Lauderdale, USA. (1995).
- [Aka'09] Akay, M. F. "Support vector machines combined with feature selection for breast cancer diagnosis". *Expert Systems with Applications*, 36(2), 3240–3247. doi:10.1016/j.eswa.2008.01.009. (2009).
- [Alb et al'02] Albrecht, A. A., Lappas, G., Vinterbo, S. A., Wong, C. K., & Ohno-Machado, L. "Two applications of the LSA machine". *Proceedings of the 9th international conference on neural information processing*, 184–189. 10.1109/ICONIP.2002.1202156. (2002).
- [Alp et al'95] Alpsan D., Towsey M., Ozdamar Z., Tsoi A.C and Ghista D.N. "Efficacy of Modified Back propagation and Optimisation Methods on a Real world Medical Problem" . *Neural Networks* , Vol .8 (6) , pp. 945-962. 1995.doi : 10.1016/0167-8655 (95) 00124 – Y.
- [Alm et Die'91] Almuallim, H. et Dietterich, T. G." Learning with many irrelevant features". In *Proceedings of the Ninth National Conference on Artificial Intelligence*, pages 547-552. AAAI Press. (1991).

[Alm et Die'92] Almuallim, H. et Dietterich, T. G. "Efficient algorithms for identifying relevant features". In Proceedings of the Ninth Canadian Conference on Artificial Intelligence, pages 38-45. Morgan Kaufmann. (1992).

[Amp et al '99] Ampazisa K., Perantonisa S.J. and Taylor J.G. "Dinamics of multilayer networks in the vicinity of temporary minima". Neural Networks , Vol. 12 (1) , pp. 43-58. 1999.doi: 10.1016/S0893-6080 (98) 00103-8.

[Ana et al'93] Anand R., Mehrotra K.G., Mohan C.K. and Ranka S. " An improved algorithm for neural network classification of imbalanced training sets ". Neural Networks , IEEE Transactions on, Vol. 4(6) , pp. 962-969. 1993.doi: 10.1109/72.286891.

[And et Phan'07] Andina, D. et Pham, D. "Computational Intelligence for Engineering and Manufacturing", Springer-Verlag, The Netherlands. (2007)

[Ash et al'14] Ashraf Osman I ; Shamsuddin, Siti Mariyam ; Ahmad, Nor Bahiah ; Salleh, Mohd Najib Mohd. "Hybrid NSGA-II of Three-Term Backpropagation network for multiclass classification problems". International Conference on Computer and Information Sciences (ICCOINS), 3-5 June 2014 Page(s):1 - 6 .Print ISBN:978-1-4799-4391-3 Conference Location :Kuala Lumpur, Malaysia .DOI:10.1109/ICCOINS.2014.6868364 Publisher: IEEE

[Bai et al'09] Bai Y., Zhang H. and Hao Y. " The performance of the back propagation algorithm with varying slope of the activation function " Chaos , Solitons & Fractals , Vol. 40(1) , pp. 69 – 77. 2009. doi:10.1016/j.chaos.2007.07.033.

[Beh et al'06] Behera L., Kumar S. and Patnaik A. " On Adaptive Learning Rate That Guarantees Convergence in Feedforward Networks ". Neural Networks , IEEE Transactions on , vol. 17(5) , pp. 1116-1125. 2006.doi:0.1109/TNN. 2006.878121.

[Ben et Chi'09] Benchaib, Y. and Chikh, M. (2009) 'A specialized learning for neural classification of cardiac arrhythmias', Journal of Theoretical and Applied Information Technology, Vol. 6, No. 1, pp.81–89.

[Bi et Poo'98] Bi GQ, Poo MM "Synaptic modifications in cultured hippocampal neurons: dependence on spike-timing, synaptic strength, and postsynaptic cell type". J Neurosci 18:10464–10472. (1998)

[Bien et al'82] Bienenstock E.L, Cooper L.N et Munro P.W , "Theory of the development of neuron selectivity :orientation specificity and binocular interaction in visual cortex". Journal of Neuroscience. Vol2(1),pp. 32-48.1982.

[Bis'06] Bishop C.M "Pattern recognition and machine learning(Information Science and Statistics)" . Springer-Verlag New York Inc.2006 . ISBN:0387310738 .

[Bli et Col'93] Bliss TV, Collingridge GL "A synaptic model of memory: long-term potentiation in the hippocampus". Nature 361:31-39.1993

[Bli et Lom'73]. Bliss TV, Lomo T "Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path". Journal of Physiology. 1973 Jul;232(2):331-56.

- [Blu et Riv'92] Blum, A. et Rivest, R. "Training a 3-node neural network is NP-complete". *Neural Networks*, 5:117–127, 1992.
- [Bra et al 2008] Bratu C.V, Muersan T et Potolea R. « Improving classification accuracy through feature selection ».In *proc.ICCP 2008, 4th international conference on intelligent computer communication and processing*.(Cluj, Napoca, Romania, 2008).25-32.10.1109/ICCP.2008.4648350.
- [Cai et al'99] Caillard O, Ben-Ari Y, Gaiarsa JL"Mechanisms of induction and expression of long-term depression at GABAergic synapses in the neonatal rat hippocampus." *J Neurosci*; 19: 7568-77 (PMID 10460263).(1999).
- [Caj'94] Cajal S.R.Y « la fine structure des centres nerveux ».Proceedings of Royal Society Vol55.pp 444-468. 1894 .
- [Che et Jai'94] Chen D.S et Jain R.C . “ A robust back propagation learning algorithm for function approximation ” . *Neural Networks , IEEE Transactions on , Vol . 5 (3) , pp . 467 – 479. 1994*.doi : 10.1109/72.286917 .
- [Che et Par'09] Cheng H.L and Park S.C. “ An efficient document classification model using an improved back propagation neural network and singular value decomposition ” *Expert Systems with Applications , Vol. 36(2) , pp. 3208 – 3215. 2009*. doi:10.1016/j.eswa.2008.01.014.
- [Che et Shi'05] Chen SM et Shie J.D “A new method for feature Subset Selection for handling classification problems”.In *Proc .FUZZ'05.14TH IEEE international conference on fuzzy systems*.(Reno, Nevada,USA,2005).183-188.doi:10.1109/FUZZY.2005.1452390.
- [Cho e Cho'99] Cho S.Y. and Chow T.W. “Training multilayer neural networks using fast global learning algorithm – least –squares and penalized optimization methods ” . *Neuro-computing , Vol . 25(1) , pp. 115-131. 1999*.doi: 10.1016/S0925-2312(99)0005-7.
- [Con et Gui'10] Conforti D and Guido R “Kernel based support vector machine via semidefinite programming : application to medical diagnosis”, *computers & operations Research*.Vol 37 ;pp1389-1394.2010.doi:10.1016/j.cor.2009.02.018.
- [Cow et al'98] Cowan RL, Wilson CJ “Spontaneous firing patterns and axonal projections of single corticostriatal neurons in the rat medial agranular cortex”. *J Neurophysiol* 71:17-32.1994.
- [Cun'85] le Cun Y. « Une procédure d'apprentissage pour réseau à seuil assymétrique. In *Cognitiva » A la Frontière de l'Intelligence Artificielle des Sciences de la Connaissance des Neurosciences*, pages599-604, Paris, 1985. (Paris 1985), CESTA.
- [D'an et Ros'98] D'Angelo,E., Rossi,P. “Integrated regulation of signal coding and plasticity by NMDA receptors at a central synapse”. *Neural plasticity*, 1: 71-79. (1998)
- [Dan et Poo'04] Dan Y, Poo MM “Spike timing-dependent plasticity of neural circuits”. *Neuron* 44:23-30. (2004).

- [Dan et Poo'06] Dan Y, Poo MM "Spike timing-dependent plasticity: from synapse to perception". *Physiol Rev* 86:1033-1048. (2006)
- [Deb et al'94] Debanne D, Gähwiler BH, Thompson SM "Asynchronous pre- and postsynaptic activity induces associative long-term depression in area CA1 of the rat hippocampus in vitro". *Proc Natl Acad Sci U S A* 91:1148-1152. (1994)
- [Deb et al'98] Debanne D, Gähwiler BH, Thompson SM "Long-term synaptic plasticity between pairs of individual CA3 pyramidal cells in rat hippocampal slice cultures". *J Physiol* 507 (Pt 1):237-247. (1998)
- [Des'05] Desbois.D , « une introduction au positionnement multidimensionnel », revue modulad numero-32(2005).
- [Dre et al'02] G Dreyfus, JM Martinez, M Samuelides, MB Gordon, F Badran, S Thiria, L.Hérault."Réseaux de neurones-Méthodologie et applications".Eyrolles .2002.Paris
- [Dre et al'02]). Dreyfuss G, Kim VN, Kataoka N "Messenger-RNA-binding proteins and the messages they carry". *Nat Rev Mol Cell Bio* 3: 195–205(2002)
- [Dud et Bea'92] Dudek SM, Bear MF "Homosynaptic long-term depression in area CA1 of hippocampus and effects of N-methyl-D-aspartate receptor blockade". *Proc Natl Acad Sci U S A* 89:4363-4367. (1992)
- [Dud et al'00], Duda, R. O., Hart, P. E. et Stork, D. G. « Pattern Classification » (2nd Edition). Wiley-Interscience. (2000).
- [Eom et al] Eom K., Jung K. and Sirisena H. "Performance improvement of back propagation algorithm by automatic activation function gain tuning using fuzzy logic ". *Neurocomputing* , Vol. 50, pp. 439-460. 2003.doi: 10.1016/S0925-2312(02)00576-3.
- [Faw'06] Fawcett, T. (2006). "An introduction to roc analysis". *Pattern recognition Letters*, 27 :861–874.
- [Fis'36] Fisher, R. A. "The use of multiple measurements in taxonomic problems". *Annals Eugen.*, 7:179-188. (1936).
- [Fra et Swi'86] Fraser, A. M. et Swinney, H. L."Independent coordinates for strange attractors from mutual information" . *Physical Review A*, 33, pp. 1134–1140. (1986).
- [Fu et Hu'96]Fu L., Hsu H.H. and Principe J.C. " Incremental back propagation learning net-works" *IEEE Transactions on Neural Networks* , Vol. 7 (3) , pp. 757-761. 1996 .doi:10.1109/72.5017321.
- [Fur et al'00] Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D.W., Schummer, M. et Haussler,D. "Support vector machine classification and validation of cancer tissue samples using microarray expression data". *Bioinformatics*, 16(10):906-914. (2000).

- [Gao et al'05] Gao, D., Madden, M., Chambers, D. and Lyons, G. "A Bayesian ANN classifier for ECG arrhythmia diagnostic system: a comparison study", Proceedings of International Joint Conference on Neural Networks, 31 July–4August, Montreal, Canada, pp. 2383–2388. (2005)
- [Gol'91] Goldberg, D.E « Algorithmes Génétiques ». Addison Wesley. (1991).
- [Got et al'11]Gothwal, H., Kedawat, S. and Kumar, R. 'Cardiac arrhythmias detection in an ECG beat signal using fast Fourier transform and artificial neural network', Journal of Biomedical Science and Engineering, Vol. 4, pp.289–296. (2011)
- [Gui et al'07] Guijarro-Berdias, B., Fontenla-Romero, O., Perez-Sanchez, B., & Fraguera, P. "A linear learning method for multilayer perceptrons using leastsquares". Lecture Notes in Computer Science, 365–374. 10.1007/978-3-540-77226-238. (2007).
- [Gus et al'87] Gustafsson B, Wigstrom H, Abraham WC, Huang YY (1987) "Long-term potentiation in the hippocampus using depolarizing current pulses as the conditioning stimulus to single volley synaptic potentials". J Neurosci 7:774-780.
- [Guy et Eli'03] : Guyon, I. et Elisseeff, A. "An introduction to variable and feature selection". J. Mach. Learn. Res., 3:1157-1182. (2003).
- [Ham et al'96]Howard J Hamilton , Ning Shan, Nick Cercone RIAC: A Rule Induction Algorithm Based on Approximate Classification. University of Regina .1996.
- [Han et McN'82] Hanley, J. and McNeil, B. "The meaning and use of the area under a receiver operating characteristic (roc). Rdiology, 43 :29–36. (1982).
- [Har'80]Harvard-MIT Division of Health Science Technology, Biomedical Health Centre MIT-BIH Arrhythmia Database, 1975–1979, 1st ed., Harvard-MIT Division of Health Science Technology, Biomedical Health Centre, Cambridge, MA, USA. (1980)
- [Heb'49]Hebb D. The organization of the behavior . Laurence Erlbaum Associates, New Jersey 2002.(reedition of the 1949 original) 1949.ISBN 0805843000, 978080584300.
- [Hoo et al'02] Hoo K.A., Sinzinger ED. And Piovoso M.J. "Improvements in the predictive capability of neural networks " . Journal of Process Control , Vol. 12 (1) , pp. 193-202.2002.doi: 10.1016/S0959-1524 (01) 00007-5.
- [Hot'33] Hotelling .H "Analysis of a complex of statistical variables into principal components". Journal of Educational Psychology, Vol 24(6), Sep 1933, 417-441.
- [Hua et al'08] Huang, C.-J., Yang, D.-X. et Chuang, Y.-T. "Application of wrapper approach and composite classifier to the stock trend prediction". Expert Syst. Appl., 34:2870-2878. (2008).
- [Ish et Nak'00]). Ishibuchi, H. et Nakashima, T."Multi-objective pattern and feature selection by a genetic algorithm". In Proc. of Genetic and Evolutionary Computation Conference (GECCO'2000, pages 1069-1076. Morgan Kaufmann.

- [Ito et al'82] Ito M, Sakurai M, Tongroach P (1982) Climbing fibre induced depression of both mossy fibre responsiveness and glutamate sensitivity of cerebellar Purkinje cells. *J Physiol* 324:113-134.
- [Jad et al'12]Jadhav, S.M, Nalbalwar, S.L. and Ghatol, A.A. (2012) 'Artificial neural network models based cardiac arrhythmia disease diagnosis from ECG signal data', *International Journal of Computer Applications*, Vol. 44, No. 15, pp.8–13.
- [Jai et Zon'97] Jain, A. et Zongker, D. "Feature selection : Evaluation, application, and small sample performance". *IEEE transactions on pattern analysis and machine intelligence*, 19:153-158. (1997).
- [Jan et Gee'13] D. Janaki Sathya, K. Geetha. "Mass classification in breast DCE-MR images using an artificial neural network trained via a bee colony optimization algorithm". *ScienceAsia* 39 (2013): 294–306.doi: 10.2306/scienceasia1513-1874.2013.39.294
- [Jed'02] Jedlicka, P. "Synaptic plasticity, metaplasticity and BCM theory", *Bratislava Medical Journal*, Vol. 103, No. 4, pp.137–143. (2002).
- [Joh'97] John, G. H. (1997). "Enhancements to the data mining process". Thèse de doctorat, Stanford,CA, USA. UMI Order No. GAX97-23376.
- [Joh et al'94] John, G. H., Kohavi, R. et Pflieger, K. "Irrelevant features and the subset selection problem". In *machine learning : proceedings of the eleventh international*, pages 121-129. Morgan Kaufmann.(1994).
- [Kac et al'10] Kachouri, R., Djemal, K. et Maaref, H. « Adaptive feature selection for heterogeneous image databases ». *Second IEEE International Conference on Image Processing Theory, Tools 38 ; Applications*, 10, Paris, France. (2010).
- [Kam et Pit'99] Kamarthi S.V and Pittne S."Accelerating neural networks training using weight extrapolations ". *Neural Networks* , Vol . 12(9) , pp. 1285-1299. 1999.doi:10.1016/S0893-6080(99)00072-6.
- [Kam'99] Kamarthi M. "Modelling with constructive back propagation" *Neural Networks* , Vol. 12 (4) , pp. 707-716 . 1999. doi:10.1016/S0893-6080(99)00018-0.
- [Kam'12] Kamath, C. (2012) 'ECG classification using morphological features derived from symbolic dynamics', *International Journal of Biomedical Engineering and Technology*, Vol. 9, No. 4,pp.325–336.
- [Kar et Cev'09] Karabatak, M., & Cevdet-Ince, M. "An expert system for detection of breast cancer based on association rules and neural network". *Expert Systems with Applications*, 36, 3465–3469. doi:10.1016/j.eswa.2008.02.064. (2009).
- [Kar et Ven'92] Karayiannis N.B. and Venetsanopoulos A. N. "Fast learning algorithms for neural networks". *Circuits and Systems II : Analog and Digital Signal Processing*, *IEE Transactions on*, Vol. 39(7) , pp. 453-474. 1992.doi: 10.1109/82.160170.
- [Kat et al'09] Kathirvalavakumar T. and Jeyaseeli – Subavathi S. "Neighborhood based modified back propagation algorithm using adaptive learning parameters for training feed forward neural networks ". *Neurocomputing*, Vol. 72 , pp. 3915 – 3921 . 2009 .doi:10.1016/j.neucom.2009.04.010.

[Kel et al'86]; Kelso SR, Ganong AH, Brown TH Hebbian synapses in hippocampus. Proc Natl Acad Sci USA 83:5326-5330. (1986).

[Kha'08] Khashman A. "A Modified back propagation Learning Algorithm With added Emotional Coefficients". Neural Networks, IEEE Transactions on, Vol. 19 (11), pp. 1896-1909.2008. doi: 10.1109/TNN. 2008. 2002913.

[Kim et Ra'91] Kim Y. K. et Ra J.B. "Weight value initialization for improving training speed in the back propagation network". In Proc. IEEE International Joint Conference on Neural Networks, IJCNN' 91, Vol. 3, pp. 2396-2401. 1991. doi: 110.1109/IJCNN.1991.170747.

[Kin et al'07] Kinto, E., Del-Moral-Hernandez, E., Marcano-Cedeño, A., Roper-Pelaez, J. "A preliminary neural model for movement direction recognition based on biologically plausible plasticity rules", (IWINAC 2007), 18–21, June, La Manga del Mar Menor, Spain, LNCS, Vol. 4528, pp.628–636.

[Kir et Ren'92] Kira, K. et Rendell, L. A. "The feature selection problem : Traditional methods and a new algorithm". In AAAI, pages 129-134, Cambridge, MA, USA. AAAI Press and MIT Press. (1992)

[Kit'78] Kittler, J. "Feature set search algorithms". Pattern Recognition and Signal Processing, pages 41-60. (1978).

[Koc et Can'11] Sabri Koçer, M. Rahmi Canal. « Classifying Epilepsy Diseases Using Artificial Neural Networks and Genetic Algorithm». Journal of Medical Systems August 2011, Volume 35, Issue 4, pp 489-498.

[Koh et Joh'97] Kohavi, R. et John, G. H. "Wrappers for feature subset selection". Artif. Intell.,97:273-324. (1997).

[Kol et Sah'96], Koller, D. et Sahami, M. "Toward optimal feature selection". In Machine Learning: Proceedings of the Thirteenth International Conference. Morgan Kaufmann .pages 284-292.(1996).

[Kon'48] Konorski J. « conditioned reflexes and neuron organization » Cambridge , UK:Cambridge Univ,press .1948.

[Kun et Jai'99], Kuncheva, L. I. et Jain, L. C."Nearest neighbor classifier : Simultaneous editing and feature selection". Pattern Recognition Letters 20 (1999) 1 149-1 156. 9.

[Lam et al'10] Karri P. Lamsa, Dimitri M. Kullmann, et Melanie A. Woodin." Spike-Timing Dependent Plasticity in Inhibitory Circuits". Front Synaptic Neurosci. 2010; 2: 8.

[Lan et al'96]. K Langley, DJ Fleet, PB Hibbard "Linear filtering precedes nonlinear processing in early vision"Current Biology 6 (7), 891-896.1996.

[Las et al'01] Mark Last , Abraham Kandel , Oded Maimon"Information-theoretic algorithm for feature selection". Pattern Recognition Letters Volume 22, Issues 6–7, May 2001, Pages 799–811.

- [Lee et al'01] Lee H.M., Chen C.M. and Huang T.C. “ Learning efficiency improvement of back – propagation algorithm by error saturation prevention method ‘ ’ . Neuro – computing , Vol . 41, pp. 125-143.2001.doi : 10.1016/S0925-2312(00)00352-0.
- [Lee et Wei'91] Lee M. et Weidman S.T. “ Expert systems for guiding back propagation training of layered perceptrons ‘ ’ . Expert systems with Applications , Vol. 2 (1) , pp .73-81 . 1991 . doi: 10.1016 / 0957 – 4174 (91) 90135-2 .
- [Leh'99] Lehtokangas.M. “Modelling with constructive backpropagation” Neural Networks. Volume 12, Issues 4–5, June 1999, Pages 707–716.
- [Leo et Kra'90] Leonard J. et Kramer M.A. “Improvement of the back propagation algorithm for training neural networks” . Computers & Chemical Engineering , Vol. 14 (3) , pp.337 – 341. 1990 .doi:10. 1016 / 0098-1354 (90) 87070-6 .
- [Li et Guo'08] Li, Y. et Guo, L.” Tcm-knn scheme for network anomaly detection using featurebased optimizations” . In Proceedings of the 2008 ACM symposium on Applied computing, SAC '08, pages 2103-2109, New York, NY, USA. ACM. (2008).
- [Lin et Con'95] Linden DJ, Connor JA “Long-term synaptic depression”. *Annu Rev Neurosci* 18:319-357. (1995)
- [Lis et Spr'05] Lisman J, Spruston N “Postsynaptic depolarization requirements for LTP and LTD: a critique of spike timing-dependent plasticity”. *Nat Neurosci* 8:839-841. (2005)
- [Liu et al'02] Liu, H., Motoda, H. et Yu, L. ”Feature selection with selective sampling”. In Proceedings of the Nineteenth International Conference on Machine Learning, ICML '02, pages 395-402, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.(2002).
- [Liu et Set'96] Liu, H. et Setiono, R. “Feature selection and classification - a probabilistic wrapper approach”. In Proceedings of the 9th International Conference on Industrial and Engineering Applications of AI and ES, pages 419-424. (1996).
- [Liu et Set'98] Liu, H. et Setiono, R. ” Incremental feature selection”. *Applied Intelligence*, 9:217-230. (1998).
- [Liu et Yu'05] Liu, H. et Yu, L. “Toward integrating feature selection algorithms for classification and clustering”. *IEEE transactions on knowledge and data engineering*, 17:491-502., (2005).
- [Lyn'04] Lynch MA “Long-term potentiation and memory”. *Physiol Rev* 84:87-136. (2004).
- [Mac et McB'96] Maccaferri, G. et McBain, C.J. “Long-term potentiation in distinct subtypes of hippocampal nonpyramidal neurons”. *J. Neurosci.*, 16: 5334–5343. (1996).
- [Mac et Men'04]. Mackay, J. et Mensah, G. “The Atlas of Heart Disease and Stroke. Organisation mondiale de la santé“ . (2004).
- [Mac et al'98] Maccaferri G., Toth K., McBain C. J. Target-specific expression of presynaptic mossy fiber plasticity. *Science* 279, 1368 1370.10.1126/ science.279.5355.1368.(1998).

- [Mag et al'97] Magoulas G.D., Vrahatis M.N. and G.S. "Effective Back propagation Training With Variable Step Size Neural Networks". *Neural Networks*, Vol. 10 (1), pp. 69 – 82. 1997. doi: 10.1016/S0893-6080(96)00052-4.
- [Mal et Nic'99]; Malenka RC, Nicoll RA "Long-term potentiation--a decade of progress?" *Science* 285:1870-1874.(1999).
- [Mal et al'14] Maliarsky Elina, Avigal Mireille and Herman Maya, "A neuro genetic system for cardiac arrhythmia classification" *Machine Learning and Data Mining in Pattern Recognition. 10th International Conference, MLDM 2014, St. Petersburg, Russia, July 21-24, 2014, Proceedings*. pp343-360. ISBN 978-3-319-08979-9.
- [Man'02] Mandische M. "A comparison of evolution strategies and back propagation for neural network training". *Neurocomputing*, Vol. 42, pp. 87-117. 2002. doi:10.1016/S0925-2312(01)00596-3.
- [Mar et al'11] Marcano-Cedeño, A., Quintanilla-Dominguez, J. et Andina, D. "Breast cancer classification applying artificial metaplasticity algorithm", *Neurocomputing*, Vol. 74, No. 8, pp.1243–1250. (2011).
- [Mar et Gre'63] Marill, T. et Green, D. M. "On the effectiveness of receptors in recognition systems". *IEEE transactions on Information Theory*, 9:11-17. (1963).
- [Mar et al'00] Martin S.J, Grimwood P.D, Morris R.G.: "Synaptic plasticity and memory: an evaluation of the hypothesis" *Annual Review of Neuroscience*, Vol 23, pp.649-711. 2000. doi:10.1146/annurev.neuro.23.1.649.
- [Mas et al'14] Masood Arbab, Khan Gul Muhammad, Sahibzada Ali Mahmud "Cardiac Arrhythmia Classification Using Cartesian Genetic Programming Evolved Artificial Neural Networks". *Experimental and Clinical Cardiology/Cardiology Academic Press* Septembre 2014.
- [McB et Mac'97] McBain C.J et Maccaferri, G; "Synaptic plasticity in hippocampal interneurons? A commentary". *Canadian journal of Physiology Pharmacology*, 75, 488-494. (1997).
- [Mes et Abe'56] Messick, S. et Abelson, R. "The additive constant problem in multidimensional scaling". *Psychometrika*, 21:1-15. (1956).
- [Met'78] Metz, C. "Basic principles of roc analysis". In *Seminars in Nuclear Medicine*, volume 3. (1978).
- [Mir et al'12] SeyedAli Mirjalili, Siti Zaiton Mohd Hashim, Hossein Moradian Sardroudi. "Training feedforward neural networks using hybrid particle swarm optimization and gravitational search algorithm". *Applied Mathematics and Computation*; Volume 218, Issue 22, 15 July 2012, Pages 11125–11137.
- [Moh et Pav'04] Mohamed-Taghi V.B and Pavesié N. "Training RBF networks with selective back propagation". *Neurocomputing*, Vol. 62, pp. 39-64. 2004. doi: 10.1016/j.neucom.2003.11.011.

- [Mon et al'08] Monteiro J.L, Lobo-Netto .M ,Andina D, Pelez J.R. "Using neural networks to simulate the Alzheimer's Disease" .in: Proc:WAC 2008,World Automation Congress. (Hawaii,HI,USA,2008) pp.1-6.2008.ISBN:978-1-889335-38-4. INSPEC Accession Number: 10411864.
- [Moo et Mar'01]Moody, G.B. and Mark, R.G. 'The impact of the MIT-BIH Arrhythmia Database', *IEEE Engineering in Medicine and Biology Magazine*, Vol. 20, No. 3, pp.45–50. (2001)
- [Mor et al'86]Morris RG, Anderson E, Lynch GS, Baudry M (1986) Selective impairment of learning and blockade of long-term potentiation by an N-methyl-D-aspartate receptor antagonist, AP5. *Nature* 319:774-776.
- [Mou'12] <http://www.algerie1.com/actualite/445-des-deces-en-algerie-sont-dus-aux-maladies-cardiaques/>(2012).
- [Mun et al'06] Muni D.P ,Pal N.R et Das J "Genetic Programming for simultaneous feature selection and classifier design" *Systems ,Man, and Cybernetics.Part B:Cybernetics ,IEEE Transactions on* Vol 36(1) pp100-1103.2006.doi:10.1109/TSMCB.2005.854499.
- [Nau et Kru'99] . Nauck, D., & Kruse, R. (1999). "Obtaining interpretable fuzzy classification rules from medical data". *Artificial Intelligence in Medicine*, 16, 149–169. doi:10.1016/S0933-3657(98)00070-0.1999.
- [Nev et al'08] Neves G, Cooke S.F ,Bliss T.V: "Synaptic plasticity ,memory and the hippocampus : a neural network approach to causality " .*Nature Rev.Neurosci*.Vol.9,pp .65-75.2008.doi:10.1146/annurev.neuro.23.1.649.
- [Oso et Lin'01] Osowski, S. and Lin, T.H. (2001) 'ECG beat recognition using fuzzy hybrid neural network', *IEEE Transactions on Biomedical Engineering*, Vol. 48, No. 11, pp.1265–1271.
- [Oua et Lac'95] Ouardouz M et Lacaille J-C . »Mechanisms of selective long term potentiation of excitatory synapses in stratum oriens/alveus interneurons of rat hippocampal slices". *J Neurophysiol* 73, 810–819. (1995).
- [Owi et al'02] Owis, M.I., Youssef, A.B.M. and Kadah, Y.M. 'Characterization of ECG signals based on blind source separation', *Medical and Biological Engineering and Computing*, Vol. 40, No. 5,pp.557–564. (2002)
- [Pea'04] Pearce, J. M. (2004). "Sir Charles Scott Sherrington (1857-1952) and the synapse". *Journal of neurology, neurosurgery, and psychiatry* 75 (4): 544. doi:10.1136/jnnp.2003.017921 .
- [Pen et Sip'99] Pena-Reyes, C. A., & Sipper, M. (1999). "A fuzzy-genetic approach to breast cancer Diagnosis". *Artificial Intelligence in Medicine*, 17, 131–155. doi:10.1016/S0933-3657(99)00019-6.
- [Pen et al'05] Peng, H., Long, F. et Ding, C."Feature selection based on mutual information : criteria of max-dependency, max-relevance, and min-redundancy" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1226-1238. (2005).

- [Pen et al'09] Peng, L., Yang, B., & Jiang, J. "A novel feature selection approach for biomedical data classification". *Journal of Biomedical Informatics*, 179(1), 809–819. doi:10.1016/j.jbi.2009.07.008. (2009).
- [Per et al'05] Pernia-Espinoza A.V., Ordieres-Meré J.B., Martínez –de – Pison F.J. and Gonzalez – Marcos A . "TAO-robust back propagation learning algorithm ". *Neural Networks* , Vol. 18(2) , pp . 191-204. 2005. doi: 10.1016/ j.neunet.2004.11.007.
- [Pol et Gun'07] Polat, K., & Günes, S. "Breast cancer diagnosis using least square support vector machine". *Digital Signal Processing*, 17(4), 694–701. doi:10.1016/j.dsp.2006.10.008. (2007).
- [Pol et al'06] Polat, K., Sahan, S. and Gnes, S. 'A new method to medical diagnosis: artificial immune recognition system (AIRS) with fuzzy weighted pre-processing and application to ECG arrhythmia', *Expert Systems with Applications*, Vol. 31, No. 2, pp.264–269. (2006).
- [Pra et Sah'03] Prasad, G.K. and Sahambi, J.S. (2003) 'Classification of ECG arrhythmias using multiresolution analysis and neural networks', *Proceedings of TENCON 2003 IEEE Conference on Convergent Technologies*, Vol. 1, 15–17 October, Bangalore, India, pp.227–231.
- [Pud et al'94] Pudil, P., Novovicova, J. et Kittler, J. "Floating search methods in feature selection". *Pattern Recogn. Lett.*, 15:1119-1125. (1994).
- [Qui'96] J R Quinlan "improved use of continuous attributes in C4.5" *Journal of Artificial Intelligence Research* . Volume 4 Issue 1, Jnauary 1996 .Pages 77-90.
- [Reb'05] Rebaine, D." Méthode de branch and bound. Cours de conception et analyse des algorithmes". Université du Quebec à Chicoutimi. 2005.
- [Reb et al'11] Nelson Rebola, Christophe Blanchet, Mario Carta, Frédéric Lanore, Christophe Mulle,"NMDA receptor–dependent metaplasticity at hippocampal mossy fiber synapses", *Nature Neuroscience* 14(6):691-693, Published online May 1, 2011, doi:10.1038/nn.2809.
- [Ric'38] Richardson, M. W. « Multidimensional psychophysics ». *Psychol. Bull.* 35: pp. 659-659(1938).
- [Rie'94] Riedmiller M. " Advanced supervised learning in multi-layer perceptrons – From back propagation to adaptive learning algorithms" . *Computer Standards & Interfaces* , Vol . 16 (3) , pp . 265-278. 1994 . doi:10.1016/0920-5489 (94) 90017-5.
- [Rop et Piq'07] Ropero-Pelaez, J. et Piqueira, J.R. "Biological clues for up-to-date artificial neurons", in *Computational intelligence for engineering and manufacturing*, Andina D and Pham D.T; Springer-Verlag; Nederlands 2007.
- [Ros'58] Rosenblatt F. "The perceptron : A probabilistic model for information storage and organization in the brain". *Psychological Review*, Vol .65(6), pp.386-408.1958.

- [Row et Sau'00]. Roweis, S. T. et Saul, L. K. "Nonlinear dimensionality reduction by locally linear embedding". *SCIENCE*, 290:2323-2326. (2000).
- [Sca et Tep'92] Scalero R.S and Tepedelenlioglu N. "A fast new algorithm for training feed forward Neural Networks ". *Signal processing , IEEE Transactions on , Vol. 40 (1) , pp. 202-210. 1992*.doi: 10.1109/78.157194.
- [Set'00] Setiono, R. "Generating concise and accurate classification rules for breast cancer diagnosis". *Artificial Intelligence in Medicine*, 18(3), 205–217. doi:10.1016/S0933-3657(99)00041-X. (2000).
- [Sex et al'98 a] Sexton R.S. , Dorsey R.E. and Johnson J.D. "Toward global optimization of neural networks : A comparison of the genetic algorithm and back propagation " *Decision Support Systems*, Vol 22(2).pp.171-185.1998.doi:10.1016/S0167-9236(97)00040-7.
- [Sex et al'98 b]Sexton R.S, Alidaee B,Dorsey R.E and Johnson J.D."Global Optimization for artificial neural networks: A tabu search application". *European Journal of Operational Research*,Vol.160(2),pp.570-584.1998.doi:10.1016/S0377-2217(97)00292-0.
- [Sha'48]. Shannon, C.E. "A mathematical theory of communication", *The Bell System Technical Journal*, Vol. 27, pp.379–423.doi:10.1145/584091.584093. (1948)
- [Sik et al'13] Sikkandar, M.Y., Akshayaa, V., Acharya, D.D. and Sree, L.D. (2013) 'Analysis of cardiac abnormalities using Hilbert-Huang transform', *International Journal of Biomedical Engineering and Technology*, Vol. 13, No. 1, pp.69–86.
- [Sil et al'08] Silva L.M., Marques de Sa J. and Alexandre L.A. "Data classification with multilayer perceptrons using a generalized error function ". *Neural Networks , Vol. 21(9) , pp. 1302 – 1310. 2008*. doi:10.1016/j.neunet.2008.04.004.
- [Sjo et al'01] Sjostrom PJ, Turrigiano GG, Nelson SB "Rate, timing, and cooperativity jointly determine cortical synaptic plasticity". *Neuron* 32:1149-1164. (2001).
- [Sk1'00] Sklansky, M. "Comparison of algorithms that the select features for pattern classifiers". In *SSPR/SPR*. (2000).
- [Sol et Moh'08] Soliman M.I., and Mohamed S.A. "A highly efficient implementation of a back propagation learning algorithm using matrix ISA . *Journal of Parallel and Distributed Computing ,Vol. 68(7) , pp. 949 – 961. 2008*. doi:10.1016/j.jpdc.2007.12.004.
- [Sol et Van'96] Solomon R. and Van –Hemmen J.L. " Accelerating back propagation through dynamic self – adaptation " . *Neural Networks , Vol . 9 (4) , pp. 589- 601. 1996*.doi: 10.1016 / 0893- 6080 (95) 00144-1 .
- [Sou et Deb'99], Sourdnet V, Debanne D."The role of dendritic filtering in associative long-term synaptic plasticity". *Learn Mem* 6:422–447. (1999)
- [Ste'06] Steil J.J. "Online stability of back propagation –decor relation recurrent learning " *Neurocomputing*, Vol. 69 (7-9) , pp. 642 – 650 . 2006 .doi:10.1016/ j.neucom.2005.12.012.

- [Sten'73] Stent GS .”A physiological mechanism for Hebb's postulate of learning”. Proc Natl Acad Sci U S A. 1973 Apr;70(4):997-1001.
- [Ste et Dob'96] Ster, B., & Dobnikar, A. ”Neural networks in medical diagnosis: Comparison with other methods”. In Proceedings of the international conference on engineering applications of neural networks (pp. 427–430). (1996).
- [Sun'10] Sun J. “ Local coupled feed forward neural network “ Neural Networks , Vol . 23 (1), pp. 108 – 113. 2010. doi:10.1016/j.neunet.2009.06.016.
- [Ten et al'00] Tenenbaum, J. B., Silva, V. d. et Langford, J. C.” A global geometric framework for nonlinear dimensionality reduction”. Science, 290(5500):2319-2323. (2000).
- [Theo et Kou'03] Theodoridis S. and Koutroumbas K. “Pattern Recognition”. Second Edition. Elsevier Academic Press, 2003.
- [Ube'07]Übeyli, E. D. “Implementing automated diagnostic systems for breast cancer Detection”. Expert Systems with Applications, 33(4), 10541062. doi:10.1016/j.eswa.2008.02.064. (2007).
- [Wan et al'07] Wang C.H. , Kao C.H. and Lee W.H. “ A new interactive model for improving the learning performance of back propagation neural network ‘. Automation in Construction , Vol. 16(6) , pp. 745 – 758. 2007. doi:10.1016/j.autcon.2006.12.007.
- [Wan et al'04] Wang X.G., Tang Z., Tamura H., Ishii M and Sun W.D. “ An improved back propagation algorithm to avoid the local minima problem” . Neurocomputing, Vol. 56 , pp. 455-460. 2004.doi: 10.1016/j.neucom.2003.08.006.
- [Whi'71] Whitney, A. W. “A direct method of nonparametric measurement selection”. IEEE Trans. Comput., 20:1100-1103. (1971).
- [Win et Swe'01], Winder DG, Sweatt JD. “Roles of serine/threonine phosphatases in hippocampal synaptic plasticity”. Nat Rev Neurosci 2001 Sep;2(9):670.
- [Wol et al'94]Wolberg, W.N. Street, and O.L. Mangasarian. “Machine learning techniques to diagnose breast cancer from fine-needle aspirates”. *Cancer Letters* Vol. 77, pages 163-171, 1994.
- [Woo et al'03] Woodin M. A., Ganguly K., Poo M. M. “Coincident pre- and postsynaptic activity modifies GABAergic synapses by postsynaptic changes in Cl⁻ transporter activity”. Neuron 39, 807–820. doi:10.1016/S0896-6273(03)00507-5.(2003).
- [WHO'11] WHO. Global Atlas on Cardiovascular Disease Prevention and Control, World Health Organization, Geneva.
- [Yam et al'97] Yam Y.F., Chow T.W. and Leung C.T. “ A new method in determining initial weights of feedforward neural networks for training enhancement ‘. Neurocomputing , Vol. 16 (1) , pp. 23-32 . 1997 .doi:10.1016/S0925-2312(96) 00058-6.

[Yam et Cho'00] Yam Y.F. and Chow T.W. "A weight initialization method for improving training speed in feedforward neural network". *Neurocomputing*, Vol. 30 (1), pp. 219-232. 2000. doi: 10.1016/S0925-2312(99)00127-7.

[Yan et al'08] Yang S.S., Siu S. and Ho C.L. "Analysis of the Initial Values in Split Complex Back propagation Algorithm". *Neural Networks, IEEE Transactions on*, Vol. 19 (9), pp. 1564-1573. 2008. doi: 10.1109/TNN. 2008. 2000805.

[Yan et Sur'96]. Yan Z, Surmeier DJ "Muscarinic (m2/m4) receptors reduce N- and P-type Ca²⁺ currents in rat neostriatal cholinergic interneurons through a fast, membrane-delimited, Gprotein pathway. *J Neurosci* 16:2592-2604. (1996)

[Yu et Cho] Yu, S.N. and Chou, K.T. 'Integration of independent component analysis and neural networks for ECG beat classification', *Expert Systems with Applications*, Vol. 34, No. 4, pp.2841–2846. (2008)

[Yu et Liu'04]. Lei Yu and Huan Liu. "Efficient Feature Selection via Analysis of Relevance and Redundancy". *J. Mach. Learn. Res.* 5 (December 2004), 1205-1224

[Zha et al'07] Zhang C., Wu W., Chen X.H. and Xiong Y. "Convergence of BP algorithm for product unit neural network with exponential weight". *Neurocomputing*, Vol. 72 (1), pp. 513 – 520. 2008. doi:10.1016/j.neucom.2007.12.004.

[Zuc et Reg'02]. Zucker RS, Regehr WG "Short-term synaptic plasticity". *Annu Rev Physiol* 64:355-405. (2002).

[Zwe et al] Zweiri Y.H., Whidborne J.F. and Seneviratne L.D. "A three-term back propagation algorithm". *Neurocomputing*, Vol. 50, pp. 305-318. 2003. doi: 10.1016/S0925-2312(02)00569-6.

Annexe

Définitions des différents intervalles de l'ECG

Intervalle PR :il se calcule du début de l'onde P jusqu'au début du QRS,il est de 0.12 à 0.23s.

Segment ST : mesurée de la fin de l'onde S ou R jusqu'au début de l'onde T.il est normalement horizontal ou légèrement oblique +/- isoélectrique.

Intervalle QT : il se mesure du début du QRS jusqu'à la fin de l'onde T.le QT est fonction de la fréquence cardiaque.

Intervalle RRp et RRs

Nous appelons RRp la distance entre le pic R du présent battement et le pic R du battement précédent. Et RRs celle entre le pic R du présent battement et le pic R du battement suivant.

Rapport des intervalles RR (RRs/RRp) :

Dans le cas d'un rythme régulier, le rapport RRs/RRp est un paramètre qui caractérise une classe donnée. Dans le cas d'un rythme régulier, ce rapport est voisin de 1, mais il peut largement dépasser cette valeur dans le cas d'une ESV avec repos compensatoire.

Complexe QRS

Ce paramètre est important pour l'identification des battements pathologiques, ces types d'arythmies sont caractérisés généralement par un large complexe QRS.

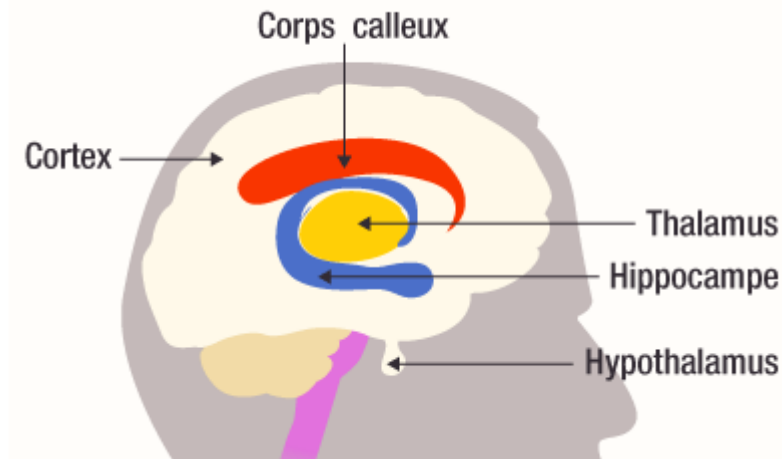
Durée de battement : lors d'une variation de rythmes par une présence d'anomalies cardiaques, cette perturbation peut s'apprécier en mesurant la durée du battement. Elle est calculée du début de l'onde P jusqu'à la fin de l'onde T.

Tableau représentant le nombre de battements de chaque enregistrement de la base MIT-BIH sélectionné.

enregistrement	Nombre des battements 'N'	Nombre des battements 'V'	Nombre des battements 'R'	Nombre des battements 'L'
----------------	---------------------------	---------------------------	---------------------------	---------------------------

100	62	0	0	0
101	5	0	0	0
103	58	0	0	0
105	10	0	0	0
106	27	34	0	0
109	0	0	0	104
111	0	0	0	41
113	6	0	0	0
115	10	0	0	0
116	45	0	0	0
118	0	0	12	0
119	50	34	0	0
122	5	0	0	0
123	5	0	0	0
124	0	0	33	0
200	0	25	0	0
203	0	15	0	0
207	0	0	0	40
208	0	152	0	0
212	5	0	26	0
214	0	50	0	50
215	103	0	0	0

Les organes spécialisés du cerveau



Le cortex constitue la majeure partie de la surface extérieure du cerveau. Sa surface est d'environ 1500 à 2000 cm², soit l'équivalent d'un à deux pages de journal.

Au centre du cerveau se situe une boucle de matière blanche appelée "**corps calleux**". Constitué de plus de 200 millions de fibres, c'est un pont entre les deux hémisphères.

Au centre du cerveau, on trouve des "noyaux" de matière grise : **le thalamus**, relais des grandes voies sensorielles de nos cinq sens, et **l'hypothalamus**, centre de nos principales pulsions (comme la faim, la soif...). Les neurones de l'hypothalamus, contrairement aux autres neurones, ne libèrent pas de neuromédiateurs lorsqu'ils sont excités, mais des hormones, substances chimiques circulant dans le sang. Du coup, l'hypothalamus contrôle le système hormonal de tout le corps humain.

L'hippocampe correspond à une aire cérébrale, appartenant au télencéphale. Il est situé dans les lobes temporaux et est impliqué dans la mémorisation et dans certaines formes de peur.

L'hippocampe a été étudié chez le rongeur dans différents domaines : la mémoire ; la navigation spatiale ; l'inhibition du comportement.

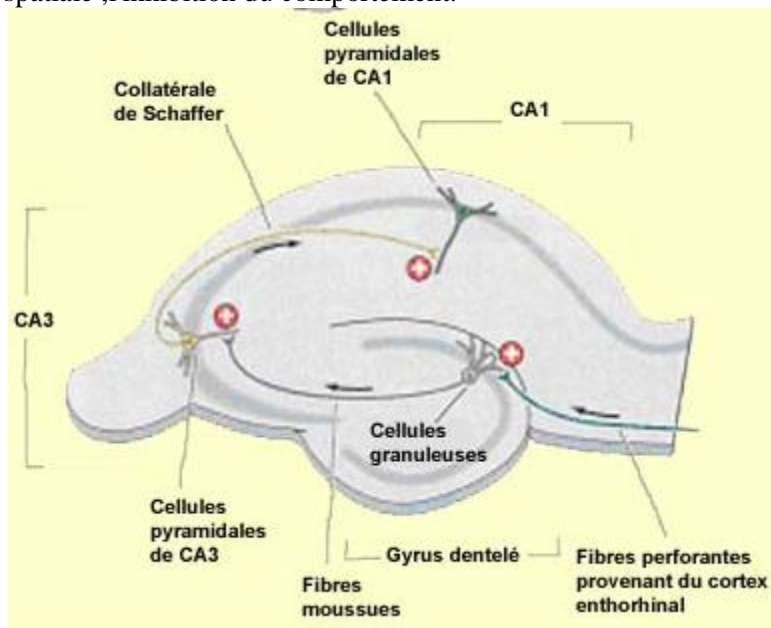


Figure : Coupe transversale de cerveau du rat au niveau des hippocampes
CA : Corne d'Ammon

