

Table des matières

Table des matières	1
Table des figures	7
Résumé	10
Abstract	11
Introduction	12
1 Notions en cardiologie	17
1.1 Introduction.....	17
1.2 Activité cardiaque.....	18
1.2.1 Anatomie du cœur.....	18
1.2.2 Activité mécanique cardiaque.....	21
1.2.3 Activité électrique cardiaque.....	23
1.3 Electrocardiographie de surface.....	25
1.3.1 Électrocardiogramme à douze dérivations.....	25
1.3.1.1 Dérivations bipolaires des membres.....	26
1.3.1.2 Dérivations unipolaires des membres.....	26
1.3.1.3 Dérivations précordiales.....	27
1.3.2 Enregistrement ambulatoire Holter.....	28
1.4 Ondes et intervalles.....	30
1.5 Troubles du rythme et de la conduction cardiaque.....	33
1.5.1 Rythme sinusal.....	33
1.5.2 Blocs cardiaques.....	34
1.5.2.1 Blocs de branche.....	34

1.5.2.1.1 Bloc de branche gauche.....	35
1.5.2.1.2 Bloc de branche droit.....	36
1.5.3 Arythmies.....	37
1.5.3.1 Arythmies supraventriculaires.....	38
1.5.3.1.1 La tachycardie supraventriculaire.....	38
1.5.3.1.2 La fibrillation auriculaire (FA).....	38
1.5.3.2 Arythmies ventriculaires.....	38
1.5.3.2.1 L'arythmie extrasystolique.....	38
1.5.3.2.2 Les bigéminismes et trigéminismes.....	39
1.5.3.2.3 Les tachycardies ventriculaires (TV).....	40
1.5.3.2.4 La fibrillation ventriculaire (FV).....	41
1.6 Diagnostic des arythmies.....	41
1.6.1 Analyse morphologique.....	41
1.7 Conclusion.....	42
2 Principes des réseaux de neurones	44
2.1 Introduction.....	44
2.2 Neurones biologiques et neurones formels.....	45
2.2.1 Neurone biologique.....	46
2.2.1.1 structure d'un neurone biologique.....	46
2.2.1.2 Création d'un potentiel d'action.....	47
2.2.2 Modèle de neurone formel.....	50
2.2.2.1 L'état des neurones.....	50
2.2.2.2 Les connexions.....	50
2.2.2.3 Les poids des connexions.....	50
2.2.3 Le neurone de Mc Culloch et Pitts.....	52
2.3 Historique des réseaux de neurones.....	55
2.4 Différentes architectures des réseaux de neurones.....	56
2.4.1 Les réseaux Feed –back.....	58
2.4.1.1 Les cartes auto – organisatrices de Kohonen.....	58
2.4.1.2 Les réseaux de Hopfield.....	58
2.4.1.3 Les ART.....	59

2.4.2 Les réseaux Feed - forward (ou réseaux non bouclés).....	59
2.4.2.1 Le Perceptron.....	59
2.4.2.2 Perceptron multi couches.....	61
2.4.2.3 Les réseaux à fonctions radiales.....	64
2.5 La rétropropagation.....	64
2.6 Propriétés des réseaux de neurones.....	65
2.6.1 La propriété d'approximation universelle.....	65
2.6.2 La propriété de parcimonie.....	66
2.6.3 La capacité d'apprentissage.....	66
2.7 Applications des réseaux de neurones.....	66
2.7.1 Intérêt des réseaux de neurones dans le diagnostic médical.....	67
2.8 Les limites des réseaux de neurones.....	68
2.9 Conclusion.....	70
3 Modèles d'apprentissage des réseaux de neurones	72
3.1 Introduction.....	72
3.2 Définition de l'apprentissage.....	73
3.3 L'apprentissage naturel.....	73
3.4 Apprentissage « artificiel » ou apprentissage « automatique ».....	74
3.5 Apprentissage des réseaux de neurones.....	75
3.5.1 Les applications de l'apprentissage.....	75
3.5.2 Les différents types d'apprentissage.....	76
3.5.2.1. Apprentissage supervisé.....	76
3.5.2.2. Apprentissage non-supervisé et auto-organisation.....	77
3.5.2.3. Apprentissage par renforcement.....	78
3.5.3 Apprentissage « en ligne » et apprentissage « hors-ligne ».....	78
3.6 Les différents algorithmes d'apprentissage.....	79
3.6.1 Choix de la fonction de coût.....	79
3.6.2 Règle de Hebb.....	80
3.6.3 Algorithme d'apprentissage par correction d'erreur ou règle du perceptron.....	81
3.6.3.1 Présentation de l'algorithme.....	81

3.6.3.2 Critiques sur la méthode par correction d'erreur.....	81
3.6.4 L'apprentissage de Widrow-Hoff ou règle Delta.....	82
3.6.5 La rétropropagation de GRADIENT (back propagation).....	83
3.6.5.1 Principe de l'optimisation par descente de gradient.....	84
3.6.5.2 Les limites de l'optimisation par descente de gradient.....	85
3.6.5.3 Problème des minima locaux.....	86
3.6.5.4 Méthodes itératives d'optimisation.....	86
3.6.6 la rétropropagation du gradient.....	87
3.6.6.1 L'apprentissage dans les réseaux multicouches.....	87
3.6.6.2 Introduction de l'algorithme.....	89
3.6.6.3 Algorithme de rétropropagation du gradient.....	93
3.6.6.4 Variantes de l'algorithme de rétropropagation du gradient de l'erreur.....	94
3.6.6.4.1 La méthode de Newton.....	94
3.6.6.4.2 La méthode de quasi-Newton.....	95
3.6.6.4.3 La méthode Levenberg-Marquardt.....	95
3.7 Problèmes d'apprentissage dans les réseaux multicouches.....	96
3.7.1 Considérations pratiques.....	96
3.7.2. Condition d'arrêt.....	97
3.7.3 Problème du surapprentissage (overfitting).....	97
3.7.3.1Utilisation d'une base de validation pendant l'apprentissage	98
3.7.3.2 Arrêt prématuré.....	98
3.8 Conclusion.....	99
4 Classification neuronale spécialisée des arythmies cardiaques	101
4.1 Introduction.....	101
4.2. Principes de la classification.....	103
4.2.1 Les phases de classification.....	104
4.2.1.1 Phase d'apprentissage.....	104
4.2.1.2 Phase de test.....	104
a. Taux de classification.....	104
b. sensibilité et spécificité.....	105

4.3 L'extraction des descripteurs et indicateurs pertinents.....	106
4.4 Intérêts des réseaux de neurones.....	106
4.4.1 Introduction.....	106
4.4.2 Choix des réseaux de neurones pour la classification	107
4.4.3 Mise en oeuvre des réseaux de neurones.....	107
4.5 le modèle de classification neuronale.....	108
4.5.1 Introduction.....	108
4.5.2 Sélection des descripteurs d'un cycle cardiaque.....	108
4.5.2.1 Introduction.....	108
4.5.2.2 Présentation des différents descripteurs.....	110
4.5.3 Sélection de la base d'exemples.....	111
4.5.4 Architecture des classifieurs neuronaux	113
4.5.4.1 Algorithme d'apprentissage.....	113
4.5.4.2 Dimensionnement du réseau pour chaque classifieur.....	114
4.6 Résultats expérimentaux de la classification.....	115
4.6.1 Apprentissage et test des différents réseaux.....	115
4.6.1.1 apprentissage et test du réseau2.....	116
4.6.1.2 apprentissage et test du réseau3.....	117
4.6.1.3 apprentissage et test du réseau4.....	118
4.7 Classification spécialisée.....	120
4.7.1 Spécialisation par patient.....	120
4.7.2 Spécialisation patient106.....	120
4.7.3 Spécialisation patient212.....	122
4.7.4 Spécialisation patient214.....	124
4.7.5 Efficacité de la spécialisation par patient.....	125
4.8 Spécialisation arythmie.....	129
4.8.1 Tests des réseaux sur des bases mono enregistrement.....	130
4.8.2 Tests des réseaux sur des bases variées.....	133
4.8.2.1 résultats de performances du réseau 2.....	133
4.8.2.2 résultats de performances du réseau 4 avec cas ciblé BBG	135
4.8.2.3 résultats de performances du réseau 4 avec cas ciblé BBD	136

4.8.2.4 résultats de performances du réseau 2 avec cas ciblé ESV	137
4.8.3 Tests des réseaux sur de grandes bases	138
4.9 Conclusion	141
Conclusion	143
Annexe	147
Bibliographie	150

Table des figures

Figure. 1.1 - Structure anatomique du cœur.....	19
Figure. 1.2 - Myocarde différencié ou circuit nodal.....	20
Figure. 1.3 - Cycle mécanique cardiaque.....	22
Figure. 1.4 - Les mouvements ioniques transmembranaires donnent naissance au potentiel d'action (myocytes ventriculaires).....	24
Figure. 1.5 - Montage d'Einthoven pour l'enregistrement des dérivations bipolaires des membres.....	26
Figure. 1.6 - Montage de Goldberger pour l'enregistrement des dérivations unipolaires des membres augmentés.....	27
Figure. 1.7 - Position des électrodes précordiales.....	28
Figure. 1.8 - Position recommandée des électrodes pour l'enregistrement Holter	29
Figure. 1.9 : ECG normal avec les notations usuelles de l'électrocardiographie clinique...	30
Figure. 1.10 - Schémas des différents aspects du complexe QRS et de leur notations...	30
Figure. 1.11: Mesure du temps d'apparition de la déflexion intrinsèque (DI)...	32
Figure. 1.12 - Exemple d'ECG normal.....	33
Figure. 1.13 : Exemple de bloc de branche gauche.....	34
Figure .1.14 : bloc de branche gauche en V1.....	35
Figure. 1.15 : bloc de branche gauche en V6.....	36
Figure .1.16 : bloc de branche droit en V1.....	37
Figure. 1.17 : bloc de branche droit en V6.....	37
Figure. 1.18 - Exemple d'extrasystole.....	39
Figure. 1.19- Exemple de doublet ventriculaire.....	39
Figure. 1.19 - Exemple de rythme de bigéminisme.....	40

Figure. 1.21 - Exemple de tachycardie ventriculaire.....	40
Figure. 1.22 - Exemple d'accès de fibrillation ventriculaire.....	41
Figure.2.1Modèle de neurone biologique.extraite de:www.lacim.uqam.ca.....	46
Figure.2.2 Un neurone biologique avec son arborisation dendritique.....	47
Figure .2.3 :potentiel d'action.....	48
Figure.2.4. Fonctionnement au niveau synaptique.....	49
Figure.2.5: Modèle général du neurone formel.....	51
Figure.2.6: Le neurone de Mc Culloch et Pitts.....	52
Figure.2.7 : les fonctions d'activation.....	54
Figure.2. 8. Mise en correspondance neurone biologique / neurone formel...	55
Figure2.9 : les grandes architectures des réseaux de neurones. Figure extraite du site : www.sm.u-bordeaux2.fr.....	58
Figure.2.10 :schéma d'un Perceptron.....	60
Figure.2.11 : le problème du « OU exclusif » en dimension2.....	61
Figure.2.12 : représentation d'un perceptron.....	61
Figure .2.13: réseau multi couches (Perceptron multi couches).....	63
Figure.3.1 :descente de gradient	83
Figure .3.2 : La méthode du gradient	85
Figure.3.3 : Limite de la Méthode du gradient	86
Figure .3.4: La fonction sigmoïde	88
Figure .3.5:notations utilisées.....	91
Figure.4.1 : schéma global de la classification.....	108
Figure .4.2 :Interface de Mesure des Paramètres de l'ECG	109
Figure .4.3: Les performances du réseau2 en utilisant les deux fonctions d'apprentissage 'trainbfg' et 'trainlm'.....	116
Figure .4.4: Les performances du réseau3 en utilisant les deux fonctions d'apprentissage 'trainbfg' et 'trainlm'.....	118
Figure .4.5: Les performances du réseau4 en utilisant les deux fonctions d'apprentissage 'trainbfg' et 'trainlm'.....	119
Figure .4.6 :taux de sensibilité des réseaux spécialisés et réseaux globaux.....	121
Figure .4.7 :taux de sensibilité des réseaux spécialisés et réseaux globaux.....	123
Figure .4.8 :taux de sensibilité des réseaux spécialisés et réseaux globaux.....	125
Figure .4.9 :taux de sensibilité des réseaux spécialisés et réseaux globaux...	127

Figure.4.10 Apprentissage du réseau1 en utilisant « trainbfg ».....	128
Figure.4.11 Apprentissage du réseau spécialisé avec 106n106v en utilisant« trainbfg ».....	128
Figure.4.12 : Apprentissage du réseau4 en utilisant « trainbfg ».....	128
Figure.4.13 :Apprentissage du réseau4 spécialisé spécialisé avec 214l214v en utilisant « trainbfg ».....	129
Figure.4.14 :Apprentissage du réseau4 avec 212n212r en utilisant « trainbfg »	129
Figure.4.15 :taux de sensibilité des réseaux spécialisés et réseaux globaux.....	131
Figure .4.16 :taux de sensibilité des réseaux spécialisés et réseaux globaux.....	132
Figure .4.17 :taux de sensibilité du réseau 2 spécialisé et global.....	134
Figure .4.18 : taux de sensibilité du réseau 2 spécialisé et global.....	137
Figure. 4.19 :taux de sensibilité des réseaux spécialisés et réseaux globaux.....	140

Résumé

L'enregistrement du signal cardiaque ECG sur de longues durées (tracé Holter) est un examen non invasif, très employé en cardiologie .

Les modifications qui apparaissent sur le tracé ECG sont souvent indicatrices de pathologies, d'où l'intérêt de l'analyse de ces signaux par un système automatique pour mieux exploiter les informations enregistrées lors d'un examen Holter et l'identification des paramètres représentatifs des différentes pathologies.

L'analyse et la caractérisation du signal ECG par des descripteurs constituent une étape primordiale pour la classification et la pose d'un diagnostic.

Nous avons développé dans ce travail un classifieur des arythmies cardiaques par application des réseaux de neurones.

A partir de ce classifieur global nous avons mis en œuvre deux classifieurs spécialisés obtenus après un deuxième apprentissage pour s'adapter à chaque patient (spécialisation par patient)et à chaque arythmie(spécialisation par arythmie).

Les résultats de performances de ces classifieurs montrent l'intérêt de la spécialisation qui ont été validés par des signaux ECG des différents patients de la base de données « MIT BIH Arrhythmia Database ».

MOTS CLES - Signal ECG, Arythmies cardiaques,Apprentissage, Classification ,MIT BIH arrhythmia Database,Réseaux de neurones,Spécialisation.

Abstract

The recording of cardiac signal ECG for long durations “Holter” is a non-invasive examination, very employed in cardiology.

The modifications which appear on the graph of the ECG are often indicating pathologies, from where interest of the analysis of these signals by an automatic system for better exploiting the information recorded during a Holter examination and the identification of the parameters representative of various pathologies.

Analysis and the characterization of signal ECG by descriptors constitute a paramount stage for the classification and for establish a diagnosis.

We developed in this work a classifier of cardiac arrhythmias by applying neural networks.

From this general classifier we implemented two specialized classifiers have been developed obtained after a second training. These classifiers are adapted respectively to each patient (specialization by patient) and to each arrhythmia (specialization by arrhythmia).

The results of performances of these classifiers show the interest of the specialization which were validated by signals ECG of the various patients of the data base “MIT BIH Arrhythmia Database”.

KEY WORDS - Signal ECG, cardiac Arrhythmias, Training, Classification, MIT-BIH arrhythmia Database, neural Networks , Specialization.

Introduction

Un cœur en bonne santé est le moteur d'une vie saine.

Le cœur est l'organe central du système cardiovasculaire : il peut être affecté de nombreuses pathologies qui peuvent soit être bénignes, comme certaines tachycardies par exemple, soit s'avérer très sérieuses. Avec l'évolution des techniques, les médecins disposent aujourd'hui d'outils performants pour observer le fonctionnement du muscle cardiaque et dresser ainsi leur diagnostic. Parmi les examens cardiologiques possibles, l'électrocardiogramme (ECG) est l'examen le plus couramment effectué, car il est rapide à mettre en place, peu coûteux et surtout non invasif donc très peu contraignant pour le patient.

Les pathologies cardiovasculaires et les arythmies cardiaques provoquent chaque année 17 millions de décès à travers le monde [Mackay et Mensah, 2004]. Elles représentent en Algérie aussi la première cause de morbidité et de mortalité avec un taux de 26%.

Toutes les 2 secondes, une personne meurt d'une maladie cardiovasculaire quelque part dans le monde ! D'après l'OMS, les maladies cardiovasculaires seront aussi la première cause de mortalité précoce d'ici 2020, c.-à-d. avant 65 ans pour les femmes et avant 55 ans pour les hommes.

Les maladies cardiovasculaires, désignent une série d'affections du cœur et des vaisseaux sanguins. Une étude récente réalisée auprès de plus de 500 patients et 750 médecins dans 10 pays, révèle que les gens ne sont pas suffisamment informés sur les risques cardiovasculaires, que ces risques ne sont pas systématiquement décelés et qu'il reste beaucoup à faire pour maintenir sous contrôle les trois principales causes des maladies cardiovasculaires : taux de cholestérol élevé, hypertension et tabagisme. Nous avons les pleins pouvoirs pour agir au niveau de ces facteurs de risque, Comment ? En

optant résolument pour un mode de vie et une alimentation plus sains, si nécessaire en association avec un traitement médicamenteux.

L'expression 'Mieux vaut prévenir que guérir' est parfaitement applicable aux maladies cardiovasculaires. La prévention des maladies cardiovasculaires a d'ailleurs tout à voir avec une bonne hygiène de vie et un mode de vie sain. Il faut manger équilibré, surveiller son poids, faire beaucoup d'exercice physique, ne pas fumer et de trouver une échappatoire pour les facteurs de stress de la vie. Il est conseillé de faire régulièrement un check-up chez votre médecin traitant, afin de déceler à temps les facteurs de risque.

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

Les spécialistes détectent les situations alarmantes à travers les signaux physiologiques recueillis, tel que l'électrocardiogramme (ECG), il est très souvent complété par un examen similaire d'une durée de 24 heures appelé « Holter », examen au cours duquel le patient peut vaquer à ses occupations habituelles. Le principal avantage de l'enregistrement Holter par rapport à l'ECG est qu'il permet la détection d'événements sporadiques qui n'interviennent pas nécessairement au cours des quelques secondes de l'enregistrement ECG lorsque le patient est au repos.

Actuellement de nombreux travaux de recherche tentent de développer de nouvelles approches d'aide à la décision, le cas sur lequel nous avons travaillé relevait 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.

En effet ,les méthodes d'électrocardiologie ,non invasives et de faible coût ,constituent un mode d'investigation privilégié et performant en vue du dépistage de la plupart des altérations pathologiques du système cardiovasculaire.

Nous proposons dans ce mémoire l'étude d'un système d'aide à la décision en cardiologie.

La démarche diagnostique pour l'aide à la décision est une application privilégiée où le raisonnement humain demeure le modèle de référence, notamment dans le domaine médical.

La finalité d'un diagnostic est de pouvoir décider, après avoir observé un objet, à quelle forme typique celui-ci est similaire, ou en d'autres termes, à quelle classe connue il peut être associé ou affecté.

L'approche neuronale appliquée au problème du diagnostic a fait ses preuves depuis déjà une décennie et présente plusieurs atouts .

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

Nous avons développé dans ce mémoire deux classifieurs spécialisés .La première étape consiste à mettre en oeuvre un classifieur général à partir d'une base de données globale et variée ensuite le soumettre à un deuxième apprentissage pour l'adapter aux données cardiologiques d'un patient spécifique (spécialisation par patient), et aussi à une arythmie précise (spécialisation de l'arythmie).ces différents classifieurs ont été basés sur des réseaux de neurones multi couches .ce choix a été dicté par leur facilité à mettre en oeuvre et leur pouvoir discriminant attesté par une littérature très abondante [Silipo1998],[Dokur2000], [Achayara2004].

Ce mémoire comprend quatre chapitres :

Dans le chapitre1 nous présentons quelques notions essentielles en cardiologie.

Le premier chapitre débute par le rappel des caractéristiques anatomiques et physiologiques du coeur, nécessaires à la compréhension de notre étude. Après nous présentons les dérivations cardiaques pour enregistrer cette activité électrique. Nous passons à la définition des ondes élémentaires d'un signal ECG ainsi que les intervalles, durées, et segments que l'on mesure.

Nous décrivons brièvement les différentes arythmies et pathologies cardiaques susceptibles d'être détectées dans un enregistrement électrocardiographique .

Le chapitre2 est une présentation générale des réseaux de neurones, c'est l'approche que nous proposons comme modèle de classification des pathologies cardiaques.

Le chapitre3 décrira La caractéristique la plus intéressante d'un RN qui est sa capacité d'apprendre,nous aborderons les différents modes d'apprentissage en particulier l'apprentissage supervisé qui sera utilisé dans notre travail .

Dans Le quatrième et dernier chapitre décrit le problème de classification spécialisée neuronale .Nous mettons en évidence la nécessité de disposer de bonnes variables descriptives d'un cycle cardiaque pour résoudre les problèmes de classification. Nous exposerons les différents classifieurs neuronaux globaux et spécialisés, ainsi que les résultats de leur test de performances sur la base de données MIT-BIH (Massachusetts Institut of Technology - Beth Israel Hospital) [Mark 1988].

Rapport-Gratuit.com

Chapitre 1

Notions en cardiologie

1.1 Introduction

Les pathologies cardiovasculaires et les arythmies cardiaques provoquent chaque année 17 millions de décès à travers le monde [Mackay et Mensah, 2004]. 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 coeur 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.

Ces arythmies se caractérisent par un fort degré d'urgence et explique qu'il est primordial de pouvoir prendre en charge très tôt les patients souffrant d'un infarctus. De plus, les arythmies mineures informent sur l'état de récupération cardiaque des patients et doivent donc être détectées notamment pour prévenir une dégénérescence possible en arythmies sévères.

La mort subite est un risque constant de l'activité cardiaque, quelle que soit la stabilité apparente de la situation cardiaque initiale, et nécessite une intervention immédiate à l'aide de moyens de ressuscitation tel qu'un défibrillateur. Certains de ces accidents cardiaques sont prévisibles mais la plupart surviennent de façon inopinée. Le traitement de certaines maladies cardiovasculaires nécessite un personnel médical aguerri et du matériel spécialisé pour une intervention thérapeutique rapide.

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. Cette analyse est effectuée par des cardiologues spécialisés appelés rythmologues, qui décèlent sur les différents tracés le mécanisme de l'activité électrique cardiaque.

Cette section est une présentation du système cardiaque et les troubles du rythme et de la conduction cardiaque.

1.2 Activité cardiaque

Le système cardio-vasculaire assure la circulation du sang qui permet les échanges respiratoires et nutritifs indispensables à la vie.

La grande circulation représente la vascularisation (ou perfusion) de toutes les cellules du corps, hormis les poumons. La petite circulation est celle qui concerne uniquement les poumons. Dans la grande circulation, l'oxygène du sang artériel est consommé par les cellules ce qui produit le sang veineux. À travers les capillaires des poumons, le sang veineux se recharge en oxygène et s'artérialise.

La circulation proprement dite est assurée par un organe ayant le rôle de « pomper » et distribuer le sang : le **cœur**. Il assure un échange régulier entre la petite et la grande circulation suivant un cycle bien précis. La régularité de ces échanges est commandée par un stimulus électrique, dépendant du système nerveux autonome, qui parcourt le cœur du nœud sinusal à l'apex pour déclencher la contraction des différentes chambres qui constituent le cœur. Ce stimulus peut être observé en mesurant les différences de potentiel de plusieurs électrodes à la surface du corps. L'interprétation de ces différences de potentiel relève du domaine de l'*électrocardiographie*.

Le cycle cardiaque ne tolère pas les interruptions car certaines cellules meurent lorsqu'elles ne sont plus alimentées en sang artériel. C'est notamment le cas des cellules du cerveau, c'est pourquoi les défaillances cardiaques sont si fatales. Par la suite, le cœur est présenté selon trois axes : l'anatomie, l'activité mécanique et l'activité électrique.

1.2.1 Anatomie du cœur

Le cœur est un organe intrathoracique situé entre les deux poumons au carrefour des grosses artères (aorte et artère pulmonaire) et grosses veines (veines caves et pulmonaires) de l'organisme. Sa structure est composée de 3 épaisseurs : l'endocarde, surface externe, où passent nerfs et vaisseaux sanguins ; l'épicarde, membrane séreuse formant la paroi interne du péricarde ; et le myocarde, partie véritablement active du cœur.

Dans les fibres du myocarde, il y a principalement deux tissus qui jouent un rôle complémentaire dans le cycle cardiaque. L'un est dédié à la mécanique musculaire, c'est le myocarde commun, l'autre engendre et conduit l'excitation (la commande de contraction), c'est le myocarde différencié (ou circuit nodal).

Myocarde commun : Le myocarde commun est essentiellement composé de cellules musculaires (ou myocytes). Il est réparti en quatre chambres creuses liées entre elles : les oreillettes droite et gauche et les ventricules droit et gauche (voir Figure 1.1). L'oreillette et ventricule gauches (resp. droits) communiquent entre eux par la valve mitrale (resp. tricuspide).

La paroi des oreillettes est très mince et celle du ventricule droit ne dépasse pas quelques millimètres. Le ventricule gauche, quant à lui, possède une paroi de plus d'un centimètre. Les deux masses auriculaire et ventriculaire sont séparées par les structures fibreuses des anneaux auriculo-ventriculaires et du septum fibreux. Le septum inter-ventriculaire, épais d'une quinzaine de millimètres, est constitué de l'adossement du ventricule gauche, en majorité, et de l'adossement du ventricule droit ;

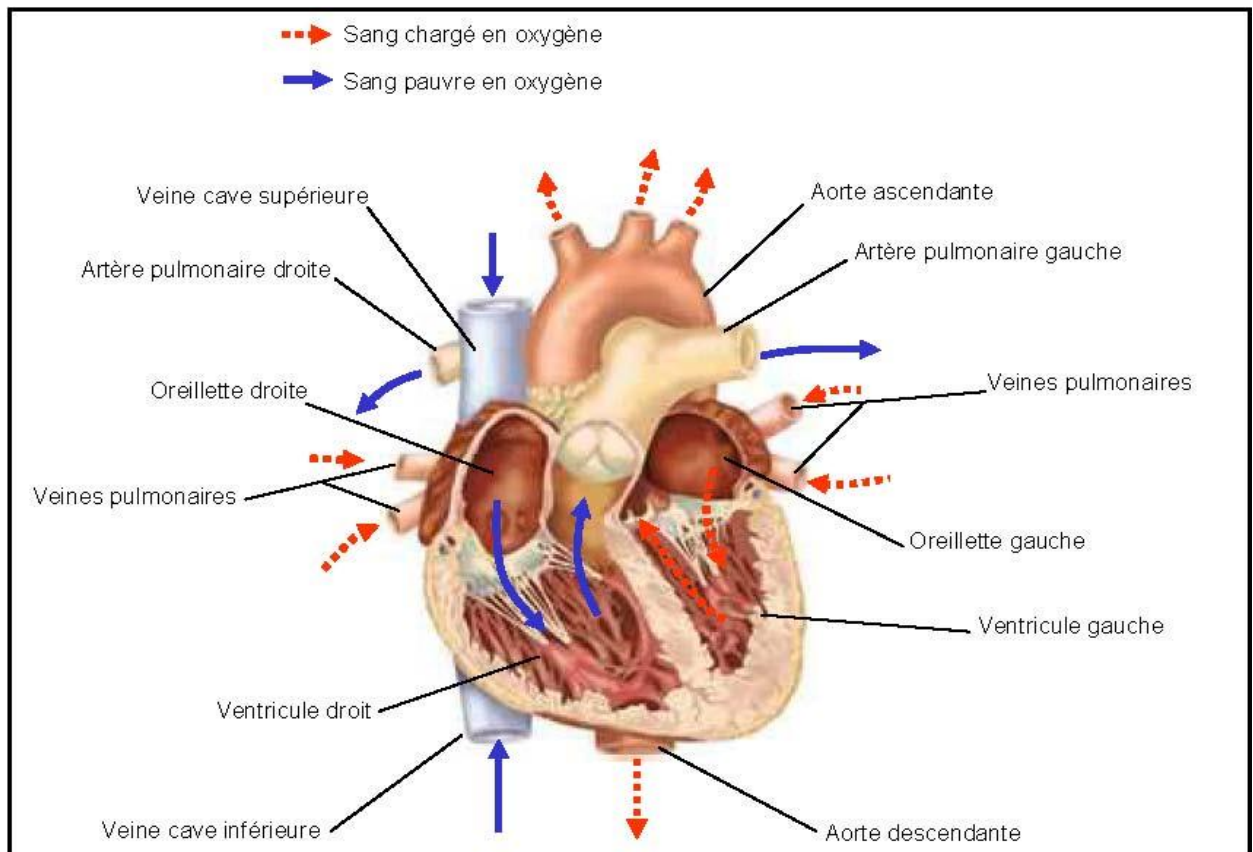


FIG. 1.1 - Structure anatomique du cœur.

Myocarde indifférencié : La contraction du myocarde est déclenchée par une excitation électrique (le potentiel d'action) conduite et générée par la structure du myocarde différencié.

Les cellules nodales du myocarde différencié sont présentes sur l'ensemble des cavités cardiaques de l'oreillette droite à l'oreillette gauche jusqu'aux ventricules (voir Figure 1.2).

Ce circuit comprend essentiellement cinq structures.

- Les cellules du nœud sinusal, ou nœud de Keith et Flack [Keith et Flack, 1906], situées dans la paroi de l'oreillette droite, génèrent le rythme cardiaque normal. L'activité du nœud sinusal est dépendante du système nerveux autonome.
- Les voies de conduction intra-auriculaire comprennent trois voies sino-nodales (antérieure, moyenne et postérieure) qui parcourent l'oreillette droite du nœud sinusal au nœud auriculo-ventriculaire et le faisceau de Bachmann qui dérive de la voie sino-nodale antérieure et se dirige vers l'oreillette gauche.
- Les cellules du nœud auriculo-ventriculaire (nœud AV), ou nœud de Tawara [Tawara, 1906; Tawara, 2000], situées dans la partie basse et antérieure de la cloison inter-auriculaire, bloquent pendant une courte pause l'activité électrique provenant des oreillettes puis la communique au faisceau de His [Kent 1914].

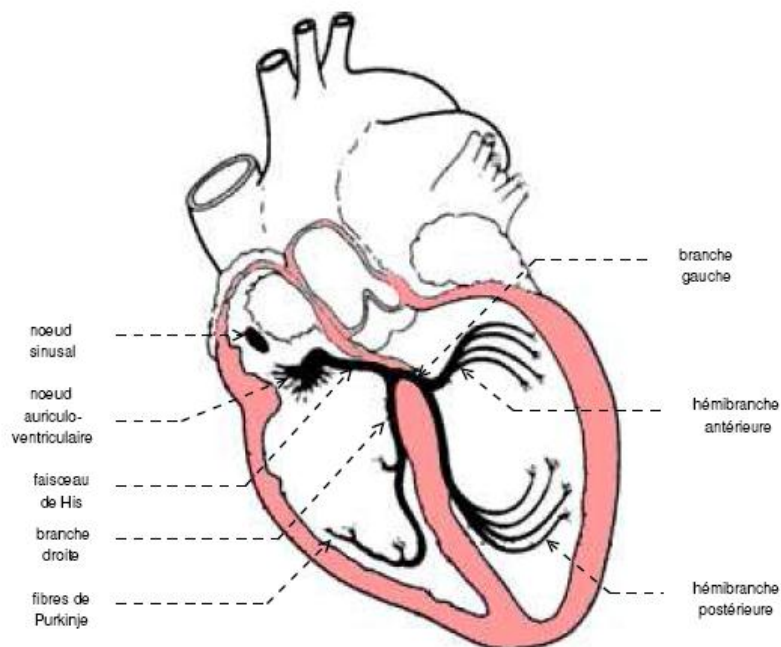


FIG. 1.2 - Myocarde différencié ou circuit nodal.

- Les cellules du faisceau de His constituent un cordon partant du nœud AV jusqu'à la jonction du septum inter-ventriculaire et du septum fibreux. Le faisceau se divise ensuite en deux branches :
 - la branche droite qui descend sur la face droite du septum inter-ventriculaire jusqu'à l'apex où elle se sépare en petites ramifications,
 - la branche gauche qui se sépare en deux hémibranches : l'hémibranche postérieure et l'hémibranche antérieure (il existe d'autres petites ramifications de moindre importance).
- Les fibres de Purkinje [Purkinje, 1845; Matousek et Posner, 1969] se situent à la terminaison des branches gauches et de la branche droite et tapissent les cavités des deux ventricules. Elles irriguent les myocytes ventriculaires.

L'activité de pompe du cœur peut être vue sous deux aspects : l'un, mécanique, mettant en jeu la contractilité des cavités cardiaques et les ouvertures et fermetures des valves des artères et veines ; l'autre, électrique, mettant en jeu la conduction électrique de la commande de contraction du muscle cardiaque à travers celui-ci. Ces deux aspects sont détaillés dans les sections suivantes.

1.2.2 Activité mécanique cardiaque

Le cycle de la circulation sanguine se répète constamment et se divise en deux périodes : la systole et la diastole. La systole est la période correspondant à l'éjection du sang dans la grande et petite circulation. Elle se décompose en trois phases : la systole auriculaire, la contraction ventriculaire isovolumique et la systole ventriculaire. La diastole est la phase de relaxation du cœur, pendant laquelle il se remplit de sang. Cette période est composée de deux phases : la relaxation ventriculaire isométrique et la phase de repos. La figure 1.3 illustre la succession des différentes phases détaillées ci-dessous.

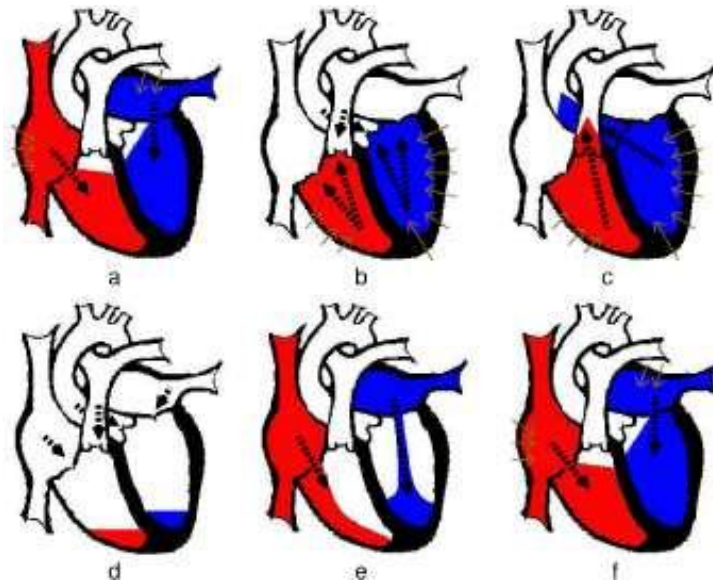


FIG. 1.3 - Cycle mécanique cardiaque. a) systole auriculaire, b) contraction isovolumique des ventricules, c) systole ventriculaire, d) relaxation des ventricules, e) repos, et f) nouvelle systole auriculaire. Les flèches pleines indiquent la contraction musculaire et les flèches en pointillés, la direction de la pression sanguine.

a et f) **La systole auriculaire** est la contraction des oreillettes lorsque celles-ci sont remplies de sang. La pression exercée par le muscle auriculaire ferme les orifices des veines caves et pulmonaires et provoque le versement du sang auriculaire dans les ventricules (Figure 1.3) par les valves tricuspide et mitrale. Cependant, la majorité du remplissage des ventricules survient passivement pendant la diastole.

b) **La contraction ventriculaire isovolumique** commence lorsque les cellules musculaires du myocarde ventriculaire se contractent. La pression ferme les valves tricuspide et mitrale et les valvules sigmoïdes restent fermées pendant quelques dizaines de centièmes de secondes. Comme il n'y a pas d'éjection ventriculaire, la pression augmente fortement dans les ventricules.

c) **La systole ventriculaire** commence lorsque les pressions dans les ventricules dépassent les pressions dans l'artère pulmonaire et l'aorte. Les valvules sigmoïdes s'ouvrent et l'éjection commence. Le sang oxygéné emprunte la crosse aortique et le sang désoxygéné le tronc pulmonaire. La pression aortique atteint un pic (pression artérielle systolique) puis redescend jusqu'à la fin de la systole.

d) **La relaxation ventriculaire** fait suite à la systole. Les ventricules se relâchent, la pression chute jusqu'à être inférieure à celle exercée dans l'aorte et l'artère pulmonaire. En conséquence, les valvules sigmoïdes se ferment et, lorsque la pression devient inférieure à celle des oreillettes, les valves tricuspide et mitrale s'ouvrent.

e) **La phase de repos** est celle pendant laquelle le sang des veines caves et pulmonaires s'écoule librement dans les ventricules via les oreillettes.

1.2.3 Activité électrique cardiaque

Le mécanisme cardiaque, qui comprend l'expulsion du sang et l'ouverture-fermeture des valves, fonctionne uniquement grâce aux contractions du myocarde. Ces contractions sont déclenchées par la propagation de proche en proche du *potentiel d'action* à travers les cellules myocardiques. Chaque cellule myocardique réagit à un stimulus électrique grâce à une membrane semi-perméable aux ions. Au repos, l'intérieur de la membrane cellulaire est chargé négativement par rapport à l'extérieur qui est pris comme référence. Dans cet état électrique stable on dit que la cellule est *polarisée*. De l'extérieur, on n'enregistre aucune activité électrique.

Lorsque la cellule est stimulée électriquement, les propriétés de la membrane se modifient et sa perméabilité aux ions augmente. Les échanges ioniques à travers la membrane des cellules myocardiques donnent naissance au potentiel d'action.

La figure 1.4 montre l'effet des échanges ioniques transmembranaires sur le potentiel d'action. La phase 0 est caractérisée par les ions sodium qui concourent à l'établissement d'un déséquilibre électrique entre le secteur extra-cellulaire et intra-cellulaire. Cette phase correspond à la *dépolarisation*. La phase 1 est le résultat de l'accroissement brutal de la perméabilité membranaire au sodium, ainsi, le potentiel de la membrane passe de -90mV à 40mV, c'est la systole électrique. Durant la phase 2, l'entrée des ions calcium à l'intérieur de la membrane permet le maintien en plateau de la dépolarisation. Durant la phase 3, l'accroissement de conductance au potassium est responsable d'une négativation des charges intra-cellulaires, et donc de la *repolarisation* cellulaire. La phase 4 correspond à la phase de repos, c'est la diastole électrique. On obtient l'équilibre avec une différence de potentiel négative.

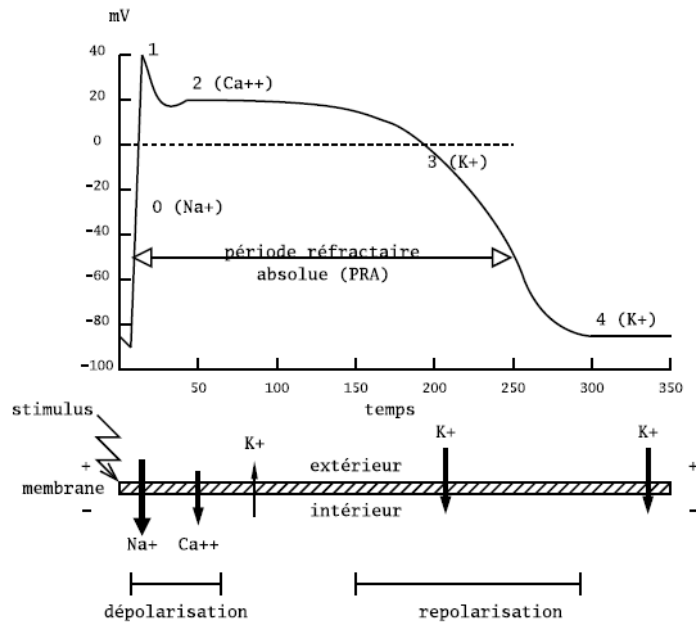


FIG. 1.4 - Les mouvements ioniques transmembranaires donnent naissance au potentiel d'action (myocytes ventriculaires).

Séquence normale d'activation cardiaque

Dans une séquence normale d'activation cardiaque telle que décrite par Tawara [Tawara, 1906; Tawara, 2000], le stimulus de départ du cycle cardiaque est généré par les cellules nodales du nœud sinusal. La propriété d'automaticité des cellules nodales du nœud génère un potentiel régulier et d'amplitude suffisante pour exciter les cellules myocardiques des oreillettes. Cette activité, communiquée de proche en proche, est stoppée au nœud AV qui le communique ensuite au faisceau de His puis à la branche droite et aux branches gauches puis aux fibres de Purkinje qui irriguent les myocytes ventriculaires. On peut ainsi suivre le cycle mécanique. Tout d'abord les cellules des oreillettes sont dépolarisées, ce qui provoque leur contraction (systole auriculaire), l'onde traverse le nœud auriculo-ventriculaire et le septum inter-ventriculaire, puis le ventricule droit puis gauche se dépolarisent (contraction ventriculaire isovolumique et systole ventriculaire). Enfin, les cellules du myocarde entre en repolarisation (relaxation ventriculaire) puis stabilisation (phase de repos).

1.3 Electrocardiographie de surface

Le corps humain étant électriquement conducteur, les potentiels d'actions générés lors de l'activité électrique cardiaque peuvent être recueillis par des électrodes placées sur la peau. L'enregistrement de cette activité électrique du cœur, sur un plan frontal (par les dérivations des membres) et sur un plan horizontal (par les dérivations précordiales), est un électrocardiogramme (ECG). L'ECG est un outil diagnostique permettant de détecter les pathologies cardiaques rythmiques, musculaires, les problèmes extracardiaques métaboliques, médicamenteux, hémodynamiques et autres.

L'électrocardiogramme (ECG) est l'examen le plus couramment effectué, car il est rapide à mettre en place, peu coûteux et surtout non invasif donc très peu contraignant pour le patient.

Les électrodes peuvent être utilisées selon deux modes, l'un dit bipolaire où le potentiel d'une électrode est soustrait à une autre, et l'autre dit unipolaire où le potentiel d'une électrode est pris par rapport à un point de référence qui est généralement une moyenne du potentiel de toutes les autres électrodes utilisées ou une électrode éloignée de toute activité électrique (la masse). Toute activité électrique se dirigeant vers l'électrode est enregistrée par une déflexion positive et toute activité s'en éloignant est enregistrée par une déflexion négative.

Dans les sections suivantes, les différentes dérivations standards de l'ECG clinique et l'enregistrement ambulatoire de Holter sont présentés. Puis, la nomenclature des ondes et intervalles utilisée pour l'analyse de l'ECG est détaillée.

1.3.1 Électrocardiogramme à douze dérivations

L'ECG est un enregistrement de surface de l'activité électrique du cœur, par des électrodes reliées à un électrocardiographe qui amplifie le signal électrique. Les tissus se trouvant entre le cœur et les électrodes parasitent le signal, le tracé électrocardiographie n'est donc qu'une estimation de l'activité électrique générée par le cœur.

L'ECG standard est enregistré sur 12 dérivations (six dérivations des membres et six précordiales), avec une vitesse de déroulement du papier à 25 mm par seconde et une amplitude de 10 mm pour 1 mV.

1.3.1.1 Dérivations bipolaires des membres

Les dérivations bipolaires des membres permettent d'étudier l'activité électrique du cœur sur le plan frontal. Elles ont été déterminées par Einthoven [Einthoven, 1906] au début du vingtième siècle et restent encore utilisées aujourd'hui. Ces trois dérivations sont déduites des trois électrodes posées sur les membres. Soit VL le potentiel sur le bras gauche, VR le potentiel sur le bras droit et VF le potentiel sur la jambe gauche, les trois dérivations sont :

- DI (dérivation I) dont la différence de potentiel respecte l'équation $DI = VL - VR$,
- DII (dérivation II) avec $DII = VF - VR$,
- DIII (dérivation III) avec $DIII = VF - VL$.

Ces trois dérivations constituent le triangle d'Einthoven et chaque dérivation est une arête du triangle (voir Figure 1.5).

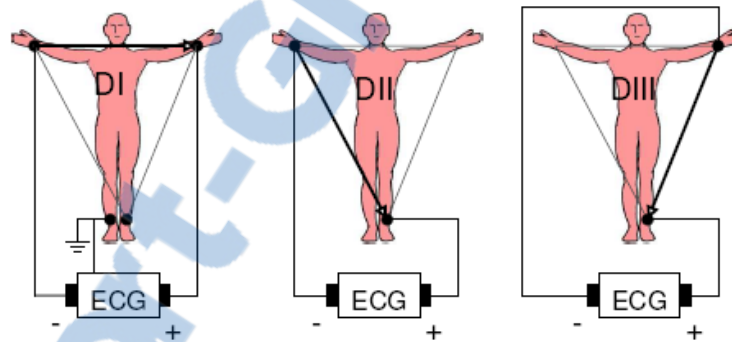


FIG. 1.5 - Montage d'Einthoven pour l'enregistrement des dérivations bipolaires des membres.

Le triangle d'Einthoven est formé par les trois électrodes posées sur le bras droit le bras gauche et la jambe gauche. Sur la jambe droite est posée une électrode qui sert de référence (la masse). Le cœur se trouve au centre du triangle et les trois dérivations bipolaires permettent l'enregistrement sous trois angles différents.

1.3.1.2 Dérivations unipolaires des membres

Les dérivations unipolaires des membres permettent d'étudier l'activité électrique du cœur sur le plan frontal. Elles ont été déterminées par Wilson [Wilson et coll., 1934] en 1934 et améliorées par Goldberger [Goldberger, 1942].

Les dérivations périphériques de Goldberger se servent des mêmes électrodes qu'Einthoven. Chaque électrode est prise comme pôle positif avec pour référence négative les 2 autres électrodes.

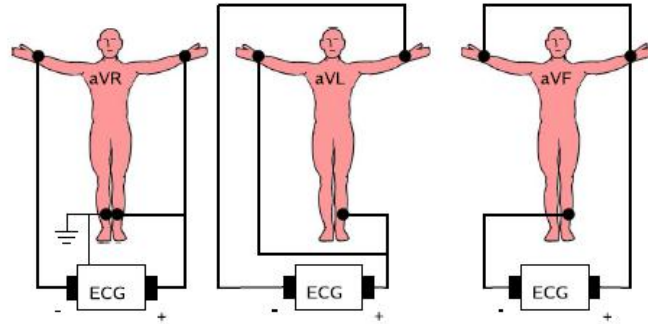


FIG. 1.6 - Montage de Goldberger pour l'enregistrement des dérivations unipolaires des membres augmentés.

Les droits dérivations sont :

- aVL (a comme *augmented*) dont la différence de potentiel respectent l'équation

$$aVL = VL - \frac{VR + VF}{2} = 1,5VL \text{ car selon la loi de Kirchhoff } VL + VR + VF = 0 ;$$

- aVR avec $aVR = 1, VR$;
- aVF avec $aVF = 1,5VF$.

Ces trois dérivations constituent trois vecteurs passant au centre du triangle d'Einthoven (voir Figure 1.6).

1.3.1.3 Dérivations précordiales

Ce sont des dérivations unipolaires, mises au point par Wilson [Wilson et coll., 1944]. Elles sont posées sur le thorax et sont désignées par la lettre V suivie du numéro de leur emplacement. Le potentiel de l'électrode exploratrice est pris par rapport à la moyenne des potentiels VL, VR et VF.

Six points, définis par Wilson, permettent d'obtenir les dérivations VI à V6. Leur emplacement est représenté sur la figure 1.7.

Les dérivations précordiales ont deux caractéristiques qui les distinguent fondamentalement des dérivations des membres : elles mesurent l'activité électrique cardiaque dans le plan horizontal et sont posées à proximité du cœur.

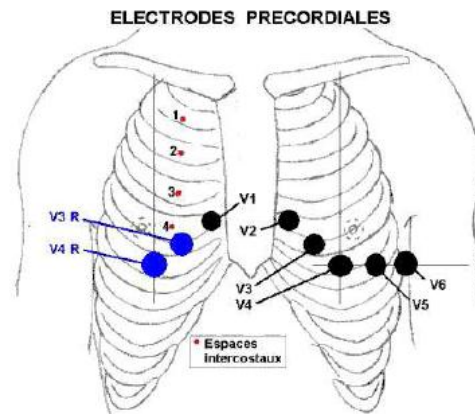


FIG. 1.7 - Position des électrodes précordiales.

1.3.2 Enregistrement ambulatoire Holter

L'ECG est très souvent complété par un examen similaire d'une durée de 24 heures appelé « Holter » [Holter 1961], examen au cours duquel le patient peut vaquer à ses occupations habituelles. Le principal avantage de l'enregistrement Holter par rapport à l'ECG est qu'il permet la détection d'événements sporadiques qui n'interviennent pas nécessairement au cours des quelques secondes de l'enregistrement ECG lorsque le patient est au repos.

C'est un examen utile pour le diagnostic des troubles de la conduction cardiaque et des troubles rythmiques, surtout s'ils sont paroxystiques (c'est-à-dire intermittents), et parfois asymptomatiques (non ressentis par le patient).

Il consiste à enregistrer un électrocardiogramme de surface modifié pendant au moins 24 heures. L'ECG Holter se différencie de l'ECG standard, de par son contexte d'observation à la fois ambulatoire et longue durée. Il est par conséquent générateur de situations souvent difficiles en termes de reconnaissance d'événements ou de diagnostic, ceci pour deux raisons :

- la diversité et le caractère aléatoire des dégradations pouvant affecter ponctuellement ou durablement les informations d'ECG,
- la modulation des caractéristiques morphologiques et temporelles du signal ECG, relative à l'activité régulatrice du système cardio-vasculaire en réponse aux besoins métaboliques du sujet.

Durant l'enregistrement, le patient doit tenir un journal d'activité. Ainsi, le cardiologue peut analyser sur une longue période (de 24 à 48 heures) l'évolution du rythme cardiaque en fonction des activités du patient. Par exemple, une augmentation du rythme sur l'enregistrement sera normale si le patient pratiquait une activité physique à ce moment là. Les dérivations utilisées ne sont pas les mêmes que pour l'ECG standard (voir Figure 1.8). Elles sont de type bipolaires thoraciques qui, contrairement aux dérivations périphériques, sont à la fois peu sensibles aux courants d'action musculaires (EMG) s'additionnant au signal ECG et très sensibles aux aspects ischémiques rencontrés en cas d'insuffisance coronaire. La position des électrodes est déterminée en partie en fonction de la pathologie suspectée.

Les électrodes sont reliées à un enregistreur à bande magnétique ou à mémoire solide (numérique) que le patient garde à la ceinture pendant son activité quotidienne. L'enregistrement est ensuite analysé par le médecin. Étant donnée la très grande quantité de données à analyser (2160 mètres d'ECG pour 24 heures), il existe trois méthodes d'analyse [Adamec et Adamec, 2000].

La lecture manuelle, inventée par N.J. Holter, consiste à visualiser l'enregistrement en vitesse accélérée sur un oscilloscope. Les battements normaux s'affichent comme une image fixe tandis que les anomalies (extrasystole, pause, tachycardie) apparaissent en dehors ou déforment l'image normale.

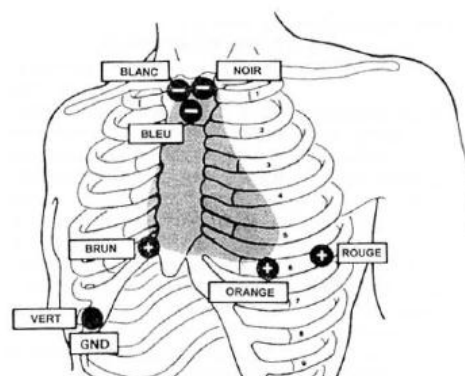


FIG. 1.8 - Position recommandée des électrodes pour l'enregistrement Holter.

La lecture semi-automatique consiste à apprendre automatiquement les éléments essentiels de l'enregistrement. Lors de la lecture, le système s'arrête sur chaque nouvelle anomalie que le médecin doit classer.

La lecture automatique utilise des techniques variables (templates matching, réseaux neuronaux, etc.) pour détecter les anomalies.

Cette méthode, très rapide, ne se passe pas d'une interprétation et vérification humaine.

1.4 Ondes et intervalles

Les différentes ondes et les intervalles les séparant ont permis de dresser le standard ECG. [J.Gay -P.Benoit] Toute interprétation de l'électrocardiogramme se réfère aux caractéristiques de forme et de largeur d'onde du signal ECG normal présentées dans la Figure 1.9, sont détaillées ci dessous.

Onde P : dépolarisation des oreillettes L'onde P correspond à la dépolarisation des oreillettes depuis le nœud sinusal vers le nœud atrio-ventriculaire. C'est l'onde qui précède le complexe QRS.

Complexe QRS : dépolarisation des ventricules Le complexe QRS est une imbrication de 3 ondes accolées qui suivent l'onde P et qui correspondent à la dépolarisation des ventricules. Par définition, l'onde Q est la première onde négative, l'onde R la première onde positive du complexe, et l'onde S la première onde négative après l'onde R. Toute onde supplémentaire, positive ou négative, sera appelée R', S', R'', etc. La figure 1.10 donne les différents aspects que peut prendre le QRS.

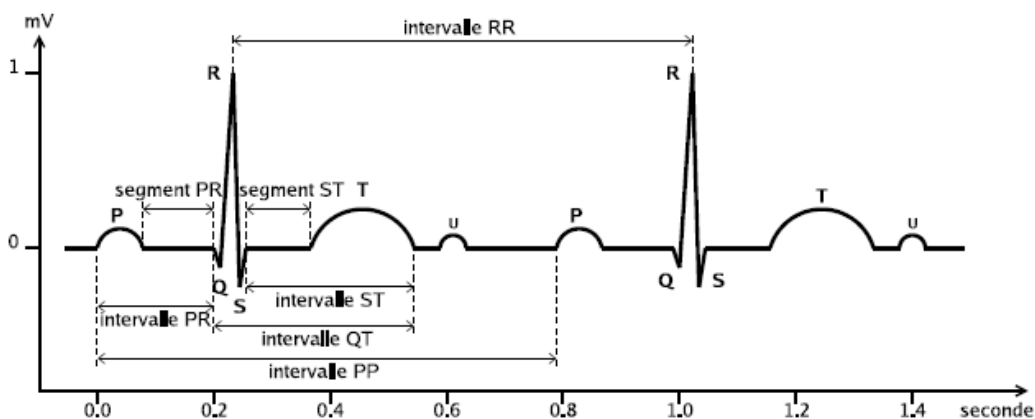


FIG. 1.9 : ECG normal avec les notations usuelles de l'électrocardiographie clinique.

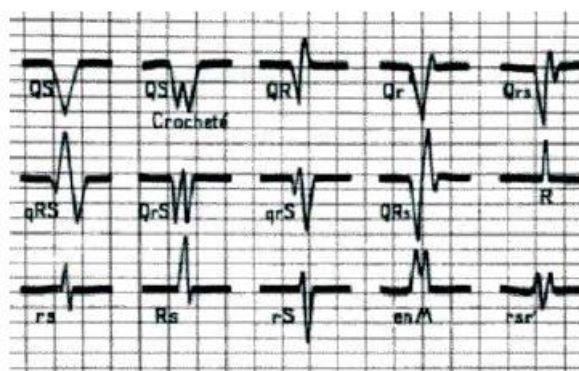


FIG. 1.10 - Schémas des différents aspects du complexe QRS et de leur notations.

Onde T : repolarisation des ventricules L'onde T correspond à la repolarisation des ventricules. Cette onde succède au complexe QRS après retour à la ligne isoélectrique (sauf pathologie particulière).

Onde U : repolarisation des fibres de Purkinje : L'onde U est une onde positive qui suit l'onde T, visible essentiellement en précordiale (surtout en V2 et V3), dont la signification est discutée (repolarisation prolongée des cellules du réseau de Purkinje ou à un facteur mécanique correspondant à la relaxation du myocarde).

Intervalle RR : fréquence des battements cardiaques L'intervalle RR correspond au délai entre deux dépolarisations des ventricules. C'est cet intervalle qui permet de calculer la fréquence cardiaque.

Intervalle PP : période de polarisation des oreillettes L'intervalle PP correspond au délai entre deux dépolarisations des oreillettes.

Segment PR : pause du nœud AV Le segment PR correspond au délai entre la fin de la dépolarisation des oreillettes et le début de celle des ventricules. C'est le temps pendant lequel l'onde de dépolarisation est bloquée au niveau du nœud AV.

Intervalle PR : durée de conduction auriculo-ventriculaire L'intervalle PR correspond au délai entre le début de la dépolarisation des oreillettes et celle des ventricules. C'est le temps de propagation de l'onde de dépolarisation jusqu'aux cellules myocardiques ventriculaires.

Intervalle QT : durée de systole ventriculaire Cet intervalle correspond au temps de systole ventriculaire, qui va du début de l'excitation des ventricules jusqu'à la fin de leur relaxation.

Segment ST : durée de stimulation complète des ventricules Le segment ST correspond à la phase pendant laquelle les cellules ventriculaires sont toutes dépolarisées, le segment est alors isoélectrique.



DI : déflexion intrinsèque : elle correspond à la fin de la dépolarisation de la paroi myocardique qui fait face à l'électrode exploratrice , donc à la rupture brutale de pente de la déflexion positive (sommet R).

Son délai d'apparition représente le temps que met l'onde de dépolarisation à parcourir le myocarde , depuis l'endocarde jusqu'à l'épicarde ;il est d'autant plus long que le myocarde est plus épais.

Cette DI se mesure sur les dérives unipolaires ,c'est-à-dire les dérives unipolaires des membres aVR, aVL,aVF ou plus habituellement les dérives précordiales qui explorent plus directement les ventricules : précordiales droites pour le calcul de la DI du ventricule droit ; précordiales gauches pour le calcul de la DI du ventricule gauche.

La mesure de DI se fait du début du QRS,c'est-à-dire de l'onde q ou du pied de l'onde R en l'absence de d'onde q,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 ;si le complexe QRS est polyphasique avec une onde R' ou R'',la mesure se fait du début de QRS au sommet de la dernière positivité R'ou R'' . (voir figure1.11)

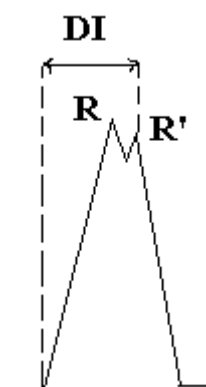


Fig 1.11: Mesure du temps d'apparition de la déflexion intrinsèque (DI)

1.5 Troubles du rythme et de la conduction cardiaque

Sous cette dénomination on regroupe les arythmies cardiaques et les blocs cardiaques. Le meilleur outil pour diagnostiquer une arythmie est l'électrocardiogramme [Lake, 1990]. Dans l'analyse de l'ECG, les pathologies ou anomalies sont détectées et classées en fonction de leur déviation par rapport au rythme idéal qu'est le rythme sinusal. Chaque déviation visible sur l'ECG peut être attribuée à une anomalie physiologique. Ainsi, les blocs cardiaques sont dus à un défaut de conduction de l'onde de dépolarisation à travers le myocarde différencié et les arythmies sont générées par un foyer ectopique prenant le relais ou supplantant le nœud sinusal. Ces pathologies ne sont pas exclusives, un patient peut être atteint d'arythmies et de blocs cardiaques. Les sections suivantes détaillent le rythme sinusal, les blocs cardiaques et certaines arythmies cardiaques.

1.5.1 Rythme sinusal

Le rythme sinusal est le rythme normal cardiaque. Il correspond à une activation physiologique des oreillettes, puis des ventricules, à partir du nœud sinusal. Son rythme est compris entre 60 à 80 battements par minute avec un intervalle régulier entre des battements normaux (figure 1.12). Le cœur s'accélère normalement lors de l'activité physique, dans les circonstances physiologiques qui exigent un surcroît de demande métabolique ou sous l'effet des émotions ou d'excitants tels que café, tabac.

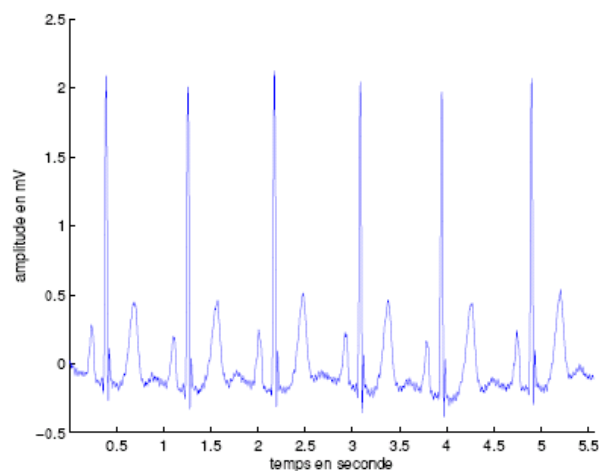


FIG. 1.12 - Exemple d'ECG normal.

1.5.2 Blocs cardiaques

Les blocs cardiaques sont dus à une rupture de conduction du myocarde indifférencié qui altère la dépolarisation du myocarde. Ces ruptures peuvent être plus ou moins sévères : freinantes (allongement du temps de parcours), intermittentes (un stimulus sur 2 ou 3 est conduit), ou complète (aucune conduction).

1.5.2.1 Blocs de branche

Le bloc de branche est dû au blocage de la dépolarisation dans une des branches du faisceau de His. Un bloc dans l'une ou l'autre branche provoque un retard dans la dépolarisation du ventricule auquel elle appartient. La dépolarisation des ventricules est désynchronisée et le complexe QRS est élargi (voir Figure 1.13).

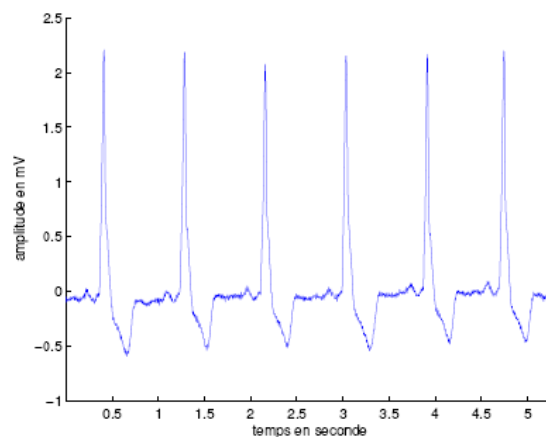


FIG. 1.13 : Exemple de bloc de branche gauche.

-Le bloc de branche est complet lorsque la largeur du QRS est supérieure à 0,12s, il est dit sinon incomplet.

Cette définition électrique n'est pas toujours corrélée avec la réalité anatomique (un bloc d'aspect complet peut en réalité être incomplet).

1.5.2.1.1 Bloc de branche gauche

Critères électrocardiographiques du BBG

- 1. Durée de QRS $> 0,12$ seconde;
- 2. Onde R élargie et crochetée : aspect en M dans les dérivations V5-V6 et I-aVL;
- 3. Absence habituelle des ondes q et s dans les mêmes dérivations;
- 4. Retard d'apparition de la déflexion intrinsécoïde ($>0,055$ secondes) en V5-V6;
- 5. Onde S profonde et large avec aspect QS ou rS dans les précordiales droites, déplacement vers la gauche de la zone de transition;
- 6. Anomalie secondaire de repolarisation : onde T inversée et parfois léger sous-décalage de ST dans les dérivations précordiales gauches et les dérivations I, aVL, onde T positive et segment ST légèrement surélevé dans les précordiales droites.

A droite (en V1), on enregistre le retard de dépolarisation du ventricule gauche, se traduisant par une onde R (positive) fine suivie d'une onde S (négative) **creuse, prolongée**, donnant l'aspect (q)RS classique.(voir figure 1.14)

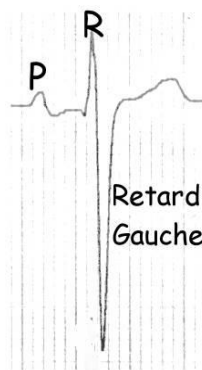


FIG1.14 : bloc de branche gauche en V1

A gauche (en V6), on enregistre un complexe QRS d'aspect RSR', l'onde R' étant la 2nde déflexion positive, correspondant au retard de dépolarisation du ventricule gauche.(voir figure 1.15).

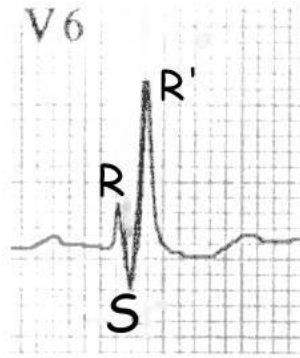


Fig 1.15 :bloc de branche gauche en V6.

1.5.2.1.2 Bloc de branche droit

Critères électrocardiographiques du BBD

- 1. Durée de QRS $> 0,12$ seconde (à mesurer dans la dérivation où le QRS paraît le plus large, ou sur l'ensemble des dérivations).
- 2. Aspect RSR', où habituellement $R' > R$, dans les dérivations précordiales droites, surtout V1.
- 3. Onde S large et empâtée (durée $> 0,04$ seconde ou $S > R$) dans les dérivations I et V6, parfois aussi en V5.
- 4. Retard de la déflexion intrinsécoïde $> 0,05$ seconde dans les précordiales droites (surtout V1).
- 5. Anomalie secondaire de repolarisation : onde T inversée dans les précordiales droites V1 et V2 et parfois segment ST légèrement sous-dénivelé.

A **droite** (en V1), on enregistre une 1ère déflexion positive, l'onde R, correspondant à la dépolarisation du ventricule gauche. Puis, on a une 2nde déflexion positive, l'onde R', correspondant à la dépolarisation retardée du ventricule droit. D'où l'aspect **RSR'** en V1 classiquement décrit dans le bloc de branche droit.(FIG1.16)



Fig 1.16 : bloc de branche droit en V1.

A **gauche** (en **V6**), on enregistre une 1^{ère} déflexion positive, l'onde R, correspondant à la dépolarisation du ventricule gauche. Puis, on a une déflexion négative **creuse, prolongée**, l'onde S, correspondant au **retard droit** (ou retard dans la dépolarisation du ventricule droit).(voir figure 1.17)

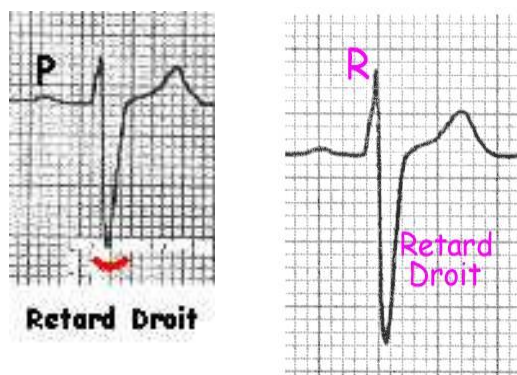


Fig 1.17 : bloc de branche droit en V6.

1.5.3 Arythmies

Les arythmies proprement dites relèvent de l'entrée en jeu d'un « foyer » ectopique qui peut se situer dans n'importe quelle portion du cœur, ou de la formation d'un circuit électrique (appelé réentrée notamment par les faisceaux de Kent), dont la localisation peut être auriculaire, jonctionnelle (entre oreillettes et ventricules) ou ventriculaire. Les paragraphes suivants présentent quelques arythmies dites supraventriculaires (auriculaires ou jonctionnelles) et des arythmies ventriculaires.

1.5.3.1 Arythmies supraventriculaires

1.5.3.1.1 La tachycardie supraventriculaire (TSV) (maladie de Bouveret) correspond à un mécanisme de réentrée qui peut soit être localisé à l'intérieur du nœud auriculo-ventriculaire (réentrée nodale), soit couvrir un large circuit empruntant dans un sens la voie de conduction normale (nœud AV et faisceau de His) et dans l'autre le faisceau de Kent. La tachycardie est rapide (entre 180 et 220 pulsations par minute) parfaitement régulière, à début et fin brusques. Les complexes QRS sont normaux et l'activité auriculaire se superpose à ces derniers ou les suit. Les accès se répètent à des intervalles variables, mais l'évolution reste bénigne car il n'existe pas, sauf association occasionnelle, d'anomalie des valves, du muscle cardiaque ou des coronaires.

1.5.3.1.2 La fibrillation auriculaire (FA) est une désorganisation totale de l'activité électrique auriculaire avec perte de la contraction mécanique rythmée des oreillettes. Les oreillettes sont parcourues de multiples ondes électriques, qui se propagent de façon anarchique à une très grande vitesse, supérieure à 350 par minute. Les ondes auriculaires de fibrillation sont en grande partie bloquées dans le nœud auriculo-ventriculaire qui joue le rôle de filtre, protégeant les ventricules d'une activité trop rapide. Le choc électrique externe est le moyen le plus efficace pour rétablir le rythme sinusal (emploi d'un défibrillateur). La FA peut être traitée par des soins médicamenteux (antiarythmiques) et s'ils se révèlent inappropriés, il devient nécessaire d'implanter un défibrillateur.

1.5.3.2 Arythmies ventriculaires

1.5.3.2.1 L'arythmie extrasystolique est la plus fréquente. Les extrasystoles sont des battements ectopiques, uniques ou répétés, provenant d'un seul ou de plusieurs foyers qui peuvent entraîner des sensations désagréables de ratés, de coups dans la poitrine, d'arrêts de cœur ou de palpitations. Les extrasystoles sont des phases systoliques en trop qui seront apparentes sur l'ECG par un complexe QRS large (Figure 1.18). Les extrasystoles ne constituent habituellement pas en elles-mêmes un facteur de gravité, leur pronostic dépend de l'état cardiaque qui peut être absolument normal (extrasystoles dites bénignes) ou pathologique. Lorsqu'il existe un double foyer ventriculaire, on parle de **doublet ventriculaire** (Figure 1.19).

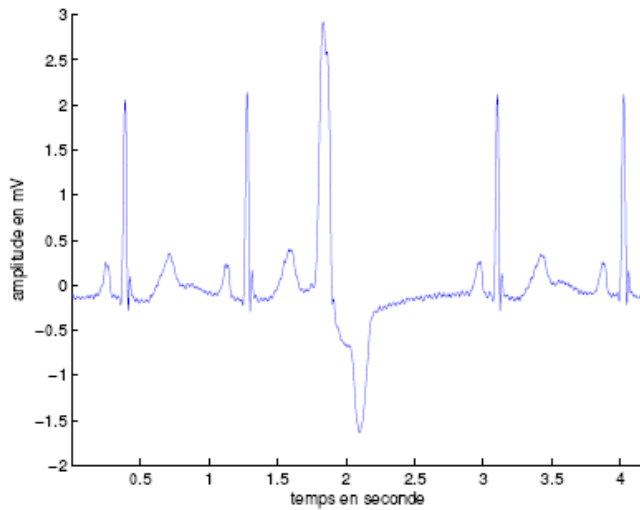


FIG. 1.18 - Exemple d'extrasystole.

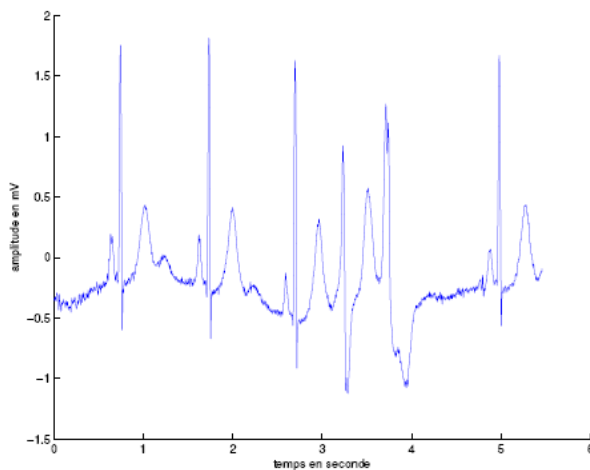


FIG. 1.19- Exemple de doublet ventriculaire.

1.5.3.2.2 Les bigéminismes et trigéminismes sont des rythmes à deux commandes. La commande de base (généralement sinusale) est interrompue par des battements d'origine ectopique. Lorsque l'on se trouve en présence d'un bigéminisme les QRS qui appartiennent au rythme de base sont suivis d'un QRS d'origine ectopique avec une succession de 1/1. On parle de trigéminisme lorsqu'on est en présence d'une succession 2/1. La figure 1.20 présente un exemple de rythme de bigéminisme dans le cas d'un rythme sinusal avec un foyer ectopique ventriculaire.

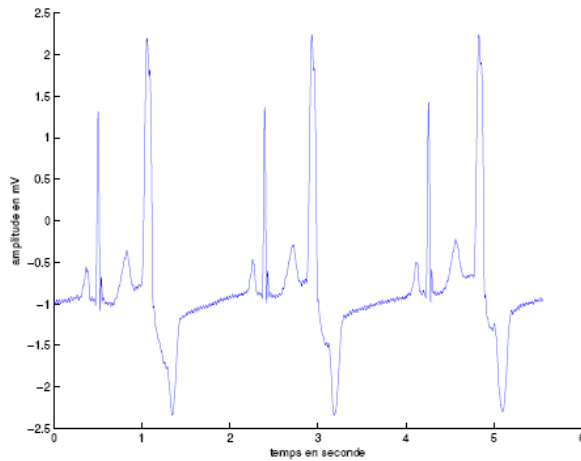


FIG. 1.19 - Exemple de rythme de bigéminisme.

1.5.3.2.3 Les tachycardies ventriculaires (TV) représentent les arythmies les plus graves, elles compromettent souvent l'hémodynamique et peuvent dégénérer en fibrillation ventriculaire létale. Les complexes QRS sont toujours élargis et le rythme est rapide. (Figure 1.21). L'accès de tachycardie ventriculaire persistante est une urgence médicale, l'arrêt de la crise pouvant être obtenu par injection intraveineuse d'un antiarythmique ou par choc électrique externe. Quand les récurrences de tachycardie ne sont pas prévenues par la médication antiarythmique, le recours à des méthodes non pharmacologiques est légitime : exérèse chirurgicale de la zone arythmogène, défibrillateur implantable.

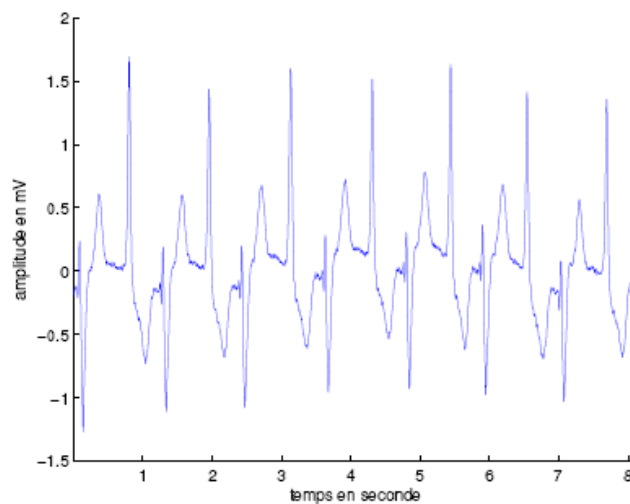


FIG. 1.21 - Exemple de tachycardie ventriculaire.

1.5.3.2.4 La fibrillation ventriculaire (FV) est la cause principale de mort subite. Elle peut apparaître d'emblée, comme manifestation d'instabilité électrique, dans les premières minutes ou heures d'un infarctus du myocarde aigu ou succéder à une autre arythmie ventriculaire chez les malades ayant une cardiopathie sévère. Les contractions complètement anarchiques des ventricules aboutissent très rapidement à une inefficacité cardio-circulatoire qui est létale en l'absence de manœuvres de réanimation (massage cardiaque, ventilation assistée, choc électrique externe). Ces dernières doivent être entreprises en quelques minutes pour éviter les complications cérébrales secondaires à une privation d'oxygène prolongée au niveau des cellules.(Figure1.22)

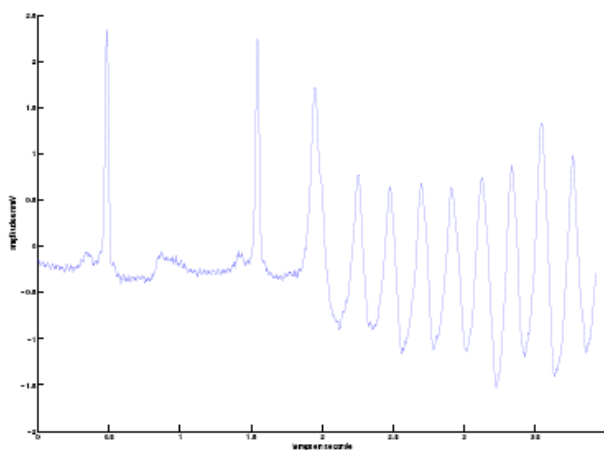


FIG. 1.22 - Exemple d'accès de fibrillation ventriculaire.

1.6 Diagnostic des arythmies

Le problème principal dans le diagnostic des arythmies cardiaques est de trouver le siège du foyer ectopique qui peut être soit supraventriculaire (auriculaire ou jonctionnel) ou ventriculaire. Pour ce faire le cardiologue décrit l'électrocardiogramme selon les ondes et intervalles définis précédemment. Nous allons présenter une des méthodes pour diagnostiquer des arythmies à partir d'ECG : l'analyse de la morphologie des QRS.

1.6.1 Analyse morphologique Les arythmies supraventriculaires, naissent au-dessus de la bifurcation du faisceau de His. Elles sont théoriquement caractérisées par des QRS fins car l'activation ventriculaire reste normale (donc rapide).

Les arythmies ventriculaires sont au contraire caractérisées par l'existence de complexes QRS élargis et déformés. L'activation naît en un point quelconque du myocarde ventriculaire et entraîne

une activation anormale de la masse ventriculaire. La durée du QRS est supérieure à 120 ms, avec un aspect de bloc de branche.

Dans certains cas, cette différenciation entre arythmie supraventriculaire ou ventriculaire peut être très difficile. C'est le cas par exemple des arythmies supraventriculaires associées à un bloc de branche préexistant, les QRS sont alors élargis quelque soit le foyer et peuvent être confondus avec une arythmie ventriculaire.[Portet 2005].

1.7 Conclusion

Ce chapitre nous a permis de définir les notions cardiologiques ainsi que le tracé utilisé dans ce mémoire qui est l'ECG et nous avons fini par présenter quelques arythmies comme les extrasystoles ventriculaires et comment apparaissent elles sur le tracé de l'électrocardiogramme pour permettre d'arriver à poser un diagnostic . Ces paramètres descriptifs seront utilisés par la suite dans la classification des différentes arythmies (extrasystoles ventriculaires, bloc de branche droit et bloc de branche gauche) par les réseaux de neurones. Le modèle neuronal sera décrit dans le chapitre suivant.

Chapitre 2

Principes des réseaux de neurones

2.1 Introduction

Dans ce chapitre nous présentons un aperçu global sur les réseaux de neurones (modèles biologiques et formels).

Les réseaux de neurones sont devenus en quelques années des outils précieux dans divers domaines tels que la reconnaissance de formes (images ou signaux), le diagnostic, le contrôle moteur, la traduction automatique, et la compréhension du langage.

Aujourd'hui de nombreux termes sont utilisés dans la littérature pour désigner le domaine des réseaux de neurones artificiels, comme connexionnisme ou neuromimétique. Connexionnisme et neuromimétique sont tous deux des domaines de recherche à part entière, qui manipulent chacun des modèles de réseaux de neurones artificiels, mais avec des objectifs différents. L'objectif poursuivi par les ingénieurs et chercheurs connexionnistes est d'améliorer les capacités de l'informatique en utilisant des modèles aux composants fortement connectés. Pour leur part, les neuromiméticiens manipulent des modèles de réseaux de neurones artificiels dans l'unique but de vérifier leurs théories biologiques du fonctionnement du système nerveux central.

La première approche a donc mené à la définition et l'étude de réseaux de neurones formels qui sont des réseaux complexes d'unités de calcul élémentaire interconnectées. Motivé par l'obtention d'algorithmes efficaces ne se préoccupant pas de la pertinence biologique. Nous nous plaçons du point de vue du premier groupe. En effet, bien que les réseaux de neurones formels aient été définis à partir de considérations biologiques, pour la plupart d'entre eux, de nombreuses caractéristiques biologiques (le temps, la mémoire, ...) ne sont pas prises en compte.

« On pense communément que “l’ordinateur de l’avenir” sera massivement parallèle et tolérera les erreurs. Toutefois la conception d’une telle machine s’étant avérée étonnamment difficile, nous aurions abandonné depuis longtemps si le cerveau n’était pas une preuve vivante que le traitement parallèle et tolérant les erreurs est possible et très efficace ».[John S. Denker, 1985]

Les réseaux de neurones artificiels sont fondés sur des modèles qui tentent d’expliquer comment les cellules du cerveau et leurs interconnexions parviennent, d’un point de vue globale, à exécuter des calculs complexes.

Ces systèmes qui stockent et retrouvent l’information de manière “similaire” au cerveau sont particulièrement adaptés aux traitements en parallèle de problèmes complexes comme la reconnaissance automatique de la parole, la reconnaissance de visages ou bien la simulation de fonctions de transfert. Ils offrent donc un nouveau moyen de traitement de l’information utilisé en reconnaissance de formes (vision, image, parole, etc).

2.2 Neurones biologiques et neurones formels

Les réseaux de neurones biologiques réalisent facilement un certain nombre d’applications telles que la reconnaissance de formes, 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 neurones formels s’inspirent de la biologie mais n’ont plus qu’une lointaine parenté avec leurs homonymes biologiques .ils en conservent cependant les caractéristiques principales, permettant ainsi de modéliser plus abstraitement les propriétés des neurones réels. Nous allons d’abord définir dans ses grandes lignes le neurone biologique. Nous aborderons ensuite la modélisation d’un neurone et de groupe de neurones.

2.2.1 Neurone biologique

2.2.1.1 structure d'un neurone biologique

Cette section décrit un modèle des neurones biologiques qui a servi à la mise en place des premiers neurones formels.(figure2.1)

Le cerveau se compose d'environ 10^{12} neurones (mille milliards), avec 1000 à 10000 synapses (connexions) par neurone.

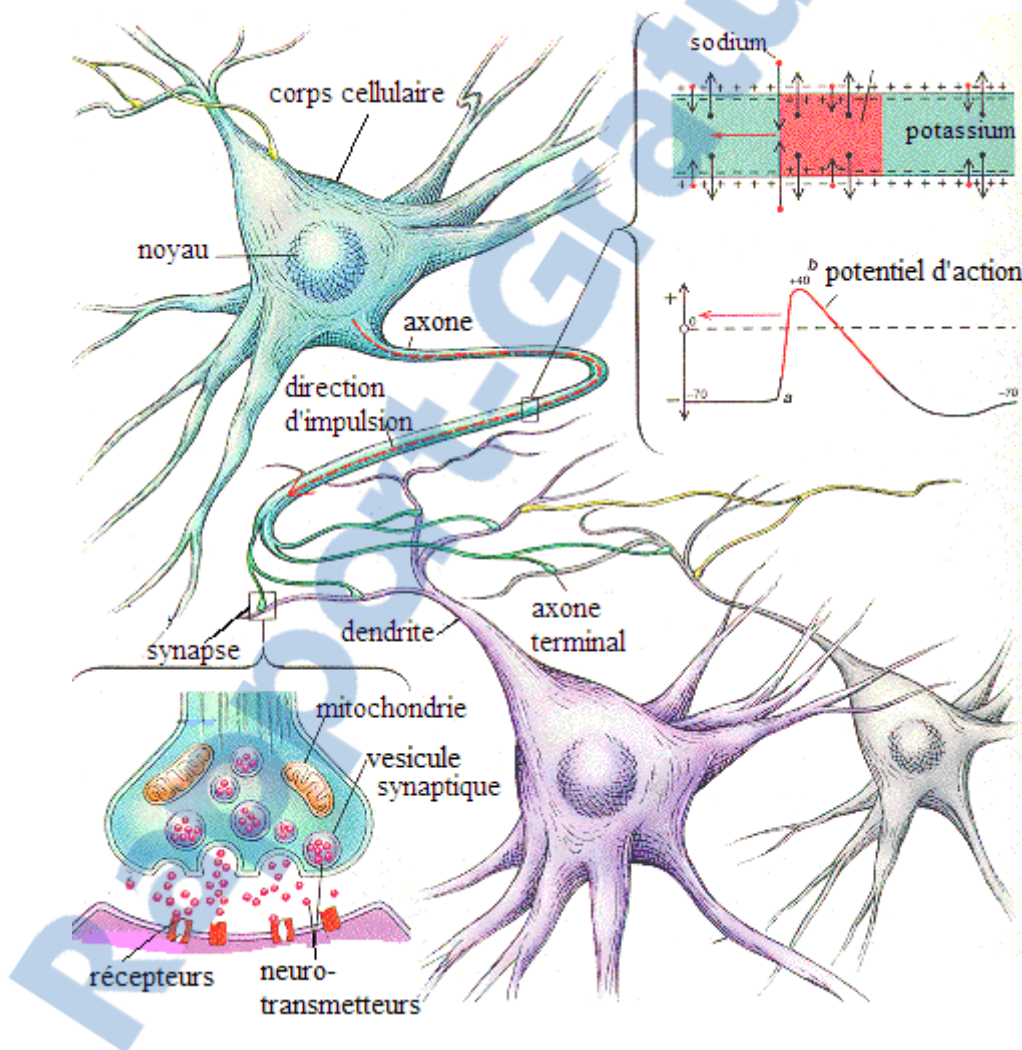


Figure2.1 *Modèle de neurone biologique.*

Figure extraite de:

<http://www.lacim.uqam.ca/~chauve/Enseignement/BIF7002/Rapports/Simon-Beaulne/neurones.htm>

Le neurone est une cellule composée d'un corps cellulaire qui contient le noyau, il renferme l'ADN possédant toutes les informations génétiques nécessaires pour l'élaboration des différentes activités biochimiques utiles au bon fonctionnement du neurone. Le corps cellulaire se ramifie pour former ce que l'on nomme les dendrites. Celles-ci sont parfois si nombreuses que l'on parle alors de chevelure dendritique ou d'arborisation dendritique (figure 2.2). C'est par les dendrites que l'information est acheminée de l'extérieur vers le soma, corps du neurone. L'information traitée par le neurone chemine ensuite le long de l'axone (unique) pour être transmise aux autres neurones. La transmission entre deux neurones n'est pas directe. En fait, il existe un espace intercellulaire de quelques dizaines d'Angstroms (10^{-9} m) entre l'axone du neurone afférent et les dendrites (on dit *une* dendrite) du neurone efférent. La jonction entre deux neurones est appelée la synapse.

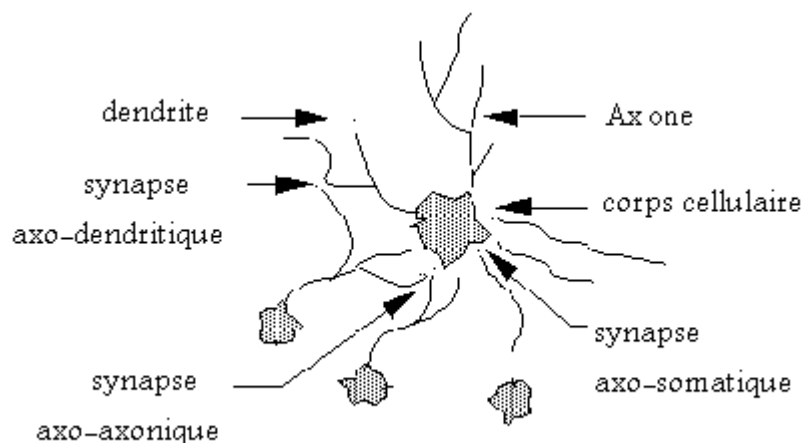


Figure 2.2 Un neurone biologique avec son arborisation dendritique

2.2.1.2 Création d'un potentiel d'action

Les neurones, comme toute cellule nerveuse vivante, possèdent un potentiel électrique intrinsèque (figure 2.3). celui-ci est dû à une différence de potentiel entre l'intérieur de la cellule et le milieu extérieur provoqué par un échange ionique (essentiellement des ions sodium et potassium) entre les deux milieux. La différence de potentiel mesurée entre le milieu intérieur de la cellule et le milieu extérieur est de -60 mV. le potassium est concentré à l'intérieur et le sodium est à l'extérieur.

Mais la différence de ces concentrations qui engendre la différence de potentiel, provoque aussi une migration des ions potassium vers le milieu extérieur, et symétriquement, des ions sodium vers l'intérieur de la cellule, dont l'effet peut être catastrophique au niveau cellulaire. Pour maintenir les concentrations à un niveau stable, et une telle différence de potentiel, la cellule fait appel à des pompes ioniques (Na^+ , K^+).

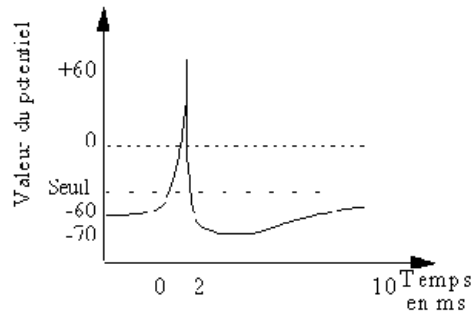


figure2.3 :potentiel d'action

La spécificité du neurone par rapport à d'autres cellules réside dans l'existence de variations brutales du potentiel de la membrane. Lorsque cette différence de potentiel atteint un certain seuil, il y a rupture des équilibres ioniques et création d'un potentiel d'action, le neurone décharge : son potentiel s'inverse soudainement (quelques millisecondes) et cette inversion se propage le long de l'axone (potentiel d'action). puis le potentiel revient à sa valeur de repos. Pendant cette récupération, le neurone ne peut plus décharger pendant une courte période de temps c'est le période réfractaire.

L'axone sert donc à conduire les potentiels d'action vers les autres neurones. Cette conduction se fait par inversion progressive et localisée de la différence de potentiel entre l'axone et le milieu extérieur.

La transmission de l'information entre l'axone d'un neurone amont et les dendrites d'un neurone aval s'effectue par l'intermédiaire de synapses. Du côté de l'axone de petites vésicules libèrent des neurotransmetteurs à l'arrivée du potentiel d'action (figure2.4). Ces neurotransmetteurs migrent du côté dendritique, où ils vont déclencher une petite variation de potentiel (le potentiel post-synaptique).

L'intensité de cette réponse traduit l'efficacité synaptique. A l'heure on suppose que la valeur de cette efficacité intervient de manière cruciale dans le stockage de

l'information, c'est-à-dire les processus de mémorisation. Une synapse peut notamment être excitatrice ou inhibitrice. Une synapse excitatrice tend à favoriser la décharge d'un neurone. Au contraire, une synapse inhibitrice diminue la propension d'un neurone à la décharge. Au niveau post synaptique, sur la membrane dendritique, se trouvent des récepteurs pour les neurotransmetteurs. suivant le type des neurotransmetteurs et le type des récepteurs, l'excitabilité du neurone va augmenter ou diminuer ce qui fera propager l'information ou non.

Les neurones collectent donc l'information provenant de leurs voisins plus ou moins lointains par les dendrites. Lorsque la somme de ces petites variations de potentiel est assez forte, le neurone décharge.

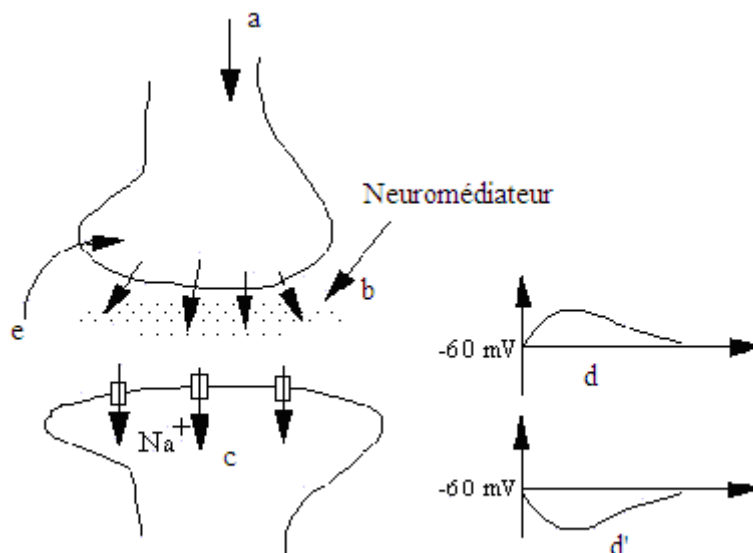


Figure2.4. Fonctionnement au niveau synaptique

- a) Arrivée d'un potentiel d'action.
- b) Libération du neuromédiateur dans l'espace synaptique.
- c) Ouvertures des canaux ioniques dues au neuromédiateur.
- d) Génération d'un potentiel évoqué excitateur.
- d') Génération d'un potentiel évoqué inhibiteur. Les synapses inhibitrices empêchent la génération de potentiel d'action.
- e) Fermeture des canaux, élimination ou recapture du neuromédiateur

2.2.2 Modèle de neurone formel

L'unité de traitement élémentaire dans un réseau connexionniste est appelée neurone formel.

C'est un modèle mathématique très simple. Ils sont appelés neurones formels pour leur similitude grossière avec les neurones biologiques.

Nous caractérisons un neurone par trois concepts : son état, ses connexions avec d'autres neurones et sa fonction de transfert.

2.2.2.1 L'état des neurones

Un neurone artificiel est un élément qui possède un état interne. Il reçoit des signaux qui lui permettent éventuellement, de changer d'état. Cet état peut être par exemple $\{0, 1\}$ où 0 sera interprété comme l'état inactif et 1 l'état actif.

Un neurone possède une fonction qui lui permet de changer d'état en fonction des signaux qu'il reçoit : c'est sa fonction de transfert.

L'état d'un neurone est fonction des états des neurones auxquels il est relié. Pour calculer l'état d'un neurone il faut donc considérer les connexions entre ce neurone et d'autres neurones et leur poids.

2.2.2.2 Les connexions

Une connexion est un lien établi explicitement entre deux neurones. Les connexions sont aussi appelées synapses, en analogie avec les neurones biologiques.

Une connexion entre deux neurones a une valeur numérique associée appelé poids de connexion.

2.2.2.3 Les poids des connexions

Le poids de connexion w_{ij} entre deux neurones j et i peut prendre des valeurs discrètes dans Z ou bien continues dans R . L'information qui traverse la connexion sera affectée par la valeur du poids correspondant. Une connexion avec un poids $w_{ij} = 0$ est équivalente à l'absence de connexion.

Le neurone formel est un automate caractérisé par un petit nombre de fonctions mathématiques. Il traite un signal d'entrée recueilli à travers ses connexions entrantes pour fournir un signal de sortie calculé par la fonction de transfert.

Son état est caractérisé par trois variables E , A et S calculées au moyen de trois fonctions successives : la fonction d'entrée f , la fonction d'activation g et la fonction de sortie h (figure2.5)

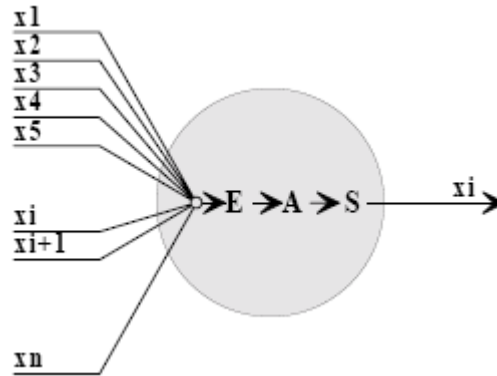


Figure2.5: *Modèle général du neurone formel.*

On a donc successivement : $E=f(x)$, $A=g(E)$ et $S=h(A)$. La définition de ses trois fonctions permet donc de fixer le comportement du neurone.

La fonction d'entrée f :

Dans un réseau de n neurones chacun d'entre eux reçoit en général de 1 à n stimulations en provenance des autres cellules. Ces valeurs, binaires ($\{0,1\}$ ou $\{-1,1\}$) ou réelles, sont en général pondérées puis additionnées par la fonction f :

$$E = f(x) = \sum w_j x_j$$

w_j est la pondération appliquée par le neurone à l'entrée j . Si le neurone dont on calcule l'entrée est le neurone i , on note w_{ij} la pondération correspondant à l'arc reliant le neurone j au neurone i . On parle alors du poids de la connexion.

La fonction d'activation g :

La fonction d'activation des neurones est généralement l'une des caractéristiques principales d'un type de réseau particulier. En effet l'état A d'un neurone peut être à valeurs binaires ou réelles et la fonction d'activation peut fortement varier d'un type de réseau à l'autre. Les deux familles principales sont la fonction signe, pour les neurones à états binaires, et la fonction sigmoïde pour les neurones à états réels. Ces deux fonctions sont généralement décalées par l'ajout d'un seuil fixe (noté θ_i ou, par analogie avec une stimulation constante, w_0). D'autres fonctions sont cependant utilisables telles que les fonctions gaussiennes ou les fonctions linéaires à seuil.

La fonction de sortie h :

L'une des plus importantes particularités de l'automate « neurone » est de n'avoir qu'une sortie pour un grand nombre d'entrées. Binaire ou réelle, la sortie S est en général du même type que les entrées. La fonction de sortie est le plus souvent la fonction identité et on a $S=A$.

2.2.3 Le neurone de Mc Culloch et Pitts

Historiquement le premier modèle de neurone formel a été défini en 1943 par Warren Mc Culloch et Walter Pitts .Ils laissent leurs noms à une modélisation du neurone biologique. Ceux sont les premiers à montrer que des réseaux de neurones formels simples peuvent réaliser des fonctions logiques, arithmétiques et symboliques complexes . Il s'agit d'un automate à état binaire ($\{0,1\}$) dont la fonction de transfert est la fonction de Heaviside appliquée à la somme des entrées pondérées auxquelles est ajouté un seuil θ_i (figure2.6)

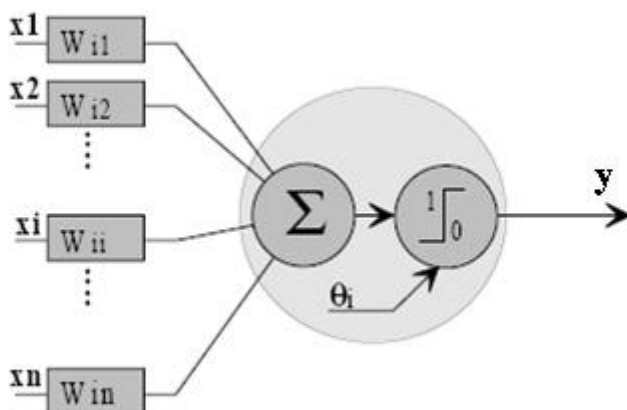


Figure2.6: *Le neurone de Mc Culloch et Pitts*

Ce type de neurone n'était pas doté de fonctions d'apprentissage. En 1957, Frank Rosenblatt a comblé cette lacune avec le Perceptron. Ce réseau est en effet capable d'apprendre une classification sur les entrées.

Le modèle neuronal de Mc Culloch et Pitts, bien que très simple, est encore utilisé pour certaines familles de réseaux.

$$y = f(\sum w_j x_j - \theta)$$

Pour un nombre compris entre $j=1\dots$ et n , L'action d'un neurone formel est de faire la somme, pondérée par les poids synaptiques (w_1, \dots, w_n) , des états des neurones qui lui sont connectés, puis de comparer cette somme à un seuil θ . Si le résultat est supérieur au seuil, alors la valeur de sortie est 1 sinon la valeur est 0. cette opération est prise en compte par une fonction de transfert.

Ce modèle peut être complexifié en prenant comme fonction de transfert une fonction sigmoïde qui est de loin la fonction la plus utilisée dans les réseaux de neurones artificiels. L'état des neurones n'est donc plus binaire $\{0, 1\}$, mais peut prendre n'importe quelle valeur dans un certain domaine.

Plus généralement, une fonction est dite sigmoïdale si elle est infiniment dérivable, non linéaire, bornée, croissante et que sa dérivée admet un et un seul maximum global. On peut également considérer qu'il s'agit d'une approximation infiniment dérivable de la fonction à seuil de Heaviside.

Dans sa première version le neurone formel était donc implémenté avec une fonction à seuil mais plusieurs versions existent. Ainsi le neurone Mc Culloch et Pitts a été généralisé de différentes manières, en choisissant d'autres fonctions de transfert ou d'activation, comme les fonctions linéaires et les sigmoïdes. (figure 2.7)

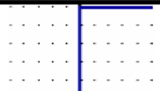
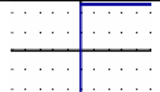
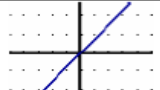


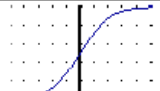
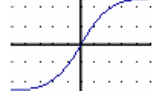
Catégorie	Type	Relation	Allure	Dérivées
Seuil	Binaire (fonction de Heaviside)	$f(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases}$		-
	Signe	$f(x) = \begin{cases} -1 & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases}$		-
	Stochastique (binaire ou signe) <i>(T est appelé température)</i>	$f(x, T) = 1$ avec la probabilité $\frac{1}{1 + e^{-\frac{x}{T}}}$	-	-
Linéaire	Identité	$f(x, k) = k.x$		$f'(x, k) = k$
	Saturé positif	$f(x, k) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x \geq \frac{1}{k} \\ k.x & \end{cases}$		$f(x, k) = \begin{cases} 0 & \text{si } x < 0 \\ 0 & \text{si } x \geq \frac{1}{k} \\ k & \end{cases}$
	Saturé symétrique	$f(x, k) = \begin{cases} -1 & \text{si } x < -\frac{1}{k} \\ 1 & \text{si } x \geq \frac{1}{k} \\ k.x & \end{cases}$		$f(x, k) = \begin{cases} 0 & \text{si } x < -\frac{1}{k} \\ 0 & \text{si } x \geq \frac{1}{k} \\ k & \end{cases}$
Sigmoïde	Positive (type logistique)	$f(x, k) = \frac{1}{1 + e^{-k.x}}$		$f'(x, k) = \frac{k}{2 + e^{-k.x} + e^{k.x}}$
	Symétrique (type tanh)	$f(x, k) = \frac{2}{1 + e^{-k.x}} - 1$		$f'(x, k) = \frac{2.k}{2 + e^{-k.x} + e^{k.x}}$

Figure2.7 : les fonctions d'activation

En résumé, un neurone formel réalise simplement une somme pondérée de ces entrées, ajoute un seuil à cette somme et fait passer le résultat par une fonction de transfert pour obtenir sa sortie.

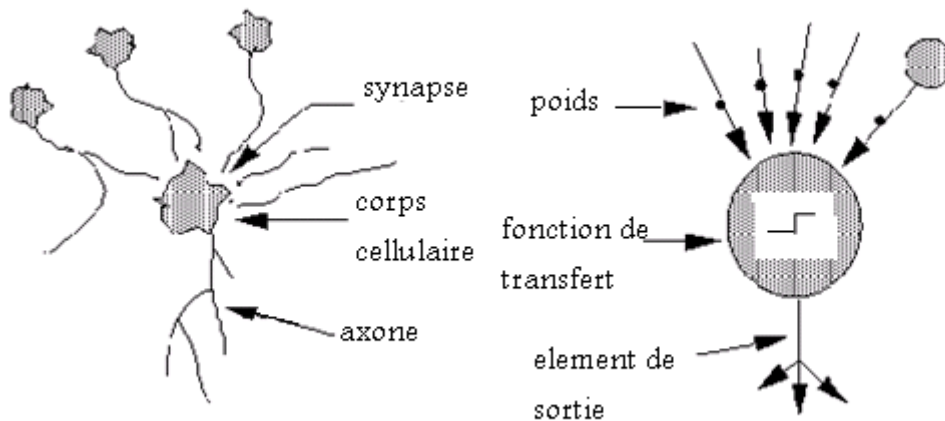


Figure2. 8. Mise en correspondance neurone biologique / neurone formel

2.3 Historique des réseaux de neurones

Une fois définis les éléments de base que sont les neurones, il faut s'attacher maintenant à construire des structures de réseaux leur permettant de résoudre certains problèmes.

Les réseaux de neurones artificiels sont des réseaux fortement connectés de processeurs élémentaires fonctionnant en parallèle. Chaque processeur élémentaire calcule une sortie unique sur la base des informations qu'il reçoit. Toute structure hiérarchique de réseaux est évidemment un réseau.

Un réseau de neurones formels est constitué d'un grand nombre de cellules de base interconnectées.

Citons quelques étapes dans la formalisation des réseaux de neurones :

Le modèle de Mc Culloch et Pitts [McCulloch 1943] a servi de base aux réseaux de neurones actuels. Parallèlement un neurophysiologiste met en avant l'importance des synapses dans l'apprentissage et a mis en oeuvre une loi de fonctionnement pour l'apprentissage sur les réseaux de neurones connue plus tard sous le nom de loi de Hebb. La première application concrète des réseaux de neurones artificiels est survenue vers la fin des années 1950 avec l'invention du réseau dit «perceptron» par un dénommé Frank Rosenblatt.

Rosenblatt et ses collègues ont construit un réseau et démontré ses habilités à reconnaître des formes.

Malheureusement, il a été démontré par la suite que ce perceptron simple ne pouvait résoudre qu'une classe limitée de problème. Environ au même moment, Bernard Widrow et Ted Hoff ont proposé un nouvel algorithme d'apprentissage pour entraîner un réseau adaptatif de neurones linéaires, le modèle de l'ADALINE (ADaptive LINear Element) dont la structure et les capacités sont similaires au perceptron. Ce sera le modèle de base des réseaux multicouches.

Vers la fin des années 1960, un livre publié par Marvin Minsky et Seymour Papert [Minsky 1988] est venu jeter beaucoup d'ombre sur le domaine des réseaux de neurones. Entre autres choses, ces deux auteurs ont démontré les limitations des réseaux développés par Rosenblatt et Widrow-Hoff. Beaucoup de gens ont été influencés par cette démonstration qu'ils ont généralement mal interprétée. Ils ont conclu à tort que le domaine des réseaux de neurones était un cul de sac et qu'il fallait cesser de s'y intéresser.

Heureusement, certains chercheurs ont persévéré en développant de nouvelles architectures et de nouveaux algorithmes plus puissants. En 1972, Teuvo Kohonen et James Anderson ont développé indépendamment et simultanément de nouveaux réseaux pouvant servir de mémoires associatives. Également, Stephen Grossberg a investigué ce qu'on appelle les réseaux auto-organisés.

Dans les années 1980, une pierre d'achoppement a été levée par l'invention de l'algorithme de rétropropagation du gradient de l'erreur. Cet algorithme est la réponse aux critiques de Minsky et Papert formulées à la fin des années 1960. C'est ce nouveau développement, généralement attribué à David Rumelhart et James McClelland, mais aussi découvert plus ou moins en même temps par Paul Werbos et par Yann LeCun, qui a littéralement ressuscité le domaine des réseaux de neurones. Depuis ce temps, c'est un domaine où bouillonne constamment de nouvelles théories, de nouvelles structures et de nouveaux algorithmes.

2.4 Différentes architectures des réseaux de neurones

Architecture est le terme le plus général pour désigner la façon dont sont disposés et connectés les différents neurones qui composent un réseau. On parle également de topologie (terme emprunté de la théorie des graphes).

Les architectures des réseaux de neurones s'inspirent de l'organisation neuronale du cerveau humain. Dans les réseaux de neurones artificiels de nombreux processeurs appelés cellules ou unités, capables de réaliser des calculs élémentaires, sont structurés en couches successives capables d'échanger des informations au moyen de connexions qui les relient. On dit de ces unités qu'elles miment les neurones biologiques. Ils en reprennent quelques grands principes :

- le **parallélisme** : les neurones sont des entités réalisant une fonction très simple, mais ils sont très fortement interconnectés entre eux, ce qui rend le traitement du signal massivement parallèle.
- les **poids synaptiques** : les connexions entre les neurones ont des poids variables, qui déterminent la force de l'interaction entre chaque paire de neurones.
- l'**apprentissage** : ces coefficients synaptiques sont modifiables lors de l'apprentissage, dans le but de faire réaliser au réseau la fonction désirée.

On peut classer les réseaux de neurones artificiels en deux grandes catégories :

- Les réseaux feed –forward ou réseaux non bouclés : ce sont des réseaux dans lesquels l'information se propage de couche en couche sans retour en arrière possible.
- Les réseaux feed –back : appelés aussi réseaux récurrents ce sont des réseaux dans lesquels il y a retour en arrière de l'information.

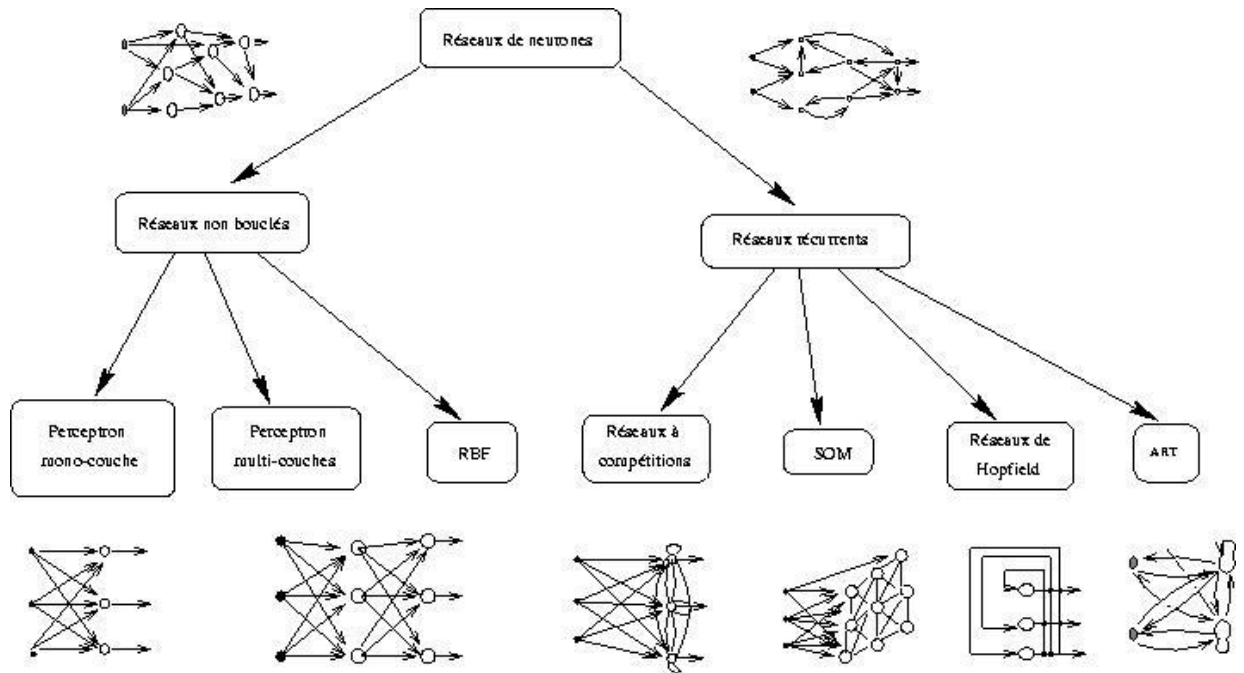


Figure 2.9 : les grandes architectures des réseaux de neurones.

Figure extraite du site : www.sm.u-bordeaux2.fr

2.4.1 Les réseaux Feed -back

Appelés aussi “réseaux récurrents”, ce sont des réseaux dans lesquels il y a retour en arrière de l’information .

2.4.1.1 Les cartes auto – organisatrices de Kohonen

Ce sont des réseaux à apprentissage non supervisé qui établissent une carte discrète , ordonnée topo logiquement , en fonction des patterns d’entrée . Le réseau forme aussi une sorte de treillis dont chaque nœud est un neurone associé à un vecteur de poids .La correspondance de chaque vecteur de poids est calculée pour chaque entrée .Par la suite le vecteur de poids ayant la meilleure corrélation,ainsi que certains de ses voisins,vont être modifiés afin d’augmenter encore cette corrélation.

2.4.1.2 Les réseaux de Hopfield

Les réseaux de Hopfield sont des réseaux récurrents et entièrement connectés. Dans ce type de réseau chaque neurone est connecté à un autre neurone et il n’y a aucune différenciation entre neurones d’entrée et de sortie. ils fonctionnent comme une mémoire associative non linéaire et sont capables de trouver un objet stocké en fonction de

représentations partielles ou bruitées .L'application principale des réseaux de Hopfield est l'entrepôt de connaissances mais aussi la résolution de problèmes d'optimisation. Le mode d'apprentissage utilisé ici est le mode non supervisé [Hopfield 1982].

2.4.1.3 Les ART

Les réseaux ART (Adaptive Resonance Theorie ,proposés par Grossberg) sont des réseaux à apprentissage par compétition. Le problème majeur qui se pose dans ce type de réseaux est le dilemme « stabilité/plasticité ».En effet dans l'apprentissage par compétition , rien ne garantit que les catégories formées vont rester stables. La seule possibilité, pour assurer la stabilité, serait que le coefficient d'apprentissage tende vers zéro, mais le réseau perdrait alors sa plasticité. Les ART ont été conçus spécifiquement pour contourner ce problème. Dans ce genre de réseau, les vecteurs de poids ne seront adaptés que si l'entrée fournie est suffisamment proche d'un prototype déjà connu par le réseau, on parlera alors de résonance. A l'inverse si l'entrée s'éloigne trop des prototypes existants, une nouvelle catégorie va alors se créer, avec pour prototype, l'entrée qui a engendré sa création. Il existe deux types d'ART : les ART-1 pour des entrées binaires et les ART-2 pour des entrées continues. Le mode d'apprentissage des ART peut être supervisé ou non.

Ils permettent un apprentissage continu et non supervisé avec création de nouvelles connexions lorsqu'une nouvelle classe est détectée. Chaque nouvelle entrée contribue à faire évoluer le réseau.

2.4.2 Les réseaux Feed - forward (ou réseaux non bouclés)

2.4.2.1 Le Perceptron

Introduit par par Frank Rosenblatt en 1958 et largement étudié par Minsky et Papert, le perceptron fut l'un des premiers systèmes adaptatifs. En effet l'idée principale de Rosenlatt a été d'incorporer au modèle de Mc Culloch et Pitts une phase d'apprentissage.

C'est un des premiers réseaux de neurones, conçu en 1958 par Rosenblatt. Il est linéaire et monocouche. Il est inspiré du système visuel. La première couche (d'entrée) représente la rétine. Les neurones de la couche suivante (unique, d'où le qualificatif de monocouche)

sont les cellules d'association, et la couche finale les cellules de décision. Les sorties des neurones ne peuvent prendre que deux états (-1 et 1 ou 0 et 1).

La structure générale du réseau se compose de trois éléments : (figure2.10).

Une rétine : constituée de plusieurs cellules ou neurones formels, sur lesquels s'inscrit le stimulus. (Les neurones d'entrée sont assimilés aux neurones de la rétine.)

Une couche de neurones d'association : chacune de ces cellules peut être connectée à des cellules de la rétine, à d'autres cellules d'association, et à des cellules de décision. Elle fait la somme des impulsions qui lui viennent des cellules auxquelles elle est connectée. La somme effectuée est comparée à un seuil. Il s'ensuit une réponse de type tout ou rien.

Une couche de neurones de sortie (ou de décision) : elles représentent la sortie du perceptron.

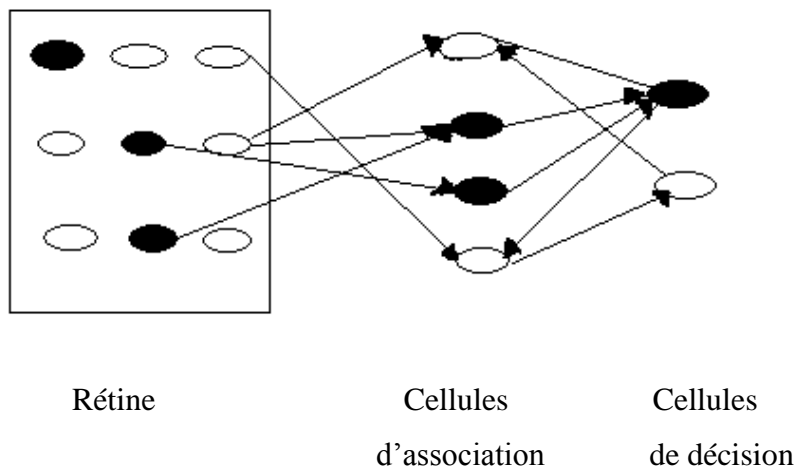


Figure2.10 :schéma d'un Perceptron.

Limites :

En 1969, Papert et Minsky ont démontré les limites du perceptron classique, incapable, par exemple de simuler la fonction ou exclusif (XOR).

Le perceptron ne converge que si les classes sont linéairement séparables. En particulier le fameux problème du OU exclusif n'est pas résolu. En effet, un seul neurone ne peut effectuer ce calcul, qui nécessite des surfaces de décisions non linéaires.(figure2.11)

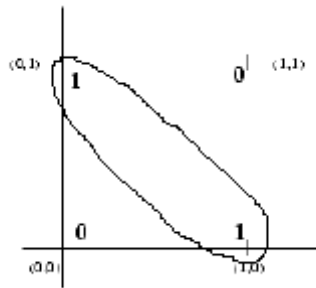


Figure 2.11 : le problème du « OU exclusif » en dimension 2. Une fonction linéaire (droite) ne peut séparer les 1 qui appartiennent à la même classe des 0 qui appartiennent à une autre classe.

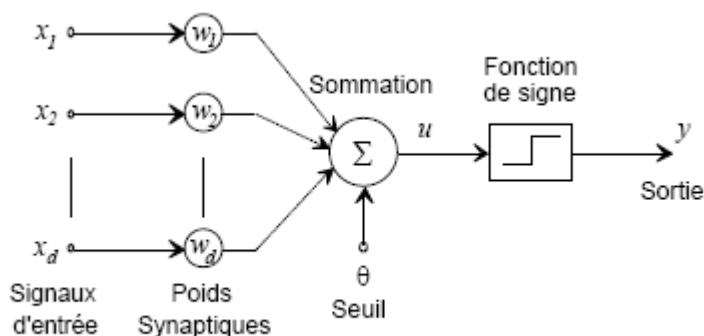


Figure 2.12 : représentation d'un perceptron

Le perceptron consiste en un seul neurone qui possède un seuil ainsi qu'un vecteur de poids synaptiques ajustables, tout comme le modèle de neurone de McCulloch & Pitts (figure 2.12).

Les paramètres du perceptron, c'est-à-dire les poids synaptiques des neurones, peuvent être déterminés grâce à un entraînement supervisé, effectué sur un ensemble de formes préclassifiées. La règle d'apprentissage du perceptron, développée originellement par Rosenblatt, est assurée de converger si les données sont linéairement séparables.

C'est cette règle de décision qui est réalisée dans le perceptron, grâce à l'utilisation de la fonction de signe comme fonction d'activation du neurone. La sortie du neurone vaut alors +1 si l'objet X appartient à la classe, et -1 dans le cas contraire.

2.4.2.2 Perceptron multi couches

L'exemple des réseaux de neurones le plus utilisé est le perceptron multi couches.

Dans ce réseau les neurones sont organisés suivant une couche d'entrée, une couche de sortie et une ou plusieurs couches cachées (figure 2.13) qui assurent la connexion entre

l'entrée et la sortie. Chaque neurone dans une couche est connecté à tous les neurones de la couche précédente et de la couche suivante (excepté pour les couches d'entrée et de sortie) et il n'y a pas de connexions entre les cellules d'une même couche.

-La couche d'entrée reçoit les données sources que l'on veut utiliser pour l'analyse. Dans le cas de l'aide au diagnostic médical cette couche recevra les symptômes. Sa taille est donc directement déterminée par le nombre de variables d'entrée.

-La couche de sortie donne le résultat obtenu. Dans le cas de l'aide au diagnostic médical, cette couche donne le diagnostic sous forme d'une probabilité par exemple entre 0 (sujet sain) et 1 (diagnostic confirmé). Sa taille est directement déterminée par le nombre de variables (ou diagnostic) qu'on veut en sortie.

-la couche intermédiaire est dite cachée dans le sens qu'elle n'a aucune utilité intrinsèque pour le réseau de neurones et n'a pas de contact direct avec l'extérieur.

Le principe de fonctionnement est le suivant : nous disposons initialement d'une base de connaissances constituée des couples de données entrées \ sorties (symptômes \ pathologie) et nous souhaitons utiliser cette base de données pour configurer et entraîner le réseau de neurones, grâce à un algorithme, à reproduire les associations constatées entre les entrées et les sorties de la base de connaissances.

Les fonctions d'activation des couches cachées sont en général non linéaires principalement les sigmoïdes. mais il n'y a pas de règles à respecter. Le choix du nombre de couches cachées et de leur taille n'est pas implicite et doit être ajusté.

L'introduction d'une couche intermédiaire permet de s'affranchir du problème de séparabilité linéaire non résolu par le perceptron.

Cette solution avait été suggérée par Rosenblatt, mais il n'avait pas réussi à trouver un algorithme pour l'apprentissage des poids. Les réseaux multi couches se sont développés à partir de 1985 sous l'impulsion de Widrow, avec la mise au point d'un algorithme de rétropropagation de l'erreur.

L'algorithme d'apprentissage par rétropropagation impose l'utilisation d'une fonction continue. Cet algorithme consiste à comparer la sortie obtenue et la sortie désirée, puis à propager la différence en sens inverse en modifiant les poids entre chaque couche. Il peut résoudre des problèmes non-linéairement séparables et des problèmes logiques compliqués, et notamment le fameux problème du XOR.

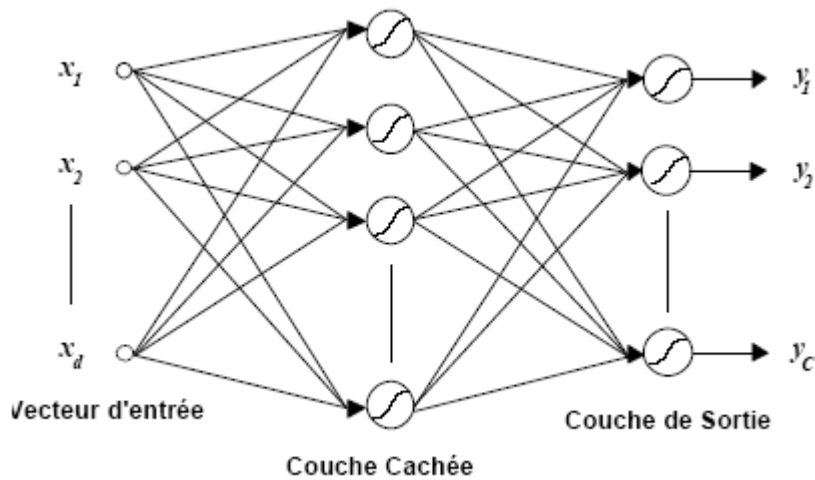


Figure 2.13: réseau multi couches (Perceptron multi couches à une couche cachée).

L'exemple de réseaux de neurones multi couches le plus utilisé reste le perceptron multi couches (Multi Layer Perceptron).

Par convention, les neurones d'entrée ont toujours une fonction d'activation «identité» laissant passer l'information sans la modifier. En ce qui concerne le neurone de sortie, on peut lui associer une fonction d'activation linéaire ou non, dérivable ou non, suivant la nature du problème à résoudre. En ce qui concerne la fonction d'activation associée aux neurones cachés, on utilise dans le cadre de cette thèse une fonction d'activation de la famille des sigmoïdes.

La fonction d'activation des neurones doit absolument être non linéaire, car sans elle, le perceptron multicouches ne ferait qu'implanter une série de transformations linéaires consécutives, qui pourraient dès lors se réduire à une seule. Grâce à l'utilisation de fonctions d'activations non linéaires, le perceptron multicouches est à même de générer des fonctions discriminantes non linéaires. L'algorithme d'apprentissage du perceptron multicouches, connu sous le nom d'*algorithme de rétro-propagation*, nécessite toutefois que les fonctions d'activations des neurones soient continues et dérivables. Cette condition mise à part, celles-ci peuvent cependant être quelconques,

offrant ainsi une grande liberté de modélisation. Les fonctions qui sont le plus couramment utilisées sont probablement les fonctions de type sigmoïdal.

Plus généralement, une fonction est dite **sigmoïdale** si elle est infiniment dérivable, non linéaire, bornée, croissante et que sa dérivée admet un et un seul maximum global. On peut également considérer qu'il s'agit d'une approximation infiniment dérivable de la fonction à seuil de Heaviside.

2.4.2.3 Les réseaux à fonctions radiales

Ce sont les réseaux que l'on nomme aussi RBF (Radial Basic Functions). L'architecture est la même que pour les PMC, Cependant les fonctions de bases utilisées sont des fonctions Gaussiennes. Les RBF seront donc employés dans les mêmes types de problèmes que les PMC à savoir, en classification et en approximation de fonctions, particulièrement. L'apprentissage le plus utilisé pour les RBF est le mode hybride et les règles sont soit, la règle de correction de l'erreur soit, la règle d'apprentissage par compétition.

2.5 La rétropropagation

L'algorithme de rétropropagation permet de faire l'apprentissage sur les réseaux multi couches. Il a été proposé simultanément par Le Cun et par Rumelhart et McClelland. Le principe de l'algorithme est de rétropropager l'erreur entre l'état des neurones de sortie et la réponse désirée sur les poids des connexions des couches antérieures.

L'algorithme d'apprentissage est le suivant :

1. Initialisation des poids du réseau
2. Présentation du vecteur d'apprentissage à l'entrée du réseau (couche d'entrée)
3. Calcul du vecteur de sortie S : le système propage les activités neuronales (état excité ou inhibé des neurones) à travers le réseau. Chaque neurone calcule la somme pondérée de ses entrées et transmet le résultat par une fonction de type sigmoïde pour produire sa

valeur de sortie. Le vecteur de sortie S est le résultat du calcul de la dernière couche (couche de sortie).

4. Calcul de l'erreur. On compare alors les valeurs de sortie actuelles S avec les valeurs désirées de R appelé vecteur de référence (ou encore vecteur cible).

On en déduit l'erreur J , définie comme la somme des carrés des différences observée sur chaque unité de la couche de sortie.

5. Rétropropagation de l'erreur : l'algorithme de rétropropagation du gradient permet récursivement de rétropropager l'erreur de la couche de sortie vers les couches cachées jusqu'à la première couche du réseau.

6. Modification des poids : les poids de chaque neurone sont modifiés, soit à chaque présentation d'un vecteur d'apprentissage (méthode du gradient stochastique), soit après cumul de l'erreur pour un certain nombre de vecteurs d'apprentissage (méthode du gradient standard).

2.6 Propriétés des réseaux de neurones

Les réseaux de neurones à couches ont la propriété générale d'être des approximateurs universels parcimonieux. Il s'agit en fait de deux propriétés distinctes détaillées ci-dessous.

2.6.1 La propriété d'approximation universelle

La propriété d'approximation universelle a été démontrée par Cybenko et Funahashi et peut s'énoncer de la façon suivante :

« Toute fonction bornée suffisamment régulière peut être approchée uniformément, avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un réseau de neurones comportant une couche de neurones cachés en nombre fini, possédant tous la même fonction d'activation, et un neurone de sortie linéaire . »

Les réseaux de neurones peuvent être considérés comme une classe d'approximateurs universels. Il a été ainsi démontré qu'un perceptron multicouches avec une seule couche cachée pourvue d'un nombre suffisant de neurones, peut approximer n'importe quelle fonction avec la précision souhaitée.

2.6.2 La propriété de parcimonie

Lorsque l'on cherche à modéliser un processus à partir des données, on s'efforce toujours d'obtenir les résultats les plus satisfaisants possibles avec un nombre minimum de paramètres ajustables.

il faut peu de paramètres pour spécifier un réseau de neurones ;les RN peuvent supporter un très grand nombre d'entrées.

2.6.3 La capacité d'apprentissage

L'un des intérêts principaux des réseaux de neurones réside dans leur capacité d'apprentissage. Il est en effet possible (par analogie avec les synapses biologiques) de stocker de l'information dans un réseau par l'intermédiaire de la modification des poids synaptiques entre les neurones. Le résultat de cette propriété est qu'il suffit de présenter une base d'apprentissage au réseau pour que, non seulement il reconnaisse les éléments de cette base, mais aussi qu'il généralise cette connaissance à des éléments non appris.

2.7 Applications des réseaux de neurones

Se trouvant à l'intersection de différents domaines (informatique, électronique, science cognitive, neurobiologie), les RN ont des applications dans plusieurs domaines :

Industrie : contrôle de qualité , diagnostic de panne , corrélations entre les données fournies par différents capteurs , analyse de signatures ou d'écriture manuscrite.

Finance :prévision et modélisation du marché (cours de monnaies),sélection d'investissements , attribution des crédits .

Télécommunications et informatique : analyse du signal , élimination du bruit , reconnaissance de formes (bruits , images , paroles) , compression de données.

Environnement : évaluation des risques , analyse chimique prévisions et modélisation météorologiques , gestion des ressources.....

2.7.1 Intérêt des réseaux de neurones dans le diagnostic médical

Le Cun a étudié l'apport de l'approche connexionniste au problème du diagnostic des douleurs abdominales , telles que l'appendicite . Il a utilisé une architecture multicouche et étudié différents modèles ,et obtenu des performances de l'ordre de 80% . L'objectif principal de la thèse de Shaltenbrand concernait l'étude et la réalisation d'un système d'analyse automatique des stades du sommeil humain par réseau de neurones multi couches. L'intérêt essentiel était de décharger les chercheurs travaillant sur le sommeil de nombreuses heures de lecture des tracés. il s'agissait de concevoir et de développer des techniques fiables d'évaluation quantitative des signaux du sommeil à l'aide des réseaux de neurones et d'outils d'analyse et de traitement d'information .

Doyle et Al. ont utilisé la structure multi couches. Ce réseau a été appliqué avec succès pour le diagnostic des maladies du foie.

Une étude comparative a été menée par l'équipe de Pesonen en vue d'évaluer les performances de différents modèles neuronaux dont les cartes de Kohonen et les réseaux multi couches . Cette étude a été appliquée au diagnostic de l'appendicite aiguë .Elle a eu le mérite de confirmer la supériorité du mode supervisé dans le domaine de l'aide à la décision .

Des travaux très actifs sont menés pour l'interprétation automatique du cancer du sein . L'étude réalisée par Kocur et Al. avait pour but la combinaison de réseaux de neurones avec l'analyse en ondelettes. La spécificité de ce travail consistait en la détermination , dans un premier temps ,des coefficients de transformée en ondelettes obtenus à partir d'images mammographiques digitalisées et , dans un deuxième temps , en l'utilisation de ces mêmes coefficients en entrée d'un réseau multi couches en vue de détecter la présence de cancers du sein .Une précision totale de 88 % a été obtenue .

Durant les deux dernières décennies, des études ont montré que lorsqu'une approche de diagnostic des pathologies cardiaques est adoptée, les RN ont donné des résultats satisfaisants. L'étude réalisée par Maglaveras et al. [Maglaveras1 2001] avait pour but d'identifier les ischémies pour un classifieur neuronal multicouche.

La spécificité de ce travail consistait en l'application d'un algorithme de rétropropagation adaptatif pour l'apprentissage, les résultats obtenus (Sensibilité=72,22%,spécificité=88,62 %) ont montré l'intérêt d'un tel modèle dans les salles de soins intensifs. En vue de la détection en temps réel de la tachycardie ventriculaire, Minami et Hirochi ont développé et implémenté un algorithme à base des RN dans un processeur à 8 bits (Hitachi H8/330) [Minami 1999]. Silopo et al [Silipo 1999a] ont mis en oeuvre une approche hybride à base de réseaux de neurones multicouches et de fonctions à bases radiales pour l'analyse du signal ECG.

Acharya et al [Acharya 2004] ont étudié l'apport de la variation de rythme dans la classification neuronale des arythmies cardiaques. Ils ont utilisé une architecture multicouche et obtenu des performances de l'ordre de 85%. Les applications des RN au problème du diagnostic cités ci-dessus ne sont que quelques exemples parmi un nombre important des travaux récents.

2.8 Les limites des réseaux de neurones

Plusieurs limites dans l'utilisation des réseaux de neurones peuvent être retenues.

L'architecture optimale. On mentionnera qu'il n'existe pas encore de théorie permettant de déterminer la structure optimale d'un réseau. En particulier la détermination du nombre de couches cachées et du nombre de neurones sur chacune d'entre elle relève encore largement de l'intuition de l'utilisateur et de sa capacité à expérimenter plusieurs architectures afin de retenir celle qui donne les résultats les meilleurs. Plus le réseau est complexe, c'est-à-dire plus il comporte de neurones cachés et de liens synaptiques, plus il est capable de reconnaître les formes qui lui sont présentées à travers l'échantillon d'apprentissage.

Mais malheureusement, l'accroissement de la complexité n'améliore pas nécessairement le taux de reconnaissance sur l'échantillon test. Il faut cependant remarquer que cette critique doit être nuancée, car plusieurs méthodes ont été proposées pour aider l'utilisateur dans cette tâche. Celles-ci s'appuient sur des techniques analytiques faisant intervenir la dimension du vecteur d'entrée, ou sur des algorithmes qui permettent soit la construction automatique pas à pas des couches

cachées, soit l'élagage des connexions les moins pertinentes dans un réseau surdimensionné au départ.

- Le pouvoir explicatif des réseaux,

Un réseau de neurones reste encore en partie aujourd'hui une « boîte noire » de laquelle il reste difficile d'extraire les relations pertinentes entre les variables.

De ce point de vue, le système n'a donc qu'un pouvoir explicatif médiocre contrairement aux systèmes experts qui sont capables de retracer le cheminement suivi pour atteindre le résultat. Pourtant, le phénomène de « boîte noire » n'est pas total puisque des analyses de sensibilité sont possibles, en faisant varier l'une des entrées pas à pas, de sa valeur minimale vers sa valeur maximale, les autres entrées du réseau restant figées à leur valeur moyenne.

- L'intervention humaine,

Comme cela a été signalé pour ce qui est du choix de l'architecture, ces systèmes font toujours trop largement appel à l'intuition de l'utilisateur. En particulier, l'apprentissage est guidé par des paramètres qu'il convient de régler manuellement. Un taux d'apprentissage (paramètre réglant la vitesse de convergence de l'algorithme) trop important peut aboutir à une oscillation du système alors qu'un taux trop faible se traduira par une convergence très lente, ce qui peut se révéler prohibitif en temps de calcul.

Par ailleurs, il revient encore à l'utilisateur de choisir le point d'arrêt de l'apprentissage de façon à ce que le réseau conserve ses capacités de généralisation.

2.9 Conclusion

A partir d'un modèle simple des neurones biologiques, nous pouvons construire un modèle plus complexe, celui des perceptrons multi-couches. Ces outils permettent de calculer des fonctions, adaptables à un ensemble d'exemples par le biais d'algorithmes d'apprentissage utilisant la technique de la rétro-propagation. De cette façon, on peut espérer faire apprendre une fonction complexe à un MLP simple, contenant assez peu de neurones.

Les réseaux de neurones sont depuis quelque temps un point de focalisation des médias, du public et des scientifiques. Le nombre croissant des travaux engagés dans le thème des réseaux de neurones contribuera sans doute à fournir toujours plus d'outils pour les domaines d'application comme la reconnaissance de formes et le diagnostic.

Nous présentons dans le chapitre suivant une des propriétés les plus intéressantes des réseaux de neurones, l'apprentissage.

Chapitre 3

Modèles d'apprentissage des réseaux de neurones

3.1 Introduction

Après une présentation des réseaux de neurones précédemment, nous présentons dans ce chapitre le principe de l'apprentissage.

Depuis bientôt un demi siècle, les chercheurs travaillent à programmer des machines capables d'effectuer des tâches qui requièrent de l'intelligence. Les ordinateurs peuvent être programmés pour leur faire exécuter des tâches considérées comme intelligentes, de multiples façons et de manière de plus en plus efficace. On s'intéresse à un aspect particulier de cette intelligence artificielle : la faculté d'apprentissage.

Parmi ces tâches l'aide au diagnostic, par exemple l'aide au diagnostic médical ; la reconnaissance de formes, par exemple la reconnaissance de la parole ou la vision artificielle ; le contrôle de processus, par exemple la conduite de procédés industriels ; la prédiction, par exemple la prédiction de consommation électrique ou la prédiction des cours boursiers ; la conduite de robots ; l'exploration de grandes bases de données (fouille de données).

L'apprentissage est vraisemblablement la propriété la plus intéressante des réseaux neuronaux.

Nous nous intéressons dans ce mémoire de magister à l'apprentissage supervisé pour la classification des arythmies cardiaques.

3.2 Définition de l'apprentissage

On peut trouver le sens de l'«apprendre» dans des dictionnaires. Par exemple, le sens de l'«apprendre» dans le dictionnaire «Le Petit Robert» se compose:

- Être rendu capable de connaître, de savoir; être avisé, informé de (qqch.)
- Acquérir (un ensemble de connaissances) par un travail intellectuel ou par l'expérience.
- Modifier la tendance comportementale par expérience.

Donc, on peut facilement reconnaître que «apprendre» est une activité concernant la connaissance.

3.3 L'apprentissage naturel

La définition classique de l'apprentissage en sciences cognitives : « capacité à améliorer les performances au fur et à mesure de l'exercice d'une activité ».cette définition s'applique en particulier au comportement d'un joueur d'échecs au fil des parties, où l'assimilation de l'expérience et la puissance de raisonnement se combinent dans sa progression. Elle est aussi pertinente pour des tâches perceptives : on s'habitue à un accent, à une écriture .on accumule de bonnes et des mauvaises expériences. A partir d'elles, on sait, consciemment ou non, en abstraire des règles pour mieux effectuer les tâches.

Il y a une autre facette de l'apprentissage c'est : la faculté à généraliser rationnellement. Si une expérience accumulée sur un certain nombre d'exemples a permis de tirer des règles, ces règles doivent s'appliquer aussi à des situations non encore rencontrées.

Dès sa naissance, un enfant apprend à reconnaître l'odeur de sa mère, puis sa voix et plus largement l'ambiance du lieu où il vit. Ensuite , il apprend à coordonner ses perceptions , comme sa vue ou son toucher , avec ses mouvements .Par des essais gratifiants ou pénalisants, il apprend plus tard à marcher , manifestant une grande capacité à intégrer des signaux différents : la vue , le sens de l'équilibre , la proprioception , la coordination motrice . Il apprend aussi la structure de sa langue maternelle et acquiert simultanément un répertoire organisé de connaissances sur le monde qui l'entoure.

Il va aussi apprendre à lire. Il apprend d'abord par cœur des mots associés à des sons et à leur signification. Plus tard, il extrait des règles permettant de distinguer des groupements syllabiques à l'intérieur des mots et de les prononcer. Cet apprentissage est long et progressif, il demande des répétitions et des séquences d'exercices bien choisies . Il est en partie supervisé par des adultes qui préparent les tâches d'apprentissage , accompagnent son cheminement et sanctionnent , par récompense ou punition , les résultats observés .

Au cours des années qui suivent , l'enfant apprend par étapes à maîtriser des concepts et des opérations de plus en plus abstraits . Finalement , cette fois sans professeur pour l'escorter , il découvrira et énoncera des points de vue personnels , des théories sur les phénomènes sociaux, sportifs , économiques , naturels , et autres .

Les modalités de l'apprentissage naturel sont donc multiples : apprentissage par cœur , par instruction , par généralisation , par découverte , apprentissage impliquant des catégorisations voire la formation de théories , apprentissage plus ou moins supervisé ou autonome , etc . [A.Cornuéjols,L.Miclet 2003]

3.4 Apprentissage « artificiel » ou apprentissage « automatique »

A quoi sert l'apprentissage artificiel ? La plupart des programmes d'intelligence artificielle possèdent aujourd'hui un module d'apprentissage et tous les programmes de reconnaissance de formes sont fondés sur des algorithmes d'apprentissage. Ces programmes sont capables de reconnaître la parole humaine et de l'interpréter. Ils assistent les experts pour prendre des décisions dans des environnements complexes et évolutifs , par exemple le marché financier ou le diagnostic médical . Ils fouillent d'immenses bases de données hétérogènes comme les millions de pages Web accessibles à tous. Ils participent aussi à des tournois : le 11 mai 1997, le tenant du titre de champion du monde du jeu d'échecs , Gary Kasparov , a été battu en match par un programme. L'apprentissage artificiel est une discipline dont les outils et les champs d'applications sont assez disparates.

L'apprentissage artificiel est la science qui cherche et établit les liens entre les principes généraux d'apprenabilité et les méthodes et outils permettant de réaliser un apprentissage.

Le terme académique le plus courant est apprentissage automatique. il semble que la notion d'apprentissage artificiel apporte quelque chose de plus profond que l'idée d'automatique qui sous entend une sorte d'activité inconsciente qui s'exécute.

L'apprentissage automatique dans le contexte informatique veut dire la capacité d'un logiciel d'améliorer son comportement sans intervention "directe" de l'extérieur.

Le but de l'apprentissage est de découvrir les règles qui gouvernent et régissent des formes. L'apprentissage est un processus calculatoire qui doit être capable d'amener à une certaine prédiction et à une certaine généralisation.

3.5 Apprentissage des réseaux de neurones

On appelle « apprentissage » des réseaux de neurones la procédure qui consiste à estimer les paramètres des neurones du réseau, afin que celui-ci remplisse au mieux la tâche qui lui est affectée

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.

3.5.1 Les applications de l'apprentissage

Il existe deux grandes catégories d'application pour l'apprentissage avec des réseaux de neurones :

- La **modélisation** : d'une fonction, c'est à dire la recherche d'une fonction qui lie les entrées et la ou les sorties. Il s'agit par exemple de déterminer la loi de commande du bras d'un robot ou une loi physique quelconque pour déterminer une variable en fonction de la mesure d'autres variables.

- La **classification** : c'est-à-dire la détermination d'une classe d'appartenance à partir de la mesure de certaines variables. La classification peut être vue comme une fonction binaire appartient / n'appartient pas à la classe. La classification est particulièrement utile dans les domaines de la reconnaissance de formes ou d'aide à la décision .dans le cadre de notre travail nous utilisons les réseaux de neurones comme un classifieur des arythmies cardiaques.

Dans les 2 cas, on dispose initialement d'un ensemble de données comprenant d'une part les mesures des variables d'entrées et d'autre part les mesures des variables de sortie. C'est la base d'exemples qui permettra de faire l'apprentissage.

L'objectif est alors de choisir la structure d'un réseau est de trouver les poids de connexion qui permettent de prédire aussi bien que possible la sortie qui sera trouvée en fonction d'un nouvel ensemble de mesures des variables d'entrées.

L'apprentissage proprement dit est la méthode qui permet de déterminer les poids de connexion du réseau.

Dans le cas de la classification, cet apprentissage est dit supervisé lorsqu'on dispose d'une base d'exemples avec pour chaque ensemble de données la sortie désirée correspondante. Puisqu'on connaît la sortie attendue (pour ces exemples), on peut en effet superviser l'apprentissage. Si on ne dispose pas de la sortie attendue, la classification est dite non supervisée.

3.5.2 Les différents types d'apprentissage

L'intérêt principal de l'utilisation des réseaux de neurones est leur capacité d'apprentissage par l'exemple. Les différentes architectures des réseaux proposées sont généralement indissociables d'algorithmes d'apprentissage permettant de modifier le comportement du réseau (généralement par une modification de poids), en fonction de différents critères (exemples de couples entrée/sortie).

On classe les différents algorithmes d'apprentissage en trois familles selon les critères utilisés pour l'entraînement du réseau et selon l'information fournie à l'algorithme .

Il existe trois types d'apprentissages principaux. Ce sont l'apprentissage supervisé, l'apprentissage non-supervisé et l'apprentissage par renforcement.

3.5.2.1. Apprentissage supervisé

C'est actuellement le mode d'apprentissage le plus couramment utilisé. Son principe est élémentaire : on soumet au réseau un grand nombre d'exemples pour lesquels l'entrée et la sortie associée sont connues et les poids sont modifiés de façon à corriger l'erreur commise par le réseau (c'est-à-dire la différence entre la sortie désirée et la

réponse du réseau à l'entrée correspondante). Le réseau a alors comme but d'approximer ces exemples aussi bien que possible et de développer à la fois la bonne représentation mathématique qui lui permet de généraliser ces exemples pour ensuite traiter des nouvelles situations (qui n'étaient pas présentes dans les exemples). Le plus répandu des algorithmes d'apprentissage supervisé est l'algorithme de rétropropagation du gradient d'erreur qui, appliqué aux réseaux en couches, utilise la dérivée de la fonction de transfert des neurones pour calculer l'erreur en amont d'une couche par rapport à l'erreur en aval . Bien que les principes théoriques de l'apprentissage non-supervisé soient antérieurs, l'apprentissage supervisé est, historiquement, le premier à avoir été utilisé pour l'entraînement des modèles de réseaux neuronaux en 1957 avec le Perceptron puis en 1960 avec l'ADALINE .

Dans le cadre de l'apprentissage supervisé nous allons étudier le type d'apprentissage « **spécialisé** » :

Le réseau de neurones reçoit directement en entrée les couples «entrée» / «sortie désiré » ou « paramètres descripteurs »/ « classes d'appartenance ». Ce réseau produit alors une action qui est appliquée à un deuxième réseau ayant préalablement été entraîné pour identifier les différentes classes. Nous verrons le rôle de cet apprentissage spécialisé dans le chapitre4.

3.5.2.2. Apprentissage non-supervisé et auto-organisation

Contrairement aux modes supervisés, seule une base d'entrées est ici fournie au réseau. Celui-ci doit donc déterminer lui-même ses sorties en fonction des similarités détectées entre les différentes entrées, c'est-à-dire en fonction d'une règle d'auto-organisation. Il n'y a donc pas là de « professeur », puisque c'est au réseau de découvrir les ressemblances entre les éléments de la base de données .Appelé aussi apprentissage sans professeur .

La règle la plus connue est celle définie par Donald O. Hebb basé sur le renforcement des connexions entre les cellules d'activités corrélées (c'est-à-dire actives simultanément), soit : $\Delta W_{ij} = \alpha x_i x_j$ Où α est une constante c'est le « facteur d'apprentissage ».

La règle de Hebb a donné naissance à de nombreux algorithmes d'apprentissage supervisé (comme dans les mémoires associatives où elle est utilisée pour apprendre des

connexions reliant les neurones d'entrée aux neurones de sortie.) ou non supervisés.

D'autres algorithmes d'apprentissage ont été proposés, généralement en association avec une topologie particulière. On citera en particulier les algorithmes de Kohonen pour les cartes auto-organisatrices et les modèles ART1 et ART2 développés par Carpenter et Grossberg .

3.5.2.3 Apprentissage par renforcement

L'apprentissage par renforcement consiste à apprendre quoi faire, comment associer des actions à des situations ,afin de maximiser quantitativement une récompense. On ne dit pas à l'apprenant quelle action faire,mais au lieu de cela,il doit découvrir quelles actions donnent le plus de récompenses en les essayant. Dans le cas le plus intéressant, des actions peuvent affecter non seulement les récompenses immédiates mais aussi la situation suivante, et par là, les récompenses à plus long terme .ces deux propriétés « recherche par essai-erreur et récompense à long terme » sont les deux caractéristiques les plus importantes de l'apprentissage par renforcement.

L'apprentissage par renforcement est différent de l'apprentissage supervisé. Ce dernier nécessite un superviseur qui dicte au réseau quelle action est correcte dans telle situation. Dans l'apprentissage par renforcement, le réseau n'a pas de superviseur à sa disposition, il interagit avec l'environnement qui lui donne un retour quantitatif sur les valeurs des ses actions.

3.5.3 Apprentissage « en ligne » et apprentissage « hors-ligne »

Il existe deux modes principaux d'apprentissage, selon la façon dont les vecteurs de poids synaptiques sont adaptés. Le premier, dit apprentissage « en-ligne », consiste à modifier les valeurs des poids synaptiques immédiatement après la présentation d'un objet ou exemple.

Dans ce cas, seul le gradient instantané de la fonction de coût est utilisé pour l'adaptation des paramètres du système.

C'est généralement le cas pour l'apprentissage par renforcement qui nécessite une réponse de l'environnement.

Dans Le second mode principal d'apprentissage : « hors-ligne » ou « off-line », Le réseau est entraîné à partir d'une base d'apprentissage (généralement une base d'exemples).

Ce type d'apprentissage consiste à accumuler les gradients instantanés consécutifs, et à n'effectuer l'adaptation des poids synaptiques que lorsque l'ensemble des objets d'apprentissage ont été présentés au réseau de neurones. On parle alors d'apprentissage « hors-ligne ». Cette dernière méthode permet de mieux estimer le gradient de la fonction de coût, puisqu'il est à présent calculé à partir d'un ensemble d'objets, plutôt qu'à partir d'un seul.

Dans l'apprentissage en-ligne chacune des formes est représenté une seule fois ce qui peut sembler donner un apprentissage plus rapide mais nécessite plus de données d'apprentissage.

3.6 Les différents algorithmes d'apprentissage

Une fois l'architecture d'un réseau de neurones choisie, il est nécessaire d'effectuer un apprentissage pour déterminer les valeurs des poids permettant à la sortie du réseau de neurones d'être aussi proche que possible de l'objectif fixé.

Cet apprentissage s'effectue grâce à la minimisation d'une fonction, appelée *fonction de coût*, calculée à partir des exemples de la base d'apprentissage et de la sortie du réseau de neurones ;cette fonction détermine l'objectif à atteindre.

3.6.1 Choix de la fonction de coût

La définition d'une fonction de coût est primordiale, car celle-ci sert à mesurer l'écart entre la sortie désirée du modèle et les mesures faites sur les exemples d'apprentissage.

La fonction la plus couramment utilisée, et dont nous nous sommes servi lors de nos travaux, est la fonction dite « Erreur quadratique ».

L'erreur quadratique sur la base d'apprentissage consiste à minimiser la somme des carrés des erreurs entre la sortie du réseau et la valeur réelle de la sortie.

L'objectif de l'apprentissage est donc de trouver un vecteur qui minimise cette fonction.

Les algorithmes utilisés nécessitent que la fonction de coût soit dérivable par rapport aux poids. Le principe de ces méthodes est de se placer en un point initial, de trouver une direction de descente du coût dans l'espace des paramètres w , puis de se déplacer d'un pas dans cette direction. On atteint un nouveau point et l'on itère la procédure jusqu'à satisfaction d'un critère d'arrêt. Ainsi, à l'itération k , on calcule :

$$w^k = w^{k-1} + \alpha_{k-1} \cdot d_{k-1}$$

α_k est le pas de la descente et d_k est la direction de descente : les différents algorithmes se distinguent par le choix de ces deux quantités

Il existe plusieurs règles d'apprentissage pour chaque type d'apprentissage. L'apprentissage supervisé est le type d'apprentissage le plus utilisé. Pour ce type d'apprentissage la règle la plus utilisée est celle de Widrow-Hoff. D'autres règles d'apprentissage sont par exemple la règle de Hebb, la règle du perceptron (procédure de correction d'erreur), et la rétropropagation de gradient qui est l'algorithme le plus répandu pour les réseaux multi couches.

3.6.2 Règle de Hebb

La règle de Hebb, du neurophysiologiste de même nom, méthode d'apprentissage la plus ancienne (1949), stipule que l'efficacité d'une synapse entre un neurone A et un neurone B est augmentée si un potentiel d'action est émis par B alors qu'il reçoit des potentiels d'action venant de A. C'est une sorte de renforcement de la causalité, car si B est actif en même temps que A alors qu'il existe une synapse pour les connecter, c'est que A a participé au déclenchement de B. Le principe est de renforcer les connexions entre deux neurones lorsque ceux-ci sont actifs simultanément. Cette règle peut être classée comme apprentissage non supervisé. Cette règle, observée sur des réseaux de neurones naturels, est également appliquée dans de nombreuses techniques d'apprentissage des réseaux artificiels. Elle pourrait être le mécanisme de base permettant la classification, ou la reconnaissance des formes.

3.6.3 Algorithme d'apprentissage par correction d'erreur ou règle du perceptron

3.6.3.1 Présentation de l'algorithme

L'algorithme d'apprentissage peut être décrit de la manière suivante. On initialise les poids du perceptron à des valeurs quelconques. A chaque fois que l'on présente un nouvel exemple, on ajuste les poids selon que le perceptron l'a correctement classé ou non. L'algorithme s'arrête lorsque tous les exemples ont été présentés sans modification d'aucun poids.

La procédure d'apprentissage du perceptron linéaire à seuil est une procédure de correction d'erreur puisque les poids ne sont pas modifiés lorsque la sortie attendue est égale à la sortie calculée par le perceptron courant. Étudions les modifications sur les poids lorsque diffère de :

si la sortie calculée=0 et sortie attendue=1, cela signifie que le perceptron n'a pas assez pris en compte les neurones actifs de l'entrée (c'est-à-dire les neurones ayant une entrée à 1) dans ce cas ;l'algorithme ajoute la valeur de la rétine aux poids synaptiques (*renforcement*).

si la sortie calculée=1 et la sortie attendue=0; l'algorithme retranche la valeur de la rétine aux poids synaptiques (*inhibition*).

Lors de la phase d'apprentissage, sachant que l'échantillon d'apprentissage est un ensemble linéairement séparable, tous les exemples sont présentés jusqu'à la convergence, c'est-à-dire jusqu'à ce qu'une présentation complète des exemples n'entraîne aucune modification des poids.

3.6.3.2 Critiques sur la méthode par correction d'erreur

Si l'échantillon est linéairement séparable, si tous les exemples sont présentés équitablement (c'est-à-dire que la procédure de choix des exemples n'en exclut aucun), et que le critère d'arrêt est la stabilité après une présentation complète de l'échantillon alors la procédure d'apprentissage par correction d'erreur converge et l'algorithme s'arrête avec un perceptron qui classe correctement l'échantillon d'apprentissage.

L'inconvénient majeur de cet algorithme est que si l'échantillon présenté n'est pas linéairement séparable, l'algorithme ne convergera pas et l'on aura aucun moyen de le savoir.

On pourrait penser qu'il suffit d'observer l'évolution des poids synaptiques pour en déduire si l'on doit arrêter ou non l'algorithme. En effet, si les poids et le seuil prennent deux fois les mêmes valeurs sans que le perceptron ait appris et alors que tous les exemples ont été présentés, cela signifie que l'échantillon n'est pas séparable.

3.6.4 L'apprentissage de Widrow-Hoff ou règle Delta

Cette règle est appelée règle delta , ou règle ADALINE , ou encore Widrow-Hoff d'après le nom de ses inventeurs.

La règle d'apprentissage de Widrow-Hoff est une règle qui permet d'ajuster les poids d'un réseau de neurones pour diminuer à chaque étape l'erreur commise par le réseau de neurones.

Son but est de faire évoluer le réseau vers le minimum de sa fonction d'erreur (erreur commise sur l'ensemble des exemples).

L'apprentissage est réalisé par itération (les poids sont modifiés après chaque exemple présenté), et on obtient le poids à l'instant $t+1$ par la formule :

$$w(t+1) = w(t) + \eta (c - o) x$$

Si w est le poids, C la sortie théorique et O la sortie réelle, x l'entrée et η un coefficient d'apprentissage (entre 0 et 1) que l'on peut diminuer au cours de l'apprentissage.

Cette règle s'interprète par le fait que l'on modifie le poids proportionnellement à l'erreur (différence entre sortie théorique et sortie donnée) et à l'état d'excitation de l'entrée gérée par le poids. Ainsi le poids ne sera pas modifié si la sortie est celle voulue, ou si le poids n'a joué aucun rôle dans le résultat.

L'avantage de l'algorithme de Widrow-Hoff par rapport à l'algorithme par correction d'erreur est que, même si l'échantillon d'entrée n'est pas linéairement séparable, l'algorithme va converger vers une solution << optimale >>.

C'est en fait un cas particulier de l'algorithme de rétropropagation du gradient pour un réseau à une couche.

3.6.5 La rétropropagation de GRADIENT (back propagation)

Il a fallu attendre le début des années 1980 pour qu'une règle efficace soit mise au point pour l'apprentissage des réseaux multicouches. Cette règle, découverte simultanément par des équipes française et américaine, est en fait une généralisation de la règle du delta.

Elle consiste simplement en une descente de gradient, qui est une méthode d'optimisation universelle.

L'algorithme le plus simple consiste à choisir comme direction de descente l'opposé du gradient de la fonction de coût ($dk = - \nabla J(w) = - \text{Grad}(J(wk))$). On cherche à minimiser une fonction d'énergie d'erreur (qui représente l'erreur entre la sortie désirée et la sortie obtenue), en suivant les lignes de plus grande pente. Une fonction d'erreur rapportée à une dimension peut se représenter ainsi (figure3.1) :

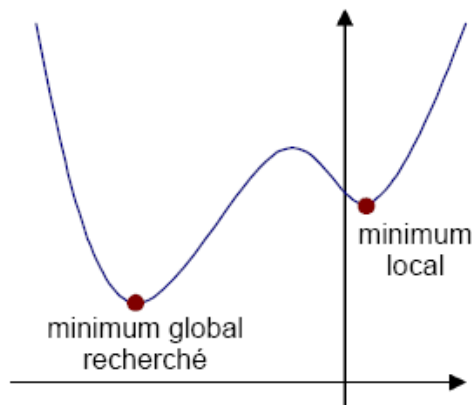


Figure3.1 :descente de gradient

On peut se représenter la descente de gradient comme une bille que l'on poserait sur la courbe, et qui descendrait logiquement la pente (le gradient représente la pente selon chaque dimension).

L'inconvénient de cette méthode est qu'elle va s'arrêter dans le premier minimum local rencontré. Pour cela diverses améliorations ont été apportées à cette méthode.

3.6.5.1 Principe de l'optimisation par descente de gradient

Le problème consiste donc à trouver un minimum global de la fonction erreur E entre la fonction $f(t)$ et les points « target ». La méthode du gradient est l'une des principales méthodes pour déterminer cette région d'optimisation.

En effet, afin de minimiser une fonction à partir d'une solution approchée, le plus simple est de suivre la ligne de plus grande pente. D'un point de vue mathématique, la pente d'une fonction correspond à la dérivée de cette dernière. Si l'on se place dans le cadre d'une fonction ayant plusieurs paramètres, la dérivée devient un vecteur : le gradient de la fonction. Chaque élément de ce vecteur correspond alors à la dérivée partielle de la fonction selon l'un de ses paramètres.

Soit f une fonction (suffisamment dérivable) dont on recherche un minimum. La méthode du gradient construit une suite x_n qui doit en principe s'approcher du minimum.

Pour cela, on part d'une valeur quelconque x_0 et l'on construit la suite $x_{n+1} = x_n - \rho f'(x_n)$ où ρ est une valeur « bien » choisie réelle non nulle.

On a $f(x_{n+1}) = f(x_n - \rho f'(x_n)) \approx f(x_n) - \rho (f'(x_n))$ d'après le théorème des approximations finies si $\rho f'(x_n)$ est "suffisamment" petit.

On voit que, sous réserve de la correction de l'approximation, $f(x_{n+1})$ est inférieur à $f(x_n)$.

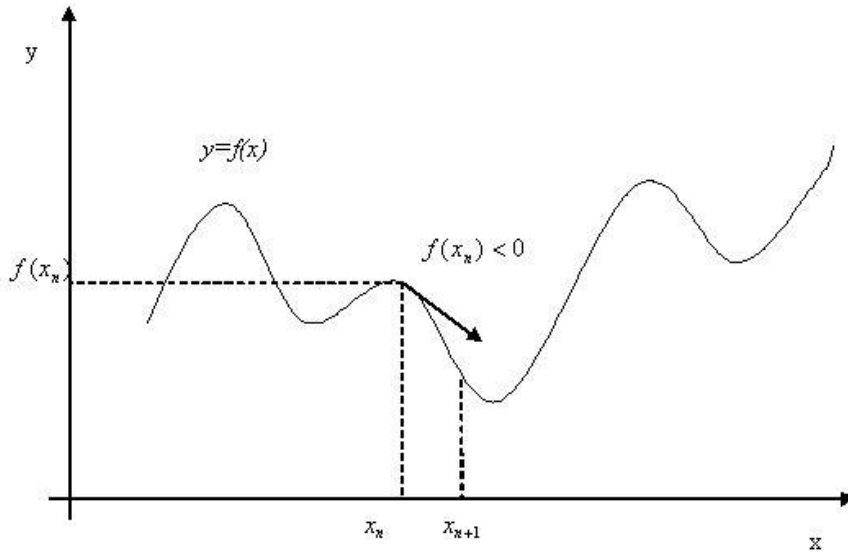


Figure 3.2 : La méthode du gradient

On remarque que x_{n+1} est d'autant plus éloigné de x_n que la pente de la courbe en x_n est grande. On peut décider d'arrêter l'itération lorsque cette pente est suffisamment faible (en dimension 2 par exemple, un minimum correspond à une pente nulle).(figure3.2).

3.6.5.2 Les limites de l'optimisation par descente de gradient

Cependant, cette méthode possède des inconvénients bien connus :

1. le choix de ρ est empirique,
2. si ρ est trop petit, le nombre d'itérations peut être très élevé,
3. si ρ est trop grand, les valeurs de la suite risquent d'osciller autour du minimum sans converger,
4. rien ne garantit que le minimum trouvé est un minimum global.

Ce dernier point est la limite la plus importante de cette méthode d'optimisation. En effet, si on considère un cas simple, comme celui de la figure suivante, souvent, le minimum global est masqué par un maximum, local. Ici, dans cet exemple, en partant du point marqué d'une croix, il est impossible avec la méthode du gradient d'aller au-delà du point marqué d'un cercle. En effet, cette méthode va s'arrêter à ce minimum et rien ne peut prouver qu'il n'existe pas un autre minimum peut être global au delà. (figure3.3).

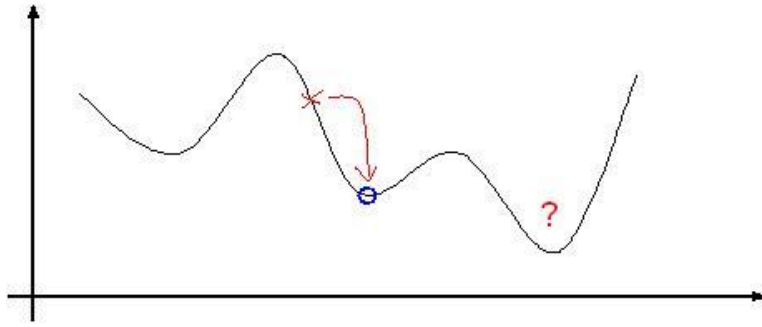


Figure 3.3 : Limite de la Méthode du gradient

3.6.5.3 Problème des minima locaux

Les minima trouvés par les algorithmes dans la plupart des cas sont des minima locaux. Le minimum trouvé dépend du point de départ de la recherche c'est-à-dire de l'initialisation des poids. En pratique, il faut effectuer plusieurs minimisations avec des initialisations différentes, pour trouver plusieurs minima et retenir le "meilleur". Il est néanmoins impossible et généralement inutile, de s'assurer que le minimum choisi est le minimum global. Les réseaux de neurones à couches présentent des symétries, si bien que l'on peut montrer que pour une architecture avec N_c neurones cachés, il existe $2^{N_c} N_c!$ minima équivalents .

3.6.5.4 Méthodes itératives d'optimisation

Nous verrons que l'apprentissage supervisé des réseaux multicouches est une simple question d'optimisation, c'est-à-dire consiste en la minimisation d'un critère d'erreur.

On va minimiser à chaque itération la valeur $f(x_n)$. grâce à des méthodes itératives numériques basées sur une approximation du critère d'erreur (son développement à plusieurs variables de Taylor) Cela est le cas si l'on a à chaque itération : $f(x_{n+1}) < f(x_n)$

Si la variation $\Delta x_n = x_{n+1} - x_n$ est suffisamment faible, on peut approcher la valeur $f(x_{n+1})$ par le développement de Taylor à plusieurs variables en x_n :

$$f(x_n + \Delta x_n) \approx f(x_n) + (\nabla f |_{x_n})^T \cdot \Delta x_n \quad \text{Où } \nabla f |_{x_n} \text{ est le gradient de } f \text{ au point } x_n$$

Notre objectif est donc d'avoir le produit scalaire entre le vecteur gradient et le vecteur des variations de x négatif :

$$(\nabla f |_{x_n})^T \cdot \Delta x_n < 0$$

On a une décroissance maximale de la fonction f si ce produit scalaire est négatif et minimal, donc si et seulement si les deux vecteurs sont de même direction et de sens opposé.

On obtient pour règle d'apprentissage :

$$x_{n+1} = x_n - \eta \cdot \nabla f |_{x_n} \quad \text{Où } \eta \text{ est un coefficient compris entre 0 et 1}$$

On constate donc qu'il faut suivre le gradient, d'où le terme de descente de gradient.

Donc : L'apprentissage d'un réseau de neurones artificiels se réduisant à un problème d'optimisation : trouver le minimum d'une fonction d'erreur, on pourra donc mettre à profit cette méthode d'optimisation universelle de descente de gradient, qui constituera la règle de rétropropagation du gradient pour les réseaux multicouches.

3.6.6 la rétropropagation du gradient

La méthode d'apprentissage la plus utilisée pour les réseaux multi couches est la rétropropagation du gradient.

3.6.6.1 L'apprentissage dans les réseaux multicouches

Rappelons qu'un perceptron multi-couches (PMC) est un réseau de neurones à couches cachées avec les cellules élémentaires ainsi définies.

Définition : Une cellule élémentaire à n entrées réelles $x = (x_1, \dots, x_n)$ est définie par les poids synaptiques réels $w = (w_1, \dots, w_n)$ et la sortie o est calculée par la formule suivante :

$$o(x) = \frac{1}{1 + e^{-y}} \quad \text{où } y = x \cdot w = \sum_{i=1}^n w_i x_i$$

Nous allons définir la fonction d'activation la plus utilisée dans l'apprentissage par rétropropagation du gradient qui est la fonction sigmoïde de paramètre $k>0$:

$$\sigma_k(x) = \frac{e^{kx}}{e^{kx} + 1} = \frac{1}{1 + e^{-kx}}$$

Cette fonction est une approximation indéfiniment dérivable de la fonction à seuil de Heaviside, d'autant meilleure que k est grand. Nous prendrons $k=1$ dans notre étude, soit la fonction σ (voir figure 3.4) définie par :

$$\sigma(x) = \frac{e^x}{e^x + 1} = \frac{1}{1 + e^{-x}} \quad (3.1)$$

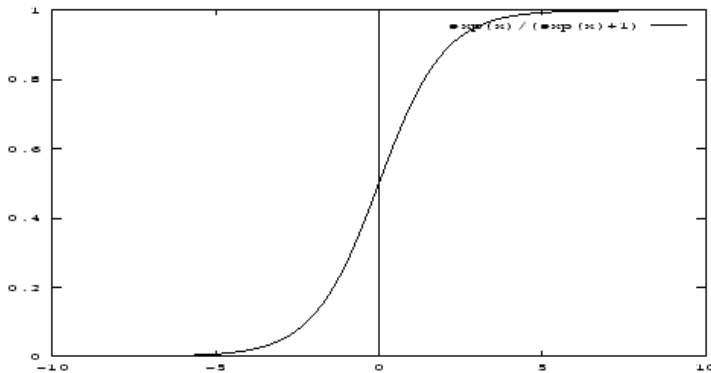


Figure 3.4: La fonction sigmoïde

On peut remarquer que la dérivée de la fonction σ est simple à calculer :

$$\sigma'(x) = \frac{e^x}{(1 + e^x)^2} = \sigma(x)(1 - \sigma(x)) \quad (3.2)$$

Il est essentiel que ce calcul soit simple car la dérivée de cette fonction sera utilisée dans la règle de mise à jour des poids par l'algorithme de rétropropagation du gradient.

Le principe de l'algorithme est, comme dans le cas du perceptron linéaire, de minimiser une fonction d'erreur. Il s'agit ensuite de calculer la contribution à cette erreur de chacun des poids synaptiques. C'est cette étape qui est difficile. En effet, chacun des poids influe sur le neurone correspondant, mais, la modification pour ce neurone va influencer sur tous les neurones des couches suivantes.

Cet algorithme fut popularisé en 1986 par Rumelhart qui l'utilisa pour la modification des poids et des biais dans les réseaux de neurones multicouches. Cet algorithme modifie les paramètres de poids et de biais, en allant de la couche de sortie vers la couche d'entrée en propageant l'erreur se trouvant à la couche de sortie vers chacune des couches précédentes, l'erreur de sortie étant obtenue par comparaison de la sortie réelle avec la sortie désirée.

L'apprentissage classique de ce réseau est conduit par une méthode de descente de gradient.

Les poids sont initialisés aléatoirement et sont modifiés selon une directive qui minimise une fonction d'erreur. Cette méthode présente néanmoins un inconvénient majeur, soit celui d'une probabilité élevée de converger vers un minimum local qui faussera l'apprentissage.

Jusqu'à maintenant, l'algorithme le plus utilisé pour l'apprentissage de ce type de réseau porte le nom de « Error back propagation » ou algorithme de rétropropagation du gradient de l'erreur.

3.6.6.2 Introduction de l'algorithme

Soit un PMC défini par une architecture à n entrées et à p sorties, soit w le vecteur des poids synaptiques associés à tous les liens du réseau. L'erreur du PMC sur un échantillon d'apprentissage S d'exemples (x^s, c^s) est définie par :

$$E(w) = \frac{1}{2} \sum_{(x^s, o^s) \in S} \sum_{k=1}^p (c_k^s - o_k^s)^2 \quad (3.3)$$

où o_k^s est la k -ième composante du vecteur de sortie o^s calculé par le PMC sur l'entrée x^s . L'erreur mesure donc l'écart entre les sorties attendue et calculée sur l'échantillon complet. On suppose S fixé, le problème est donc de déterminer un vecteur w qui minimise $E(w)$. Cependant, de la même façon que pour le perceptron avec la règle de Widrow-Hoff, plutôt que de chercher à minimiser l'erreur globale sur l'échantillon complet, on cherche à minimiser l'erreur sur chaque présentation individuelle d'exemple.

L'erreur pour un exemple est :

$$E_{(x, c)}(w) = \frac{1}{2} \sum_{k=1}^p (c_k - o_k)^2 \quad (3.4)$$

Nous notons E la fonction $E_{(x,c)}$, E est une fonction des poids synaptiques, pour appliquer la méthode du gradient, il nous faut évaluer les dérivées partielles de cette fonction E par rapport aux poids synaptiques. Les calculs qui suivent sont faciles. La seule complication provient de la complexité des notations et des indices utilisés, complication due à la structure du PMC. Nous utilisons les notations suivantes (voir figure 3.5) :

- chaque cellule est définie par un indice , le réseau comporte p cellules de sortie.
- si i est l'indice d'une cellule de sortie, c_i est la sortie attendue pour cette cellule sur l'entrée x .
- w_{ij} est le poids synaptique associé au lien entre cellule j vers la cellule i , ce qui implique qu'elles se trouvent sur deux couches successives par définition de l'architecture.
- x_{ij} est l'entrée associée au lien entre cellule j vers cellule i .
- $Pred(i)$ est l'ensemble des cellules dont la sortie est une entrée de la cellule i ; ceci implique que la cellule n'est pas une cellule d'entrée et que tous les éléments de $Pred(i)$ appartiennent à la couche précédente de celle à laquelle appartient la cellule i ,
- y_i l'entrée totale de la cellule i , soit $y_i = \sum_{j \in pred(i)} w_{ij} x_{ij}$
- o_i est la sortie de la cellule i , soit $o_i = \sigma(y_i)$.
- $Succ(i)$ est l'ensemble des cellules qui prennent comme entrée la sortie de la cellule i , ceci implique la cellule n'est pas une cellule de sortie et que tous les éléments de $Succ(i)$ appartiennent à la couche suivante de celle à laquelle appartient la cellule i .

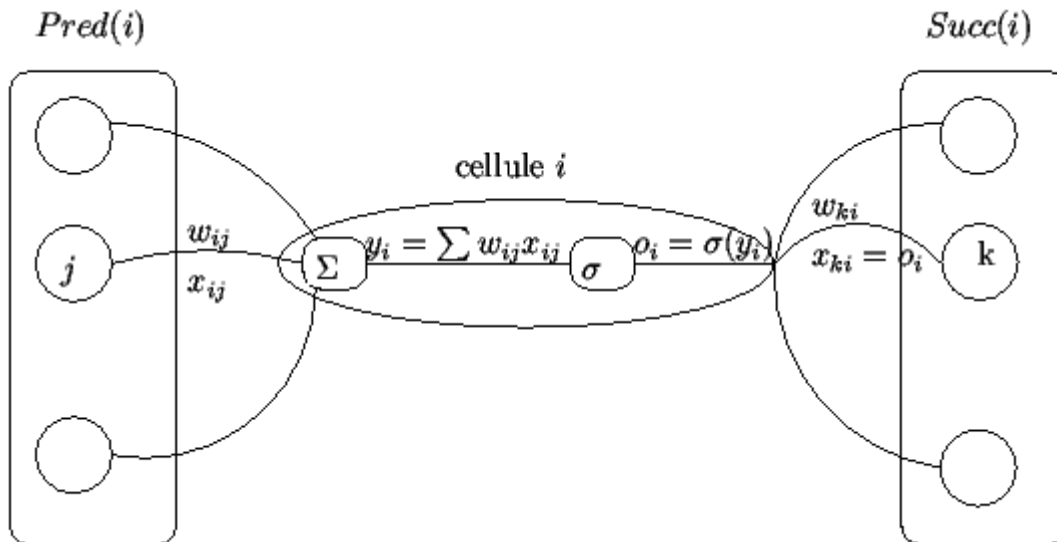


Figure 3.5: notations utilisées

Il nous reste maintenant à évaluer $\partial E(w)/\partial w_{ij}$ que nous noterons $\partial E/\partial w_{ij}$. Tout d'abord remarquons que w_{ij} ne peut influencer la sortie du réseau qu'à travers le calcul de la quantité y_i , ce qui nous autorise à écrire que :

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial y_i} \frac{\partial y_i}{\partial w_{ij}} = \frac{\partial E}{\partial y_i} x_{ij} \quad (3.5)$$

Il nous suffit donc de calculer $\partial E/\partial y_i$, pour cela, nous allons distinguer deux cas : le cas où la cellule i est une cellule de sortie et le cas où c'est une cellule interne.

- La cellule i est une cellule de sortie

Dans ce cas, la quantité y_i ne peut influencer la sortie du réseau que par le calcul de o_i .

Nous avons donc :

$$\frac{\partial E}{\partial y_i} = \frac{\partial E}{\partial o_i} \frac{\partial o_i}{\partial y_i} \quad (3.6)$$

Nous allons maintenant calculer chacune des deux dérivées partielles apparaissant dans cette équation.

Pour la première de ces deux dérivées nous avons :

$$\frac{\partial E}{\partial o_i} = \frac{\partial}{\partial o_i} \frac{1}{2} \sum_{k=1}^p (c_k - o_k)^2 \quad (3.7)$$

Seul le terme correspondant à $k=i$ a une dérivée non nulle, ce qui nous donne finalement :

$$\frac{\partial E}{\partial o_i} = \frac{\partial}{\partial o_i} \frac{1}{2} (c_i - o_i)^2 = -(c_i - o_i) \quad (3.8)$$

Pour la seconde des deux dérivées de l'équation 3.6, en utilisant la définition du calcul de la sortie d'une cellule élémentaire et la formule de calcul de la dérivée de la fonction sigmoïde, nous avons :

$$\frac{\partial o_i}{\partial y_i} = \frac{\partial \sigma(y_i)}{\partial y_i} = \sigma(y_i)(1 - \sigma(y_i)) = o_i(1 - o_i) \quad (3.9)$$

En substituant les résultats obtenus par les équations 3.8 et 3.9 dans l'équation 3.6, nous obtenons

$$\frac{\partial E}{\partial y_i} = -(c_i - o_i) o_i (1 - o_i) \quad (3.10)$$

La cellule i est une cellule interne

Dans ce cas, la quantité y_i va influencer le réseau par tous les calculs des cellules de l'ensemble $Succ(i)$. Nous avons alors :

$$\frac{\partial E}{\partial y_i} = \sum_{k \in Succ(i)} \frac{\partial E \partial y_k}{\partial y_k \partial y_i} = \sum_{k \in Succ(i)} \frac{\partial E \partial y_k \partial o_i}{\partial y_k \partial o_i \partial y_i} = \sum_{k \in Succ(i)} \frac{\partial E}{\partial y_k} \times w_{ki} \times o_i (1 - o_i) \quad (3.11)$$

Soit encore :

$$\frac{\partial E}{\partial y_i} = o_i (1 - o_i) \sum_{k \in Succ(i)} \frac{\partial E}{\partial y_k} \times w_{ki} \quad (3.12)$$

Par l'étude de ces deux cas, nous avons obtenu deux équations 3.10 et 3.11 qui nous permettent de calculer les dérivées partielles $\frac{\partial E}{\partial y_i}$ pour toute cellule i . Le calcul devra être fait pour les cellules de sortie puis des cellules de l'avant-dernière couche jusqu'aux cellules de la première couche. C'est pour cette raison que l'on parle de « rétropropagation ». Grâce à l'équation 3.5, nous pouvons calculer toutes les dérivées partielles $\frac{\partial E(w)}{\partial w_{ij}}$. Enfin, pour en déduire la modification à effectuer sur les poids synaptiques, il nous reste simplement à rappeler que la méthode du gradient nous indique que :

$$\Delta w_{ij} = -\eta \frac{\partial E(w)}{\partial w_{ij}} \quad (3.13)$$

Tous les éléments sont donc en place pour nous permettre de définir l'algorithme de rétropropagation du gradient.

3.6.6.3 Algorithme de rétropropagation du gradient.

Pour écrire l'algorithme, nous allons simplifier quelques notations. Nous appelons δ_i la quantité $-\frac{\partial E}{\partial y_i}$. En utilisant les équations 3.5, 3.10, 3.12 et 3.13, nous obtenons les formules suivantes:

Pour une cellule i de sortie, nous avons :

$$\delta_i = o_i(1 - o_i)(c_i - o_i)$$

Pour une cellule i interne, nous avons :

$$\delta_i = o_i(1 - o_i) \sum_{k \in \text{succ}(i)} \delta_k w_{ki}$$

La modification du poids w_{ij} est alors définie par :

$$\Delta w_{ij} = \eta x_{ij} \delta_i$$

On peut faire les remarques suivantes :

- la règle de modification des poids pour le perceptron linéaire est :

$$w_i \longrightarrow w_i + \eta (c - o) x_i$$

Dans le cas du PMC, cette règle est : $w_{ij} \longrightarrow w_{ij} + \eta \delta_i x_{ij}$

Ces deux règles sont très similaires, le terme d'erreur $c-o$ est remplacé par un terme plus compliqué δ_i

- pour une cellule i de sortie, la quantité δ_i correspond à l'erreur usuelle c_i-o_i multipliée par la dérivée de la fonction sigmoïde.
- pour une cellule i interne, le calcul de δ_i dépend de la somme pondérée des erreurs des cellules de la couche suivante.
- après présentation de l'entrée x et calcul de la sortie o , le calcul des erreurs δ_i sera effectué de la couche de sortie vers la couche d'entrée.

3.6.6.4 Variantes de l'algorithme de rétropropagation du gradient de l'erreur :

Nous allons présenter deux variantes de cet algorithme que nous avons utilisé dans cette étude qui sont l'algorithme de Quasi Newton et l'algorithme de Levenberg-Marquardt .

3.6.6.4.1 La méthode de Newton

La méthode de Newton utilise la courbure (dérivée seconde) de la fonction de coût pour atteindre le minimum. La modification des paramètres s'écrit ainsi :

$$w^k = w^{k-1} - H^{-1}_{k-1} \cdot \nabla J (w^{k-1})$$

La direction de descente est $H^{-1}_{k-1} \cdot \nabla J (w^{k-1})$ où H_{k-1} est l'inverse du hessien de la fonction de coût, la matrice hessienne est la matrice des dérivées secondes de l'indice de performances de poids et de biais.

Dans la pratique, le calcul du hessien et surtout de son inverse est à la fois complexe et source d'instabilités numériques ; on utilise de préférence une méthode de "quasi-Newton".

3.6.6.4.2 La méthode de quasi-Newton

Les méthodes de quasi-Newton consistent à approcher l'inverse du hessien plutôt que de calculer sa valeur exacte.

La modification des paramètres s'écrit : $w_k = w_{k-1} - \alpha_{k-1} M_{k-1} \nabla J(w_{k-1})$

La suite M_k est construite de façon à converger vers l'inverse du hessien avec M_0 égale à la matrice identité. Cette suite est construite grâce à la méthode dite BFGS (Broyden, Fletcher, Goldfarb, Shanno 1970) dont la vitesse de convergence est beaucoup plus grande que celle de la méthode du gradient.

Dans Matlab on retrouve ces algorithmes dans la fonction « **trainbfg** ».

3.6.6.4.3 La méthode Levenberg-Marquardt

Si la descente de gradient est trop lente pour réaliser l'apprentissage, il est également possible d'utiliser un algorithme du second ordre, tel que celui de Levenberg Marquardt. Dans ce cas chaque itération demande plus de calculs, mais dans la plupart des cas le nombre d'itérations nécessaires pour converger est bien moindre.

Une autre manière de diminuer le nombre d'itérations d'un algorithme d'optimisation est d'utiliser les dérivées secondes.

Mais le calcul des dérivées secondes peut être très long, tout d'abord parce que le nombre de dérivées secondes est le carré de celui des dérivées premières, et également parce que la dérivée seconde peut être assez complexe. De nombreux algorithmes, peut-être abusivement appelés algorithmes d'ordre 2, utilisent en fait une approximation des dérivées secondes calculées à partir de dérivées premières. Cependant ils gardent l'avantage de nécessiter beaucoup moins d'itérations qu'une descente de gradient.

L'algorithme de Levenberg Marquardt fait partie de ces algorithmes, et s'applique au cas où la fonction est une erreur quadratique moyenne.

En pratique cet algorithme, en particulier dans le cas des réseaux de neurones, permet de converger avec beaucoup moins d'itérations. Mais chaque itération demande plus de calculs, en particulier pour l'inversion de la matrice \mathbf{H} , et son utilisation se limite donc aux cas où le nombre de paramètres à optimiser n'est pas très élevé. En effet le nombre d'opérations nécessaires à l'inversion d'une matrice est proportionnel à N^3 , N étant la taille de la matrice.

Dans Matlab on retrouve ces algorithmes dans la fonction « **trainlm** ».

3.7 Problèmes d'apprentissage dans les réseaux multicouches

Il ne faut pas perdre de vue que le but d'un bon apprentissage n'est pas d'obtenir un excellent résultat sur la base d'apprentissage, mais de permettre au réseau de traiter correctement l'ensemble de la base de test. Un entraînement trop intensif du réseau tend à lui faire perdre ses capacités de généralisation, car il finit par se spécialiser sur les données de la base d'apprentissage : il les apprend par coeur.

Le problème de la généralisation est fortement dépendant de la topologie du réseau. Lorsqu'on augmente le nombre de neurones cachés, le nombre d'itérations nécessaire pour obtenir une précision donnée diminue, mais la généralisation se fait moins bien. Cela tient au fait que le réseau est surdimensionné : les paramètres (les poids) en surnombre évoluent plus facilement vers une des multiples solutions, mais le réseau apprend moins bien les caractéristiques des données qui lui permettraient de généraliser correctement.

Le choix du nombre de neurones cachés est donc d'une grande importance.

3.7.1 Considérations pratiques

Quelques choix s'imposant pour l'entraînement du réseau multi couches :

- Les poids du réseau doivent être initialisés à de petites valeurs aléatoires
- La valeur du taux d'apprentissage μ a un effet significatif sur les performances du réseau, si ce taux est petit l'algorithme converge lentement, par contre s'il est grand l'algorithme risque de générer des oscillations.

- Généralement, η doit être compris entre 0 et 1 pour assurer la convergence de l'algorithme vers une solution optimale.
- Il n'existe pas de règles permettant de déterminer le nombre de couches cachées dans un réseau donné ni le nombre de neurones dans chacune d'elles.
- Théoriquement, l'algorithme ne doit se terminer dès que le minimum de l'erreur commise par le réseau sera atteint, correspondant à un gradient nul, ce qui n'est jamais rencontré en pratique. C'est pourquoi un seuil est fixé à priori afin d'arrêter l'apprentissage.

3.7.2. Condition d'arrêt

Le nombre d'itérations est important car:

- Si trop faible, l'erreur n'est pas suffisamment réduite.
- Si trop grand, le réseau devient trop spécifique aux données d'entraînement.

Il y a plusieurs conditions d'arrêt possible:

- Après un certain nombre fixe d'itérations.
- Lorsque l'erreur dans les sorties des exemples d'entraînement descend en dessous d'une certaine borne.
- Lorsque l'erreur avec les exemples de validation descend en dessous d'une certaine borne.

3.7.3 Problème du surapprentissage (overfitting)

A trop vouloir ajuster le réseau sur les valeurs, on risque de trouver une généralisation trop compliquée. Pour les réseaux de neurones, il faut trouver un compromis entre une fonction simple qui donne quelques erreurs sur les exemples d'apprentissage mais qui généralise bien, en comparaison d'une fonction plus complexe qui donne moins d'erreurs sur les exemples mais qui généralise sans doute de façon inappropriée. Un problème qui apparaît lors d'un apprentissage est le problème du sur

apprentissage. Si le réseau de neurone apprend par coeur, il donnera de mauvais résultats quand on lui présentera des données un peu différentes .

Pour éviter le surapprentissage, on procède typiquement de la façon suivante : on divise les exemples en 2 ensembles. Le 1er sert à l'apprentissage du réseau, le 2ème sert de test. Si la généralisation est parfaite, il ne doit pas y avoir plus d'erreur avec les exemples du 2ème ensemble qu'avec le 1er. Une technique simple pour définir le critère d'arrêt consiste ainsi à arrêter l'apprentissage lorsque l'erreur sur l'ensemble de test commence à augmenter.

Des méthodes existent pour optimiser la phase d'apprentissage afin que le phénomène de sur ou sous apprentissage disparaisse.

3.7.3.1 Utilisation d'une base de validation pendant l'apprentissage

Le principe consiste à mesurer les performances pendant l'apprentissage sur une base de validation qui est différente de la base d'apprentissage. Lorsque le modèle n'est pas trop ajusté aux données de l'apprentissage, les fonctions de coût sur la base de validation et d'apprentissage diminuent ensemble. Lorsque le modèle commence à être surajusté, la fonction de coût sur la base d'apprentissage continue de diminuer, alors que la fonction de coût sur la base de validation augmente.

3.7.3.2 Arrêt prématuré

Comme nous l'avons vu précédemment, l'apprentissage consiste à minimiser, grâce à un algorithme itératif, une fonction de coût calculée sur la base d'apprentissage. La méthode de l'arrêt prématuré (*early stopping*) consiste à arrêter les itérations avant la convergence de l'algorithme. Si la convergence n'est pas menée à son terme, le modèle ne s'ajuste pas trop finement aux données d'apprentissage : le surajustement est limité.

Pour mettre en oeuvre cette méthode, il faut déterminer le nombre d'itérations à utiliser pendant l'apprentissage. La méthode la plus classique consiste à suivre l'évolution de la fonction de coût sur une base de validation, et à arrêter les itérations lorsque le coût calculé sur cette base commence à croître.

Cependant cette méthode peut être inapplicable, car il est difficile de déterminer avec précision le moment exact où il faut arrêter l'apprentissage puisque les performances sur la base de validation ne se dégradent pas nettement.

3.8 Conclusion

L'apprentissage à l'aide de réseaux de neurones est tolérant au bruit et aux erreurs. Le temps d'apprentissage peut être long, par contre, après apprentissage, le calcul des sorties à partir d'un vecteur d'entrée est rapide. La critique principale est que le résultat de l'apprentissage, c'est-à-dire le réseau de neurones calculé par l'algorithme d'apprentissage, n'est pas interprétable par l'utilisateur : on ne peut pas donner d'explication au calcul d'une sortie sur un vecteur d'entrée. On parle de << boîte noire >>. Ceci est la principale différence entre réseaux de neurones et arbres de décision. Si l'utilisateur a besoin de pouvoir interpréter

le résultat de l'apprentissage, il choisira un système basé sur les arbres de décision, sinon les deux méthodes sont concurrentes.

Dans le chapitre suivant nous allons essayer de contourner ces problèmes d'apprentissage dans les réseaux de neurones en les adaptant pour chaque situation (apprentissage spécialisé par patient et pour chaque arythmie cardiaque)

Chapitre 4

Classification neuronale spécialisée des arythmies cardiaques

4.1 introduction

Dans certains domaines tels que le médical, le principal obstacle auquel est confronté le concepteur lors de la mise en oeuvre d'un système de classification est la caractérisation des données par des paramètres fiables et pertinents.

Dans ce chapitre, nous aborderons le problème de reconnaissance des extrasystoles ventriculaires (ESV 'V') et des blocs de branche droit (BBD 'R') et gauche (BBG 'L'). De telles arythmies conduisent généralement à un élargissement anormal du complexe QRS. Le choix des descripteurs représentatifs est indispensable lors de la conception d'un classifieur pour identifier tels troubles cardiaques. Il convient de remarquer que de nombreuses approches citées dans la littérature ont porté sur la difficulté que représentent la mesure et le choix des paramètres pertinents du signal ECG et leur classification. Lagerholm et Person [Lagerholm 2000] ont caractérisé chaque cycle cardiaque par son intervalle RR et les coefficients résultant de la décomposition du complexe QRS en fonctions d'Hermite de base. Ces paramètres ont été utilisés pour séparer entre les arythmies disponibles dans la base de données MIT-BIH, en utilisant une carte de Kohonen. Krishna et Sohambi [Krishna 2003] ont employé un perceptron multicouche pour identifier les battements ectopiques, avec un vecteur d'entrée composé des coefficients d'ondelettes et le rythme RR. Minami et al [Minami 1999] ont calculé le spectre de chaque complexe QRS pour la reconnaissance des tachycardies ventriculaires via un réseau de neurones multicouches. Fredric et Soowhan [Fredric 1996] ont calculé deux coefficients par la méthode de la prédiction linéaire à chaque complexe QRS pour la classification des VEBs.

Nichola et al [Nicholas 2004] ont segmenté le signal ECG en un ensemble de valeurs caractéristiques, en utilisant les modèles de Markov cachés. D'autres techniques ont été testées pour la bonne représentation d'un cycle cardiaque comme la transformée de Fourier discrète et la prédiction linéaire [Chikh2005].

En ce qui nous concerne, nous avons testé et comparé plusieurs classifieurs neuronaux. Nous avons employé un perceptron multicouche pour classer trois arythmies cardiaques (V,R,L) et les différencier du cas normal.

Un autre problème, qui affecte énormément les résultats de la classification des arythmies cardiaques, est lié à la grande ressemblance des signaux ECG de classes différentes d'une part et, d'autre part, à leur disparité au sein d'une même classe. Dans ce cas une séparation très fine des différentes classes d'arythmies doit être réalisée. Compte tenu de cette difficulté, nous proposons dans ce travail deux classifieurs neuronaux, un classifieur pour la spécialisation de l'arythmie (entraîné par une base d'arythmies variée) et un autre adapté au patient que nous l'appellerons classifieur de spécialisation patient.

Nous commençons par rappeler brièvement les différentes phases de classification ensuite nous présentons l'approche neuronale qui est utilisée comme modèle de classification dans ce mémoire.

L'objectif de notre application est l'adaptation en ligne et hors ligne des classifieurs c'est ce que nous appelons la spécialisation. L'intérêt de la spécialisation n'est pas d'affecter la généralisation car le fait de spécialiser n'induit pas à un sur-apprentissage mais par contre à une amélioration de la sensibilité des classifieurs dans le cas de la spécialisation de l'arythmie, et à une amélioration du taux de classification correcte dans le cas de la spécialisation patient. Ces deux types de spécialisation sont le centre de la problématique de ce mémoire de magister.

4.2. Principes de la classification

Résoudre un problème de classification, c'est trouver une application de l'ensemble des objets à classer, décrits par les variables descriptives choisies, dans l'ensemble des classes. L'algorithme ou la procédure qui réalise cette application est appelé classifieur.

Dans certains cas, il est possible de décrire complètement, de manière linguistique, la démarche de classification; dans ce cas, un algorithme reproduisant cette démarche peut être construit et le problème est résolu. Dans d'autres cas, il est impossible de décrire précisément la classification; une solution consiste alors à demander à un professeur (expert) de classer un échantillon d'objets. Des méthodes de résolution qui apprennent par l'exemple (ici un exemple est un objet déjà classé par le professeur) sont capables de reproduire la classification de l'expert et, ensuite, de classer automatiquement de nouveaux exemples inconnus. Nous utilisons ce deuxième type qui se base sur l'apprentissage par l'exemple.

La tâche de l'algorithme de classification est d'autant plus aisée que la représentation de l'objet est pertinente. Nous abordons dans ce chapitre une étape primordiale dans la classification qui est l'extraction des descripteurs. Ce qui va nous aider à la préparation de la base de données .cette base d'exemples a un effet direct sur l'apprentissage et sur la qualité des performances des classifieurs .

Par exemple pour la classification des extrasystoles ventriculaires (ESV), chaque cycle cardiaque est entièrement défini par les paramètres de l'ECG. Si les modules d'acquisition (électrodes) ne fournissent pas des descripteurs pertinents à l'algorithme de classification, celui-ci ne pourra pas faire de miracle et distinguer les différentes arythmies.

Une bonne caractérisation de l'ECG conditionne la réussite de la phase de classification, pour cette raison nous étudions dans la section suivante comment arrive-t-on à définir les caractéristiques des différentes arythmies cardiaques que nous avons à classer dans ce travail de magister.

4.2.1 les phases de classification

La classe reçoit une définition qui peut être une définition purement descriptive ou une interprétation, par exemple du type diagnostic médical. Le classifieur qui réalisera le classement des formes doit passer par deux phases, une phase d'apprentissage et une phase de test [Leschi 1991].

4.2.1.1 Phase d'apprentissage

Le but de l'apprentissage comme nous l'avons décrit au chapitre précédent est de découvrir les règles qui gouvernent et régissent des formes. L'apprentissage est un processus calculatoire qui doit être capable d'amener à une certaine prédiction et à une certaine généralisation. Il existe principalement deux types d'apprentissage, supervisé et non supervisé [Milgram 1993]. Dans le premier cas, on doit apprendre des associations (individus, classes); dans le deuxième cas on ne fournit pas d'indications sur les classes.

4.2.1.2 Phase de test

Cette phase doit permettre l'affectation d'un nouvel objet à l'une des classes, au moyen d'une règle de décision intégrant les résultats de la phase d'apprentissage [Leschi 1991]. L'objectif est d'obtenir une estimation la plus fidèle possible du comportement du classifieur dans des conditions réelles d'utilisation. Pour cela, des critères classiques comme les taux de classification et les taux d'erreur sont presque systématiquement utilisés. Mais d'autres critères, comme la spécificité et la sensibilité, apportent aussi des informations utiles.

a. Taux de classification

Les taux de classification et d'erreurs permettent d'évaluer la qualité du classifieur par rapport au problème pour lequel il a été conçu. Ces taux sont évalués grâce à une base de test qui contient des formes étiquetées par leur classe réelle d'appartenance comme celles utilisées pour l'apprentissage afin de pouvoir vérifier les réponses du classifieur .

Pour que l'estimation du taux de reconnaissance soit la plus fiable possible, il est important que le classifieur n'ait jamais utilisé les échantillons de cette base pour faire son apprentissage, de plus, cette base de test doit être suffisamment représentative du problème de classification [Guyon 1998].

En général, quand les échantillons étiquetés à disposition sont suffisamment nombreux, ils sont séparés en deux parties disjointes et en respectant les proportions par classes de la base initiale. Une partie sert pour former la base d'apprentissage et l'autre pour former la base de test.

Les performances en terme de taux de classification sont alors déterminées en présentant au classifieur chacun des exemples de la base de test et en comparant la classe donnée en résultat à la vraie classe.

Le taux de classification correcte est défini par : $CC = \frac{vp(i) + vn(i)}{vp(i) + vn(i) + fp(i) + fn(i)}$

b. sensibilité et spécificité

L'évaluation des performances d'un classifieur peut être réalisée par l'appréciation de deux lois statistiques, qui sont la sensibilité et la spécificité.

La sensibilité $Se(i)$ représente la probabilité de bonne classification de la classe i et la spécificité $Sp(i)$ est une mesure indirecte de la probabilité de fausse alarme égale à $1 - Sp(i)$.

Pour rappel, ces deux quantités sont définies par :

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

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

où les grandeurs $VP(i)$, $FN(i)$, $VN(i)$, $FP(i)$ sont définies dans le tableau 4.1 :

	Présence d'événement de classe i	Absence d'événement de classe i
Classification Positive	Vrai Positif $VP(i)$	Faux Positif $FP(i)$
Classification Négative	Faux Négatif $FN(i)$	Vrai Négatif $VN(i)$

Tab 4.1 : les définitions des grandeurs VP , VN , FP et FN

4.3 L'extraction des descripteurs et indicateurs pertinents

La classification se fait à partir de descripteurs ou paramètres caractérisants les différentes formes des signaux ECG .

Il existe deux approches pour caractériser une forme :

_ **l'approche purement mathématique** (analyse en composantes principales, prédiction linéaire, transformée de Fourier, analyse de données, etc ..), qui consiste à retenir comme paramètres certains coefficients pertinents.

_ **l'approche intuitive**, qui laisse au spécialiste le soin de définir les descripteurs qui lui semblent importants. Cette approche donne souvent de meilleurs résultats, car les paramètres choisis résultent d'une grande expérience et peuvent être plus discriminants.

4.4 Intérêts des réseaux de neurones

4.4.1 Introduction

Les réseaux de neurones connaissent depuis quelques années un succès croissant dans divers domaines des sciences de l'Ingénieur; celui du génie biomédical n'échappe pas à cette règle [Suzuki 1995], [Maglaveras 1998], [Osowski 2001], [Acharya 2004]. La résolution d'un modèle de classification est considérée comme une approche d'optimisation, qui consiste à trouver les coefficients du réseau minimisant une fonction de coût par calcul du gradient, en appliquant des méthodes comme la rétropropagation. Les réseaux de neurones sont des approximateurs universels .Notons aussi que la propriété d'approximation universelle n'est pas spécifique aux réseaux de neurones : les polynômes, les séries de Fourier etc, possèdent cette même particularité. Ce qui différencie les réseaux de neurones des autres approximateurs universels usuels, c'est leur parcimonie : pour obtenir une approximation d'une précision donnée, les réseaux de neurones utilisent moins de paramètres que les approximateurs usuels. En pratique, les RN sont donc avantageux, par rapport aux approximateurs conventionnels, lorsque l'on cherche à résoudre un problème de grande dimension (par exemple, plus de 10 ou 20 variables).[Chikh2005].

4.4.2 Choix des réseaux de neurones pour la classification

Notre choix pour ce travail s'est porté sur l'utilisation d'un réseau multi couches avec l'algorithme d'apprentissage le plus répandu qui est la rétro-propagation du gradient de l'erreur, nous expliquons dans ce qui suit les raisons qui nous ont mené à prendre ce de ce choix.

La découverte en 1985 par différentes équipes, [Le Cun 1984], d'un algorithme d'apprentissage pour les réseaux multi-couches (rétro-propagation du gradient de l'erreur) a ouvert de nouveaux horizons aux modèles connexionnistes. Depuis le premier congrès international sur les réseaux de neurones en 1987, on assiste à une explosion incroyable des colloques et des congrès autour de ce sujet. Les raisons de cette explosion peuvent être résumées ainsi :

1. plusieurs architectures, mettant en coopération un certain nombre de neurones formels, ont permis d'apporter des solutions au moins équivalentes aux solutions déjà existantes;
2. l'utilisation du principe de l'apprentissage d'une base contenant des exemples triés d'un problème réel réduit considérablement les difficultés rencontrées pour la formalisation de tels problèmes;
3. la facilité de développement d'une solution neuronale dans la résolution d'un problème est très attractive et permet à des non spécialistes d'intégrer de telles solutions très rapidement.

4.4.3 Mise en oeuvre des réseaux de neurones

La mise en oeuvre des réseaux de neurones comporte à la fois une partie conception, dont l'objectif est de permettre de choisir la meilleure architecture possible, et une partie de calcul numérique, pour réaliser l'apprentissage.

L'apprentissage d'un réseau de neurones nécessite différents éléments, que nous avons déjà décrits au chapitre 3 qui sont :

- 1-La fonction de coût : nous avons pris comme fonction de coût dans notre travail l'erreur quadratique, où l'objectif de l'apprentissage est donc de trouver un vecteur qui minimise cette fonction.

2-Le calcul du gradient : Sauf dans le cas d'un modèle linéaire par rapport aux paramètres, la minimisation de la fonction de coût s'effectue de manière itérative, par une descente de gradient qui est une méthode du premier ordre. Il existe aussi des méthodes du second ordre, Levenbeg-Marquardt ou quasi-Newton, qui sont souvent plus efficaces mais plus complexes à mettre en oeuvre.

3-l'algorithme d'optimisation

L'algorithme d'optimisation permet d'ajuster de manière itérative les poids du modèle de façon à converger vers un minimum local de la fonction de coût.

4.5 le modèle de classification neuronale

4.5.1 Introduction

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 des algorithmes d'apprentissage de type supervisé. La stratégie que nous envisageons de mettre en oeuvre est présentée dans la figure4.1

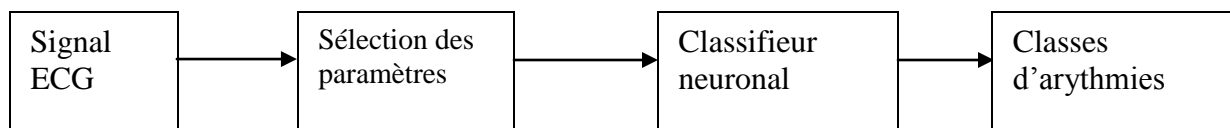


Fig.4.1 : schéma global de la classification

4.5.2 Sélection des descripteurs d'un cycle cardiaque

4.5.2.1 Introduction

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

Pour ce travail nous avons utilisé dix paramètres .notre choix est basé sur une approche intuitive, et avec l'aide de spécialistes en cardiologie ,nous avons pris les descripteurs qui semblent 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 » [Bechar,Chikh 2003] développée sous matlab. Cette interface a été modifiée et développée afin de pouvoir détecter tous les rythmes sur l'ensemble des enregistrements (figure 4.2).

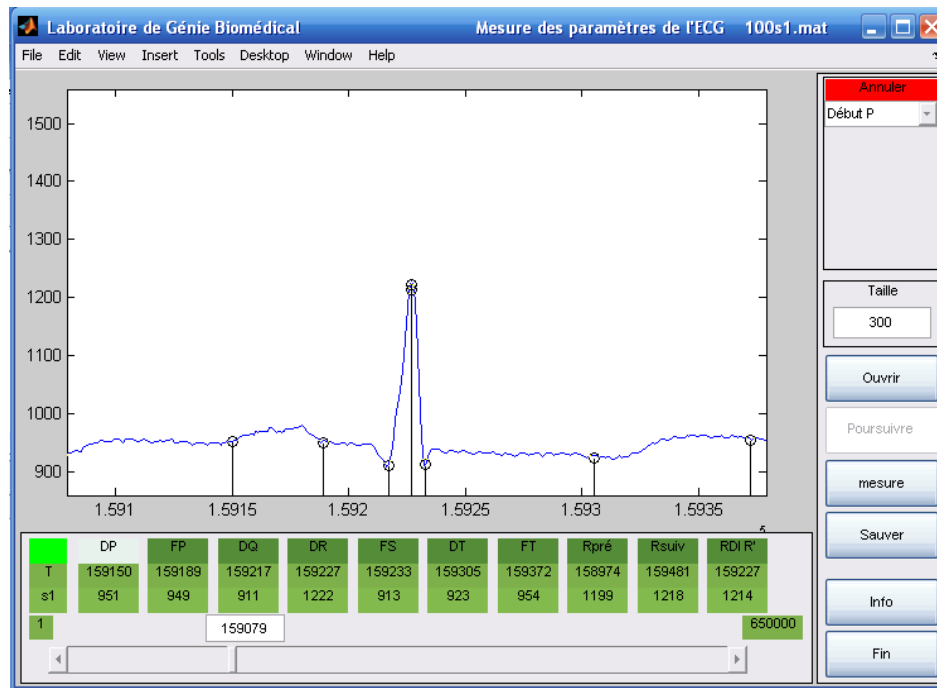


Figure 4.2 :Interface de Mesure des Paramètres de l'ECG .

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, S et T.

Ces descripteurs permettent de caractériser un battement cardiaque. Ils sont comparables aux paramètres sur lesquels sont construites les connaissances médicales.

4.5.2.2 Présentation des différents descripteurs

Les critères électrocardiographiques des différentes arythmies nous ont mené à retenir des descripteurs pertinents.

Les Extrasystoles ventriculaires se caractérisent par :

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

Le diagnostic des blocs de branche gauche et blocs de branche droit repose sur des critères :

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

Une description plus approfondie de ces arythmies peut être trouvée dans le chapitre1.

Les descripteurs retenus pour l'entrée du classifieur neuronal sont :Durée P ; intervalle PR ; complexe QRS ; segment ST ; intervalle QT ; RR précédent ; RR suivant ; RDI (retard de la déflexion intrinsecoïde); durée battement ; rapport RRsuivant \ RRprécédent.

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 : on le mesure 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.

4.5.3 Sélection de la base d'exemples

Nous avons utilisé, dans le cadre de ce mémoire de magister, 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 4.2.

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.

Classification neuronale spécialisée des arythmies cardiaques

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

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

Tab.4.2 Les enregistrements choisis de la base de données MIT-BIH

Remarque : Pour l'enregistrement 109, nous l'avons divisé en deux : 10911 contient tous les cycles (104 cycles annotés L); par contre 1091 ne contient que les 25 premiers cycles annotés L.

4.5.4 Architecture des classifieurs neuronaux

4.5.4.1 Algorithme d'apprentissage

L'architecture que nous avons retenue est un Perceptron multicouche. Pour l'apprentissage, il existe généralement trois protocoles utilisés par les réseaux multicouches :

- _ apprentissage stochastique, hors-ligne,
- _ apprentissage par lot,
- _ apprentissage en ligne.

L'apprentissage en ligne est utile lorsque l'ensemble d'apprentissage est très grand ou que la mémoire devient prohibitive, pour la sauvegarde des données. Aujourd'hui, les ordinateurs ne présentent plus de problèmes de limitation en mémoire et en coût. Dans l'apprentissage en ligne, chacune des formes est présentée une seule fois, ce qui peut sembler donner un apprentissage rapide mais nécessite plus de données d'apprentissage.

L'apprentissage par lot est très utilisé de nos jours. Dans Matlab, on retrouve les trois protocoles d'apprentissage mais nous utiliserons seulement le protocole d'apprentissage par lot. Au cours de ce travail nous avons utilisé principalement deux algorithmes très classiques en optimisation non linéaire que nous avons décrits au chapitre 3.

1. la méthode de Quasi-Newton : qui est une alternative aux méthodes du gradient conjugué pour atteindre rapidement l'optimisation des paramètres de poids. Dans Matlab, on retrouve cet algorithme dans la fonction « `trainbfg` ».

2. la méthode de Levenberg-Marquardt : Dans Matlab, on retrouve cet algorithme dans la fonction « `trainlm` ».

En les confrontant sur plusieurs problèmes, il apparaît qu'aucune de ces méthodes ne prend un avantage considérable sur l'autre. Chaque méthode d'optimisation possède des avantages

et des inconvénients. Bien entendu, aucune méthode ne conduit à coup sûr au minimum global. Il convient donc de se placer dans des conditions où les minima locaux sont aussi peu nombreux que possible ou correspondent à une erreur suffisamment faible.

4.5.4.2 Dimensionnement du réseau pour chaque classifieur

Plusieurs paramètres sont à déterminer au niveau de l'architecture des réseaux de neurones : nombre de neurones dans la couche d'entrée 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 au paragraphe 4.5.2.2:

le nombre de neurones sur la couche d'entrée est liée directement au vecteur de caractéristiques.

_ Pour le premier classifieur, que nous appelons classifieur neuronal global LM.

l'algorithme d'apprentissage est celui de Levenberg Marquardt. Avec la fonction « trailm ».

_ Pour le deuxième classifieur, que nous appelons classifieur neuronal global BFG.

l'algorithme d'apprentissage est celui de Quasi-Newton. Avec la fonction « trainbfg ».

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 utilisé une seule couche cachée composée de trois neurones lors de ce travail. Les fonctions d'activation des neurones cachés sont des sigmoïdes « tansig ».

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

La fonction d'activation est une fonction linéaire (purelin) qui assure les sorties suivantes.

$N=1, V=2, R=3, L=4$.

Le nombre d'itérations est fixé à 500, ce qui a permis de réaliser un apprentissage satisfaisant.

Ce paramètre est 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 0.001 pendant ces essais.

4.6 Résultats expérimentaux de la classification

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

Nous prenons comme annotation :103n pour les cycles que nous avons mesuré par l'IMPE et classés N de l'enregistrement 103 de MIT.

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

Nous appelons Réseau 1 le réseau dont la base d'apprentissage est formée à partir des mesures prises sur les enregistrements suivants :103n,200v,118r et 207l.

L'ordre de présentation des exemples des différentes classes est sans importance.

Nous adoptons la même configuration pour les autres réseaux (exemples classés 'N' exemples classés 'V' exemples classés 'R' exemples classés 'L')

Le Réseau2 est composé de 119n,214v,212r,109l.

Le Réseau3 est composé de 116n,203v,124r,111l.

Le Réseau4 est composé de 215n,208v,124r,109l.

Dans ce qui suit, nous présentons les résultats des deux classifieurs obtenus. Les performances sont représentées par la sensibilité 'Se' et la spécificité Sp' et le taux de classification, déjà définis plus haut.

4.6.1 Apprentissage et test des différents réseaux

L'apprentissage des différents réseaux sera effectué par deux fonctions « trainbfg » et « trainlm » implémentée sous Matlab.

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

Dans ce qui suit nous présentons quelques résultats choisis parmi plusieurs essais.

4.6.1.1 apprentissage et test du réseau2

Nous avons testé le réseau2 dans les deux cas avec une base composée d'une seule matrice 208v formée à partir de 152 cycles classés ESV de l'enregistrement 208 de MIT et une autre base plus grande et plus variée composée des différentes arythmies ESV,BBD et BBG et du cas normal.

L'extrasystole ventriculaire 'V' est l'arythmie ciblée dans ces deux tests.

Les résultats sont présentés dans le tableau 4.3 et figure 4.3.

	vp	vn	fp	fn	se	sp	cc	fonction
Réseau2 testé avec 208v	132	0	0	20	86.8421	NaN	86.8421	Trainbfg
	18	0	0	134	11.8421	NaN	11.8421	Trainlm
Réseau2 testé avec 215n208v124r109l1	132	117	0	20	86.8421	100	92.5651	Trainbfg
	18	218	22	134	11.8421	90.8333	60.2041	Trainlm

Tab 4.3 :résultats des tests de performances du réseau2

Remarque : Puisque $Sp = \frac{vn}{vn + fp}$ donc le résultat NaN veut dire la division impossible par zéro (vn=fp=0).

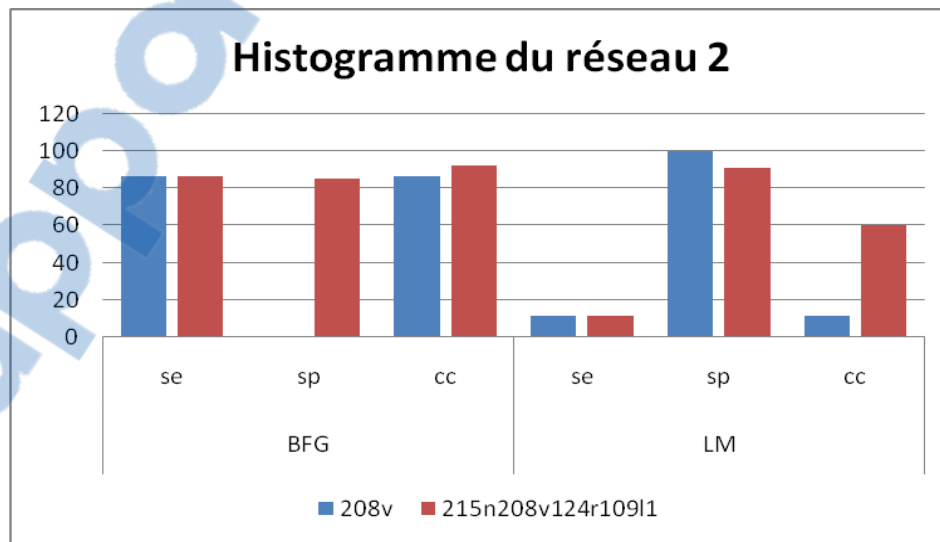


Fig 4.3: Les performances du réseau2 en utilisant les deux fonctions d'apprentissage 'trainbfg' et 'trainlm'.

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

Nous avons testé le réseau2 dans les deux cas avec une base composée d'une seule matrice 208v formée à partir de 152 cycles annotés V de l'enregistrement 208 de MIT et une base plus grande et plus variée composée des différentes arythmies ESV,BBD et BBG et du cas normal.

Notons que l'extrasystole ventriculaire 'V' est l'arythmie ciblée.

4.6.1.2 apprentissage et test du réseau3

Nous présentons dans le tableau 4.4 et la figure 4.4 les résultats des tests du réseau3 sur différentes bases. Comme pour le cas du réseau 2,nous avons effectué l'apprentissage par les deux méthodes celle de Quasi-Newton et celle de Levenberg-Marquardt.

	tp	tn	fp	fn	se	sp	cc	Fonction
Réseau3 testé avec 106n106v	28	25	0	6	82.3529	100	89.8305	Trainbfg
	9	27	0	24	27.2727	100	60	Trainlm
Réseau3 testé avec 214l214v	47	13	37	3	94	26	60	Trainbfg
	5	48	0	42	10.6383	100	55.7895	Trainlm
Réseau3 testé avec 208v	151	0	0	1	99.3421	NaN	99.3421	Trainbfg
	0	0	0	150	0	NaN	0	Trainlm
Réseau3 testé avec 215n208v124r10911	151	184	56	1	99.3421	77.0833	85.4592	Trainbfg
	0	237	3	150	0	98.75	60.7692	Trainlm
Réseau3 testé avec 207l	40	0	0	0	100	NaN	100	Trainbfg
	0	0	0	40	40	NaN	0	Trainlm
Réseau3 testé avec 103n200v118r207l	40	87	8	0	100	91.5789	94.0741	Trainbfg
	0	95	0	40	0	100	70.3704	Trainlm

Tab 4.4 :résultats des tests de performances du réseau3

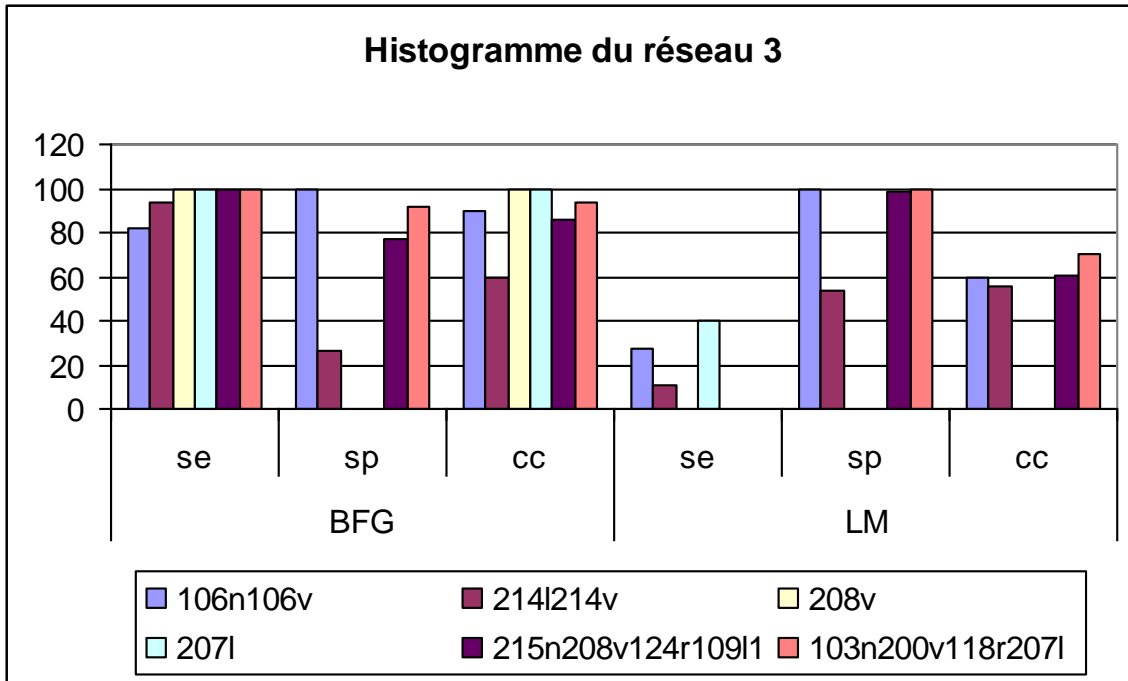


Fig 4.4: Les performances du réseau3 en utilisant les deux fonctions d'apprentissage 'trainbfg' et 'trainlm'

4.6.1.3 apprentissage et test du réseau4

Lors du test du réseau 4 avec la base 119n119v(l'arythmie ciblée est l'ESV) les résultats obtenus ;présentés dans le tableau 4.5 et la figure 4.5 ; 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 réseau 4 avec la base 212n212r et l'anomalie BBG 'L' lors du test du réseau 4 avec la base 214l214v.

	vp	vn	fp	fn	se	sp	cc	fonction
Réseau4 testé avec 119n119v	34	50	0	0	100	100	100	Trainbfg
	4	27	23	30	11.7647	54	36.9048	Trainlm
Réseau4 testé avec 212n212r	18	0	5	8	69.2308	0	58.0645	Trainbfg
	0	5	0	26	0	100	16.1290	Trainlm
Réseau4 testé avec 214l214v	40	49	0	10	80	100	89.8990	Trainbfg
	3	48	2	47	6	96	51	Trainlm

Tab 4.5 :résultats des tests de performances du réseau4

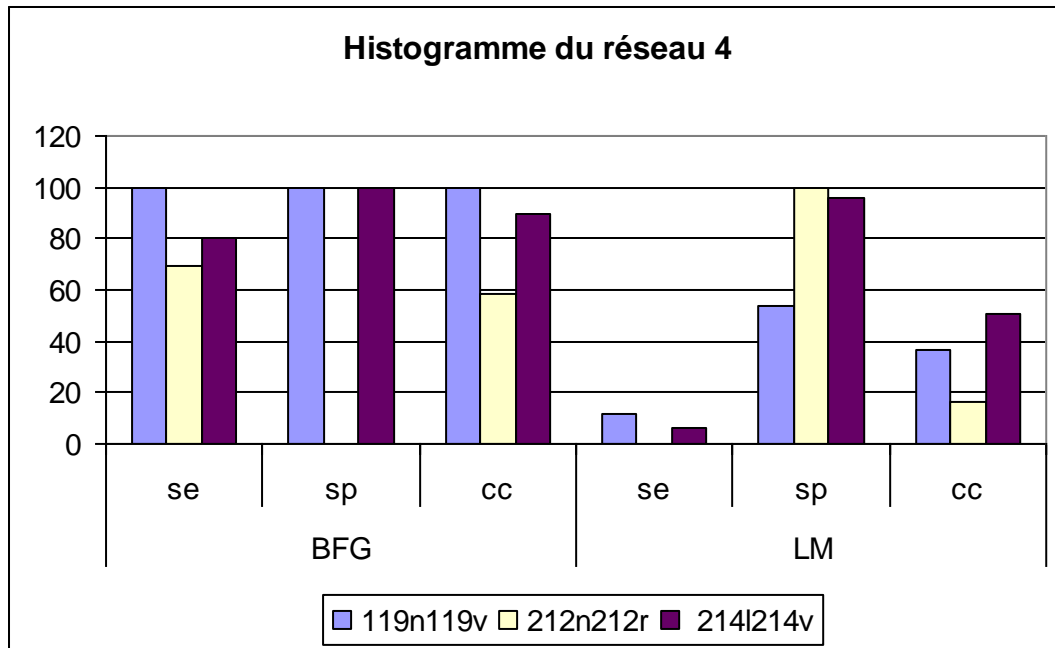


Fig 4.5: Les performances du réseau4 en utilisant les deux fonctions d'apprentissage 'trainbfg' et 'trainlm'

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 réseaux sur multiples bases de test nous ont mené à choisir la fonction « trainbfg » pour l'apprentissage des réseaux spécialisés.

Dans notre étude, nous avons développé des classifieurs spécialisés pour chaque patient et pour chaque arythmie. Nous avons utilisé le meilleur classifieur, celui qui a obtenu les meilleurs performances lors des expérimentations précédentes, c'est-à-dire le réseau qui utilise comme fonction d'apprentissage la méthode de Quasi-Newton (BFGs), pour élaborer cette tâche.

Le classifieur choisi (classifieur neuronal global) sera mis sous un deuxième apprentissage pour le spécialiser.

Nous présentons dans la section suivante les résultats de cette méthode.

4.7 Classification spécialisée

4.7.1 Spécialisation par patient

Nous avons conçu un classifieur global avec une base d'apprentissage très variée, et englobant toutes les pathologies ciblées dans ce travail (N,ESV,BBD et BBG).

On attend d'un classifieur global qu'il reconnaisse des anomalies cardiaques similaires sur différents patients, mais des fois il peut y avoir des échecs pour cela nous adaptons le classifieur global sur chaque patient.

Ce classifieur global va maintenant être spécialisé, au cas par cas, sur des patients particuliers.

Les résultats des tests de performances des classifieurs neuronaux spécialisés sont nettement meilleurs que ceux des classifieurs globaux.

Pour la spécialisation d'un patient nous prenons comme base d'apprentissage quelques cycles de son tracé électrocardiographique et nous procédons à un deuxième entraînement du réseau global qui a déjà été optimisé hors-ligne.

Donc le réseau spécialisé est initialisé par le réseau global et ne nécessite que peu de temps pour s'adapter à un patient.

4.7.2 Spécialisation patient106

Pour le patient 106 nous avons pris les 15 premiers cycles classés 'N'et'V' pour spécialiser les différents réseaux globaux définis dans la section précédente, et nous effectuons le test de ce classifieur spécialisé sur le même patient 106 .la base de test est formée à partir de cycles différents de ceux utilisés dans l'apprentissage spécialisé (106n106v).

Les résultats sont présentés dans les tableaux 4.6 ,4.7 et figure 4.6 .

Réseaux	vp	vn	fp	fn	Se	Sp	CC
Spécialisés avec 106n106v testés sur 106n106v							
Réseau1	26	19	0	1	96.2963	100	97.8261
Réseau2	27	19	0	0	100	100	100
Réseau3	26	19	0	1	96.2963	100	97.8261
Réseau4	22	19	0	5	81.4815	100	89.1304

Tab 4.6 : résultats des tests de performances des réseaux spécialisés

Réseaux	vp	vn	fp	fn	Se	Sp	CC
globaux testés sur 106n106v							
Réseau1	29	27	0	5	85.2941	100	91.8033
Réseau2	17	25	2	5	77.2727	92.5926	85.7143
Réseau3	28	25	0	6	82.3529	100	89.8305
Réseau4	4	27	0	30	11.7647	100	50.8197

Tab 4.7 : résultats des tests de performances des réseaux globaux

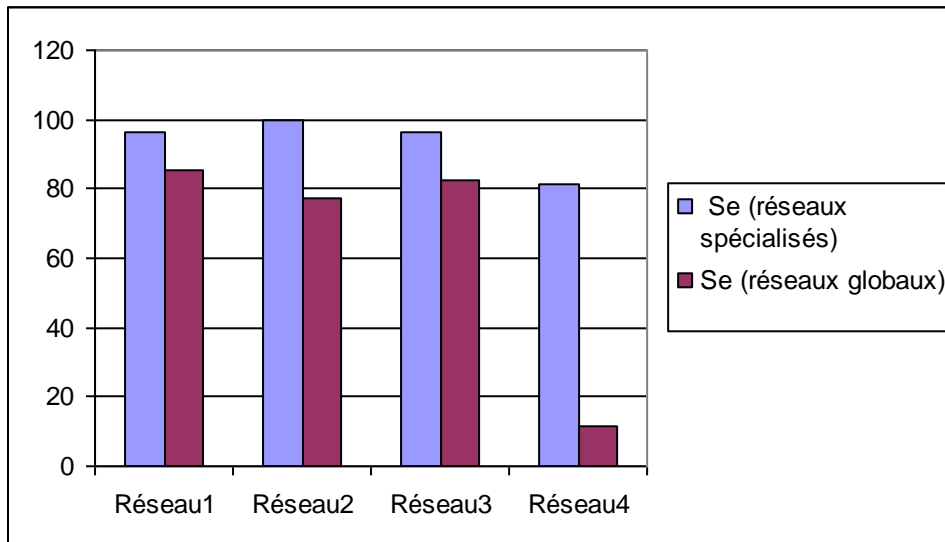


Fig 4.6 :taux de sensibilité des réseaux spécialisés et réseaux globaux.

La base de réapprentissage des réseaux globaux est composée de 8 cycles N et 7 cycles V du patient 106, donc la base de test des réseaux spécialisés contient 19 cycles N et 27 cycles V restants.

Les réseaux globaux sont testés sur la totalité des cycles 106n106v, donc 27 cycles N et 34 cycles V.

D'après les résultats présentés dans les tableaux, les classifieurs spécialisés ont correctement classés les cas ciblés qui sont les extrasystoles ventriculaires d'où une sensibilité maximale.

Nous remarquons une amélioration du taux de classification et de la sensibilité par rapport à ceux des réseaux globaux.

Presque tous les cas pathologiques ont été reconnus et bien classés.

Les résultats des tests de performances des classifieurs spécialisés nous donne une Probabilité de fausse alarme égale à zéro ($1 - Sp$) et une nette diminution des faux négatifs (pathologies ciblées confondues avec d'autres cas).

4.7.3 Spécialisation patient 212

Dans le cas précédent la pathologie ciblée était l'extrasystole ventriculaire. Nous avons aussi effectué une spécialisation du patient 212 et l'anomalie ciblée dans ce cas est le bloc de branche droit 'R'. La base d'apprentissage spécialisée est formée de 5 cycles 'N' et 8 cycles 'R' la spécialisation a permis au classifieur d'améliorer ses performances surtout pour la reconnaissance des cas ciblés. Le réseau global avait une sensibilité nulle, après sa spécialisation avec juste quelques cycles de cas classés 'R' dans la base de données MIT, la sensibilité est passée à un taux de 100% (une bonne adaptation).

Les résultats sont présentés dans les tableaux 4.8, 4.9 et figure 4.7.

Réseaux Spécialisés avec 212n212r testés sur 212n212r	vp	vn	fp	fn	Se	Sp	CC
Réseau3	18	5	0	0	100	100	100
Réseau4	18	5	0	0	100	100	100

Tab 4.8 : résultats des tests de performances des réseaux spécialisés

Réseaux globaux testés sur 212n212r	vp	vn	fp	fn	Se	Sp	CC
Réseau3	0	5	0	26	0	100	16.1290
Réseau4	18	0	5	8	69.2308	0	58.0645

Tab 4.9 : résultats des tests de performances des réseaux globaux

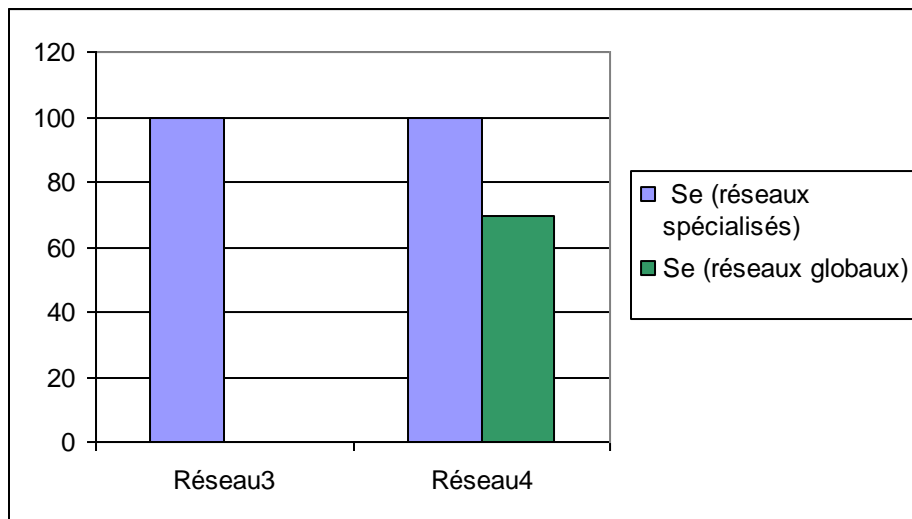


Fig 4.7 :taux de sensibilité des réseaux spécialisés et réseaux globaux.

La spécificité ,dans ce cas où nous présentons dans la base de test que deux classes N et R, représente le taux de bonne classification du cas non ciblé c à d le cas normal. Les réseaux 3 et 4 spécialisés ont reconnus et correctement classés les cas normaux et les cas pathologiques.

Un classifieur spécialisé idéal est défini par ces deux caractéristiques :différencier entre les cas présentés et les classer correctement avec une bonne généralisation et éviter le surapprentissage.

4.7.4 Spécialisation patient214

Nous avons appliqué la spécialisation sur un autre patient le 214 avec comme pathologie ciblée le bloc de branche gauche 'L'.

La base d'apprentissage spécialisée est formée de 10 cycles 'L'et 8 cycles 'V' et la base de test 214l214v de 40 cycles 'L' et 42 cycles 'V'.

Les résultats sont présentés dans les tableaux 4.10 ,4.11 et figure 4.8 .

Réseaux Spécialisés avec 214l214v testés sur 214l214v	vp	vn	fp	fn	Se	Sp	CC
réseau3	40	38	4	0	100	90.4762	95.1220
réseau4	40	38	4	0	100	90.4762	95.1220

Tab 4.10 : résultats des tests de performances des réseaux spécialisés

Réseaux globaux testés sur 214l214v(cible 'L')	vp	vn	fp	fn	Se	Sp	CC
Réseau3	0	49	1	50	0	98	49
Réseau4	40	49	0	10	80	100	89.8990

Tab 4.11 : résultats des tests de performances des réseaux globaux

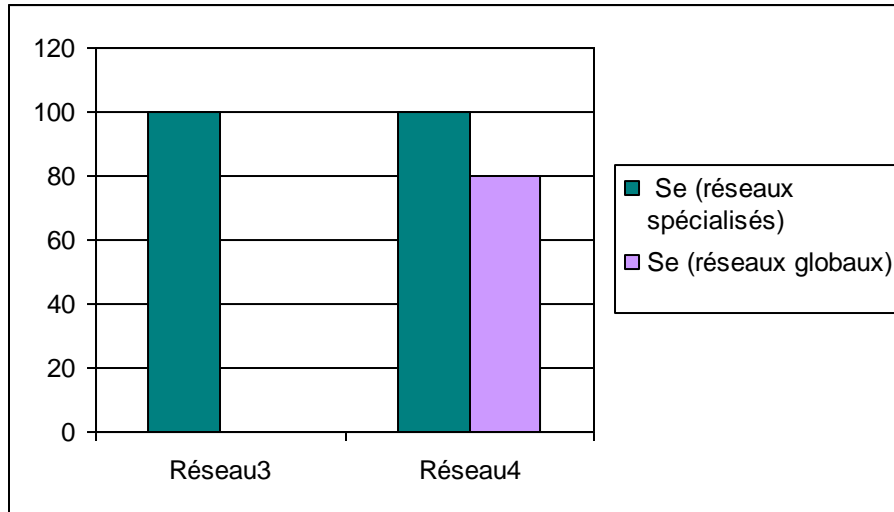


Fig 4.8 :taux de sensibilité des réseaux spécialisés et réseaux globaux.

La base de test des réseaux globaux est composée de 50 cycles L et 50 cycles V. Le réseaux 3 global a mal classé tous les cas 'L',mais après l'apprentissage spécialisé ,le classifieur a donné de très bons résultats ,la sensibilité ou probabilité de bonne classification est passée de 0 à 100% .

4.7.5 Efficacité de la spécialisation par patient

Nous savons que le réseau global déjà optimisé hors ligne est mis à disposition du spécialiste qui procèdera à l'annotation en ligne de quelques premiers cycles de l'ECG du patient.

Donc la présence d'un cardiologue pour annoter en ligne l'enregistrement ne dure q'un cours moment,ça représente un des avantages de la spécialisation(le médecin n'aura pas à annoter tous les cycles mais juste quelques uns).

Si les premiers cycles sont des cas normaux alors est ce que le classifieur sera en mesure d'identifier les cas pathologiques.

Dans cette expérimentation nous allons spécialiser notre classifieur global uniquement sur des cycles normaux et les tester sur d'autres cycles (V,R,L).

Les réseaux 1 et 2 sont entraînés sur 10 battements classés N du patient 106 et sont testés sur une base formée par des cycles non présentés à la base d'apprentissage, 17 cycles N et 34 cycles V du même patient.

Les résultats des tests obtenus sont très satisfaisants avec un taux de classification de classification de l'arythmie ciblée de 85%. (l'arythmie ciblée est l'ESV).

Tous les cas normaux présentés dans la base de test ont été classés correctement, le taux de spécificité est de 100%.

Nous pouvons voir le rôle de l'apprentissage spécialisé en comparant les résultats des réseaux spécialisés et globaux effectués sur la même base de test et du même patient présentés dans les tableaux 4.12, 4.13 et figure 4.9.

Réseaux	vp	vn	fp	fn	Se	Sp	CC
Spécialisés avec 106n testés sur 106n106v							
Réseau1	29	17	0	5	85.2941	100	90.1961
Réseau2	29	17	0	5	85.2941	100	90.1961

Tab 4.12 : résultats des tests de performances des réseaux spécialisés

Réseaux	vp	vn	fp	fn	Se	Sp	CC
globaux testés sur 106n106v							
Réseau1	29	27	0	5	85.2941	100	91.8033
Réseau2	17	25	2	5	77.2727	92.5926	85.7143

Tab 4.13 : résultats des tests de performances des réseaux globaux

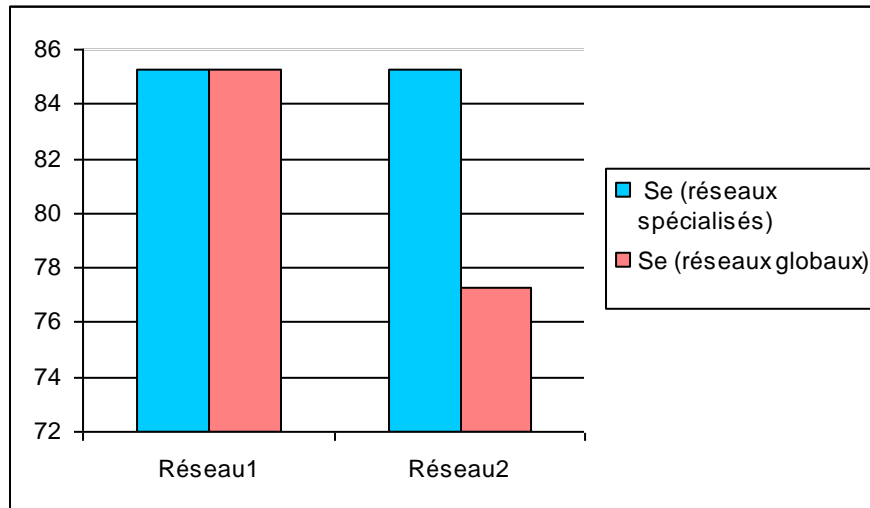


Fig 4.9 :taux de sensibilité des réseaux spécialisés et réseaux globaux.

Nous remarquons que le réseau 2 spécialisé a reconnu plus d'extrasystoles ventriculaires 'ESV' ($vp=29$) que le réseau 2 global ($vp=17$).

Donc l'apprentissage spécialisé nous a permis d'améliorer la sensibilité.

La spécialisation patient ;comme nous l'avons nommée ;nous permet de proposer un système automatique, qui nécessiterait la présence d'un spécialiste uniquement pendant les premières minutes, nécessaires pour la spécialisation, puis qui serait totalement autonome par la suite.

Les réseaux spécialisés n'ont besoin que de peu de temps pour s'adapter à un patient. Lors de l'apprentissage spécialisé l'erreur quadratique est de 10^{-6} tandis que lors du premier apprentissage dans les réseaux globaux, elle était de 10^{-3} .

Les figures (4.10),(4.11),(4.12),(4.13) et (4.14) nous aident à apprécier la différence qu'il y a entre le temps de convergence des réseaux globaux et spécialisés.

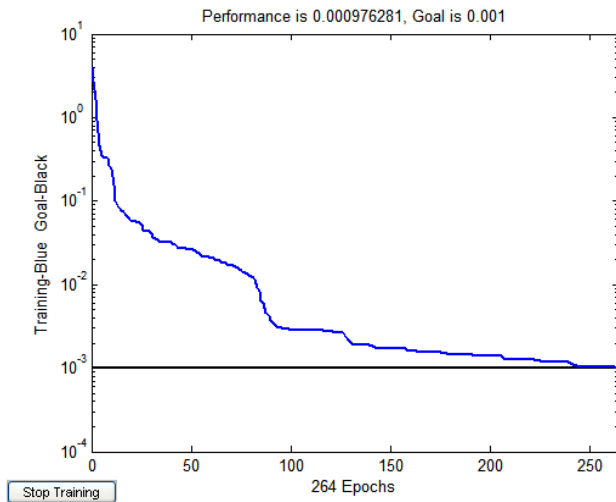


Fig4.10 Apprentissage du réseau1 en utilisant « trainbfg »

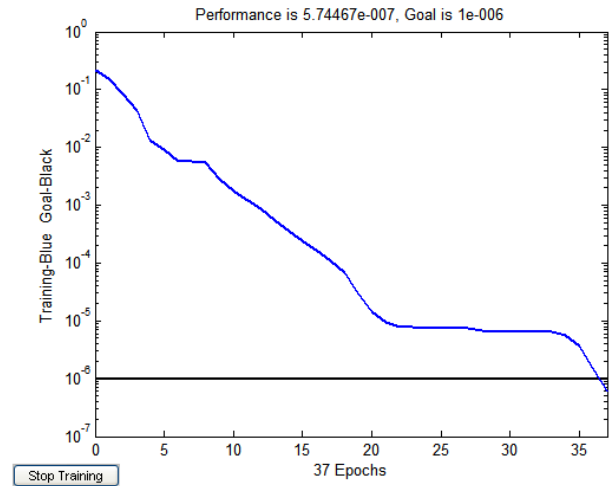


Fig4.11 Apprentissage du réseau spécialisé avec 106n106v en utilisant « trainbfg »

L'apprentissage du réseau 1 global a convergé au bout de 260 itérations pour un seuil d'erreur de 10^{-3} alors que le réseau spécialisé a pris moins de temps à converger (40 itérations) pour un seuil d'erreur de 10^{-6} .

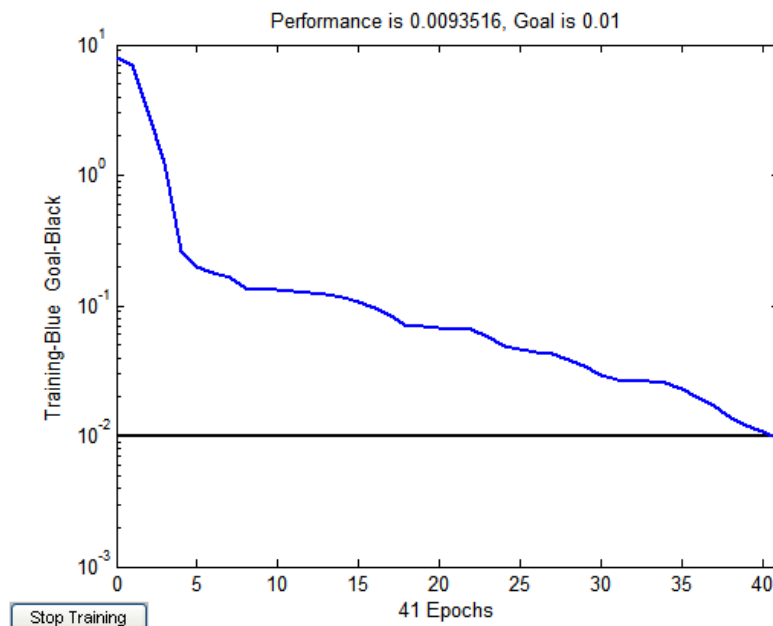


Fig4.12 : Apprentissage du réseau4 en utilisant « trainbfg »



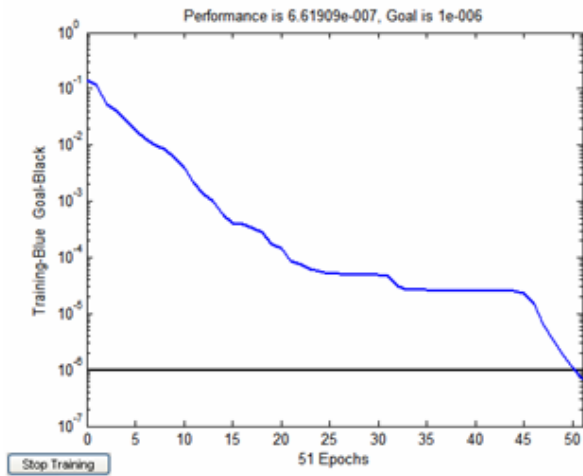


Fig4.13 :Apprentissage du réseau4 spécialisé avec 214l214v en utilisant « trainbfg »

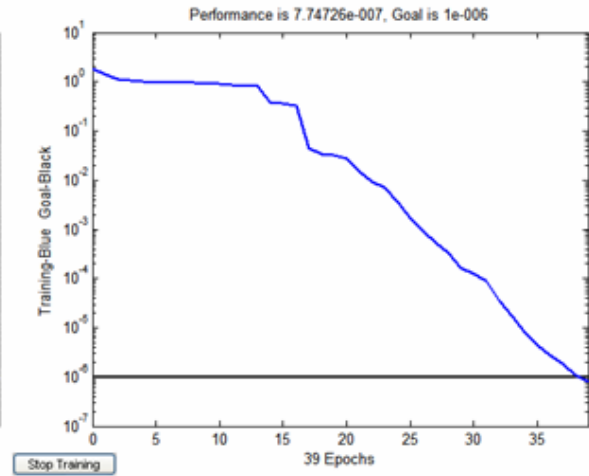


Fig4.14 :Apprentissage du réseau4 avec 212n212r en utilisant « trainbfg »

Le réseau 4 global a convergé pour un seuil d'erreur de 10^{-2} (fig 4.12) alors que les réseaux spécialisés avec les deux bases (124l124v et 212n212r) ont convergé pour un seuil d'erreur quadratique de 10^{-6} .(fig 4.13 et fig 4.14).

L'apprentissage spécialisé ne prend pas beaucoup de temps pour converger et permet au réseau de diminuer le seuil d'erreur quadratique.

Le classifieur spécialisé patient est optimisé en-ligne, nous aborderons dans la section suivante la spécialisation d'arythmie qui se base sur un apprentissage hors-ligne.

4.8 Spécialisation arythmie

Comme déjà défini dans la spécialisation patient le réseau global est déjà optimisé et les poids sont enregistrés,un deuxième apprentissage est effectué sur ce réseau,nous allons mettre en œuvre un classifieur spécialisé pour chaque arythmie .nous n'allons pas prendre en compte l'origine des cycles de la base de réapprentissage.(prendre des enregistrements de différents patients).

Pour voir l'intérêt de cette spécialisation nous allons tester ces classifieurs spécialisés sur des bases différentes de celle de l'apprentissage et du réapprentissage et nous les comparons aux résultats obtenus par les réseaux globaux.

Dans le cas de la spécialisation d'arythmie nous n'allons pas tenir compte de la complexité d'apprentissage et du nombre d'exemples présents dans la base de réapprentissage car l'apprentissage et le réapprentissage sont faits en hors-ligne.

Chaque réseau global va être spécialisé par les trois anomalies cardiaques ciblées (V,R,L).

Nous présentons les résultats des tests de performances des différents réseaux dans les tableaux ci-dessous .

Plusieurs bases d'apprentissage et de test ont été constituées.

Nos bases de données d'apprentissage et de test sont issues de la base MIT-BIH composée d'enregistrements ECG à partir de la dérivation des membres DII. D'après les caractéristiques électrocardiographiques des différentes anomalies cardiaques (ESV, BBD, BBG) que nous avons présenté au chapitre 1, l'extrasystole apparaît dans toutes les dérivations que ce soit des membres(DI,DII,DIII) ou précordiales(V1,V2,V3,V4,V5,V6) mais pour les blocs de branches ils n'apparaissent que dans les dérivations V1 et V6. C'est pour ces raisons que les réseaux globaux n'ont pas donné de résultats satisfaisants lors de la reconnaissance des blocs de branche, mais de meilleurs résultats pour l'ESV.

Cette difficulté de classification des BBD et BBG a été palliée par la spécialisation de l'arythmie.

4.8.1 Tests des réseaux sur des bases mono enregistrement

Les tableaux suivants montrent que les réseaux spécialisés de l'arythmie donnent de meilleurs résultats que les réseaux globaux.

Les réseaux sont spécialisés avec 20 cycles de l'enregistrement choisi.

La base de test est composée que des cas que nous avons ciblé l'extrasystole ventriculaire, nous l'appelons base mono enregistrement.

Les résultats sont présentés dans les tableaux 4.14 et figure 4.15 .

	vp	vn	fp	fn	Se	Sp	CC
Réseau3 global testé sur 214v	47	0	0	3	94	NaN	94
Réseau3 spécialisé avec 106v testé sur 214v	50	0	0	0	100	NaN	100
Réseau1 global testé sur 203v	13	0	0	2	86.6667	NaN	86.6667
Réseau1 spécialisé avec 106v testé sur 203v	15	0	0	0	100	NaN	100

Tab 4.14 : résultats des tests de performances des réseaux spécialisés et réseaux globaux.

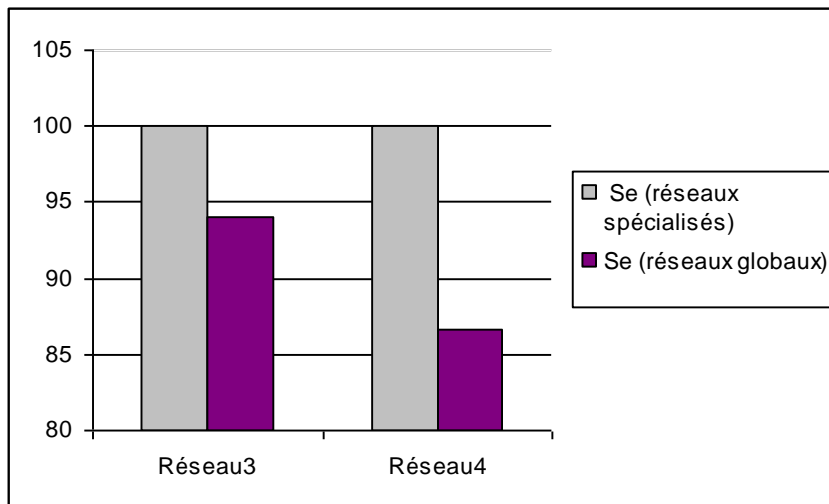


Fig 4.15 :taux de sensibilité des réseaux spécialisés et réseaux globaux.

Comme nous l'avons déjà expliqué la reconnaissance des extrasystoles ventriculaires est plus facile que la reconnaissance des blocs de branche.

Les résultats des tests de performances des réseaux 1 et 3 ne diffèrent pas trop de ceux des réseaux 1 et 3 spécialisés. Il y a une amélioration de la sensibilité qui passe des 90% à 100%(tous les cas rythmiques correctement classés).

Dans le tableau 4.14,nous ciblons la reconnaissance des 'V' ,donc nous allons avoir uniquement les grandeurs vp (classification positive de la classe) et /ou fn (classification négative de la classe).

Les grandeurs vn et fp seront nulles d'où $Sp = \frac{vn}{vn + fp} = NaN$ (division impossible par zéro).

Pour voir l'efficacité de la spécialisation de l'anomalie nous avons testés différents réseaux en ciblant les blocs de branches.

Les résultats sont présentés dans le tableau 4.15 et figure 4.16 .

	vp	vn	fp	fn	Se	Sp	CC
Réseau3 global testé sur 118r	2	0	0	10	16.6667	NaN	16.6667
Réseau3 spécialisé avec 212r testé sur 118r	12	0	0	0	100	NaN	100
Réseau1 global testé sur 1091l	0	0	0	104	0	NaN	0
Réseau1 spécialisé avec 214l testé sur 1091l	104	0	0	0	100	NaN	100
Réseau4 global testé sur 111l	0	0	0	41	0	NaN	0
Réseau4 spécialisé avec 214l testé sur 111l	8	0	0	33	19.5122	NaN	19.5122

Tab 4.15 : résultats des tests de performances des réseaux spécialisés et réseaux globaux.

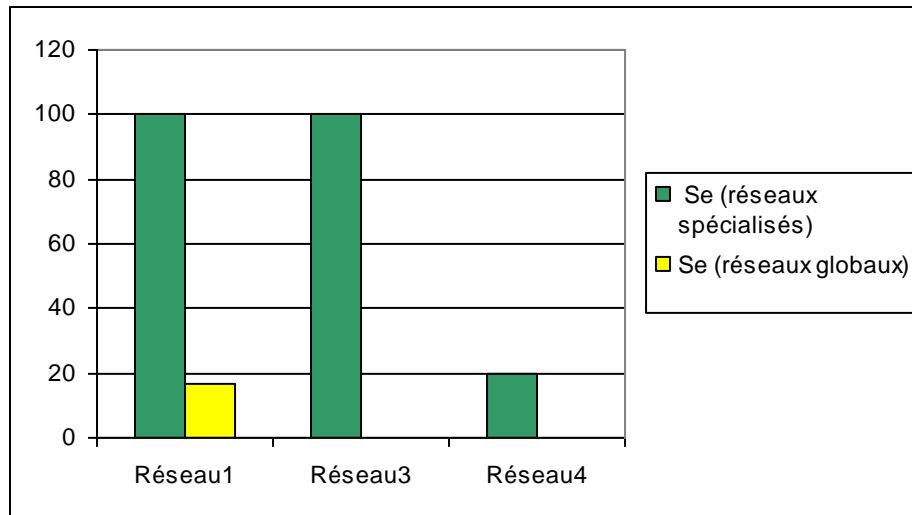


Fig 4.16 :taux de sensibilité des réseaux spécialisés et réseaux globaux.

Le réseau3 a été spécialisé avec la base 212r qui est formée de 26 cycles annotés Bloc de Branche Droit 'R'. Le test de performances de ce réseau a donné de très bons résultats .La sensibilité est passée de 16% à 100% (La sensibilité nous donne le taux de bonne classification de la classe ciblée).

Le réseau global a classé que 2 cas R présentés dans la base de test ,tandis que le réseau spécialisé a détecté tous les cas ciblés. (la base de test est composée de 12 cas R).

D'autres tests ont été effectués mais en ciblant le Bloc de Branche Gauche 'L'.
Le réseau 1 global a une sensibilité égale à zéro mais après spécialisation avec une base composée que de cas L, le réseau a classé tous les cas pathologiques.
Même si pour le réseau 4 spécialisé le taux de classification correcte n'est pas très grand mais il y a une amélioration par rapport aux performances du réseau 4 global.

Donc la spécialisation des cas difficiles à détecter comme les blocs de branche droit et gauche est très efficace. Il y a une grande différence entre les résultats de sensibilité des réseaux initiaux et des réseaux spécialisés.

Notons que nous avons présenté uniquement une partie des expérimentations parmi plusieurs,sachant qu'ils donnent des résultats identiques.

4.8.2 Tests des réseaux sur des bases variées

Nous avons testé les performances de ces réseaux spécialisés pour l'arythmie sur des bases plus larges et différentes des bases d'apprentissage et de réapprentissage, elles sont constituées à partir de cycles annotés N ,V ,R et L.

La spécialisation permet d'améliorer la sensibilité de détection d'une arythmie.

Les résultats obtenus sont présentés dans les tableaux suivants :

4.8.2.1 résultats de performances du réseau 2

Pour montrer l'intérêt de la spécialisation d'arythmie nous avons choisi de mener le réapprentissage d'un réseau qui a donné de mauvais résultats de reconnaissance des extrasystoles ventriculaires lors de son test sur différentes bases variées,i.e le réseau 2.

Nous avons spécialisé le réseau 2 avec 20 cycles annotés 'V' de l'enregistrement 106 et nous l'avons testé sur différentes bases.

L'anomalie ciblée dans ces tests est l'extrasystole ventriculaire.

Les résultats des tests du réseau 2 spécialisé et global sont présentés dans les tableaux 4.16 et 4.17 et figure 4.17.

Réseau2 spécialisé avec 106v

Bases de test	vp	vn	fp	fn	Se	Sp	CC
103n200v118r207l	25	83	27	0	100	75.4545	80
116n203v124r111l	15	55	64	0	100	46.2185	52.2388
115n119v118r111l	34	23	40	0	100	36.5079	58.7629

Tab 4.16 : résultats des tests de performances du réseau 2 spécialisé

Réseau2 global

Bases de test	vp	vn	fp	fn	Se	Sp	CC
103n200v118r207l	0	58	0	0	NaN	100	100
116n203v124r111l	0	40	5	0	NaN	88.8889	88.8889
115n119v118r111l	2	7	3	32	5.8824	70	20.4545

Tab 4.17 : résultats des tests de performances du réseau 2global

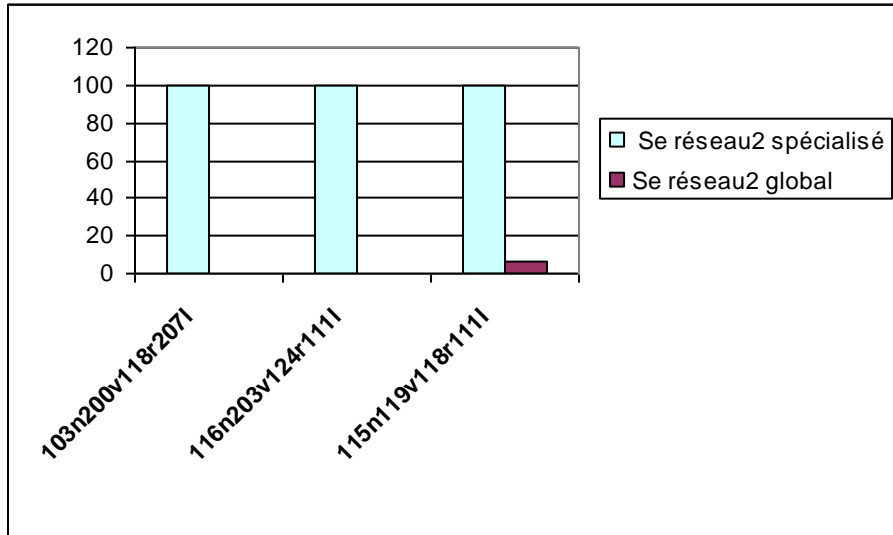


Fig 4.17 :taux de sensibilité du réseau 2 spécialisé et global.

Les tests de performances de ce classifieur ont été effectués sur des bases formées de différents cas N,V,R et L.

Pour bien voir l'intérêt de la spécialisation nous comparons les résultats des tests des réseaux globaux et spécialisés sur les mêmes bases.

Dans ces deux tableaux nous nous sommes intéressés à la reconnaissance des cas 'V' parmi plusieurs cas.

Le réseau2 global a donné de mauvais résultats de classification de l'arythmie ciblée.

Après spécialisation le réseau est capable de détecter toutes les extrasystoles dans une base de test variée (Se=100).

4.8.2.2 résultats de performances du réseau 4 avec cas ciblé BBG

Nous avons spécialisé le Réseau4 avec 40cycles de 207l.

Les résultats des tests du réseau 4 global et spécialisé sont présentés dans les tableaux 4.18 et 4.19 .

Réseau 4 spécialisé avec 207l

Bases de test	vp	vn	fp	fn	Se	Sp	CC
115n119v118r111l	41	34	10	0	100	77.2727	88.2353
116n203v124r111l	41	0	14	0	100	0	74.5455

Tab 4.18 : résultats des tests de performances du réseau 4 spécialisé

Réseau 4 global

Bases de test	vp	vn	fp	fn	Se	Sp	CC
115n119v118r111l	0	48	8	41	0	85.7143	49.4845
116n203v124r111l	0	81	12	41	0	87.0968	60.4478

Tab 4.19 : résultats des tests de performances du réseau 4 global

4.8.2.3 résultats de performances du réseau 4 avec cas ciblé BBD

Nous avons Spécialisé le Réseau4 avec 26cycles de 212r.

Les résultats des tests du réseau 4 global et spécialisé sont présentés dans les tableaux 4.20 et 4.21.

Réseau 4 Spécialisé avec 212r

Bases de test	vp	vn	fp	fn	Se	Sp	CC
103n200v118r207l	12	28	94	0	100	22.9508	29.8507
115n119v118r111l	12	31	54	0	100	36.4706	44.3299

Tab 4.20 : résultats des tests de performances du réseau 4 spécialisé.

Réseau 4 global

Bases de test	vp	vn	fp	fn	Se	Sp	CC
103n200v118r207l	0	82	41	12	0	66.6667	60.7407
115n119v118r111l	0	83	2	12	0	97.6471	85.5670

Tab 4.21 : résultats des tests de performances du réseau 4 global

Les résultats présentés au 4.8.2.2 et 4.8.2.3 nous montrent l'intérêt de la spécialisation de l'anomalie cardiaque .nous avons testé le réseau4 global sur différentes bases en ciblant une fois la reconnaissance des Blocs de branche gauche et l'autre fois la reconnaissance des blocs de branche droit. Le réseau 4 global a mal classé les cas ciblés d'où $vp=0$ et $Se=0$. Les BBG et BBD sont plus difficile à détecter que les extrasystoles ventriculaires en raison de l'origine des données d'apprentissage et de test qui sont des mesures faites sur un ECG à partir de la dérivation DII .

Mais après la spécialisation le réseau a donné de très bons résultats lors de la reconnaissance de BBG et BBD ,la sensibilité est maximale.

4.8.2.4 résultats de performances du réseau 2 avec cas ciblé ESV

Nous avons Spécialisé le Réseau2 avec 33 cycles BBD de l'enregistrement 124. nous avons ciblé dans le test les extrasystoles ventriculaires.

Les résultats des tests du réseau 2spécialisé et global sont présentés dans les tableaux 4.22 et 4.23 et figure 4.18.

Réseau 2 Spécialisé avec 124r

Bases de test	vp	vn	fp	fn	Se	Sp	CC
115n119v118r111l	34	63	0	0	100	100	100
103n200v118r207l	25	110	0	0	100	100	100

Tab 4.22 : résultats des tests de performances du réseau 2 spécialisé

Réseau 2 global

Bases de test	vp	vn	fp	fn	Se	Sp	CC
115n119v118r111l	2	7	3	32	5.8823	70	20.4545
103n200v118r207l	0	58	0	0	NaN	100	100

Tab 4.23 : résultats des tests de performances du réseau 2 global

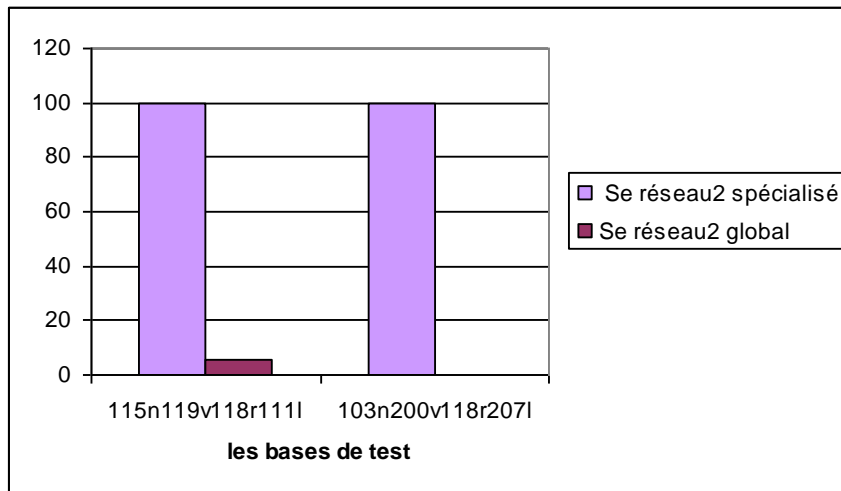


Fig 4.18 : taux de sensibilité du réseau 2 spécialisé et global.

Le réseau 2 spécialisé a bien classé tous les cas présentés dans la base de test .il a non seulement reconnu les ESV, mais aussi les blocs de branche droit et gauche et les cycles normaux.

Le taux de classification correcte est égale à 100%.Sensibilité et spécificité maximale.

A la différence du réseau 2 global qui a eu de mauvais résultats, avec une probabilité de classification positive de l'anomalie ciblée égale à zéro.

Après cette spécialisation avec peu de cycles, nous sommes arrivés à mettre en œuvre un classifieur idéal avec un taux de classification correcte maximal.

4.8.3 Tests des réseaux sur de grandes bases

Nous avons mis en œuvre deux nouveaux réseaux, en utilisant les mêmes architectures des autres réseaux globaux, en utilisant la fonction d'apprentissage « trainbfg ».

Nous respectons les mêmes notations des premiers réseaux :

Le réseau 5 :100n,106v,212r,214l.

Le réseau6 :115n,119v,118r,111l.

Nous avons construit aussi une grande base d'exemples pour le test, elle est formée à partir de tous les enregistrements que nous avons sélectionné dans le tableau 4.2.

Nous testons les réseaux initiaux et spécialisés sur la même base de test.

La base de test est formée de tous les enregistrements sauf ceux utilisés dans l'apprentissage du réseau global ou réseau spécialisé.

Les résultats de tests de plusieurs réseaux sur les grandes bases de test sont présentés dans les tableau suivant (4.24) et la figure 4.19.

Classification neuronale spécialisée des arythmies cardiaques

	vp	vn	fp	fn	Se	Sp	CC
réseau1	58	380	144	193	23.1076	72.5191	56.5161
réseau1 spécialisé avec 119v	198	297	227	53	78.8845	56.6794	63.8710
réseau2	0	555	1	117	0	99.8201	82.4666
réseau2 spécialisé avec 207l	117	254	302	0	100	45.6835	55.1263
réseau3	2	700	68	10	16.6667	91.1458	90
réseau3 spécialisé avec 212r	12	16	728	0	100	2.1505	3.7037
réseau4	0	328	185	12	0	63.9376	62.4762
réseau4 spécialisé avec 212r	12	64	449	0	100	12.4775	14.4762
réseau5	1	464	101	143	0.6944	82.1238	65.5853
réseau5 spécialisé avec 111l	94	149	438	50	65.2778	25.3833	33.2421
réseau6	167	426	133	71	70.1681	76.2075	74.4040
réseau6 spécialisé avec 214v	238	0	559	0	100	0	29.8620
réseau6	40	545	108	104	27.7778	83.4609	73.4003
réseau6 spécialisé avec 214l	144	0	653	0	100	0	18.0678

Tab 4.24 : résultats des tests de performances des réseaux spécialisés et globaux

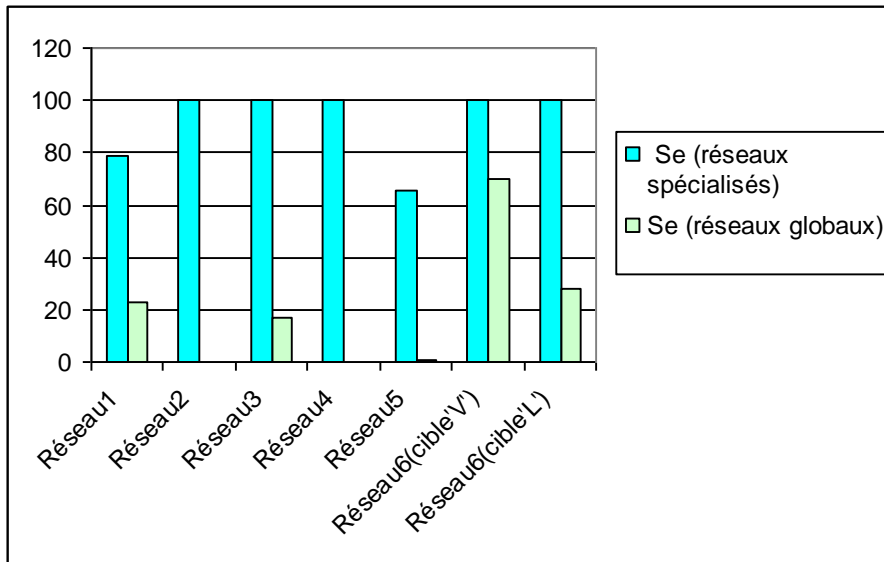


Fig 4.19 :taux de sensibilité des réseaux spécialisés et réseaux globaux.

Les réseaux spécialisés ont donné de meilleurs résultats pour la reconnaissance des arythmies ciblées. Il y a une nette amélioration de la sensibilité par rapport aux réseaux initiaux.

Même si la spécificité=0 notre but est d'arriver à concevoir un classifieur qui reconnaît parfaitement la pathologie ciblée parmi des centaines de cycles différents, d'où l'appellation de cette section spécialisation de l'arythmie.

Nous pouvons proposer un système automatique d'aide à la détection d'une pathologie précise ou système d'annotation automatique (alarme arythmie) quelque soit la taille de la base de test.

4.9 Conclusion

Les résultats des tests de performances obtenues par les différents classifieurs réalisés, ont montré que les réseaux de neurones ont un pouvoir discriminant fort.

Dans ce chapitre nous avons présenté la classification neuronale spécialisée des arythmies cardiaques. La classification réalisée sépare bien les battements normaux (N) des battements arythmiques (V,R ,L).

Nous avons mis en œuvre deux classifieurs spécialisés : l'un pour chaque patient et l'autre pour chaque arythmie. Les résultats des tests de performances de la spécialisation sont nettement meilleurs que les résultats obtenus par les classifieurs initiaux.

la spécialisation des réseaux de neurones pour la classification des cas difficiles à détecter comme les blocs de branche droit et gauche est très efficace. Nous avons remarqué une grande différence entre les résultats de sensibilité des réseaux initiaux et des réseaux spécialisés.



Conclusion

Dans ce mémoire nous avons présenté une méthode de classification des pathologies cardiaques à partir des enregistrements longue durée du signal ECG.

Les réseaux de neurones présentent un grand avantage dans la classification surtout pour leur précision et leur adaptation c'est le cas de la spécialisation dans notre application.

Les signaux physiologiques constituent une source d'information qui prend une part de plus en plus importante dans la démarche diagnostique en médecine et qui deviennent de plus en plus facilement accessible du fait du développement de systèmes d'acquisition de signaux.

Dans ce mémoire de magister nous nous sommes intéressés à un signal physiologique particulier, le signal électrocardiogramme (ECG). Le signal ECG a un potentiel diagnostique important et son utilisation fait partie intégrante de l'examen clinique d'un patient. Son contenu informationnel est reconnu et exploité essentiellement dans le cadre de l'auscultation cardiaque.

Construire un classifieur spécialisé représente une mission très intéressante surtout pour les praticiens du fait qu'elle offre un avantage dans la détection des arythmies avec un taux de bonne classification maximal.

Nous avons choisi d'utiliser les réseaux de neurones artificiels à la fois pour leur robustesse ainsi que pour leur facilité d'implémentation.

Les réseaux présentés dans ce mémoire sont des réseaux multi couches qui utilisent l'algorithme de rétropropagation du gradient.

Parmi les algorithmes d'apprentissage que nous avons utilisé pour l'apprentissage des réseaux :la méthode de Quasi Newton 'BFGs' et la méthode de Levenberg Marquardt 'LM'.

L'intérêt de la spécialisation entre dans le cadre de l'amélioration des performances du classifieur d'une part ,d'autre part la spécialisation est utilisée dans des situations alarmantes (surveillance des malades en risque d'une attaque cardiaque ou présence d'arythmies sévères, qui sont des facteurs prédicteurs de la mort subite). Pour cela nous avons prévu deux types de classifieurs :

1- Un classifieur spécialisé pour chaque patient bien défini « spécialisation par patient ». dans cette partie nous avons réussi à montrer l'intérêt de la spécialisation en comparant les résultats de tests de performances des réseaux globaux et des réseaux spécialisés.

Les réseaux spécialisés ont donné d'excellents résultats (réseaux 3 et 4) avec des taux de classification correcte de 100% et en même temps une sensibilité et une spécificité de 100%.

2- Un classifieur spécialisé pour la détection d'arythmies : « spécialisation arythmie » . Ce type de classifieur est très utile pour la détection d'une arythmie ciblée ,par exemple pour la classification du bloc de branche droit ou gauche qui sont des anomalies plus difficile à détecter que les extrasystoles ventriculaires dans notre travail.

La raison de la difficulté revient à l'origine du signal ECG issu de la base MIT-BIH les enregistrements des différents patients que nous avons choisis dans ce mémoire proviennent de la dérivation DII alors que les blocs de branches apparaissent plus clairement dans les dérivations précordiales(V1,V5 et V6).

Malgré cet inconvénient les réseaux de neurones se sont bien adaptés aux différentes arythmies .

Les réseaux spécialisés (1, 3 et 4) ont donné de très bons résultats lors des tests de performances par rapport aux résultats des réseaux globaux.

La sensibilité de reconnaissance de l'anomalie ciblée est nettement améliorée elle est passée de 0% pour le réseau4 global à 100% pour le réseau4 spécialisé lors de la reconnaissance des BBD et BBG

Les résultats des taux de classification des réseaux spécialisés pour l'arythmie sont nettement meilleurs que les résultats des réseaux globaux (Se,Sp,CC).

Dans un milieu hospitalier plusieurs systèmes de monitoring continu en cardiologie sont très utilisés dans les salles de soins intensifs .ces systèmes sont munis de logiciels de détection automatique des pathologies et des arythmies graves telles que les extrasystoles ventriculaires polymorphes .introduire ces systèmes de spécialisation reste l'un de nos perspectives majeures :spécialisation par patient dans le cas on line et spécialisation pour arythmies dans le cas off line.

La classification neuronale spécialisée peut être encore plus performante si nous arrivons à élargir notre base de données en intégrant d'autres sources d'information par exemple la pression artérielle ,et d'autres dérivations.

Une autre perspective très intéressante est la fusion de classifieurs (les modèles hybrides)en utilisant les modèles flous tels que les systèmes d'inférence flous (SIF) et les arbres de décision flous (ADDFs).

