

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ IBN-KHALDOUN DE TIARET

FACULTÉ DES SCIENCES APPLIQUEES
DÉPARTEMENT DE GENIE ELECTRIQUE



MEMOIRE DE FIN D'ETUDES

Pour l'obtention du diplôme de Master

Domaine : Sciences et Technologie

Filière : Génie Electrique

Spécialité : Informatique Industrielle

THÈME

**APPROCHE NEURONALE POUR LA CLASSIFICATION MORPHOLOGIQUE
DES SIGNAUX ECG**

Préparé par : SAFI Adila

Devant le Jury :

Nom et prénoms	Grade	Qualité
Mr. BENAATIA Aada	MAA	Président
Mr. BENAABID Houari	MAA	Examineur 1
Mr. MAASKRI Mustapha	MAA	Examineur 2
Mr. GOISMI Mohamed	MAA	Encadreur

PROMOTION 2016 /2017

Dédicace

*À la plus belle créature que Dieu a créée sur terre,,
À cet source de tendresse, de patience et de générosité,,*

À ma mère !

*À mes chères sœurs : Amína, Soumia, Fatíma Zohra, Hiba, Síhem
et Hadjer*

À mes chers grands-parents et à toute ma famille

À tous mes amis et collègues

*À tous les étudiants de la promotion 2016/2017 Option :
Informatique Industrielle*

À tous ceux qui, par un mot, m'ont donné la force de continuer

Remerciements

Je tiens à remercier tout d'abord mon encadreur **Mr Goismi Mohamed**, pour sa patience, et surtout pour sa confiance, ses remarques et ses conseils, sa disponibilité et sa bienveillance. Qu'il trouve ici le témoignage de ma profonde gratitude.

Je voudrais également remercier les membres du jury pour avoir accepté d'évaluer ce travail et pour toutes leurs remarques et critiques, ainsi que le personnel et les enseignants de l'université d'**Ibn Khaldoun de Tiaret**, qui ont contribué à ma formation.

À tous mes enseignants qui m'ont initié aux valeurs authentiques, en signe d'un profond respect et d'un profond amour !!! Merci à vous tous

Table des matières

Liste des figures	i
Liste des tableaux	iii
Liste des abréviations	iv
Introduction générale.....	1
 Chapitre I: Le signal ElectroCardioGramme 	
I.1. Introduction.....	5
I.2. Le cœur	5
I.2.1. Anatomie du cœur	5
I.2.2. Activité mécanique du cœur	6
I.2.3. Activité électrique du cœur	7
I.2.4. La conduction électrique du cœur.....	8
I.3. Les signaux ECG.....	9
I.3.1. Historique de l'électrocardiogramme	10
I.3.2. Description du signal électrocardiogramme.....	10
I.3.3. Caractéristiques du signal ECG	10
I.3.4. Détection du signal ECG.....	12
I.4. Les pathologies cardiaques	14
I.4.1. Les maladies cardio-vasculaires	14
I.4.2. Les facteurs de risque cardio-vasculaire.....	15
I.5. Trouble du rythme et de la conduction cardiaque.....	15
I.5.1. Rythme sinusal	15
I.5.2. Blocs cardiaques	15
I.6. Les arythmies cardiaques	16
I.6.1. Extrasystoles supra-ventriculaires	17
I.6.2. Contraction ventriculaire prématurée (extrasystole ventriculaire).....	17
I.6.3. Tachycardie supra- ventriculaire	17

I.6.4.	La fibrillation auriculaire.....	18
I.6.5.	La tachycardie ventriculaire.....	18
I.6.6.	Fibrillation ventriculaire.....	19
I.7.	Artefacts visibles sur l'électrocardiogramme.....	20
I.7.1.	Bruits techniques	20
I.7.2.	Artefacts physiques	21
I.8.	Conclusion.....	23

Chapitre II: La classification en intelligence artificielle

II.1.	Introduction.....	25
II.2.	La classification en intelligence artificielle	25
II.2.1.	Chaîne de classification d'objets	26
II.2.1.1.	Les principaux modules de classification d'objets	26
II.2.1.2.	L'extraction des descripteurs	27
II.2.1.3.	Procédure de résolution par apprentissage	27
II.2.1.3.1.	Phase d'apprentissage	27
II.2.1.3.2.	Phase de test.....	28
II.2.2.	Formalisation mathématique d'un problème de classification.....	29
II.3.	Les méthodes de classification.....	30
II.3.1.	La logique floue en classification	30
II.3.1.1.	Conception de classificateur flou.....	30
II.3.1.1.1.	Principe de base	30
II.3.1.1.2.	Base des règles floues.....	31
II.3.1.2.	Avantages et Inconvénients	31
II.3.2.	Les arbres de décision	32
II.3.2.1.	Historique.....	32
II.3.2.2.	Exemple	32
II.3.2.3.	Avantages et Inconvénients	34

II.3.3.	Réseaux bayésiens.....	34
II.3.3.1.	Construction des réseaux bayésiens	35
II.3.3.2.	Exemple	35
II.3.3.3.	Avantages et Inconvénients	35
II.3.4.	Machine à vecteurs de support.....	36
II.3.4.1.	Historique.....	36
II.3.4.2.	Principe de base des SVM	37
II.3.4.3.	Avantages et Inconvénients :	37
II.3.5.	Les cartes de Kohonen	38
II.3.5.1.	Principe des cartes auto-adaptives de Kohonen.....	38
II.3.5.2.	Avantages et Inconvénients	39
II.4.	Conclusion :.....	40

Chapitre III: Les réseaux de neurones

III.1.	Introduction.....	42
III.2.	Historique sur les réseaux de neurones.....	42
III.3.	Neurone biologique	43
III.4.	Neurone artificiel (formel)	44
III.5.	Le perceptron	45
III.5.1.	Apprentissage du perceptron	47
III.6.	Architecture des réseaux de neurones.....	47
III.6.1.	Les réseaux de neurones non-bouclés	47
III.6.2.	Les réseaux de neurones bouclés.....	48
III.7.	Apprentissage des réseaux de neurones.....	48
III.7.1.	Les réseaux à apprentissage supervisé	48
III.7.2.	Les réseaux à apprentissage non-supervisé.....	49
III.8.	Les différentes applications des réseaux de neurones artificiels.....	49
III.9.	Conception et mise œuvre des réseaux de neurones.....	50

III.9.1. Détermination des entrées/sorties du réseau de neurones	50
III.9.2. Choix et préparation des échantillons.....	50
III.9.3. Elaboration de la structure du réseau.....	51
III.9.4. Apprentissage	51
III.9.5. Validation et tests	52
III.10. Les réseaux de neurones les plus utilisés	52
III.11. Les réseaux de neurones multicouches	53
III.11.1. L'algorithme de rétro-propagation du gradient.....	54
III.12. Avantages et Inconvénients	55
III.13. Conclusion.....	57
Chapitre IV: Contribution et Réalisation	
IV.1. Introduction.....	59
IV.2. Classification neuronale des arythmies cardiaques	59
IV.2.1. Sélection de la base d'exemples	60
IV.2.1.1. Sélection des descripteurs temporaires d'un cycle cardiaque	62
IV.2.2. Architecture du réseau de neurones.....	63
IV.2.2.1. Algorithme d'apprentissage	63
IV.2.2.2. Dimensionnement du réseau pour chaque classifieur.....	63
IV.2.2.2.1. Réduction par l'ACP.....	64
IV.1.2 Résultats expérimentaux des classifications	65
IV.3. Architecture d'un classifieur à base des descripteurs morphologiques	66
IV.4. Présentation de l'application.....	67
IV.5. Conclusion.....	70
Conclusion générale	72
Annexe A.....	74
Annexe B	79
Références bibliographiques.....	83

Liste des figures

Introduction Générale

Figure a: Les différentes techniques de l'IA entre 1995 et 2007 dans la médecine..... 2

Chapitre I

Figure I.1: L'anatomie du cœur..... 6

Figure I.2: les phases du potentiel d'action d'une cellule du muscle cardiaque..... 7

Figure I.3: Activité électrique du cœur..... 9

Figure I.4: ECG normal. 11

Figure I.5: Les déviations bipolaires. 12

Figure I.6: Triangle d'Eithoven. 12

Figure I.7: Dérivation unipolaire. 13

Figure I.8: Le triaxe des dérivations frontales..... 13

Figure I.9: Exemples d'ECG. (a) ECG normal; (b) bloc de branche. 16

Figure I.10: Tachycardie supra-ventriculaire..... 17

Figure I.11: Fibrillation auriculaire..... 18

Figure I.12: Tachycardie ventriculaire. 19

Figure I.13: Fibrillation ventriculaire. 19

Figure I.14: Interface secteur 50HZ. 20

Figure I.15: Bruits dû au mouvement des électrodes. 21

Figure I.16: Dérivé de la ligne de base..... 22

Figure I.17: Bruit myoélectrique ou troublement somatique. 22

Chapitre II

Figure II.1: Diagnostic basé sur le modèle..... 25

Figure II.2: Schéma synoptique d'un classifieur supervisé. 31

Figure II.3: Arbre de décision correspondant. 33

Figure II.4: Structure de causalité..... 35

Figure II.5: La carte auto-adaptive de Kohonen 20*20..... 39

Chapitre III

Figure III.1: Neurone biologique..... 44

Figure III.2: Fonction d'activation. 44

Figure III.3: Neurone formel..... 45

Figure III.4: Des exemples linéairement séparables par une droite. 46

Figure III.5: les points XOR pas des droite séparatrice. 46

Figure III.6: Organigramme de conception..... 52
Figure III.7: Schéma synoptique des réseaux de neurones les plus utilisés. 53
Figure III.8: Réseaux de neurones multicouches avec une couche cachée. 53

Chapitre IV

Figure IV.1: Schéma global de la classification..... 60
Figure IV.2: Battements normaux et intervalles d'intérêt. 62
Figure IV.3: Variabilité des données de la base d'apprentissage (l'ACP)..... 64
Figure IV.4: Architecture du classifieur à base des descripteurs morphologiques. 66
Figure IV.5: Interface principale du classifieur morphologique. 67
Figure IV.6: Paramétrage du réseau de neurone. 68
Figure IV.7: Interface d'apprentissage du réseau de neurone..... 68
Figure IV.8: Graphe de performance. 69
Figure IV.9: Les graphes du gradient, moyenne d'erreurs et validation du test. 69
Figure IV.10: Mesure des paramètres de l'ECG du signal 210s1. 70

Liste des tableaux

Introduction générale

Tableau a: Le nombre d'application des différentes techniques de l'IA entre 1995 et 2007 dans la médecine 2

Chapitre II

Tableau II.1: Les définitions des grandeurs VP, VN, FP et FN..... 29
 Tableau II.2: Base des règles floues. 31
 Tableau II.3: Exemple de données. 33

Chapitre III

Tableau III.1: Analogie entre les neurones biologiques et artificiels..... 45

Chapitre IV

Tableau IV.1: Les enregistrements choisis de la base de données MIT-BIH. 61
 Tableau IV.2: Nombre de composantes principales retenues pour chaque ECG. 65
 Tableau IV.3: Performances du classifieur neuronal à données morphologiques réduites. 66

Annexes

Tableau B.1: les 48 enregistrements de la base de données MIT-BIH..... 80

Liste des abréviations

- MNT** Maladies Non Transmissibles
OMS Organisation Mondiale de Santé
AVC Accidents Vasculaires Cérébraux
IA Intelligence Artificielle
LF Logique Floue
RN Réseaux de Neurones
AG Algorithmes Génétiques
FL Fuzzy Logic
NN Neural Networks
GA Genetic Algorithms
ECG ElectroCardioGramme
P.R.A Période Réfractaire Absolue
P.R.E Période Réfractaire Effective
P.R.R Période Réfractaire Relative
NS Nœud Sinusal
NAV Nœud Auriculo-ventriculaire
SVM Support Vector Machines
MLP Multi-Layer Perceptrons
R.N.A Réseaux de Neurones Artificiels
ACP Analyse en Composantes Principales

Introduction Générale

Introduction générale

Les maladies cardiovasculaires constituent un problème majeur de santé publique. Elles sont responsables de la plupart des décès par les Maladies Non Transmissibles (MNT), à savoir 17 millions de personnes chaque année à travers le monde. Elles sont suivies par le cancer (7,6 millions), les maladies respiratoires (4,2 millions) et le diabète (1,3 millions). Ces quatre groupes de maladies représentent près de 80% de l'ensemble des décès par MNT et ont en commun des facteurs de risque telles que : le tabagisme, la sédentarité, l'usage nocif de l'alcool, l'obésité, hypertension artérielle et des facteurs génétiques.

Le premier rapport de l'Organisation Mondiale de Santé (OMS) sur les MNT rendu public le 27 avril 2011, révèle que les pays en voie de développement sont les plus touchés par les MNT.

Selon un communiqué publié à l'OMS en 2008, 36,1 millions de personnes sont décédés d'affections telles que : les cardiopathies, les Accidents Vasculaires Cérébraux (AVC), les pneumopathies chroniques, les cancers et le diabète. Plus de 80% de ces décès sont survenus dans les pays à revenu faible ou moyen.

Le cœur c'est l'organe central du système cardiovasculaire, il peut être affecté par de nombreuses pathologies qui peuvent être soit bénignes, comme certaines tachycardies par exemple, soit très sérieuses, comme l'infarctus du myocarde qui cause 10% de décès dans le monde. En raison de l'ampleur du problème, le suivi des patients à risques devient primordial. Les arythmies mineures informent le médecin sur l'état cardiaque du patient. Elles doivent être détectées notamment pour prévenir une dégénérescence possible en arythmies sévères.

Actuellement, le domaine médical exige de nouvelles techniques et technologies, afin d'évaluer l'information d'une manière objective. Cela est du aux développements récents dans l'électronique qui a poussé l'informatique à un stade de plus en plus avancé. Ceci a permis d'avoir des machines de plus en plus performantes permettant d'exécuter des algorithmes complexes et de tester de nouvelles approches de l'Intelligence Artificielle (IA) qui s'avérait impossible auparavant.

Ainsi la médecine comme aussi plusieurs autres domaines a bénéficié de cette révolution en informatique particulièrement l'IA. Dans ce contexte, l'informatique est devenue un outil incontournable dans la pratique médicale moderne en générale et comme support d'aide au diagnostic en particulier[1]. Plusieurs techniques de l'IA, de la logique floue (LF), réseaux de neurones (RN), les algorithmes génétiques (AG), et la fusion entre eux (LF-RN), (AG-RN) (LF-AG) sont couramment utilisées aux applications médicales en vue d'améliorer la performance des systèmes d'aide au diagnostic médical. La figure (a) montre la quantité de travaux publiés depuis 1995 à 2007 exploitant ces techniques dans le domaine médical.

Année de publication	1995-1999	2000	2001	2002	2003	2004	2005	2006	2007	Total
*FL	184	41	81	44	45	58	42	44	34	573
NN	641	160	171	172	192	239	194	211	186	2166
GA	43	20	18	17	14	28	29	40	36	245
FL-NN	29	6	23	13	14	8	14	21	16	144
NN-GA	17	2	5	5	6	6	8	11	9	56
FL-GA	3	1	-	1	1	1	-	4	-	11
FL-NN-GA	1	-	-	1	1	1	-	1	-	5

Tableau a: Le nombre d'application des différentes techniques de l'IA entre 1995 et 2007 dans la médecine. [1]

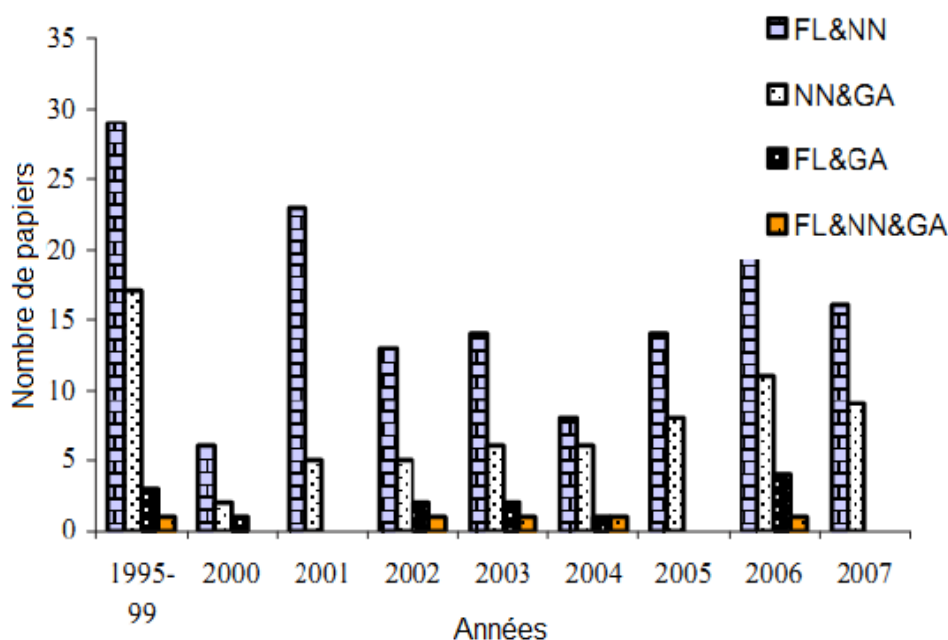


Figure a: Les différentes techniques de l'IA entre 1995 et 2007 dans la médecine. [1]

Ce qui est constaté du (Tableau a) et de la (Figure a), c'est la prédominance des techniques neuronales. Ceci pourra être expliqué par leurs propriétés remarquables: tels que leur capacité d'adaptation, apprentissage à partir d'exemples et généralisation, robustesse face au bruit. Aussi, compte tenu qu'ils sont intégrables dans des processus temps réel et qu'ils sont remarquablement efficaces pour l'extraction des règles à partir d'exemples pour des problèmes difficilement modélisables. En plus la nature des données médicales, qui sont généralement bruitées et incomplètes, renforce leur application dans ce domaine[2].

Parmi ces données on trouve l'électrocardiogramme (ECG) de surface, qui 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.

L'électrocardiogramme est une représentation graphique du potentiel électrique qui commande l'activité musculaire du cœur. Ce potentiel est recueilli par des électrodes posées à la surface du corps. Il se présente comme une suite de déflexions (ondes) répétitives représentant chacune une phase de fonctionnement du cœur. Chaque déformation visible sur ces ondes peut être attribuée à un dysfonctionnement cardiaque ou arythmie.

Les troubles du rythme cardiaque trouvent principalement leurs origines dans la naissance du stimulus cardiaque ou encore dans la conduction de l'onde de dépolarisation à travers le myocarde (chemin suivi par l'onde de dépolarisation à partir de son point d'activation électrique).

Une manière de détecter des troubles cardiaques, consiste à recueillir le signal électrique cardiaque (ECG) par des capteurs, puis l'analyse. Cette analyse présente des enjeux à la fois pratiques et théoriques pour les recherches actuelles en reconnaissance de formes et en médecine. L'objectif poursuivi à travers cette thèse et de proposer de nouvelles méthodes de reconnaissance d'arythmies cardiaques afin d'aider le médecin à lire les enregistrements de longue durée.

Le but de ce mémoire est d'appliquer des techniques permettant d'interpréter de manière automatique un signal ECG en basant sur une voie essentielle consiste en premier lieu sur la morphologie du signal ECG.

Le manuscrite de ce mémoire est divisé en 4 chapitres :

Le premier chapitre est responsable de décrire l'anatomie et la biologie du cœur, pour passer par la suite au signal ECG, son historique et ses caractéristiques (les ondes élémentaires et les intervalles d'un signal ECG), en fin la dernière partie est conservée pour citer les différentes arythmies et pathologies cardiaques susceptibles d'être détectées dans un enregistrement ECG.

Le deuxième chapitre permet d'expliquer la fonctionnalité de la classification dans l'IA, en citant les différentes méthodes de classification avec une analyse détaillée de leurs avantages et inconvénients, telles que (la logique floue, les arbres de décisions, les réseaux bayésiens, les SVM et les cartes de Kohonen).

Le troisième chapitre décrit les réseaux de neurones artificiels (le modèle de classification utilisé). En donnant leur historique, leur algorithme d'apprentissage et leurs avantages et inconvénients.

Le quatrième chapitre sert à démontrer notre application, le modèle neuronal à base des caractéristiques morphologiques extraites depuis le signal ECG.

Chapitre I

Le signal

ElectroCardioGramme

I.1. Introduction

Le cœur est un organe musculaire creux en forme de poire situé entre les poumons, au milieu de la poitrine, il est responsable de pomper le sang qu'il fait circuler dans tous l'organisme, permettant aux cellules de recevoir oxygène et nutriments.

L'enregistrement de l'activité électrique du cœur se fait sur un plan frontal ou sur un plan horizontal selon l'emplacement des électrodes. La représentation graphique d'une telle activité est appelée ElectroCardioGramme (ECG).

Dans ce chapitre on va présenter les notions de base de l'électrocardiographie, tout d'abord on va présenter l'anatomie du cœur, son activité mécanique et son activité électrique. Ensuite on va détailler l'électrocardiogramme et ses caractéristiques. On passe par la suite à citer les pathologies cardiaques. En fin on va décrire les artefacts des signaux ECGs.

I.2. Le cœur

I.2.1. Anatomie du cœur

Le cœur est l'élément principal du système cardiovasculaire, il est fixé au sternum par des tissus conjonctifs particuliers appelés ligaments. La taille d'un cœur adulte est comparable à celle du poing. Chez un individu moyen, il mesure environ de 13 centimètre de long sur 8 centimètres de large, et pèse moins de 500 grammes. Chaque jour, 100000 battements cardiaques pour pomper 8000 litres de sang pour répondre aux besoins énergétiques du corps. Il assure la circulation sanguine dans les veines et les artères grâce aux contractions de son tissu musculaire appelé myocarde, c'est un muscle strié creux avec une épaisse cloison qui le divise en deux moitiés indépendantes (cœur droit/ cœur gauche). Sa partie droite contient du sang pauvre en oxygène et garantit la circulation pulmonaire, sa partie gauche renferme du sang riche en oxygène et propulse dans tous les tissus. Chacune des moitiés comporte deux cavités : l'oreillette et le ventricule, qui communiquent par des valves d'admission qui, laissent passer le sang uniquement de l'oreillette vers le ventricule à l'état normal. Il y'en a aussi des valves d'échappement qui sont responsables à la communication entre le ventricule droit et l'artère pulmonaire (valve pulmonaire), ainsi qu'entre le ventricule gauche et l'artère aorte (valve aortique)[3].

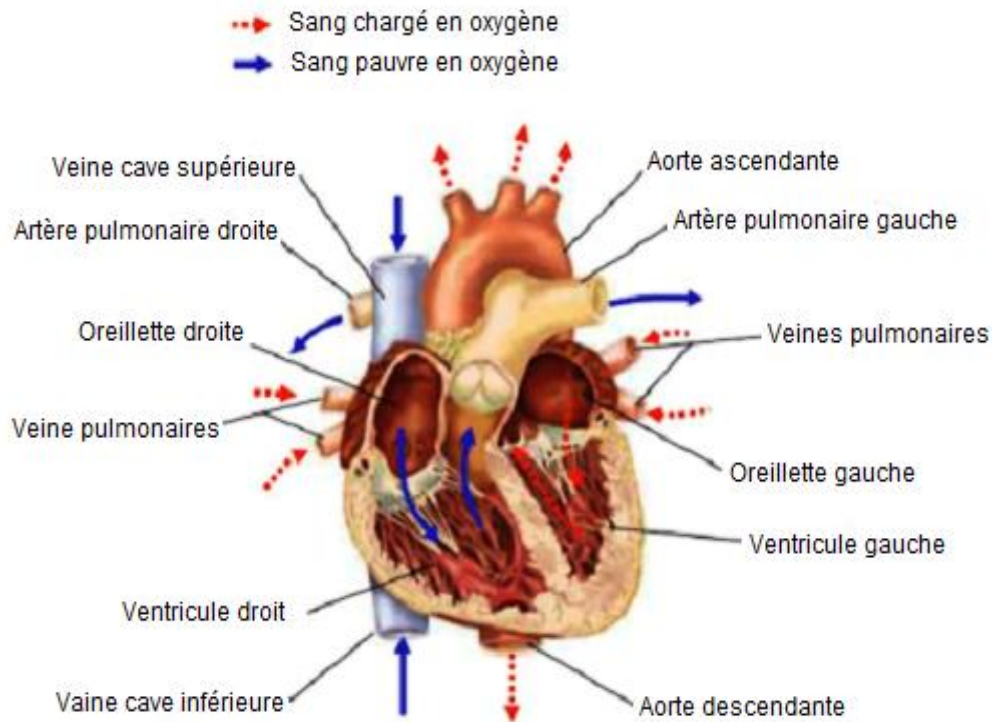


Figure I.1: L'anatomie du cœur.

I.2.2. Activité mécanique du cœur

Le cycle de la circulation sanguine se répète constamment et comprend deux périodes[3]:

- ❖ **La systole:** Est la période correspond à l'éjection du sang dans la grande et la petite circulation. Elle est déphasée en trois phases :
 - **La systole auriculaire:** Est la contraction des oreillettes quand elles sont remplies de sang.
 - **La systole ventriculaire iso-volumique :** Est la contraction des cellules musculaires du myocarde ventriculaire.
 - **La systole ventriculaire :** Est le dépassement des pressions dans les ventricules aux celles dans l'artère pulmonaire et l'aorte.
- ❖ **La diastole :** est la période correspondant au remplissage du sang dans le cœur. Elle est déphasée en deux phases :
 - **La relaxation ventriculaire :** Suivi la systole, ce pendant les ventricules se relâchent, les pressions dans l'artère pulmonaire et l'aorte dépassent les pressions dans les ventricules.
 - **La phase de repos :** C'est le moment où le sang des veines caves et pulmonaires se remplit dans les ventricules via les oreillettes.

I.2.3. Activité électrique du cœur

Les contractions du myocarde sont responsables au mécanisme cardiaque, qui comprend l'expulsion du sang et le mouvement des valves. Ces contractions sont déclenchées par la diffusion d'un potentiel d'action. Chaque cellule myocardique réagit à un stimulus électrique grâce à une membrane permettant une perméabilité sélective 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.

Lorsque la cellule est stimulée électriquement, la perméabilité aux ions de la membrane augmente, ces échanges ioniques doivent la naissance au potentiel d'action.

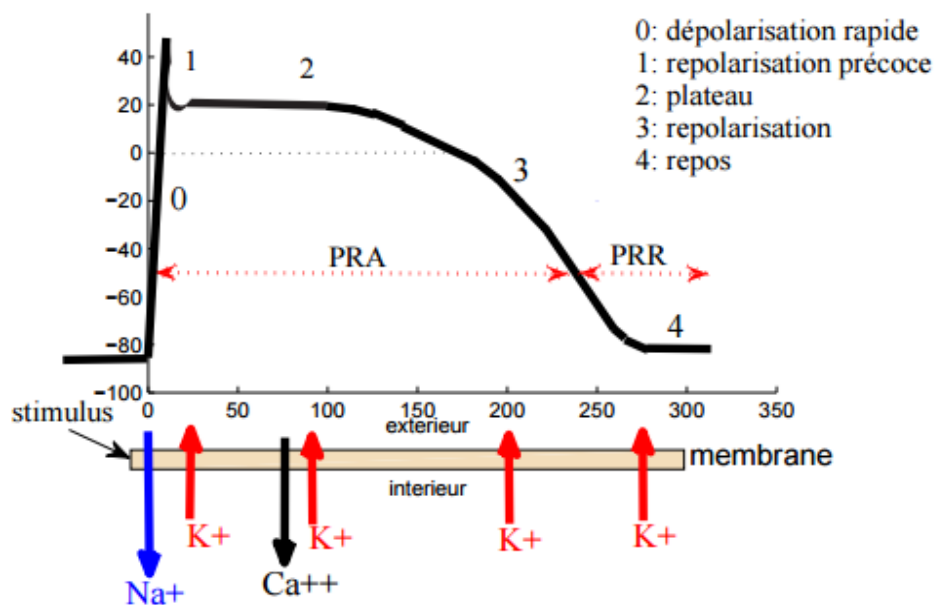


Figure I.2: les phases du potentiel d'action d'une cellule du muscle cardiaque.

- ❖ **La phase 0:** C'est la phase de dépolariation rapide, elle est caractérisée par l'irruption des ions de sodium Na^+ à l'intérieur de la cellule et concourent à l'établissement d'un déséquilibre électrique entre l'intérieur et l'extérieur.
- ❖ **La phase 1:** C'est la phase de dépolariation précoce, elle correspond à la fermeture des canaux sodium et l'ouverture des canaux potassium K^+ de suivre le gradient électrostatique.
- ❖ **La phase 2:** La fuite de potassium se poursuit mais l'entrée des ions de calcium Ca^{++} à l'intérieur de la membrane permet le maintien de dépolariation en plateau car il y'a un équilibre entre charges entrantes Ca^{++} et sortantes K^+ .
- ❖ **La phase 3 :** L'accroissement de conductance au potassium est responsable des charges négatives intracellulaires, et donc de la dépolariation finale.
- ❖ **La phase 4 :** C'est la phase de repos. On obtient l'équilibre avec une différence de potentiel négative.

L'excitabilité de la cellule myocardique, sa capacité à générer un potentiel d'action en réponse à une stimulation se décompose en trois phases :

- ❖ **La période réfractaire absolue (P.R.A)** : période pendant laquelle tout stimulus externe n'a aucun effet sur la cellule (aucune excitation possible).
- ❖ **La période réfractaire effective (P.R.E)** : période incluant la P.R.A, on y ajoute une phase pendant laquelle la cellule se stimule mais ne conduit pas.
- ❖ **La période réfractaire relative (P.R.R)** : période pendant laquelle seul un stimulus puissant peut générer un potentiel d'action.

I.2.4. La conduction électrique du cœur

La contraction du muscle cardiaque qui s'appelle myocarde est provoquée par la propagation d'une onde électrique qui excite les cellules musculaires dans une onde bien établie pour augmenter l'efficacité de la contraction le plus possible. Le système spécialisé d'excitation et conduction électrique suit la séquence d'activation suivante :

- ❖ **Le nœud sinusal (NS)** : L'activation électrique est générée normalement du (NS), qui est situé dans la partie haute de la paroi intérieure de l'oreillette droite, au niveau où débouche la veine cave supérieure. L'impulsion cardiaque initiée dans le nœud sinusal est transmise aux deux oreillettes. Cette activation est facilitée au moyen des voies spécialisées intermodales qui relient le nœud sinusal au nœud auriculo-ventriculaire.
- ❖ **Le nœud auriculo-ventriculaire (NAV)** : Il est situé en bas de l'oreillette droite, et il est constitué de cellules qui présentent une conduction électrique lente. L'activation électrique qui arrive au NAV est physiologiquement ralentie (100ms) avant d'arriver au faisceau de His. Cette propriété physiologique du NAV permet de protéger les ventricules d'un nombre excessif d'activations du NAV et d'activations auriculaires et donne aux oreillettes un temps de vidange plus grand, ce qui optimise la contraction ventriculaire.
- ❖ **Le faisceau de His** : Il est situé dans la partie haute du septum inter-ventriculaire et ses fibres traversent le tissu connectif (non excitable) qui sépare électriquement les oreillettes des ventricules. Dans les cas normaux, le NAV et le faisceau de His constituent la seule voie de propagation de l'activité électrique cardiaque entre les oreillettes et les ventricules. L'ensemble de ces deux structures est souvent appelé la jonction auriculo-ventriculaire. Le faisceau de His comprend un tronc initial qui se divise en deux parties, droite pour le ventricule droit et gauche pour le ventricule gauche.

❖ **Les fibres de Purkinje** : les branches du faisceau de His finissent dans un réseau de fibres qui arrivent dans les parois ventriculaires. Les fibres de Purkinje terminent en anastomoses avec les fibres myocardiques musculaires. Facilitant leur excitation.

L'impulsion électrique se propage dans le muscle cardiaque et induit sa contraction. Elle prend naissance dans le sinus (a) puis se propage dans les oreillettes (b) entraînant leurs contractions (systole auriculaire). L'impulsion arrive alors au nœud auriculo-ventriculaire (NAV) seul point de passage électrique entre les oreillettes et les ventricules une courte pause est alors induite (c) juste avant la propagation dans les fibres constituant le faisceau de His. Au passage de l'impulsion électrique (d) les ventricules se contractent à leurs tour (e) (systole ventriculaire). Après la diastole (décontraction du muscle) les cellules se repolarisent (f). Le cycle du battement cardiaque est alors terminé et le cœur est prêt pour un nouveau battement[4].

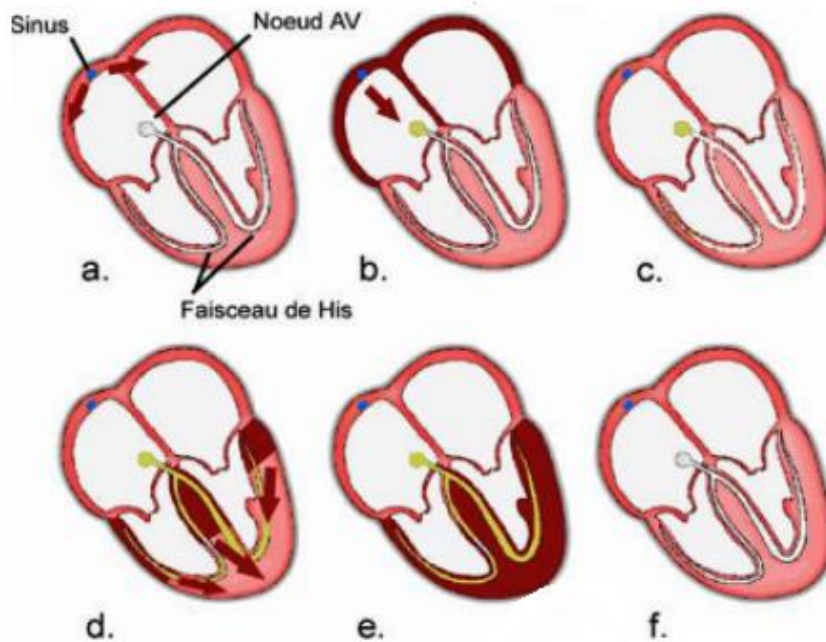


Figure I.3: Activité électrique du cœur.

I.3. Les signaux ECG

L'ECG est un outil de diagnostic permettant de détecter toute pathologie liée à l'activité du cœur. Il permet de mesurer des grandeurs anatomiques et fonctionnelles du cœur :

- Orientation anatomique.
- Taille.
- Variation de conduction.
- Arythmies et atteintes ischémiques.
- Effet de variation des électrolytes.
- Influence des thérapeutiques.

I.3.1. Historique de l'électrocardiogramme

En 1842 un physicien Italien Carlo Matteucci, a indiqué que chaque battement cardiaque entraîne un courant électrique. Puis en 1887 un physiologiste Anglais John Burden, a annoncé le premier électrocardiogramme d'un humain. Ensuite en 1897 un ingénieur électronicien Clément Ader, a proposé un système d'amplification qui s'appelle "galvanomètre à cord", jusqu'au présent utilisé pour les communications télégraphiques sous-marines. En 1903 Einthoven a rassemblé ces courants d'une manière satisfaisante en utilisant un montage électrique qui s'appelle "le pont de Wheatstone", en 1924 il a obtenu un prix Nobel pour ces innovations sur l'électrocardiographie. Les dérivations précordiales sont utilisées pour le diagnostic médical à partir de 1932 et les dérivations frontales unipolaires à partir de 1942, ce qui a aidé Emanuel Goldberger de réaliser le premier tracé sur 12 voies.

Aujourd'hui l'électrocardiographie est une technique relativement peu coûteuse, permettant à l'aide d'un examen indolore et sans danger, de surveiller l'appareil cardiocirculatoire, notamment pour la détection des troubles du rythme et la prévention de l'infarctus du myocarde[5].

I.3.2. Description du signal électrocardiogramme

L'électrocardiographie explore l'activité électrique du cœur par enregistrement des électrocardiogrammes, tracés bidimensionnels qui inscrivent en fonction du temps les variations du potentiel électrique induites dans les différents points du corps par le cœur en activité. Les innombrables cellules musculaires qui le constituent sont dotées de propriétés spéciales dont les deux plus importantes sont le pouvoir mécanique de contraction et l'activité électrique rythmique, elle-même liée à des déplacements ioniques à travers la membrane des cellules. La dépolarisation très brusque, se maintient environ durant 0,3 secondes, puis elle est suivie aussi tôt de la repolarisation qui rétablira les charges électriques initiales. Elle se propage rapidement de proche en proche, aux cellules voisines et finalement au cœur tout entier en 5 centièmes de secondes environ. Cependant, comme la repolarisation est beaucoup plus lente, la durée totale de l'activation de la masse cardiaque est de l'ordre de 40 centièmes de seconde. L'état de repos électrique dure[6].

I.3.3. Caractéristiques du signal ECG

L'analyse de l'électrocardiogramme comprend la mesure des amplitudes et durées ainsi que l'examen de la morphologie de l'onde P, du complexe QRS, de l'onde T, de l'intervalle PR, du segment ST, de l'intervalle QT.

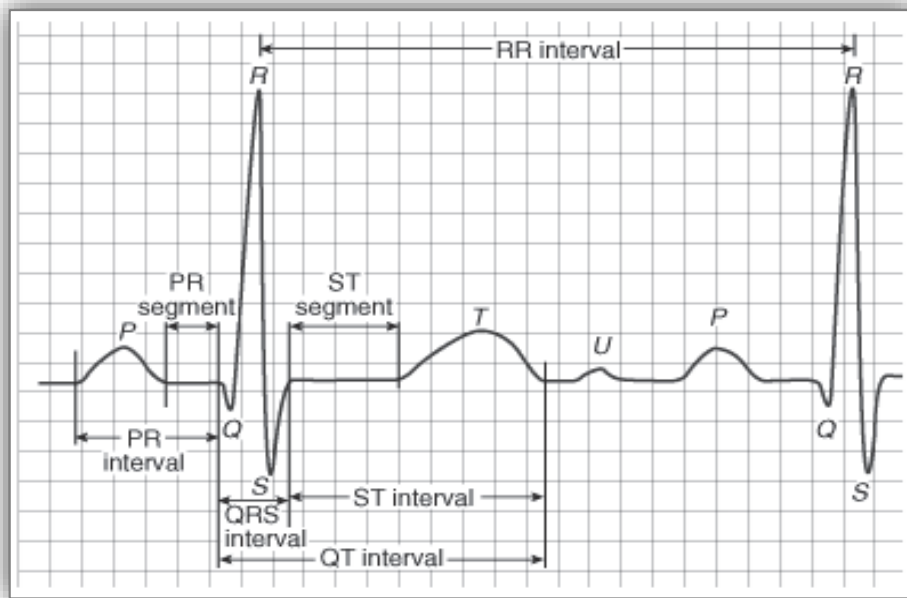


Figure I.4: ECG normal.

Ces composants réunies formant le cycle cardiaque complet, elles sont définies comme suit [7]:

- ❖ **L'onde P** : Représente la dépolarisation née du nœud sinusal, qui se propage dans les oreillettes et provoque leur contraction. Sa durée est inférieure à 120 ms, son amplitude inférieure à 2mm, elle est monophasique.
- ❖ **L'onde Q** : Elle est la première déflexion négative qui suit l'onde P souvent, il n'existe pas d'onde Q. Sa durée peut atteindre 0,2 secondes.
- ❖ **L'onde R** : Représente la première déflexion négative qui suit l'onde P.
- ❖ **L'onde S** : Représente la déflexion négative qui suit l'onde R. Le complexe QRS (appelé ainsi par extension, même s'il ne comprend pas toutes les ondes) correspond à la dépolarisation des ventricules, et à leur contraction. Sa durée est inférieure à 100ms.
- ❖ **Le segment ST** : Fait suite au QRS, il est normalement isoélectrique.
- ❖ **L'onde T** : Suit le complexe QRS et correspond à la repolarisation des ventricules. Elle survient pendant la diastole, et ne témoigne d'aucun événement mécanique. C'est un phénomène purement électrique, pendant lequel les ventricules redeviennent stimulables. Elle est habituellement dissymétrique.
- ❖ **L'interval QT** : Il s'agit de la distance entre le début du complexe QRS et la fin de l'onde T, englobant la dépolarisation et la repolarisation ventriculaires.
- ❖ **L'onde U** : Est une déflexion positive de faible amplitude qui est parfois observée après l'onde T et presque uniquement visible dans les précordiales. Sa signification exacte reste discutée.

I.3.4. Détection du signal ECG

Le signal ECG est détecté sur le corps en utilisant des électrodes, celles-ci peuvent être disposées de différentes manières standard qui sont connues par dérivations.

Il existe deux types de dérivations : Les dérivations frontales et les dérivations précordiales.

Ce sont "les dérivations des membres". Connues par D1, D2, D3, AVR, AVL et AVF.

Les dérivations D1, D2 et D3 sont des dérivations bipolaires. Elles traduisent la différence de potentiel entre deux membres :

- **D1** : Entre bras droit (pole-) et bras gauche (pole+).
- **D2** : Entre bras droit (pole-) et jambe gauche (pole+).
- **D3** : Entre bras gauche (pole-) et jambe gauche (pole+).

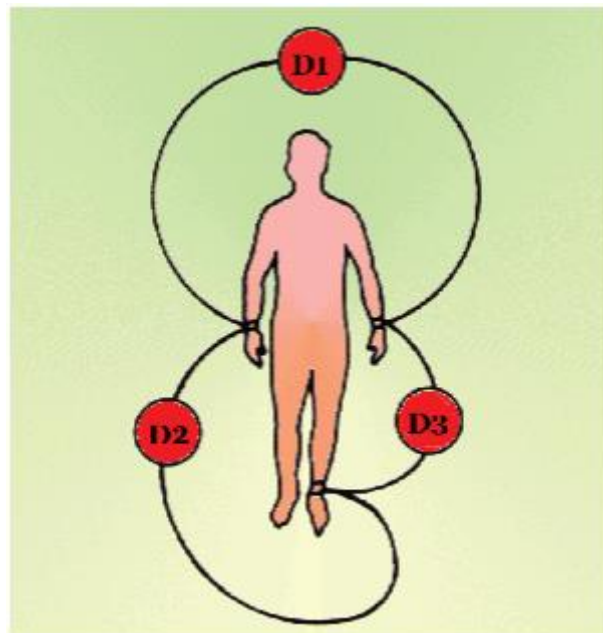


Figure I.5: Les dérivations bipolaires.

Dans la figure(I.6), D1, D2 et D3 décrivent le triangle d'Eithoven :

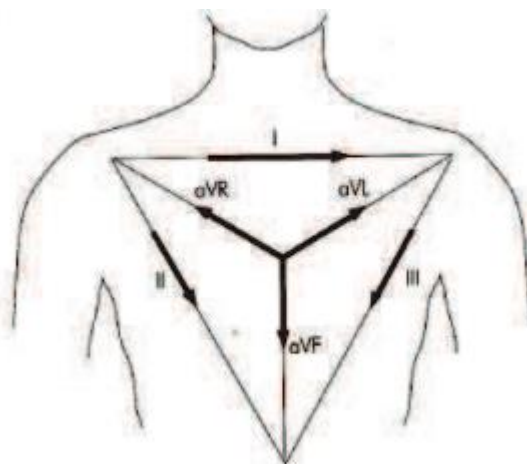


Figure I.6: Triangle d'Eithoven.

Les dérivations AVR, AVL et AVF sont des dérivations unipolaires.

L'ECG est détecté par rapport à une électrode de référence, celle-ci est connectée à la borne de Wilson (B.W)[8]. Cette borne correspond à la moyenne des potentiels au niveau de chaque électrode. Le potentiel ECG est alors mesuré dans chaque cas en plaçant l'électrode egolovatrice sur les membres bras droit (AVR), le bras gauche (AVL) en pied gauche (AVF).

Le voltage est alors amplifié (d'où le préfixe A) pour obtenir un tracé de même amplitude que D1, D2, D3.

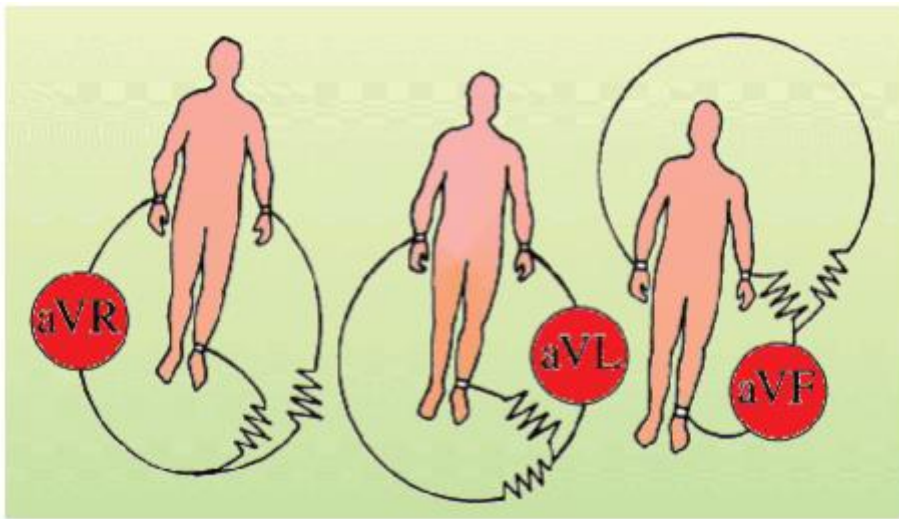


Figure I.7: Dérivation unipolaire.

L'ensemble des dérivations unipolaires et bipolaires projetées géométriquement représente un double triaxe avec un centre schématisé : le cœur.

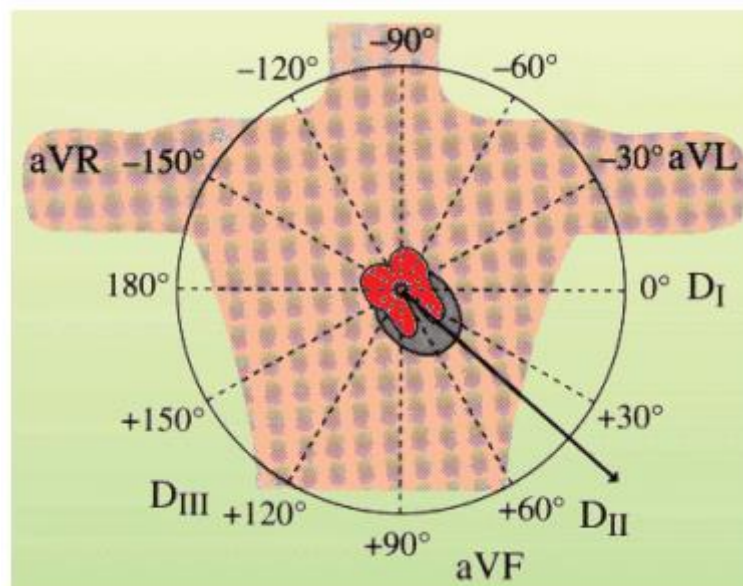


Figure I.8: Le triaxe des dérivations frontales.

On peut déjà apercevoir que les régions explorées par ces dérivations périphériques seront :

- **D1, AVL** : Paroi latérale du ventricule gauche.
- **D2, D3, AVF** : Paroi inférieure.
- **AVR** : Intérieur des cavités du cœur[9].

I.4. Les pathologies cardiaques

Les maladies cardio-vasculaires sont les premières causes de mortalité dans le monde. Dans les pays occidentaux, l'expression la plus courante est la maladie coronarienne, responsable de l'angine de poitrine ou encore des infarctus. Ces maladies touchent plus certaines catégories de populations et leur prévalence régionale. Elles courent souvent parmi les facteurs qui diminuent le plus l'espérance de vie d'une population et semblent être un facteur de risque de dépression.

I.4.1. Les maladies cardio-vasculaires

On trouve plusieurs types de maladies cardio-vasculaires :

- ❖ Les maladies coronariennes :
 - Angor (ou angine de poitrine).
 - Infarctus du myocarde.
- ❖ Les maladies du muscle cardiaque (myocarde) :
 - Cardiomyopathie.
 - Insuffisance cardiaque.
- ❖ Les maladies des valves cardiaques :
 - Endocardie.
 - Valvulopathies cardiaques.
- ❖ La maladie du péricarde :
 - Précarité.
- ❖ Les maladies du rythme ou de la conduction cardiaque (arithmologie) :
 - Syncope d'origine cardio-vasculaire :
 - Troubles de la conduction cardiaque.
 - Troubles du rythme cardiaque.
- ❖ Les maladies des vaisseaux :
 - Anévrisme.
 - Artériopathie oblitérante des membres inférieurs.
 - Dissection aortique aiguë.
 - Hypertension artérielle pulmonaire.
 - Maladie thromboembolique.

Comme il y'a d'autres types :

- Arrêt cardio-vasculaire.
- Cardiopathie congénitale.
- Hypertension artérielle.
- Hypotension artérielle.
- Tumeurs du cœur (Myxomes cardiaques).
- Syndrome d'Adams-Stokes.

I.4.2. Les facteurs de risque cardio-vasculaire

Les principaux facteurs de risque cardio-vasculaire héréditaires sont :

- L'hypercholestérolémie.
- L'hypertension.
- Les diabètes.
- L'excès de poids.
- Le tabagisme qu'il soit actif ou passif.

Les maladies cardio-vasculaires résultent d'une accumulation de ces facteurs de risque qui entraînent des altérations du code génétique et du métabolisme des graisses dans l'organisme. Ils accélèrent considérablement le phénomène physiologique d'artériosclérose entraînant un remaniement de la paroi des artères : l'athérome.

I.5. Trouble 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. Dans l'analyse de l'ECG, les pathologies ou anomalies sont détectées et classées en fonction de leur dérivation par rapport au rythme idéal qu'est le rythme sinusal. Chaque dérivation visible sur l'ECG peut être attribuée à une anomalie physiologique[10].

I.5.1. Rythme sinusal

C'est le rythme 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. 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 et alcool.

I.5.2. Blocs cardiaques

Les blocs cardiaques sont dus à une rupture de conduction du myocarde qui altère la dépolarisation. Ces ruptures peuvent être plus ou moins sévères : freinantes (allongement du temps de

parcours), intermittentes (le blocage de la conduction se fait aléatoirement), ou complètement (aucune conduction).

- ❖ **Bloc Sino-Auriculaires (BSA)** : Le nœud sinusal ne peut pas transmettre de stimulus aux cellules des oreillettes. La conséquence est qu'au moins un cycle reprend normalement si aucun autre foyer ectopique n'a déclenché la contraction.
- ❖ **Blocs Auriculo-ventriculaires (BAV)** : On appelle BAV l'altération de la conduction du stimulus de dépolarisation entre les oreillettes et les ventricules.
- ❖ **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 par rapport à l'ECG normal comme montré sur la figure(I.9).

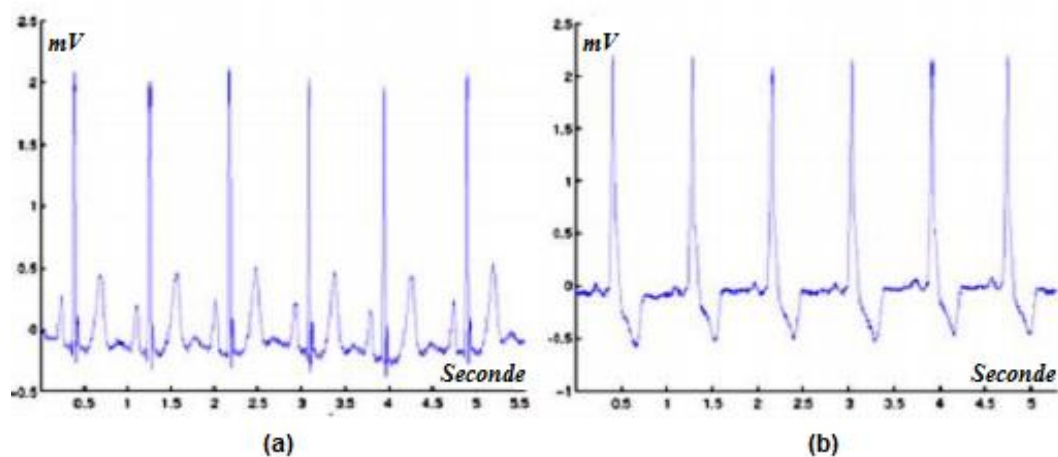


Figure I.9: Exemples d'ECG. (a) ECG normal; (b) bloc de branche.

I.6. Les arythmies cardiaques

Les arythmies cardiaques sont une cause principale de mortalité, elles forment une grande partie de la problématique des maladies cardio-vasculaires. L'arythmie survient lorsque l'excitation électrique naît hors du nœud sinusal, dans le nœud atrio-ventriculaire ou les ventricules par exemple, ou que l'onde électrique ne suit plus les voies préférentielles de propagation. La conduction cardiaque qui en résulte s'écarte du rythme normal. Le battement supplémentaire qu'elle provoque se nomme extrasystole, par opposition à la contraction normale nommée systole. On classe les arythmies en fonction de leur lieu de formation et de leurs effets sur le rythme cardiaque. Si le trouble se déclare dans les oreillettes ou le nœud atrio-ventriculaire, on parlera d'arythmie ventriculaire. Si une arythmie accélère le rythme cardiaque, on parlera alors de tachycardie (tachy= rapide) ; si elle le ralentit, on parlera alors de bradycardie (brady= lent)[11].

I.6.1. Extrasystoles supra-ventriculaires

L'extrasystole supra-ventriculaire est caractérisée par une stimulation cardiaque prématurée par rapport au rythme de base. Selon son origine, elle pourra être auriculaire (oreillettes) ou jonctionnelle (nœud atrio-ventriculaire). Souvent l'hypertension artielle, en raison de sa surcharge pour le cœur, provoque une distension des oreillettes et leur contraction prématurée. La faiblesse du muscle cardiaque peut constituer une autre cause de ce type d'extrasystole.

I.6.2. Contraction ventriculaire prématurée (extrasystole ventriculaire)

L'extrasystole ventriculaire est due à la présence d'une zone irritative dans le ventricule qui ne se souvent pas à l'harmonie de l'activation cardiaque. L'émotion, un déséquilibre de la teneur en sels minéraux, un état fébrile ou une infection peuvent démasquer une propension à cette arythmie. Dans ces cas, celle-ci sera considérée comme bénigne, tandis que l'apparition d'extrasystole accompagnant une maladie cardiaque peut dévoiler un problème qui nécessite un traitement. Si les extrasystoles se multiplient et que la stimulation électrique suit le mouvement, la fréquence cardiaque augmentera, on parlera alors de tachycardie.

I.6.3. Tachycardie supra-ventriculaire

Lors de tachycardie paroxystique, le cœur s'emballé soudainement et bat très rapidement, jusqu'à 150 à 220 fois par minute. Ces épisodes peuvent durer quelques secondes ou se prolonger pendant plusieurs heures. Ils s'arrêtent aussi soudainement qu'ils sont apparus. Les tachycardies supra-ventriculaires peuvent apparaître dès le très jeune âge et persister toute la vie. Après ces crises, le sujet se sent fatigué, mais ce type d'arythmie est rarement dangereux. Elles nécessitent néanmoins l'attention du médecin, car elles peuvent être gênantes et diminuer la qualité de vie.

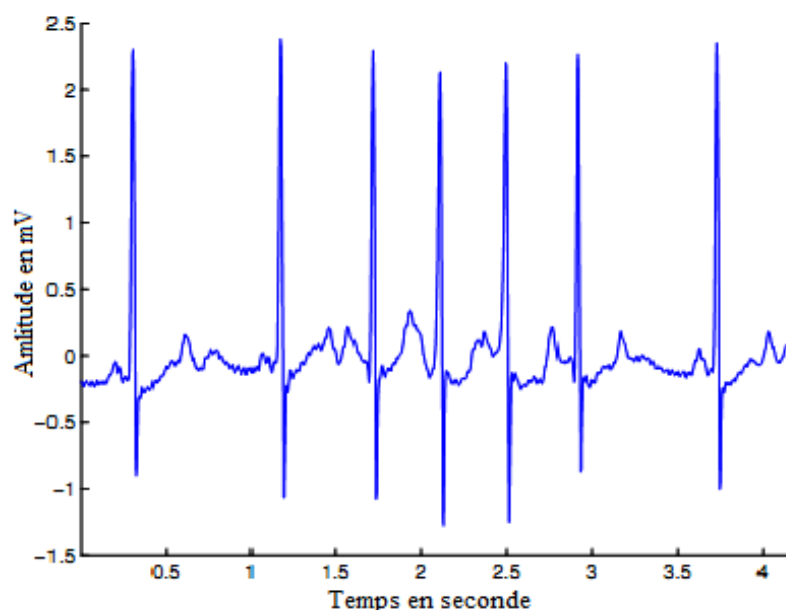


Figure I.10: Tachycardie supra-ventriculaire.

I.6.4. La fibrillation auriculaire

Elle est définie comme la contraction anarchique des oreillettes, qui entraîne une contraction rapide et irrégulière des ventricules. Il s'agit d'un trouble du rythme du cœur très fréquent, en particulier chez les sujets de plus de 60 ans. Le diagnostic est difficile car les symptômes de fibrillation auriculaire ne sont absolument pas ressentis.

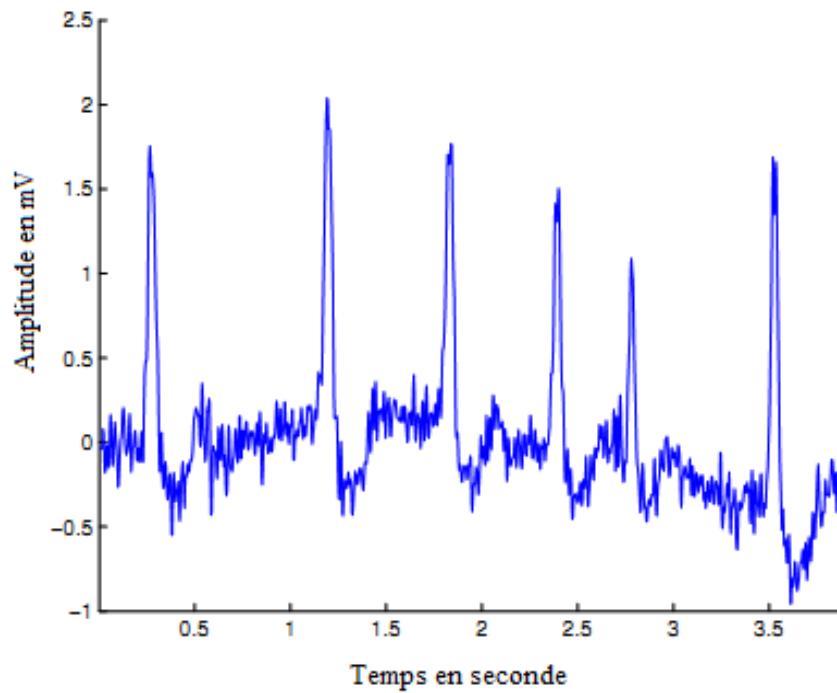


Figure I.11: Fibrillation auriculaire.

I.6.5. La tachycardie ventriculaire

C'est la manifestation d'un dysfonctionnement du muscle cardiaque. Elle peut apparaître à tout âge, mais elle est fréquemment la conséquence d'un infarctus du myocarde. Elle se manifeste par la désynchronisation des battements ventriculaires qui conduit à une diminution du pompage du sang.

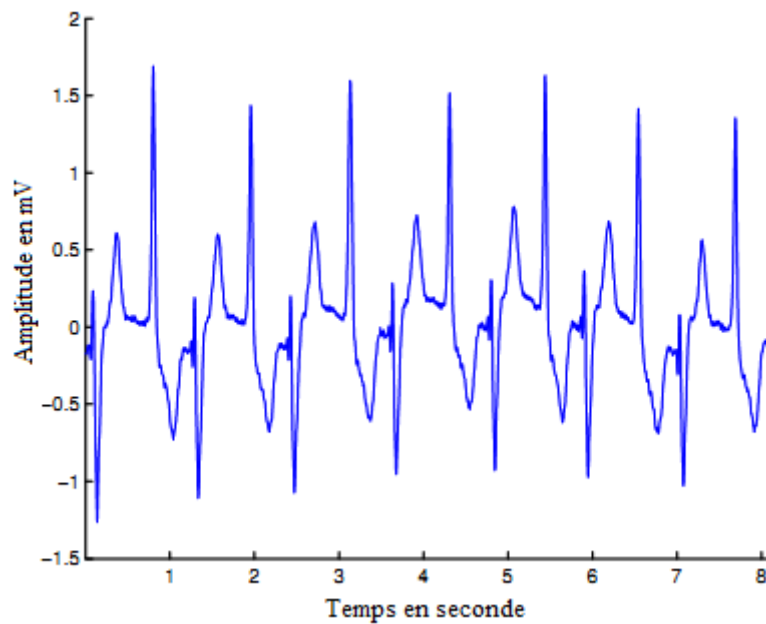


Figure I.12: Tachycardie ventriculaire.

I.6.6. Fibrillation ventriculaire

La fibrillation ventriculaire est parfois inaugurale (mort subite), précoce (fibrillation ventriculaire secondaire) accompagnant alors un anévrisme du ventricule gauche suite à un mauvais diagnostic. La fibrillation ne peut pas se produire dans un milieu homogène. En fait, la période réfractaire, l'excitabilité, ainsi que la vitesse de conduction n'ont pas des propriétés constantes partout dans le tissu ventriculaire.

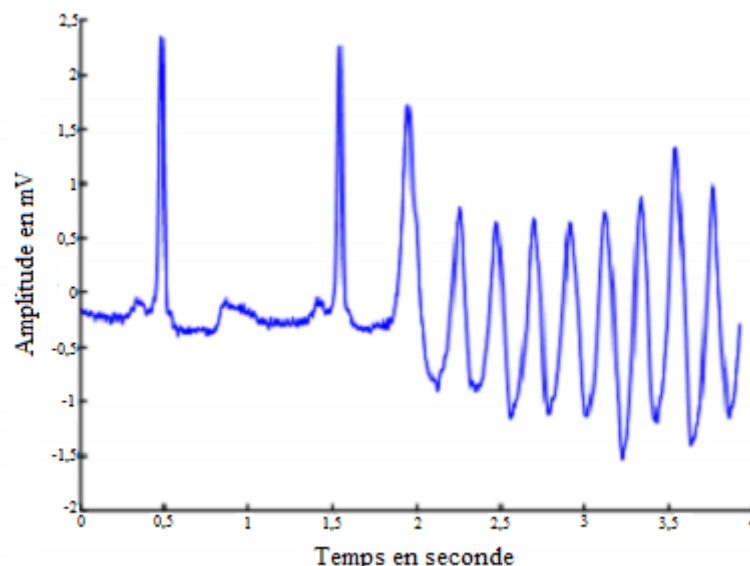


Figure I.13: Fibrillation ventriculaire.

I.7. Artefacts visibles sur l'électrocardiogramme

Sur tout enregistrement électro-cardiographique il peut apparaître des événements indésirables pouvant brouiller le tracé et parfois induire en erreur le diagnostic final. Ces bruits sont reconnaissables par l'œil expérimenté qui les identifie avant d'effectuer son diagnostic. Les effets indésirables peuvent avoir plusieurs sources: techniques, physiques ou pathologiques. Nous allons surtout développer l'aspect technique et physique des bruits et artefacts présents sur les tracés électro-cardiographiques. En partant du principe que les bruits fréquents en électrocardiographie sont des bruits additifs, les caractéristiques de ces bruits auxquelles nous allons nous attacher sont l'amplitude, la période et la bande spectrale. Les artefacts prennent une place particulière puisqu'ils induisent des modifications des performances des algorithmes d'analyse automatique de l'électrocardiogramme[11].

I.7.1. Bruits techniques

Le matériel utilisé lors de l'enregistrement doit être manipulé avec précaution car il peut être source de bruits lors de l'enregistrement. Les bruits les plus courants sont présentés comme suit :

- ❖ **Bruit dû au secteur :** Le réseau de distribution électrique peut parfois brouiller le signal électro-cardiographique avec une onde dont l'harmonique principale est à 50HZ. Ce type de bruit apparaît sur tout l'enregistrement et peut être assez fort mais il s'élimine facilement avec un filtre sélectif car c'est un bruit haute fréquence a bande étroite.

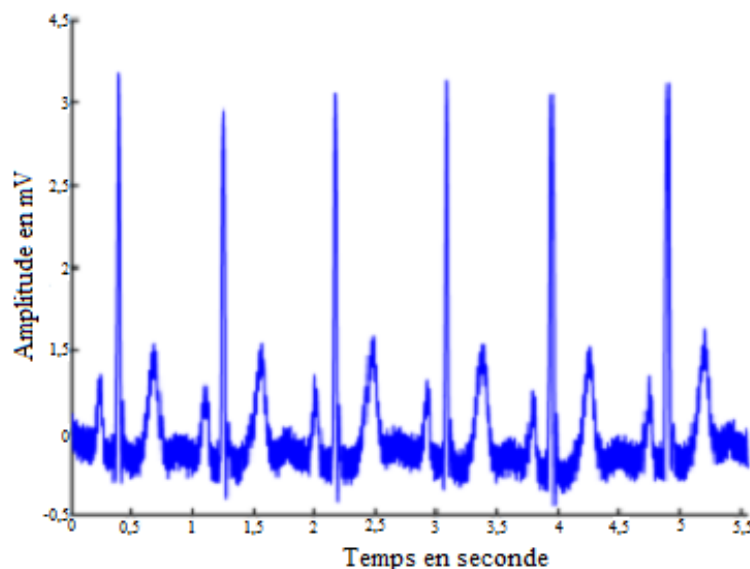


Figure I.14: Interface secteur 50HZ.

- ❖ **Bruit dû aux mouvements d'électrodes :** Lorsque les électrodes sont connectées incorrectement, des sauts brusques de la ligne de base apparaissent. L'effet sur le tracé peut aller de la simple diminution d'amplitude à l'apparition de pics lorsque les électrodes sont en contact

intermittent avec la peau. Ces pics peuvent parfois être confondus avec les ondes du tracé normal. Ce type de bruit intermittent à bande spectrale large s'élimine difficilement car son énergie se trouve dans la même gamme de fréquence que le complexe QRS.

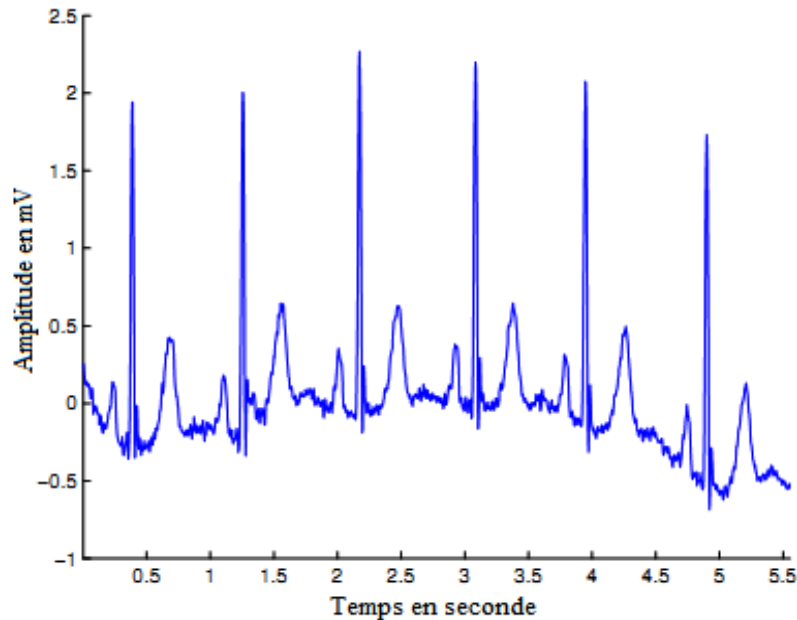


Figure I.15: Bruits dû au mouvement des électrodes.

Parmi les bruits courants on peut citer encore les artefacts dus aux mouvements des câbles électriques, la saturation des instruments de mesure, les mauvais câblages, les artefacts dus au port de vêtements synthétiques...etc.

I.7.2. Artefacts physiques

Les artefacts physiques sont dus aux activités électriques du corps humain telles que les commandes de contraction des muscles ou la respiration.

- ❖ **Mouvements de la ligne de base :** Lors de l'enregistrement de l'ECG, l'activité respiratoire peut faire osciller la ligne de base de l'ECG à un rythme régulier. D'autres perturbations peuvent avoir pour effet de déplacer temporairement le ligne de base comme par exemple : les mauvais contacts entre la peau et les électrodes. Ces perturbations sont généralement peu gênantes pour l'analyse de l'ECG et peuvent être en grande partie filtrée car leur énergie se situe dans une bande de fréquence basse, qui empiète peu sur celle de l'ECG normal.

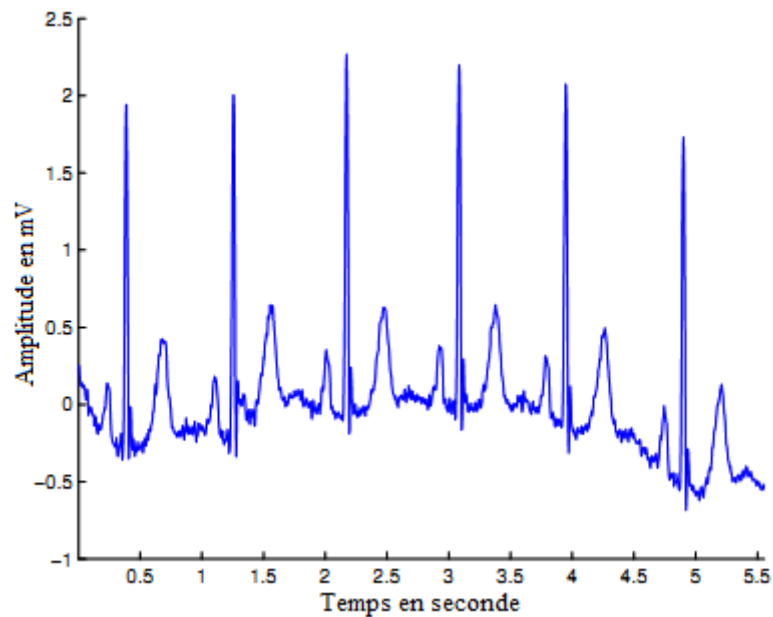


Figure I.16: Dérivé de la ligne de base.

- ❖ **Bruit myo-électrique ou tremblement somatique** : La contraction d'un muscle est commandée par une dépolarisation de cellules musculaires et bien que les électrocardiographes soient construits pour être surtout sensibles aux fréquences du myocarde, l'ECG enregistre les contractions des muscles squelettiques. L'aspect le plus courant est une oscillation haute fréquence liée à la tension musculaire d'un sujet qui n'est pas convenablement détendu. Ces perturbations sont assez gênantes lorsque le patient bouge beaucoup ou lorsqu'il frissonne, elles peuvent noyer les ondes P et T et empêcher un diagnostic fiable. L'apparition de ces perturbations dépend de l'état du patient, s'il est très tendu ou atteint de maladie de Parkinson, l'enregistrement peut être de mauvaise qualité sur toutes les voies de l'ECG.

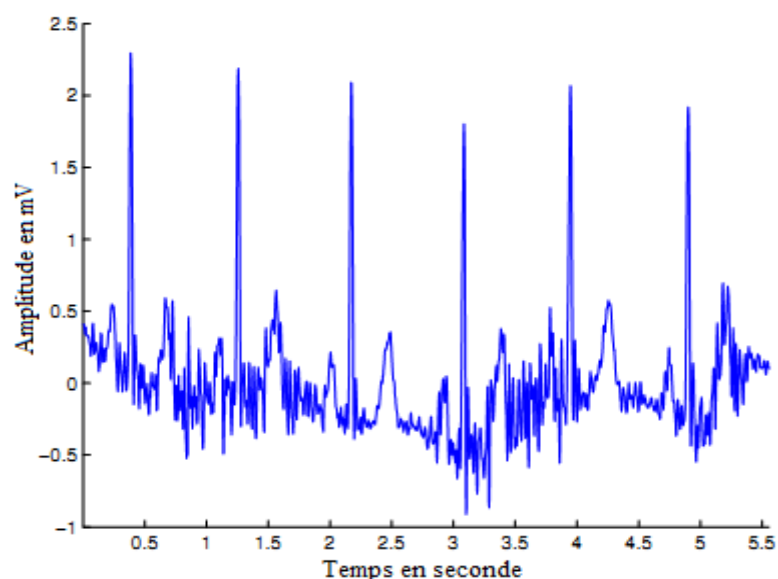


Figure I.17: Bruit myoélectrique ou troublement somatique.

I.8. Conclusion

A travers ce chapitre, nous avons décrit des éléments de base de l'électrocardiographie du cœur, de son fonctionnement et des différents aspects positifs et négatifs de l'électrocardiogramme.

Après une introduction sur l'anatomie du cœur, ses fonctionnements électriques et mécanique et le lien entre eux.

On a passé dans la deuxième tranche de ce chapitre, à exposer les notions sur l'électrocardiographie, les pathologies cardiaques et des facteurs de risque cardio-vasculaire et les arythmies cardiaques, illustrées chacune par un enregistrement ECG correspondant.

En fin dans la troisième tranche de ce chapitre, on a cité des artefacts visibles sur l'électrocardiogramme.

On a réservé le chapitre deux pour expliquer les modèles de classification en citant des exemples évalués pour juger le choix des réseaux de neurones artificiels et les systèmes experts dans ce mémoire.

Chapitre II

La classification en intelligence artificiel

II.1. Introduction

Un diagnostic médical est le résultat du raisonnement d'un médecin, décision très souvent prise à partir d'informations incertaines et/ou incomplètes. De nombreuses techniques d'intelligence artificielle ont été appliquées pour essayer de modéliser ce raisonnement[12]. Ainsi présente l'utilisation détaillée de plusieurs systèmes experts en médecine.

En amont de ce raisonnement, il faut aussi être capable de modéliser ces informations incertaines et/ou incomplètes. Les performances des ordinateurs sont de plus en plus remarquables, et avec elles, la nécessité et la possibilité de développer des méthodes de classification automatiques capables et robustes augmentent également. Dans ce contexte, de nombreuses méthodes de classification automatique sont apparues. Les plus connues : les méthodes de la logique floue les réseaux de neurones bayésiens, les arbres de décision, les machines à vecteurs de support, les cartes de Kohonen...etc.

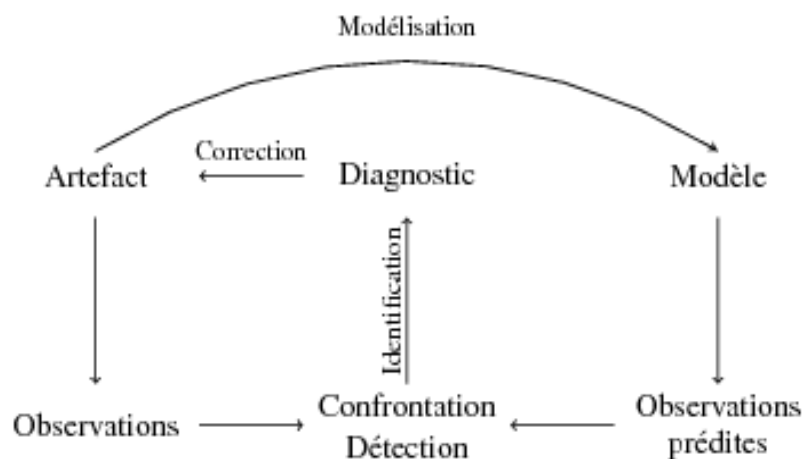


Figure II.1: Diagnostic basé sur le modèle.

D'après ce chapitre on va expliquer la fonctionnalité de la classification dans l'intelligence artificielle, dans la première partie on va démontrer le fonctionnement en général.

Puis, dans la deuxième partie on va citer la chaîne de classification d'objets et ces principaux modules.

En fin, dans la troisième partie on va détailler les méthodes de classification.

II.2. La classification en intelligence artificielle

La classification est l'une des techniques les plus utilisées dans l'apprentissage automatique, avec un large éventail d'applications, y compris l'analyse des sentiments, le ciblage des annonces, la détection des spams, l'évaluation des signaux, le diagnostic médical et la classification des images. L'objectif principal de la classification est de prédire une catégorie ou une classe Y à partir de certaines entrées X .

La classification rassemble une famille de méthodes permettant d'automatiser le processus de reconnaissance, elle peut être considérée comme un domaine permettant de définir des algorithmes et des techniques susceptibles de classer des objets dont l'aspect est variable par rapport à un objet type. Le but de la classification est donc de classer des observations sur la base d'une série de caractéristiques prédéfinies par apprentissage.

Aujourd'hui on ne peut pas parler d'une théorie de la classification d'objets, mais d'approches différentes dont le but commun est de mettre en évidence, de façon automatique, une structure connue à l'ensemble des objets en présence, par recherche de similarités. Le champ d'application de la classification est particulièrement vaste. Il couvre principalement le domaine du traitement des signaux et d'images. On distingue habituellement deux types de classification :

- ❖ **Classification en mode supervisé :** C'est la classification avec ensemble d'apprentissage connu et parfaitement expertisé.
- ❖ **Classification en mode non supervisé :** C'est le cas où on ne dispose pas d'un ensemble d'apprentissage préalablement expertisé.

II.2.1. Chaîne de classification d'objets

II.2.1.1. Les principaux modules de classification d'objets

Un dispositif de classification ou de reconnaissance automatique de formes est généralement conçu comme une chaîne de modules de traitement[13].

Un système de reconnaissance de formes comporte habituellement les modules suivants :

- ❖ **Un module d'acquisition :** Des capteurs mesurent des grandeurs caractéristiques de l'objet à classer (signal, image...etc.). Cet ensemble de grandeurs constitue la première présentation de l'objet.
- ❖ **Un module de prétraitement :** Il peut être judicieux de modifier des grandeurs brutes issues des capteurs par un algorithme afin de tenir compte des connaissances qui peuvent être disponible à priori sur le problème. Par exemple à partir de la réponse d'un capteur on peut appliquer un ensemble de filtres destinés à éliminer des bruits indésirables. Ainsi on obtient une nouvelle représentation de l'objet, plus adéquate pour la classification envisagée. D'autres modules de traitement peuvent élaborer des représentations successives de l'objet, ces différentes représentations ont généralement pour objectif de réduire la dimension de représentation, c'est-à-dire de diminuer le nombre de descripteurs de l'objet et d'élaborer des descripteurs de plus en plus pertinents pour la tâche de discrimination à accomplir.
- ❖ **Un module de classification :** L'algorithme de classification considère la dernière représentation de l'objet et décide d'affecter celui-ci à une classe. Cet algorithme peut fournir

soit une réponse binaire aux valeurs discrètes (appartenance à une classe) soit une réponse à valeurs continues.

II.2.1.2. L'extraction des descripteurs

Habituellement une classification de formes ne se fait pas directement sur des formes brutes (morphologie de l'ECG par exemple), mais plutôt à partir de descripteurs ou paramètres caractérisant les formes. Chacune des N formes F_i est représentée par un point X_i qui est l'espace des paramètres. Il existe deux approches pour caractériser une forme [14]:

- ❖ **L'approche purement mathématique** : (Analyse en composants principales, prédiction linéaire, transformée de Fourier...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 semble importants. Cette approche donne souvent de meilleurs résultats, car les paramètres choisis résultent d'une grande expérience et peuvent plus discriminants.

II.2.1.3. Procédure de résolution par apprentissage

La classe reçoit une définition qui peut être une définition purement descriptive ou une interprétation, par exemple du type diagnostic médical. La classification qui réalise le classement des formes doit passer par deux phases, une phase d'apprentissage et une phase de test[15].

II.2.1.3.1. Phase d'apprentissage

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 une certaine généralisation. Il existe principalement deux types d'apprentissage, supervisé et non supervisé[16]. Dans le premier cas, on doit apprendre des associations (individus, classes), dans le deuxième cas on ne fournit pas d'indications sur les classes.

- ❖ **Apprentissage supervisé** : L'apprentissage est dit supervisé si les différentes familles de formes, ou classes, sont connues a priori ainsi que l'affectation de chaque forme à telle ou telle famille. Le processus d'apprentissage en mode supervisé peut se résumer comme suit :

On dispose d'un ensemble de formes ou d'objets, qui est un ensemble d'apprentissage A d'objets parfaitement classés, ensemble que l'on se propose de reconnaître, il peut être présenté sous la forme de paires (X_i, C_j) , $1 \leq i \leq N$, $1 \leq j \leq m$, $m < N$, et interprété : « l'objet X_i appartient la classe C_j ».

On construit dans un premier temps une « machine » sachant classer correctement les objets de l'ensemble A . la description de chaque forme est prise en compte par la tâche d'apprentissage. Le travail d'apprentissage consiste à analyser les ressemblances entre les formes d'une même

classe et les dissemblances entre les formes de classes différentes pour déduire les meilleures séparations possibles entre classes. Ensuite, on fait en sorte que cette « machine » ait une bonne capacité à prédire la classe d'une forme qui n'est pas dans l'ensemble d'apprentissage A. D'où l'exigence de tester cette machine sur un ensemble de validation et d'objets parfaitement classés. Cette phase de validation est nécessaire car elle aussi de savoir si la machine élaborée n'a pas été entraînée à apprendre « par cœur » sans possibilités de généralisation (sur-apprentissage ou over fitting)[13] [15].

❖ **Apprentissage non supervisé :** On l'appelle aussi, suivant l'approche utilisée, classification automatique, ou apprentissage sans professeur ou encore apprentissage par corrélation[15]. Ce type d'apprentissage est utilisé dans le cas où on dispose d'une base d'apprentissage dont les classes ne sont pas définies à l'avance : soit constitué d'une part d'une liste d'objets X_i et d'autre part de la donnée du nombre de catégories, soit enfin uniquement composé d'un ensemble d'objets X_i , $1 \leq i \leq N$.

Il s'agit alors de regrouper les différentes formes en classes en fonction d'un critère de similarité choisi a priori. Ce type d'apprentissage permet la construction automatique des classes sans intervention d'un opérateur. On distingue essentiellement deux types de méthodes : les processus de coalescence et la classification hiérarchique[13]. Cette approche nécessite cependant d'avoir une bonne estimation du nombre de classes.

II.2.1.3.2. 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[15]. 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.

❖ **Taux de classification et taux d'erreurs :** Les taux de classification et d'erreurs permettent d'évaluer la qualité du classifieur C par rapport au problème pour lequel il a été conçu. Ces taux sont évalués grâce à une base de test qui contient des formes décrites pour l'apprentissage.

Elles sont aussi étiquetées par leur classe réelle d'appartenance afin de pouvoir vérifier les réponses du classifieur. Pour que l'estimation du taux de reconnaissance soit la plus faible possible, il est important que le classifieur n'ait jamais utilisé les échantillons de cette base pour faire son apprentissage (la base de test ne doit avoir aucun objet en commun avec la base

d'apprentissage et les éventuelles bases de validation). De plus, cette base de test doit être suffisamment représentative du problème de classification[17].

En générale, quand les échantillons étiquetés à disposition sont suffisamment nombreux, ils sont séparés en deux parties disjointes et en respectent 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. Le découpage le plus courant est de 2/3 pour l'apprentissage et le 1/3 restant pour la base de test. Les performances en termes de taux de classification sont alors déterminées en présentant la classe donnée en résultat $C(e_j) = s$ à la vraie classe de e_j . En considérant que la base de test contient N objets et que sur ceux-ci $\tau_{corrects}$ sont biens classés par le système, le taux de classification τ_{clas} est simplement défini par :

$$\tau_{clas} = \frac{N_{corrects} * 100}{N} \dots\dots\dots \text{II.1}$$

Le taux d'erreur τ_{err} est défini partir du nombre d'objets N_{err} mal classés [18]:

$$\tau_{err} = \frac{N_{err} * 100}{N} \dots\dots\dots \text{II.2}$$

Sensibilité spécialité : L'évaluation des performances d'un classifieur peut être réalisée par l'appréciation d deux lois statistique, qui sont la sensibilité $S_e(i)$ et la spécialité $S_p(i)$:

$$S_e(i) = \frac{V_p(i)}{V_p(i) + F_N(i)} \dots\dots\dots \text{II.3}$$

$$S_p(i) = \frac{V_N(i)}{V_N(i) + F_P(i)} \dots\dots\dots \text{II.4}$$

Où les grandeurs $VP(i)$, $FN(i)$, $VN(i)$ et $FP(i)$ sont définies dans le tableau suivant :

	Présence d'évènement de classe i	Absence d'évènement de classe i
Classification positive	Vrai positif $VP(i)$	Faux positif $FP(i)$
Classification négative	Faux négatif $FN(i)$	Vrai négatif $VN(i)$

Tableau II.1: Les définitions des grandeurs VP , VN , FP et FN . [14]

La sensibilité $S_e(i)$ représente la probabilité de bonne classification de la classe i et la spécificité $S_p(i)$ est une mesure indirecte de la probabilité de fausse alarme puisque cette dernière n'est rien que $1-S_p(i)$.

II.2.2. Formalisation mathématique d'un problème de classification

La classification apparait comme une tâche qui consiste à ranger des formes ou objets décrits par un ensemble de variables descriptives en un certain nombre de catégories ou classes définies a priori. Traduit en termes mathématiques, un problème de classification comporte les éléments suivants :

- Une population de N objets $\Theta_i, 1 \leq i \leq N$.

- P variables descriptives X_i^d , qui permettent de décrire les objets, elles sont aussi appelées plus simplement descripteurs, $1 \leq d \leq P$.
- C classes C_k dans lesquelles on cherche à ranger les objets (k variant de 1 à C). 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 appelée classifieur. Les variables descriptives considérées ici sont celles qui sont fournies à l'algorithme de classification. Comme indiqué plus haut, elles peuvent être le résultat d'un prétraitement des variables initiales[14].

II.3. Les méthodes de classification

Il existe de nombreuses méthodes de classification en IA appliquées dans le domaine médical, parmi celles-ci on trouve la logique floue, les arbres de décision, les réseaux de neurones, les réseaux bayésiens, les systèmes experts, les machines à vecteurs de support...etc.

II.3.1. La logique floue en classification

L'approche par la logique floue permet l'introduction du concept de degré d'appartenance, qui détermine les « forces » avec lesquelles un individu appartient à une classe, mais qu'elle raisonne plutôt sur la base d'un intervalle de valeurs. L'idée qui soutient l'approche est la possibilité d'appartenance à la fois à plusieurs classes. Partant dans ce sens, ce sont exclus toutes les méthodes de classification « dure » contraignant les individus à être membre d'une, et une seule, classe. Quoique la probabilité d'appartenance des objets aux classes ne soit pas évidemment une exclusivité des techniques de classification floue, il nous semble que les résultats obtenus des approches floues sont plus pertinents en ce sens qu'ils donnent une matrice des degrés d'appartenance de chaque individu à chaque classe, ce qui n'est pas tout à fait le cas dans les autres analyses « classiques » de classification[19].

II.3.1.1. Conception de classificateur flou

II.3.1.1.1. Principe de base

Le problème de diagnostic médical peut être formulé comme suit :

Soit $C = \{C_1; C_2; \dots; C_M\}$ un ensemble de M diagnostics possibles dans le contexte d'un certain problème médical. C peut être un ensemble d'anomalies du cœur, des types de tissus dans une Image par Résonance Magnétique (IRM) du cerveau ou des types de cellules sanguines, etc. Soit x une description d'un objet (par exemple, un patient, un morceau de tissu cérébral, une cellule, etc.) sous forme d'un vecteur de réels de Q éléments : $x = [P_1; P_2; \dots; P_Q] \in \mathbb{R}^Q$.

Les composantes du vecteur x encodent des paramètres tels que des mesures cliniques, des détails physiologiques d'un patient, des résultats de tests, des paramètres d'imagerie telle que l'intensité du niveau de gris, l'aspect circulaire d'une cellule, des électrocardiogrammes...etc.

Un classificateur flou est toute application : $D : \mathbb{R}^Q \rightarrow [0,1]^M$, le résultat de la classification est alors donné par : $D(x) = [\mu_1(x) ; \dots ; \mu_M(x)]$ où $\mu_i(x)$ représente le degré d'appartenance de x à la classe C_i [20].

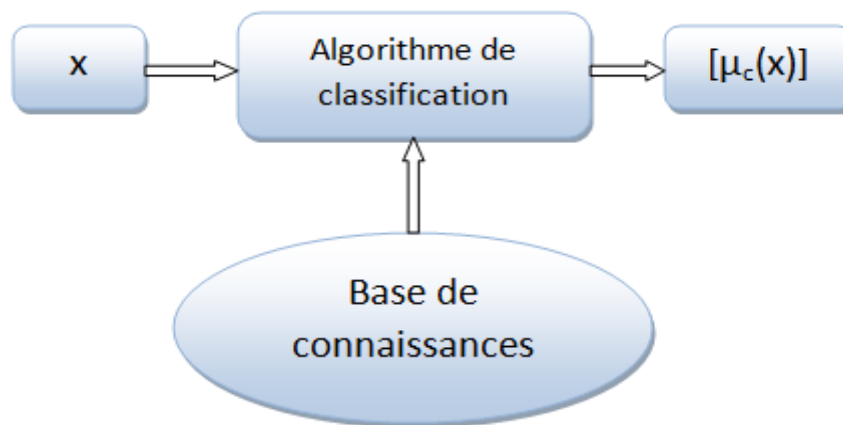


Figure II.2: Schéma synoptique d'un classificateur supervisé.

II.3.1.1.2. Base des règles floues

La base des règles floues définie comporte 4 règles (Tableau II.2). Ces règles ont été élaborées manuellement. Une fois encore, nous avons adopté l'approche intuitive pour la construction de la base des connaissances bien que plusieurs méthodes de génération et d'apprentissage des règles à partir des données numériques aient été proposées.

Règle	F	RR	P	QRS	Diagnostic
r1	Faible	Irrégulier	Absente	Fin	FL
r2	Moyenne	Irrégulier	Absente	Fin	ACFA
r3	Elevée	irrégulier	Absente	Fin	TACFA
r4	Moyenne	Régulier	Présente	Moyen	N

Tableau II.2: Base des règles floues.

II.3.1.2. Avantages et Inconvénients

Les systèmes à base de logique floue ont des formalismes robustes qui ne nécessitent aucun modèle mathématique, ils permettent l'intégration de connaissances a priori sous forme de règles, ils garantissent le suivi et l'interprétation des processus et des résultats, ils sont facilement implémentables et permettent l'extraction des règles.

Par contre, Ils n'ont aucun apprentissage possible, ils ont une adaptation relativement difficile aux modifications des données, ils nécessitent un processus de déffusification et l'inconvénient majeur pour nous est qu'ils sont sensibles aux données bruités, le cas de notre base d'apprentissage.

II.3.2. Les arbres de décision

Un arbre de décision est un outil d'aide la décision représentant un ensemble de choix sous la forme graphique d'un arbre. Les différentes décisions possibles sont situées aux extrémités des branches (les feuilles de l'arbre), et sont atteints en fonction de décisions prises à chaque étape.

L'apprentissage par arbre de décision désigne une méthode basée sur l'utilisation d'un arbre de décision comme modèle prédictif. On utilise notamment en fouille de données et en apprentissage automatique. C'est une technique d'apprentissage supervisé, on utilise un ensemble de données pour lesquelles on connaît la valeur de la variable-cible afin de construire l'arbre, puis on extrapole les résultats à l'ensemble des données de test.

II.3.2.1. Historique

La construction des arbres de décision à partir de données est une discipline déjà ancienne. Les statisticiens en attribuent la paternité à Morgan et Sonquist (1963), les premiers qui ont utilisé les arbres de régression dans un processus de prédiction et d'explication (AID Automatic Interaction Detection). Il s'en est suivi toute une famille de méthodes étendus, jusqu'aux problèmes de discrimination et classement, qui s'appuyaient sur le même paradigme de la représentation par arbres (THAID Morgan et Messenger, 1973, CHAID Kass, 1980). On considère généralement que cette approche a connu son apogée avec la méthode CART (Classification and Regression Tree) de Breiman et AL (1984) décrite en détail dans une monographie qui fait encore référence aujourd'hui.

II.3.2.2. Exemple

Les arbres de décision sont utilisés dans des domaines d'aide à la décision ou l'exploitation de données. Ils décrivent comment répartir une population d'individus en groupes homogènes selon un ensemble de variables discriminantes et en fonction d'un objectif fixé.

Pour mieux appréhender la démarche, nous allons reprendre et découler un exemple qui est présenté dans l'ouvrage de Quinlan (1993). Il s'agit de prédire le comportement des individus par rapport à un jeu {jouer, ne pas jouer} à partir des prévisions météorologiques[21].

Numéro	Ensoleillement	Température (°F)	Humidité(%)	Vent	Jouer
1	Soleil	75	70	OUI	OUI
2	Soleil	80	90	OUI	NON
3	Soleil	85	85	NON	NON
4	Soleil	72	95	NON	NON

5	Soleil	69	70	NON	OUI
6	Soleil	72	90	OUI	OUI
7	Couvert	83	78	NON	OUI
8	Couvert	64	65	OUI	OUI
9	Couvert	81	75	NON	OUI
10	Pluie	71	80	OUI	NON
11	Pluie	65	70	OUI	NON
12	Pluie	75	80	NON	OUI
13	Pluie	68	80	NON	OUI
14	Pluie	70	96	NON	OUI

Tableau II.3: Exemple de données.

L'arbre de décision correspondant est décrit comme suit :

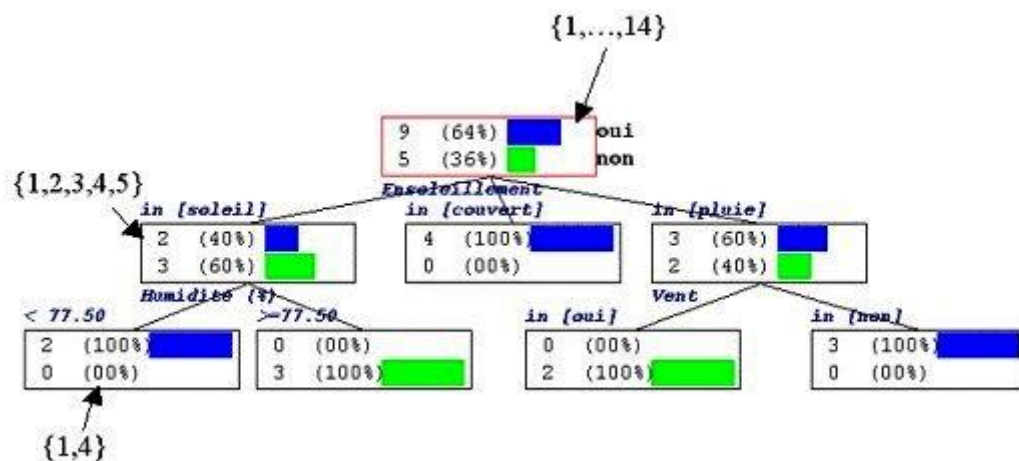


Figure II.3: Arbre de décision correspondant.

Chaque nœud de l'arbre décrit la distribution de la variable Jouer à prédire. Dans le cas du premier nœud, la racine de l'arbre, nous constatons qu'il y a 14 observations dans notre fichier : 9 cas où une partie a eu lieu (Jouer=oui) et 5 où aucune partie n'a eu lieu (Jouer=non).

Ce premier nœud a plusieurs fils construits en utilisant la variable Ensoleillement :

Le plus à gauche (Ensoleillement = Soleil) comporte 5 observations, le suivant (Ensoleillement=couvert) en comporte 4, et ainsi de suite. La suite de décisions continue jusqu'à ce que, dans l'idéal, les observations dans un nœud soient toutes « oui » ou toutes « non ». On dit alors que le nœud est homogène.

Le processus de décision s'arrête aux feuilles de l'arbre. Dans l'arbre ci-dessus, toutes les feuilles sont homogènes, c'est-à-dire que les variables prédictives utilisées permettent de prédire complètement si une partie va avoir lieu ou non.

L'arbre se lit intuitivement de haut en bas, ce qui se traduit en termes de règles logiques sans perte d'informations : par exemple, la feuille la plus à gauche se lit : « si ensoleillement = soleil et humidité < 77,5 % alors jouer = oui » [21].

II.3.2.3. Avantages et Inconvénients

Comparativement à d'autres méthodes de fouille de données, les arbres de décision présentent plusieurs avantages :

- La simplicité de compréhension et d'interprétation.
- Peu de préparation des données.
- Le modèle peut gérer à la fois des valeurs numériques et des catégories. D'autres techniques sont souvent spécialisées sur un certain type de variables.
- Il est possible de valider un modèle à l'aide de tests statistiques, et ainsi de rendre compte de la fiabilité du modèle.
- Performant sur de grands jeux de données.

En revanche, elle présente certains inconvénients :

- L'apprentissage de l'arbre de décision optimal est NP-complet concernant plusieurs aspects de l'optimalité. En conséquence, les algorithmes d'apprentissage par arbre de décision sont basés sur des heuristiques telles que les algorithmes gloutons cherchant à optimiser le partage à chaque nœud, et de tels algorithmes ne garantissent pas d'obtenir l'optimum global.
- L'apprentissage par arbre de décision peut amener des arbres de décision très complexes, qui généralisent mal l'ensemble d'apprentissage.
- Certains concepts sont difficiles à exprimer à l'aide d'arbres de décision. Dans ces cas, les arbres de décision deviennent extrêmement larges.
- Lorsque les données incluent des attributs ayant plusieurs niveaux, le gain d'information dans l'arbre est biaisé en faveur de ces attributs.

II.3.3. Réseaux bayésiens

Un réseau bayésien est en informatique et en statistique un modèle graphique probabiliste représentant des variables aléatoires sous la forme d'un graphe orienté acyclique. Intuitivement, ils sont à la fois :

- Des modèles de représentation des connaissances.
- Des machines à calculer les probabilités conditionnelles.
- Une base pour des Système d'aide à la décision.

Les réseaux bayésiens constituent la technologie la plus puissante de ces 10 dernières années en IA et en apprentissage automatique. Les RBs constituent un langage graphique et une méthodologie,

simple et corrects, pour exprimer pratiquement ce de quoi on est certain ou incertain. Ils reposent sur la formule de Bayes reliant des probabilités conditionnelles avec des probabilités jointes[21].

II.3.3.1. Construction des réseaux bayésiens

Construire un réseau bayésien, c'est donc :

- ✓ Définir le graphe du modèle.
- ✓ Définir les tables de probabilité de chaque variable, conditionnellement à ses causes.

Le graphe est aussi appelé la « structure » du modèle, et les tables de probabilités ses « paramètres ». Structure et paramètres peuvent être fournis par des experts, ou calculés à partir de données, même si en général, la structure est définie par des experts et les paramètres calculés à partir de données expérimentales.

II.3.3.2. Exemple

Un opérateur travaillant sur une machine risque de se blesser s'il l'utilise mal. Ce risque dépend de l'expérience de l'opérateur et de la complexité de la machine. « Expérience » et « Complexité » sont deux facteurs déterminants de ce risque.

Bien sûr, ces facteurs ne permettent pas de créer un modèle déterministe. Quand bien même l'opérateur serait expérimenté et la machine simple, un accident reste possible. D'autres facteurs peuvent jouer : l'opérateur peut être fatigué, dérangé, etc. La survenance du risque est toujours aléatoire, mais la probabilité de survenance dépend des facteurs identifiés.

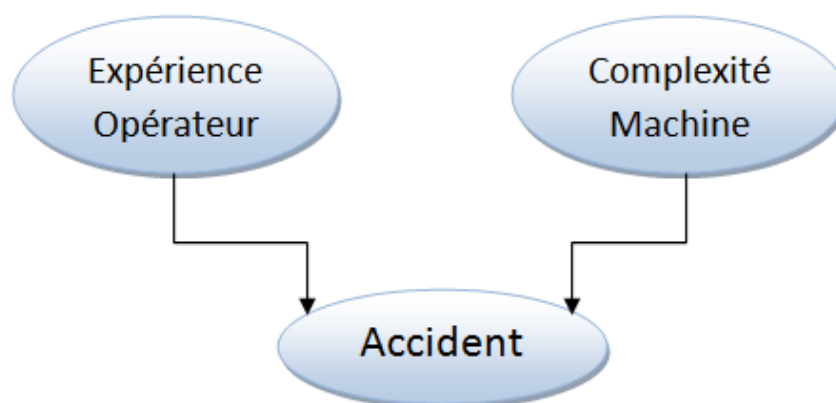


Figure II.4: Structure de causalité.

II.3.3.3. Avantages et Inconvénients

Les aspects suivants des réseaux bayésiens les rendent, dans de nombreux cas, préférables à d'autres modèles :

- ❖ **Acquisition des connaissances** : Les réseaux bayésiens donnent la possibilité de rassembler et de fusionner des connaissances de diverses natures dans un même modèle.

- ❖ **Représentation des connaissances** : La représentation graphique d'un réseau bayésien est explicite, intuitive et compréhensible.
- ❖ **Utilisation des connaissances** : Un réseau bayésien est polyvalent : on peut se servir du même modèle pour évaluer, prévoir, diagnostiquer, ou optimiser des décisions. Mais les réseaux bayésiens ont pas mal d'inconvénients surtout la complexité des algorithmes.
- ❖ **Les variables continues** : La plupart des algorithmes développés pour l'inférence et l'apprentissage dans les réseaux bayésiens, aussi bien que les outils disponibles sur le marché pour mettre en œuvre ces algorithmes utilisent des variables discrètes.
- ❖ **La complexité des algorithmes** : La généralité du formalisme des réseaux bayésiens aussi bien en termes de représentation que d'utilisation les rend difficiles à manipuler à partir d'une certaine taille.

II.3.4. Machine à vecteurs de support

Les machines à vecteurs de support ou « séparateurs à vaste marge », en anglais « support vector machines, SVM », sont un ensemble de techniques d'apprentissage supervisé destinées à résoudre des problèmes de discrimination et de régression. Les SVM sont une généralisation des classifieurs linéaires.

Les SVMs ont été développés dans les années 1990 à partir des considérations théoriques de Vladimir Vapnik sur le développement d'une théorie statistique de l'apprentissage : la théorie de Vapnik-Chervonenkis. Ils ont rapidement été adoptés pour leur capacité à travailler avec des données de grandes dimensions, le faible nombre d'hyper-paramètres, leurs garanties théoriques, et leurs bons résultats en pratique.

Les SVM ont été appliqués à de très nombreux domaines : bioinformatique, recherche d'information, vision par ordinateur, finance...etc. [22]. Selon les données, la performance des machines à vecteurs de support est de même ordre, ou même supérieure, à celle d'un réseau de neurones ou d'un modèle de mélanges gaussiens.

II.3.4.1. Historique

Les séparateurs à vastes marges reposent sur deux idées clés : la notion de marge maximale et la notion de fonction noyau. Ces deux notions existaient depuis plusieurs années avant qu'elles ne soient mises en commun pour construire les SVM.

L'idée des hyperplans à marge maximale a été explorée dès 1963 par Vladimir Vapnik et A. Lerner[23], et en 1973 par Richard Duda et Peter Hart dans leur livre Pattern Classification[24]. Les fondations théoriques des SVM ont été explorées par Vapnik et ses collègues dans les années 70 avec

le développement de la théorie de Vapnik-Chervonenkis, et par Valiant et la théorie de l'apprentissage PAC[25].

L'idée des fonctions noyaux n'est pas non plus nouvelle : le théorème de Mercer date de 1909[26], et l'utilité des fonctions noyaux dans le contexte de l'apprentissage artificiel a été montrée dès 1964 par Aizermann, Bravermann et Rozenner[27].

Ce n'est toutefois qu'en 1992 que ces idées seront bien comprises et rassemblées par Boser, Guyon et Vapnik dans un article, qui est l'article fondateur des séparateurs à vaste marge[28].

L'idée des variables ressorts, qui permet de résoudre certaines limitations pratiques importantes, ne sera introduite qu'en 1995. À partir de cette date, qui correspond à la publication du livre de Vapnik[29], les SVM gagnent en popularité et sont utilisés dans de nombreuses applications.

Un brevet américain sur les SVM est déposé en 1997 par les inventeurs originels[23].

II.3.4.2. Principe de base des SVM

Les SVM peuvent être utilisés pour résoudre des problèmes de discrimination, c'est-à-dire décider à quelle classe appartient un échantillon, ou de régression, c'est-à-dire prédire la valeur numérique d'une variable. La résolution de ces deux problèmes passe par la construction d'une fonction h qui à un vecteur d'entrée x fait correspondre une sortie y :

$$y = h(x) \dots \dots \dots \text{II.5}$$

On se limite pour l'instant à un problème de discrimination à deux classes (binaire), c'est-à-dire $Y \in \{-1, 1\}$, le vecteur d'entrée x étant dans un espace X muni d'un produit scalaire. On peut prendre par exemple :

$$X = \mathbf{R}^N \dots \dots \dots \text{II.6}$$

II.3.4.3. Avantages et Inconvénients :

Les SVM ont comme avantages :

- Absence d'optimum local.
- Contrôle explicite du compromis entre la complexité du classifieur et l'erreur.
- Possibilité d'utilisation de structure de données comme les chaînes de caractère et arbres comme des entrées.
- Traitement des données à grandes dimensions.

Comme ils ont des inconvénients :

- Demande des données négatives et positives en même temps.
- Besoin d'une bonne fonction Kernel.
- Problèmes de stabilité des calculs dans la résolution de certain programme quadratique à contraintes.

II.3.5. Les cartes de Kohonen

Les cartes auto-adaptatives, cartes auto-organisatrices ou cartes topologiques forment une classe de réseau de neurones artificiels fondée sur des méthodes d'apprentissage non-supervisées.

Elles sont souvent désignées par le terme anglais self organizing maps (SOM), ou encore cartes de Kohonen du nom du statisticien ayant développé le concept en 1984. La littérature utilise aussi les dénominations : « réseau de Kohonen », « réseau auto-adaptatif » ou « réseau auto-organisé ».

Elles sont utilisées pour cartographier un espace réel, c'est-à-dire pour étudier la répartition de données dans un espace à grande dimension. En pratique, cette cartographie peut servir à réaliser des tâches de discrétisation, quantification vectorielle ou classification.

II.3.5.1. Principe des cartes auto-adaptatives de Kohonen

Les cartes auto-adaptatives de Kohonen sont des réseaux neuro-mimétiques qui s'inspirent des modèles biologiques de perception. Le réseau simulé pour la classification des battements cardiaque est une SOM bidimensionnelle composée de $N \times N$ neurones artificiels.

- Le vecteur d'entrée $X = [x_1, x_2, \dots, x_{12}]$ représente les 12 paramètres caractérisant un battement cardiaque, et il est lié à tous les neurones.
- Les connexions extérieures du neurone i avec le vecteur d'entrée X se matérialise par un vecteur poids synaptiques $M_i = [m_{i1}, m_{i2}, \dots, m_{i12}]$.
- D'autre part, chaque neurone i de la carte est lié à tous les autres neurones de la carte : interaction entre neurones. Les poids des connexions internes w_{ik} du neurone i avec ses voisins sont donnés par la fonction dite du « chapeau mexicain ». Ces connexions dépendent de la distance entre neurones et sont invariants dans le temps.
- La sortie S_i du neurone i à l'instant t s'exprime par :

$$S_i(t) = \sigma[\sum_j m_{ij} X_j + \sum_k w_{ik} S_k(t-1)] \dots \dots \dots \text{II.7}$$

La sortie S_i à l'instant t du neurone i est donc fonction de deux termes, une entrée totale $(\sum_j m_{ij} X_j)$ et une entrée interne résultante d'une somme pondérée (par l'importance des interactions latérales) des signaux émis aux sorties des neurones voisins à l'instant $(t-1)$ $(\sum_k w_{ik} S_k(t-1))$. La somme globale est évaluée par la fonction d'activation sigmoïde σ [30].

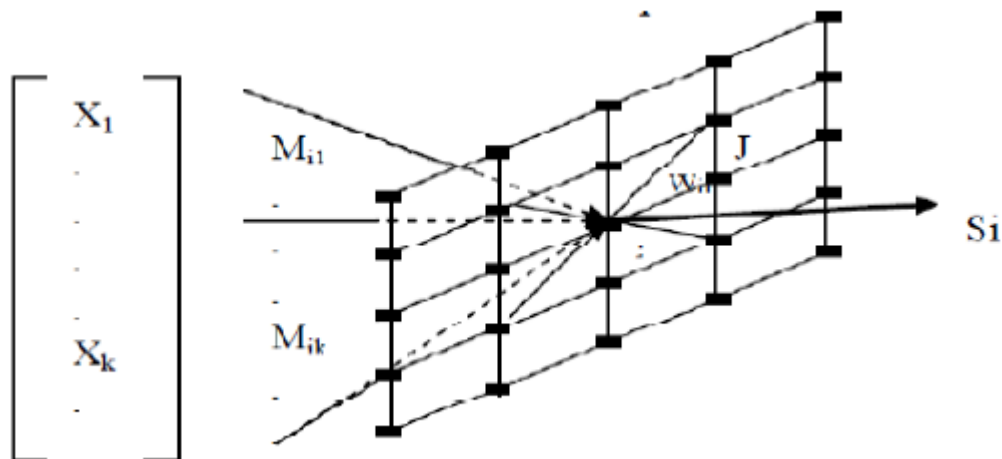


Figure II.5: La carte auto-adaptive de Kohonen 20*20.

Ils ont commencé tout d'abord par localiser les ondes QRS, P, et T et par calculer les paramètres temporelles et morphologiques qui caractérisent un battement en utilisant principalement la technique des ondelettes.

Ensuite, ils ont implémenté une carte auto-organisatrice de Kohonen qui est chargé de déterminer le type du battement en fonction de ses caractéristiques. Le système a été validé sur des enregistrements extraits de la base de données MIT-BIH.

II.3.5.2. Avantages et Inconvénients

Les ancêtres des cartes de Kohonen, les algorithmes comme "moyennes", réalisent la discrétisation de l'espace d'entrée en ne modifiant à chaque cycle d'adaptation qu'un seul vecteur référent. Leur processus d'apprentissage est donc très long. L'algorithme de Kohonen profite des relations de voisinage dans la grille pour réaliser une discrétisation dans un temps très court. On suppose que l'espace n'est pas constitué de zones isolées, mais de sous-ensembles compacts. Donc en déplaçant un vecteur référent vers une zone, on peut se dire qu'il y a probablement d'autres zones dans la même direction qui doivent être représentées par des vecteurs référents. Cela justifie le fait de déplacer les neurones proches du vainqueur dans la grille dans cette même direction, avec une amplitude de déplacement moins importante. L'algorithme présente des opérations simples ; il est donc très léger en termes de coût de calculs. Par contre, le voisinage dans les cartes auto adaptatives est malheureusement fixe, et une liaison entre neurones ne peut être cassée même pour mieux représenter des données discontinues.

II.4. Conclusion :

Grâce à ce chapitre, nous avons expliqué le fonctionnement de la classification dans l'intelligence artificielle, son rôle important dans le domaine médical.

Après une introduction sur la classification en intelligence artificielle, son rôle, son champ d'application et ses différents types.

On a passé par la suite, à la classification des objets, on a détaillé ces principaux modules, comme on a détaillé les types d'apprentissage et la formalisation mathématique d'un problème de classification.

En fin, nous avons cité les différentes méthodes de classification existantes (la logique floue, les arbres de décision, les réseaux bayésiens, les SVM et les cartes de Kohonen...), chacune est accompagnée avec un exemple pour montrer ces avantages et ces inconvénients, afin de juger le choix des réseaux de neurones artificiels pour traiter notre problème.

Le chapitre suivant est responsable de montrer une description des réseaux de neurones.

Chapitre III

Les réseaux de neurones

III.1. Introduction

Les réseaux de neurones sont des composantes importantes du domaine de l'intelligence artificielle. Un réseau de neurones artificiels, est un ensemble d'algorithmes dont la conception est à l'origine très schématiquement inspirée du fonctionnement des neurones biologiques, et qui par la suite s'est rapproché des méthodes statistiques.

Les réseaux de neurones sont généralement optimisés par des méthodes d'apprentissage de type probabilité, en particulier bayésien. Ils sont placés d'une part dans la famille des applications statistiques, qu'ils enrichissent avec un ensemble de paradigmes permettent de créer des classifications rapides, et d'autre part dans la famille des méthodes de l'intelligence artificielles auxquelles ils fournissent un mécanisme perceptif indépendant des idées propres de l'implémenteur, et fournissent des informations d'entrée au raisonnement logique formel.

En modélisation des circuits biologiques, ils permettent de tester quelques hypothèses fonctionnelles issues de la neurophysiologie, ou encore les conséquences de ces hypothèses pour les comparer au réel.

Ce chapitre est consacré pour une explication détaillée des réseaux de neurones artificiels :

Tout d'abord, on va parler sur leur historique, les deux types du neurone (biologique et artificiel), l'architecture des RNAs, leur apprentissage et leur conception et mise en œuvre.

Puis, on va discuter sur les réseaux de neurones multicouches.

En fin, on va détailler leurs avantages et Inconvénients.

III.2. Historique sur les réseaux de neurones

Les premiers travaux sur les réseaux de neurones ont été menés par Mc Culloch et Pitts en 1943. Leur but était de comprendre les propriétés des systèmes nerveux à partir de composants élémentaires.

Grâce à des modèles à base neurones simplifiés, les neurones formels, ils montrent qu'il est possible de construire des systèmes vérifiant la définition de Turing pour les machines à calculer à usage général et donc capables de calculer des fonctions logiques.

En 1949, Donald Hebb s'attaque au problème de l'apprentissage : la loi de Hebb décrit la manière dont les cellules apprennent à modifier l'intensité des connexions qui les relient. Dans les années 50, des chercheurs ont commencé à développer des modèles software mais aussi hardware pour simuler les fonctions de bas niveaux du système nerveux.

En 1956, a lieu une grande conférence à Darmouth sur le thème de l'intelligence artificielle et de l'apprentissage : elle sera le point de départ de l'âge d'or des R.N. et d'I.A. Peu de temps après, Rosenblatt crée le premier Perceptron d'un réseau de neurones artificiels capable de classer des individus en catégories de façon robuste capable de conserver ses performances face à des données bruitées et à des dégradations de sa structure (pertes de cellules, modifications d'une partie des connexions). De son côté, Widrow propose l'Adaline ou élément linéaire adaptatif : le premier réseau

de neurones destiné à résoudre un problème du monde réel : la suppression d'échos sur une ligne téléphonique.

En 1965, Nilsson publie "Machine Learning" qui donne les fondements mathématiques de l'apprentissage automatique pour la reconnaissance des formes.

Après des débuts prometteurs, les réseaux de neurones subissent la concurrence des techniques de calculs plus classiques et tombent en désuétude suite notamment aux critiques de Minsky et Papert en 1969. Ceux-ci mettent en évidence l'incapacité des réseaux à une couche à calculer un simple XOR et les difficultés théoriques posées par l'apprentissage dans les réseaux multicouches.

Les autres techniques d'intelligence artificielle vont avoir les faveurs des chercheurs pendant dix ans avant de montrer elles-mêmes leurs limites. Les réseaux de neurone vont renaître dans les années 80 avec les travaux de Hopfield en 82 sur les réseaux entièrement interconnectés, McClelland, Rumelhart et Hinton qui introduisent l'algorithme de rétropropagation en 86, Kohonen 82 qui introduit les cartes auto-organisatrices à apprentissage non supervisé.

Actuellement l'utilisation des réseaux de neurones est importante et la recherche continue à mobiliser les énergies. Plusieurs articles ont été publiés sur les apports réciproques des algorithmes génétiques, la logique floue et des R.N[31].

III.3. Neurone biologique

Le neurone biologique est un corps cellulaire composé entre autres de dendrites et d'un axone. Des informations électriques (provenant du système nerveux) arrivent par les dendrites qui sont les entrées du neurone. Un signal électrique (de type impulsion) peut être émis le long de l'axone si les signaux électriques arrivant par les dendrites excitent assez le neurone (c'est-à-dire passage d'un seuil d'excitation). Ainsi les axones des neurones (sorties) sont connectés aux dendrites d'autres neurones par les synapses, ce qui fait de notre cerveau un réseau extrêmement complexe de neurones, compte tenu que nous possédons plusieurs milliards de neurones et que chaque neurone peut avoir plusieurs milliers de dendrites. Les synapses jouent un rôle important car elles sont le lien entre la sortie d'un neurone et l'entrée d'un autre neurone. L'influx nerveux passe par ces synapses, qui peuvent inhiber ou laisser passer le signal nerveux. Si l'amplitude du signal arrivant à la synapse est assez élevée, alors elle laissera passer le signal, sinon il sera inhibé. La synapse est capable de s'adapter pour laisser passer le signal ou empêcher son passage: ce mécanisme d'adaptation est à la base de l'apprentissage de notre cerveau[32] [33].

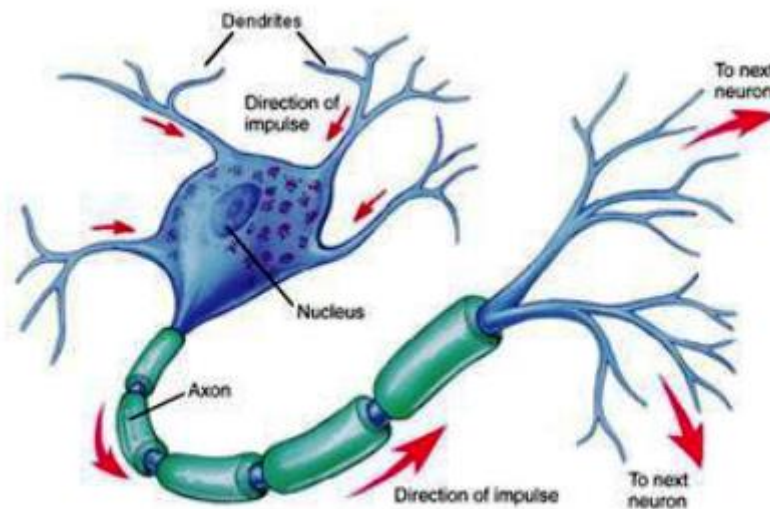


Figure III.1: Neurone biologique.

III.4. Neurone artificiel (formel)

Un neurone formel est une minuscule structure qui traite les influx nerveux qui arrivent (inputs), chacun selon son importance relative, et qui émet un signal de sortie (output). Les neurones artificiels reproduisent le même procédé, recevant chaque signal d'entrée (input) pondéré par un poids (weight). Ces poids sont aussi appelés poids synaptiques, par analogie. Les intrants pondérés, habituellement (mais pas toujours) sommés, sont ensuite comparés à un seuil d'activation θ^1 puis transmis dans la fonction du neurone (habituellement une fonction très simple). Plusieurs possibilités existent pour spécifier la nature de la fonction d'activation du modèle. Les fonctions d'activation les plus souvent utilisées font appel aux fonctions mathématiques telles que les fonctions linéaires, sigmoïdes et gaussiennes.

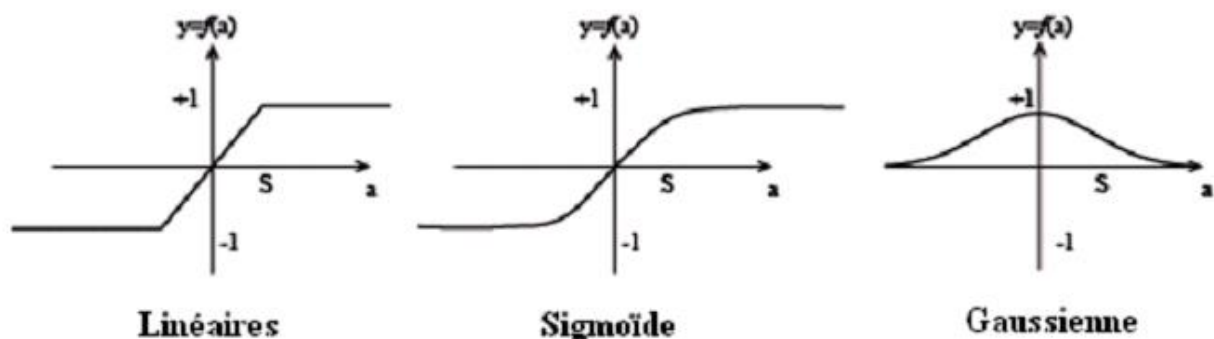


Figure III.2: Fonction d'activation.

Elles produisent l'extrant (output) désiré. Dans les neurones biologiques, les intrants et extrants sont des influx électriques, représentés artificiellement par des valeurs numériques. Habituellement,

les neurones sont reliés ensemble en réseaux, les neurones d'un niveau (les réseaux de neurones sont habituellement bâtis en niveaux hiérarchiques) passant leurs extrants aux neurones du niveau suivant (devenant les intrants des neurones suivants). D'une manière générale, un neurone constitue un minuscule processeur, capable de gérer une fonction simple[33].

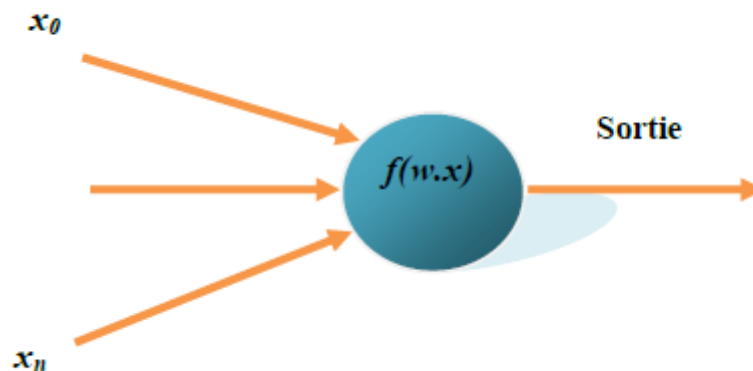


Figure III.3: Neurone formel.

Le tableau (1) illustre l'analogie de celui-ci avec le neurone biologique.

Neurone biologique	Neurone artificiel	Représentation symbolique
Soma	Neurone	
Dendrite	Intrant (<i>Input</i>)	X_i, X (intranet net)
Axone	Extrant (<i>Output</i>)	Y
Synapse	Poids	W_i

Tableau III.1: Analogie entre les neurones biologiques et artificiels.

III.5. Le perceptron

Le perceptron est le premier modèle des réseaux de neurones inventé en 1957 par Frank Rosenblatt[34]. C'est un réseau simple, puisqu'il ne se compose que d'une couche d'entrée et d'une couche de sortie. Ainsi ce perceptron est capable d'apprendre plusieurs fonctions booléennes. On peut par exemple lui faire apprendre la fonction AND ou bien la fonction OR. En fait un perceptron est capable d'apprendre un ensemble de points linéairement séparables. Si l'on se situe dans un cas en deux dimensions, le perceptron pourra apprendre à reconnaître deux types de points différents (valeurs de sortie : 1 et -1) si ces types de points peuvent être séparés dans le plan par une droite.

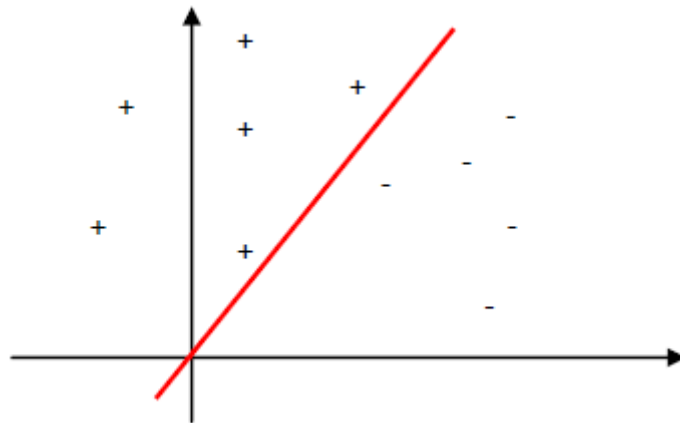


Figure III.4: Des exemples linéairement séparables par une droite.

On peut étendre ce cas en deux dimensions au cas à n dimensions. Le perceptron pourra apprendre dans ce cas un hyper-plan séparant les exemples. Revenons sur le cas de points non linéairement séparables, par exemple le cas du XOR (OU exclusif).

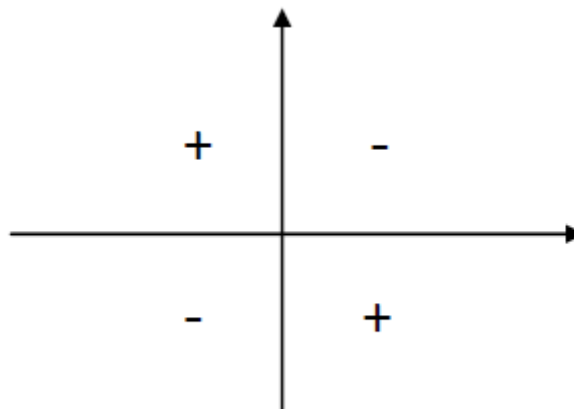


Figure III.5: les points XOR pas des droite séparatrice.

Un perceptron seul est incapable d'apprendre à reconnaître cet ensemble de points, mais il est possible en combinant plusieurs perceptrons d'apprendre un XOR : on dit dans ce cas que l'on a un réseau de perceptrons (MLP). Nous entamons plus loin ce type de réseaux, mais pour avoir une idée de la manière dont on peut représenter un XOR avec plusieurs neurones, on peut voir chaque neurone du réseau comme une porte logique. Avec les portes logiques AND, OR et NOT on peut représenter un XOR, donc en combinant des neurones tels que la sortie des uns soit l'entrée des autres, on peut réaliser un XOR (comme avec des portes logiques).

III.5.1. Apprentissage du perceptron

Faire apprendre un neurone, c'est régler ses poids de manière à ce que la sortie du neurone évolue dans le sens que l'on souhaite sans bien sûr changer les entrées. Il s'agit dans ce cas d'un apprentissage supervisé. La règle d'apprentissage du perceptron va donc faire évoluer les poids du neurone vers une cible que l'on a spécifiée, l'algorithme simple suivant nous montre comment :

```
Initialiser les poids du neurone  $w_i$  aléatoirement
Fixer le coefficient d'apprentissage :  $\eta=0.05$  (coefficient d'apprentissage ou learning rate).
Répéter :
    Pour chaque x dans exemple :
        Calculer la sortie du neurone  $S = f(x)$  (f fonction d'activation)
        Calculer la correction des poids :  $\Delta w_i = \eta * (Cible - s) * x_i$ 
    Fin pour
    Pour chaque poids du réseau  $w_i$  :
        Mettre à jour les poids :  $w_i = w_i + \Delta w_i$ 
    Fin pour
Fin répéter
```

III.6. Architecture des réseaux de neurones

On distingue deux grandes familles de réseaux de neurones[33] :

III.6.1. Les réseaux de neurones non-bouclés

Un réseau de neurones non-bouclé (en anglais feed forward), réalise une ou plusieurs fonctions algébriques de ses entrées par composition des fonctions réalisées par chacun des neurones le composant. L'entrée au réseau désigne l'information reçue par ce dernier de l'environnement extérieur. Dans un tel réseau, l'information circule dans une seule direction, de l'entrée vers la sortie. Ce type de réseaux n'autorise pas le retour en arrière de l'information. Le temps ne joue aucun rôle fonctionnel dans un réseau de neurones non bouclé. Pour cette raison, ces réseaux sont aussi qualifiés de réseaux statiques, par opposition aux réseaux dynamiques. Ce type de réseau comprend deux groupes d'architectures: les réseaux monocouches et les réseaux multicouches. Ils diffèrent par l'existence ou non de neurones intermédiaires appelés neurones cachés entre les unités d'entrées et les unités de sorties appelées nœuds sources ou nœuds d'entrée et nœuds de sortie respectivement. Un réseau non-bouclé monocouche possède une couche d'entrée recevant les entrées à traiter par l'intermédiaire des nœuds sources. Cette couche se projette en une couche de sortie composée de neurones (nœuds de calcul) transmettant les résultats du traitement au milieu extérieur. Un réseau non-

bouclé multicouche se caractérise par la présence d'une ou de plusieurs couches cachées, dont les nœuds de calcul correspondants s'appellent neurones cachés. Les couches cachées s'interposent entre l'entrée du réseau et sa sortie. Leur rôle est d'effectuer un prétraitement des entrées, reçus par la couche d'entrée en provenance du milieu extérieur et de transmettre les résultats correspondant à la couche de sortie où sera déterminée la réponse finale du réseau avant qu'elle soit transmise au milieu extérieur. Les réseaux les plus fréquemment utilisés de cette catégorie sont les perceptrons multicouches (Multi-layer Perceptrons, MLP).

III.6.2. Les réseaux de neurones bouclés

Dans un réseau bouclé, appelé aussi réseau récurrent ou dynamique, il est possible de trouver au moins un cycle où l'information peut revenir à son point de départ. Cela suppose que la sortie d'un neurone peut être fonction d'elle-même. Or cela n'est possible que si la notion de temps est prise en compte. Ce type de réseau est notamment utilisé pour la modélisation dynamique de processus non linéaires.

Cependant qu'un réseau soit non bouclé ou bouclé, dans la phase de son développement, une phase d'apprentissage est nécessaire. Différents types d'apprentissages sont connus dans la littérature.

III.7. Apprentissage des réseaux de neurones

L'apprentissage est une phase du développement du réseau de neurones durant laquelle on calcule les poids des neurones de telle manière que les sorties du réseau soient aussi proches que possible des sorties désirées. Les procédures d'apprentissage sont divisées, elles aussi, en deux grandes catégories [33] [34]:

III.7.1. Les réseaux à apprentissage supervisé

Les réseaux à apprentissage supervisé qui sont généralement destinés à reproduire un processus quelconque (chimique, mécanique, financier...etc.), dont on connaît seulement quelques variables et les résultats correspondants.

L'apprentissage supervisé implique l'existence d'un professeur qui a pour rôle d'évaluer le succès (ou l'échec) du réseau quand il lui est présenté un exemple appartenant à la base d'apprentissage. Cette supervision consiste à renvoyer au réseau une information lui permettant de faire évoluer ses connexions afin de faire diminuer son taux d'échec. L'information peut être explicite, sous la forme d'une mesure de l'erreur commise, par exemple, ou globalement sur l'ensemble des exemples de la base. C'est ce qui se passe dans le cas du Perceptron. Mais elle peut être plus implicite (apprentissage par renforcement), sous forme d'une simple appréciation (bon ou mauvais, punition ou récompense), sans mesure d'erreur, et même être globale, sur l'ensemble des tâches que le réseau doit exécuter. Dans ce dernier cas, il est facile de comprendre que l'apprentissage est le plus difficile, la difficulté

majeure consistant pour le réseau à identifier les étapes du processus qui sont responsables de l'échec ou du succès.

III.7.2. Les réseaux à apprentissage non-supervisé

Les réseaux à apprentissage non-supervisé utilisés par exemple en classification lorsque les classes auxquelles doivent appartenir les données ne sont pas connus a priori. Les données à l'apprentissage sont donc différentes dans les deux cas, puisqu'elles consistent en une série de couples (entrée, sortie désirée correspondante) pour les apprentissages supervisés, tandis que pour les apprentissages non supervisés, les données sont uniquement des entrées.

L'apprentissage non supervisé implique la fourniture à un réseau autonome d'une quantité suffisante d'exemples contenant des répétitions, telles que celui-ci en dégage les régularités automatiquement. Ces réseaux sont souvent appelés auto-organiseurs, ou encore à apprentissage compétitif. Dans l'apprentissage non supervisé, les données ne contiennent pas d'informations sur une sortie désirée, il n'y a pas de superviseur ou expert humain. Il s'agit de déterminer les paramètres du réseau de neurones suivant un critère à définir. Dans ce cas, les exemples présentés à l'entrée provoquent une auto adaptation du réseau afin de produire des valeurs de sortie qui soient proches en réponse pour des valeurs d'entrées similaires. Ce type d'apprentissage possède souvent moins de complexité dans le calcul par rapport à l'apprentissage supervisé. Bien entendu, l'architecture du réseau, préalablement définie par son utilisateur, est une forme de supervision.

III.8. Les différentes applications des réseaux de neurones artificiels

Les réseaux de neurones sont utilisés dans plusieurs domaines, citons la classification, la reconnaissance de formes, l'identification et la commande de processus. Le choix d'utiliser tel ou tel type de réseau de neurones dépend de l'application mais aussi des capacités de traitement du processeur sur lequel ils s'exécutent sans tarder voici quelques domaines [34]:

- Reconnaissance des formes.
- Modélisation.
- Traitement de la parole.
- Détection d'anomalies en médecine (qui est notre axe de recherche).
- Traitements dépendant du temps (prédiction, d'identification et de commande de processus).

III.9. Conception et mise œuvre des réseaux de neurones

Les réseaux de neurones réalisent des fonctions non-linéaires paramétrées. Leur mise en œuvre nécessite :

III.9.1. Détermination des entrées/sorties du réseau de neurones

Pour toute conception de modèle la sélection des entrées doit prendre en compte deux points essentiels :

- ❖ **Premièrement**, la dimension intrinsèque du vecteur des entrées doit être aussi petite le plus que possible, en d'autre terme, la représentation des entrées doit être la plus compacte possible, tout en conservant pour l'essentiel la même quantité d'information, et en gardant à l'esprit que les différentes entrées doivent être indépendantes.
- ❖ **Deuxièmement**, toutes les informations présentées dans les entrées doivent être pertinentes pour la grandeur que l'on cherche à modéliser : elles doivent donc avoir une influence réelle sur la valeur de la sortie.

III.9.2. Choix et préparation des échantillons

Le processus d'élaboration d'un réseau de neurones commence toujours par le choix et la préparation des échantillons de données. La façon dont se présente l'échantillon conditionne le type de réseau, le nombre de cellules d'entrées, le nombre de cellules de sortie et la façon dont il faudra mener l'apprentissage, les tests et la validation. Il faut donc déterminer les grandeurs qui ont une influence significative sur le phénomène que l'on cherche à modéliser[35].

Lorsque la grandeur que l'on veut modéliser dépend de nombreux facteurs, c'est-à-dire lorsque le modèle possède de nombreuses entrées, il n'est pas possible de réaliser un « pavage » régulier dans tout le domaine de variation des entrées : il faut donc trouver une méthode permettant de réaliser uniquement des expériences qui apportent une information significative pour l'apprentissage du modèle. Cet objectif peut être obtenu en mettant en œuvre un plan d'expériences. Pour les modèles linéaires, l'élaboration de plans d'expérience est bien maîtrisée, par ailleurs, ce n'est pas le cas pour les modèles non linéaires.

Afin de développer une application à base de réseaux de neurones, il est nécessaire de disposer de deux bases de données, une pour effectuer l'apprentissage et l'autre pour tester le réseau obtenu et déterminer ses performances.

Notons qu'il n'y a pas de règle pour déterminer ce partage d'une manière quantitative, néanmoins chaque base doit satisfaire aux contraintes de représentativité de chaque classe de données et doit généralement refléter la distribution réelle, c'est-à-dire la probabilité d'occurrence des diverses classes[36].

III.9.3. Elaboration de la structure du réseau

La structure du réseau dépend étroitement du type des échantillons. Il faut d'abord choisir le type de réseau : un perceptron standard, un réseau de Hopfield, un réseau à décalage temporel, un réseau de Kohonen, un perceptron Multicouches...etc.

Par exemple, dans le cas de ce dernier, il faudra aussi bien choisir le nombre de couches cachées que le nombre de neurones dans cette couche.

❖ **Nombre de couches cachées** : Mis à part les couches d'entrée et de sortie, il faut décider du nombre de couches intermédiaires ou cachées. Sans couche cachée, le réseau n'offre que de faibles possibilités d'adaptation. Néanmoins, il a été 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[37].

❖ **Nombre de neurones cachés** : Chaque neurone peut prendre en compte des profils spécifiques de neurones d'entrées. Un nombre plus important permet donc de mieux « coller » aux données présentées mais diminue la capacité de généralisation du réseau. Il faut alors trouver le nombre le nombre adéquat de neurones cachés nécessaire pour obtenir une approximation satisfaisante.

Il n'existe pas, à ce jour, de résultat théorique permettant de prévoir le nombre de neurones cachés nécessaires pour obtenir une performance spécifique du modèle, compte tenu des modèles disponibles. Il faut donc nécessairement mettre en œuvre une procédure numérique de conception de modèle.

III.9.4. Apprentissage

L'apprentissage est un problème numérique d'optimisation. Il consiste à calculer les pondérations optimales des différentes liaisons, en utilisant un échantillon.

La méthode la plus utilisée est la rétro-propagation, qui est généralement plus économe que les autres en termes de nombres d'opérations arithmétiques à effectuer pour évaluer le gradient.

Pour rendre l'optimisation plus performante, on peut utiliser des méthodes de second ordre. Le calcul est très efficace, mais lourd. Elles ont de nombreuses limitations, quant aux conditions de convergence, sur les dérivées secondes. Des corrections sont proposées pour éviter le problème, et sont prises en compte par les méthodes dites de Quasi-Newton ou de Newton modifiée.

III.9.5. Validation et tests

Alors que les tests concernent la vérification des performances d'un réseau de neurones hors échantillon et sa capacité de généralisation, la validation est parfois utilisée hors de l'apprentissage. Une fois le réseau de neurones développé, des tests s'imposent afin de vérifier la qualité des prévisions du modèle neuronal.

Cette dernière étape doit permettre d'estimer la qualité du réseau obtenu en lui présentant des exemples qui ne font pas partie de l'ensemble d'apprentissage. Une validation rigoureuse du modèle développé se traduit par une proportion importante de prédictions exactes sur l'ensemble de la validation. Si les performances du réseau ne sont pas satisfaisantes, il faudra, soit modifier l'architecture du réseau, soit modifier la base d'apprentissage.

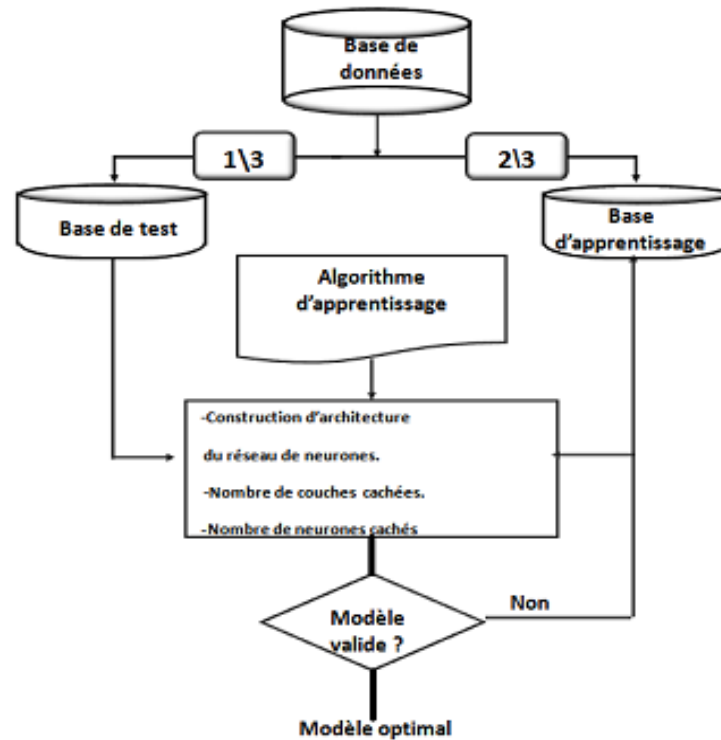


Figure III.6: Organigramme de conception.

III.10. Les réseaux de neurones les plus utilisés

Les réseaux de neurones les plus utilisés dans la littérature sont représentés sur la figure (6) ci-dessous. Cependant, dans ce qui suit, nous allons faire une description détaillée que des MLPs compte tenu que ce sont les réseaux qui vont être appliqués dans ce travail.

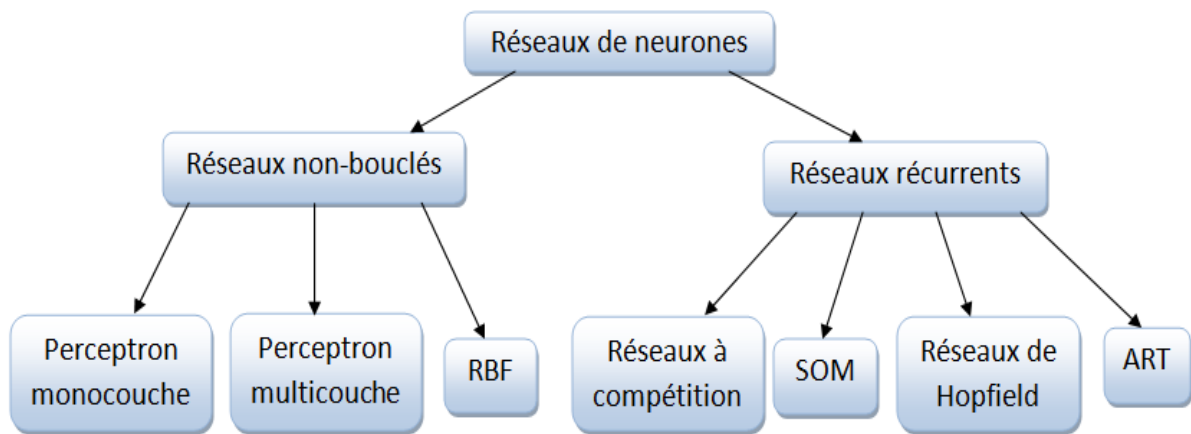


Figure III.7: Schéma synoptique des réseaux de neurones les plus utilisés.

III.11. Les réseaux de neurones multicouches

Un réseau de neurones multicouches MLP (perceptron multicouches PMC), est un réseau de neurones de type non bouclé constitué d'une couche d'entrée, d'une ou de plusieurs couches et d'une couche de sortie. Seuls les neurones de deux couches consécutives sont connectés.

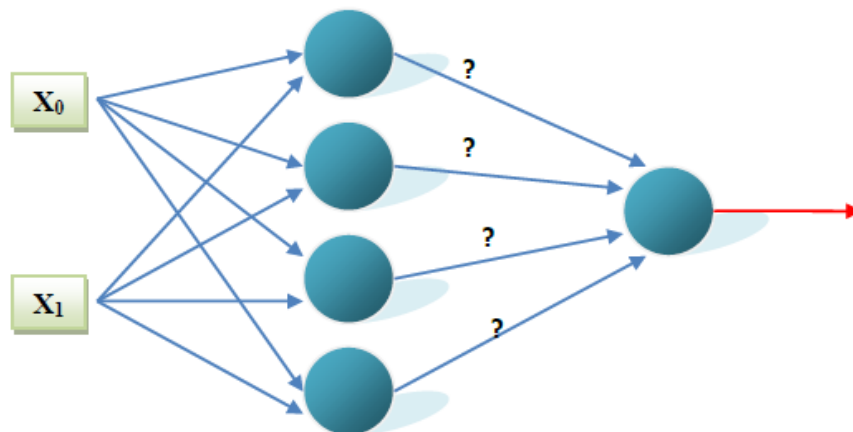


Figure III.8: Réseaux de neurones multicouches avec une couche cachée.

Un réseau de neurones multicouches définit une famille de fonctions. Considérons d'abord le cas le plus classique, celui d'un réseau de neurones à une couche cachée, qui associe à un vecteur de dimension d un vecteur de longueur m (par exemple, pour effectuer une régression) :

$$g(x) = b + W * \tanh(c + V_x) \dots \dots \dots \text{III.1}$$

Où :

- x est vecteur de longueur d (l'entrée).
- V est une matrice k*d (les poids de la couche d'entrée vers la couche cachée).
- c est un vecteur de longueur k (les biais de la couche cachée).
- b est un vecteur est un vecteur de longueur m (les biais de la couche de sortie).
- W est une matrice m*h (les poids de la couche cachée vers la couche de sortie).

La fonction :

$$h(x) = \tanh(x + V_x) \dots \dots \dots \text{III.2}$$

qui retourne un vecteur, est appelée la sortie de la **couche cachée**. Notons que dans le réseau ci-dessus, la sortie est une transformation affine de la couche cachée. Une non-linéarité peut être appliquée par dessus dans certaines architectures. Les éléments de la couche cachée sont appelés *unités cachées*.

Il est possible d'appliquer à nouveau le même genre de transformation à $h(x)$ elle-même, avec des paramètres (biais et poids) différents. Cela donnerait un réseau de neurones multicouches (à propagation avant) avec deux couches cachées. De manière plus générale, on peut construire un réseau de neurones profond en empilant plusieurs de ces couches. Chacune des couches peut avoir une taille différente (le k utilisé ci-dessus). Une variante commune est l'ajout de connexions directes, c'est-à-dire qu'une couche peut prendre comme entrée à la fois la couche précédente et d'autres couches plus basses (plus proches de l'entrée).

III.11.1. L'algorithme de rétro-propagation du gradient

C'est un algorithme qui permet à un réseau de neurones multicouches d'apprendre par mises à jour des poids des différentes couches. Comme on va considérer des réseaux avec plus d'une unité de sortie, on va redéfinir l'erreur E :

$$E(\vec{W}) = \frac{1}{2} \sum_{d \in \text{Exemples}} \sum_{k \in \text{Sorties}} (\text{Cible}_{kd} - \text{Sortie}_{kd})^2 \dots \dots \dots \text{III.3}$$

De la même manière que nous avons considérée E comme une surface d'erreur pour les unités linéaires, nous allons chercher dans le cas de la rétro-propagation à minimiser la surface d'erreur avec notre nouvelle définition. Dans le cas de l'unité linéaire, nous n'avions qu'un seul extremum qui était le minimum global. Avec des unités sigmoïdes, nous allons avoir plusieurs extrema. Malheureusement, on n'est pas sûr que l'algorithme converge vers le minimum global, il peut converger vers un minimum local. En pratique cependant, la rétro-propagation donne de bons résultats.

Nous allons décrire l'algorithme pour un réseau de sigmoïdes à 2 couches, mais il est possible de l'étendre à un nombre quelconque de couches.

```

Créer un réseau à 2 couches de neurones.
Initialiser les poids aléatoirement avec de petites valeurs.
L'entrée de l'unité i dans l'unité j est notée  $x_{ji}$  et les poids de l'unité i à l'unité j sont notés  $w_{ji}$ .
Répéter :
    Pour chaque e dans exemple :
        Calculer la sortie de chaque neurone du réseau.
    Fin pour.
    Pour chaque unité k dans les neurones de sortie :
        Calculer le terme d'erreur  $\sigma_k = Sortie_k * (1 - Sortie_k * (Cible_k - Sortie_k))$ 
    Fin pour
    Pour chaque unité h dans les neurones cachés :
        Calculer le terme d'erreur  $\sigma_h = Sortie_h * (1 - Sortie_h) * \sum_{k \in Sorties} W_{kh} * \sigma_k$ 
    Fin pour
    Pour chaque poids du réseau  $w_{ji}$  :
        Mettre à jour les poids  $W_{ji} = W_{ji} + \eta * \sigma_j * X_{ji}$ 
    Fin pour
Fin répéter.

```

Cet algorithme est spécifique aux réseaux de neurones à 1 couche cachée avec des unités sigmoïdes comme nous les avons décrites précédemment. On peut facilement l'adapter à un nombre quelconque de couches cachées et pour des unités à fonction d'activation différentiables.

On remarque que l'algorithme n'utilise que des données que nous possédons : il calcule en premier toutes les sorties (nous avons les entrées x), puis calcule l'erreur sur la dernière couche, comme nous l'avons fait pour un neurone simple, et seulement après il s'occupe de la couche cachée. Si l'on étendait l'algorithme à plusieurs couches cachées, il faudrait procéder de la même façon et calculer les termes d'erreur en partant de la couche de sortie. Pour les couches cachées intermédiaires, il faudrait se servir des résultats de la couche suivante à chaque fois.

Il existe des améliorations à l'algorithme de rétro-propagation qui permettent d'éviter certains minima locaux ou d'accélérer la convergence[33] [34].

III.12. Avantages et Inconvénients

Les réseaux de neurones artificiels ont comme avantages :

- ❖ **Temps de réponse** : C'est l'un des avantages principaux du réseau de neurones : en effet une fois que le réseau a appris, il peut sortir quasi-instantanément la réponse. En fait, les opérations que fait un réseau de neurones sont très simples du point de vue informatique, et peu gourmandes en CPU.

- ❖ **Large panel de fonctions** : Un réseau de neurones est capable, s'il est doté d'une bonne structure, d'apprendre un très large panel de fonctions, ce qui présente un intérêt dans les domaines où les fonctions cibles sont un peu compliquées. De plus le réseau de neurones ayant une bonne capacité de généralisation, on peut l'utiliser pour résoudre des problèmes réels.
- ❖ **Apprentissage résistant à l'erreur** : Le réseau de neurones a une résistance naturelle aux données bruitées lors de l'apprentissage. En effet, si la base d'exemples est assez grande, une erreur ne faussera pas beaucoup la mise à jour des poids.

Comme ils ont comme inconvénients :

- ❖ **Trouver une bonne structure** : Avant de passer des exemples à un réseau de neurones, il faut trouver une structure permettant au réseau de bien apprendre les exemples. On ne sait malheureusement pas prévoir le nombre de neurones cachés nécessaires pour un problème donné. Il existe pour pallier ce manque des algorithmes qui peuvent apprendre une structure, en ajoutant au fur et à mesure de l'apprentissage des neurones cachés. On peut aussi utiliser des algorithmes génétiques pour obtenir des résultats meilleurs.
- ❖ **Temps d'apprentissage** : Un réseau doit parfois apprendre les exemples plusieurs dizaines de milliers de fois. Si la base d'exemples est énorme (bases de données industrielles par exemple), le temps d'apprentissage risque d'être démesuré.
- ❖ **Structure statique** : Un réseau de neurones a une structure statique : une fois que la structure est fixée (nombre de neurones d'entrée, sortie, neurones cachés) et que le réseau a appris, il est impossible de lui faire apprendre de nouvelles données sans recommencer son apprentissage au début, contrairement aux algorithmes de type paresseux k-plus proche voisins, classificateur naïf de Bayes.
- ❖ **Over-fitting (sur apprentissage)** : Un réseau de neurones qui apprend a généralement une bonne capacité de généralisation : quand on demande à un humain de reconnaître une couleur il peut dire rouge pour une couleur rouge-orangée; un bon réseau de neurones doit pouvoir le faire aussi. Si l'on gère mal l'apprentissage du réseau, par exemple si l'on arrive à trop minimiser l'erreur, le réseau aura appris par cœur les exemples, et si la couleur n'est pas exactement celle qu'il aura apprise, il la classera mal. On appelle ce type de sur-apprentissage de l'over-fitting. Le réseau se borne à ressortir exactement ce qu'on lui a appris sans généraliser.
- ❖ **Réseau boîte noire** : Une fois qu'il a appris, on peut voir un réseau de neurones comme une boîte noire : on lui passe des entrées et il ressort un résultat. C'est parfois suffisant, mais dans certaines applications, on a besoin de savoir ce qu'il se passe dans cette boîte noire, notamment dans des applications de marketing, afin de comprendre les modes de pensée du

client. Parfois, on préférera des techniques d'apprentissage comme les arbres de décision pour pouvoir comprendre les actions du client plutôt que d'utiliser un réseau de neurones qui se bornera à ne sortir qu'un résultat.

III.13. Conclusion

D'après ce chapitre nous avons donné une explication bien détaillée sur les réseaux de neurones artificiels.

Dans un premier lieu, nous avons parlé de leur historique, des deux types du neurone (biologique et artificiel). Nous avons parlé du perceptron et son apprentissage comme nous avons donné un exemple d'algorithme d'apprentissage. Comme nous avons discuté sur l'architecture des RNAs (les RNAs non-bouclés et bouclés), leur apprentissage (supervisé et non-supervisé), leurs différentes applications et leur conception et mise en œuvre.

Dans la troisième partie, nous avons parlé sur les réseaux de neurones multicouches et son algorithme d'apprentissage, comme on a donné un exemple.

Dans la dernière tranche, nous avons détaillé leurs avantages et inconvénients.

Le quatrième chapitre est réservé pour la contribution et la réalisation.

Chapitre IV

Contribution et Réalisation

IV.1. Introduction

La classification est une des techniques les plus utilisées en médecine. Elle constitue le principe de base de plusieurs systèmes d'aide au diagnostic. Il s'agit d'affecter un ensemble d'objets (enregistrements des électrocardiogrammes, encéphalogrammes, etc...) à une classe selon un descripteur de celle-ci. Cette description est effectuée grâce à des propriétés ou des conditions typiques propres aux classes. Les objets sont alors classés suivant qu'ils vérifient ou non ces conditions ou propriétés.

Ce chapitre a pour but de présenter le travail réalisé pour donner une solution de la problématique posée « Approche Neuronale pour la classification Morphologique des signaux ECG ».

Dans la première partie, on va parler sur la classification neuronale des arythmies cardiaques, en donnant un schéma global de la classification.

Puis, dans la deuxième partie, on va expliquer la sélection de la base d'exemples, et la sélection des descripteurs temporaires d'un cycle cardiaque.

Ensuite, dans la troisième partie, on va schématiser l'architecture d'un agent classifieur à base des descripteurs morphologiques.

En fin, dans la dernière partie, on va discuter sur notre application, son interface principale, en donnant des figures explicatives de ses touches.

IV.2. Classification neuronale des arythmies cardiaques

Comme tout modèle de traitement, les classifieurs utilisés en reconnaissance des formes ou d'objets sont amenés à effectuer des tâches complexes (notamment lors de l'apprentissage), à partir d'une base de données réelles.

Pour que le classifieur soit opérationnel en phase d'exploitation, il est primordial que la nature des données et notamment leurs variations soit correctement prises en compte. Le problème se profile à plusieurs niveaux, comportant chacun son lot de difficultés théoriques et pratiques. Il faut disposer premièrement d'un système d'acquisition efficace, qui nous fournira des données de qualité (avec un minimum de bruit). Il faut ensuite pouvoir représenter cette information (possiblement la réduire et la transformer) de manière efficace et compatible avec les étapes de traitement ultérieures.

Cette représentation et ses modalités constituent le modèle avec lequel le mécanisme de reconnaissance devra travailler, afin de produire le résultat désiré.

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

L'analyse de la littérature présentée dans le chapitre des modèles de classification montre que, lorsqu'une approche de reconnaissance des arythmies cardiaques est adoptée, les réseaux de neurones artificiels sont une bonne technique de classification.

Pour notre étude, le réseau de neurones utilisé est optimisé par des algorithmes de type supervisé. La stratégie que nous envisageons en œuvre est présentée dans la figure (IV.1).

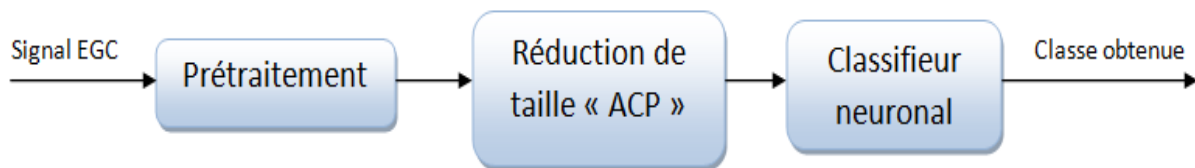


Figure IV.1: Schéma global de la classification.

Les étapes : « Prétraitement », « Réduction de taille (ACP) » et « Classifieur neuronal » font partie de la classification neuronale en basant sur la morphologie du signal ECG en question. Puisque ce dernier contient trop de données (c'est une base de données volumineuse), ce qui nécessite une réduction de taille, donc, il faut avoir une méthode d'analyse numérique qui a les performances nécessaires en termes de minimisation de perte de données et d'optimisation de réduction de taille (compression de données), ACP répond à nos besoins et critères posés.

Notons que vous trouverez les détails nécessaires de la méthode d'Analyse en Composantes Principales « ACP » dans « l'annexe B » dans la fin de ce manuscrite.

Et comme nous l'avons précédemment mentionné dans le chapitre 3, dans le paragraphe 3.9, la mise en œuvre de chaque réseau de neurones nécessite :

- La détermination des entrées/sorties du réseau de neurones.
- Le choix et la préparation des échantillons.
- L'élaboration de la structure du réseau.
- La réalisation et l'apprentissage.
- Validation et Testes.

IV.2.1. Sélection de la base d'exemples

Nous avons utilisé, dans le cadre de ce mémoire, la base de données MIT-BIH. Cette dernière 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 (pour plus d'information, voir annexe B). Les enregistrements comprenant suffisamment de battements normaux ont été choisis pour construire la base d'exemple.

Nous avons rassemblé les battements de chaque enregistrement en deux groupes (classes):

- Le groupe N est composé des battements différents des « Normal ».
- Le groupe A est composé des battements autres que normal.

Il est important de remarquer que tous les enregistrements choisis pour faire l'expérimentation sont des battements normaux, mais ils contiennent des cycles cardiaques anormaux.

Toutes ces informations sont récapitulées dans le tableau suivant :

Enregistrement	Nombre de battements de type « N »	Nombre de battements de type « A »
100	2239	34
101	1860	05
103	2082	02
105	2526	46
112	2537	02
116	2302	110
203	2529	511
205	2571	85
209	2621	384
210	2423	205
213	2641	610
215	3195	168
230	2255	01

Tableau IV.1: Les enregistrements choisis de la base de données MIT-BIH.

La base d'apprentissage affecte la performance des classifieurs neuronaux. Le critère général à respecter est de sélectionner des exemples représentatifs de toutes les classes. Dans notre cas, nous avons préparé deux bases, une d'apprentissage et l'autre de validation ou de test.

Notre expérimentation est composée de 60 cycles de Type N (battement normal) et 36 cycles de Type «A» (mélange entre les BVP, BBDR, BBG), qui vaut 96 cycles pour chaque base choisis parmi les 13 enregistrements sélectionnés.

IV.2.1.1. Sélection des descripteurs temporaires d'un cycle cardiaque

Pour obtenir une meilleure performance d'un classifieur, il vous faut choisir des données pertinentes permettant de décrire de manière assez complète l'échantillon à classer. Ces données sont appelées caractéristiques ou descripteurs.

Dans notre cas, en plus le complexe QRS, des descripteurs permettant de caractériser un battement cardiaque sont retenus. Ils sont comparables aux paramètres sur lesquels sont construites les connaissances médicales.

- ❖ **Intervalle RRp** : Nous appelons RRp la distance entre le pic R du présent battement et le pic R du battement précédent. Nous avons divisé la valeur de la distance RRp par la moyenne des distances des huit intervalles RR précédents. Ce rapport est au voisinage de la valeur 1 pour un battement Normal et un peu plus petit que 1 pour un battement de type « A ».
- ❖ **Rapport des intervalles RR** : Nous appelons RRs la distance entre le pic R du présent battement et le pic R du battement suivant. 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 « N », ce rapport est voisin de 1, mais il peut largement dépasser cette valeur dans le cas « A ».
- ❖ **Largeur du complexe QRS** : Ce paramètre est important pour l'identification des battements cardiaques, la mesure de ce paramètre a été réalisée en utilisant un algorithme implémenté sous matlab, au laboratoire L.T.S.I à l'université de Rennes 1.
- ❖ **Distance PP** : C'est la distance entre deux points minimum et maximum du QRS. Ce paramètre a été aussi retenu comme valeur informationnelle potentiellement pertinente pour chaque cycle cardiaque, sachant que cette distance varie énormément avec la déformation du complexe QRS.

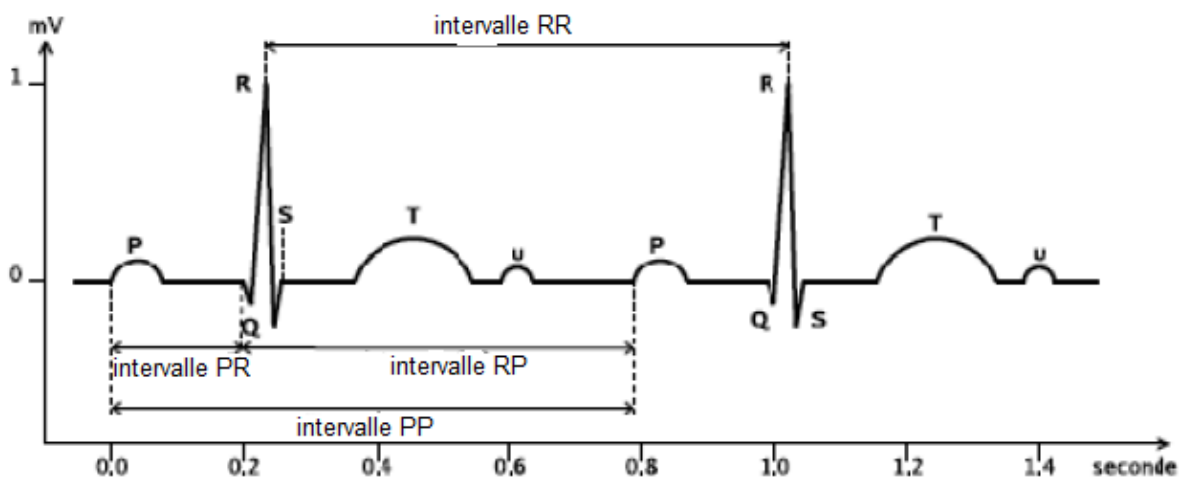


Figure IV.2: Battements normaux et intervalles d'intérêt.

IV.2.2. Architecture du réseau de neurones

IV.2.2.1. Algorithme d'apprentissage

On a développé nos réseaux de neurones (Par des descripteurs morphologiques et d'autres temporels) sous Matlab version R2010a. Au cours de ce travail nous avons utilisé principalement trois algorithmes très classiques en optimisation non linéaire.

- Quasi-Newton.
- Levenberg-Marquardt.
- Méthode du gradient à pas.

En les confrontant sur plusieurs problèmes, il apparait qu'aucune de ces trois méthodes ne prend un avantage considérable sur l'autre. Chaque méthode d'optimisation possède des avantages et des inconvénients. La méthode de Levenberg-Marquardt présente un intérêt pratique car elle peut être utilisée sans avoir à choisir le pas.

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[14].

IV.2.2.2. Dimensionnement du réseau pour chaque classifieur

La structure neuronale est déterminée par le nombre de neurones de la couche d'entrée, N_e et celui de la couche cachée, N_c , tous les deux sont variables.

- ❖ **Couche d'entrée** : le nombre de neurones sur cette couche est lié directement au vecteur de caractéristiques :
 - Le classifieur temporaire, le nombre de neurones dans la couche d'entrée est 96 neurones, puisque, nous avons choisi des descripteurs donnant un vecteur de taille est égale à ce nombre (96).
 - Le classifieur morphologique, le nombre de neurones dans la couche d'entrée est lié fortement au nombre de composants principales retenues et qui contribuent à plus de 99% à l'inertie totale du nuage. Le but, est d'avoir une perte de données minimale (moins de 01%).
- ❖ **Couche cachée** : le choix de la taille de la couche cachée est réalisée souvent de façon heuristique et expérimentale, car aucune formule théorique ne peut nous aider à l'optimiser. Donc, nous avons mené une série d'essais pour arrêter le nombre N_c à mettre sur la couche cachée afin d'avoir un meilleur résultat possible. On a constaté que plus le nombre de neurones cachés est important, plus le nombre d'exemples bien classés est grand, car le réseau, grâce au nombre de paramètres croissant, se spécialise sur la base d'apprentissage. Les fonctions d'activation des neurones cachés sont des sigmoïdes. Le nombre d'itérations, fixé dans ces deux

essais à 30000, a permis de réaliser un apprentissage satisfaisant. Par ailleurs, il est clair que ce dernier 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.

IV.2.2.2.1. Réduction par l'ACP

Nous avons appliqué la méthode d'analyse en composantes principales sur la base d'apprentissage et sur des enregistrements de la base de données MIT-BIH., pour voir la variabilité des données par rapports au nombre de composantes principales. En effet, comme nous pouvons le constater dans la figure IV.3, (La base d'apprentissage), la première composante n'exprime pas plus de 45% de la variabilité totale des données, alors que les 15 premières composantes expriment 99% de la variabilité.

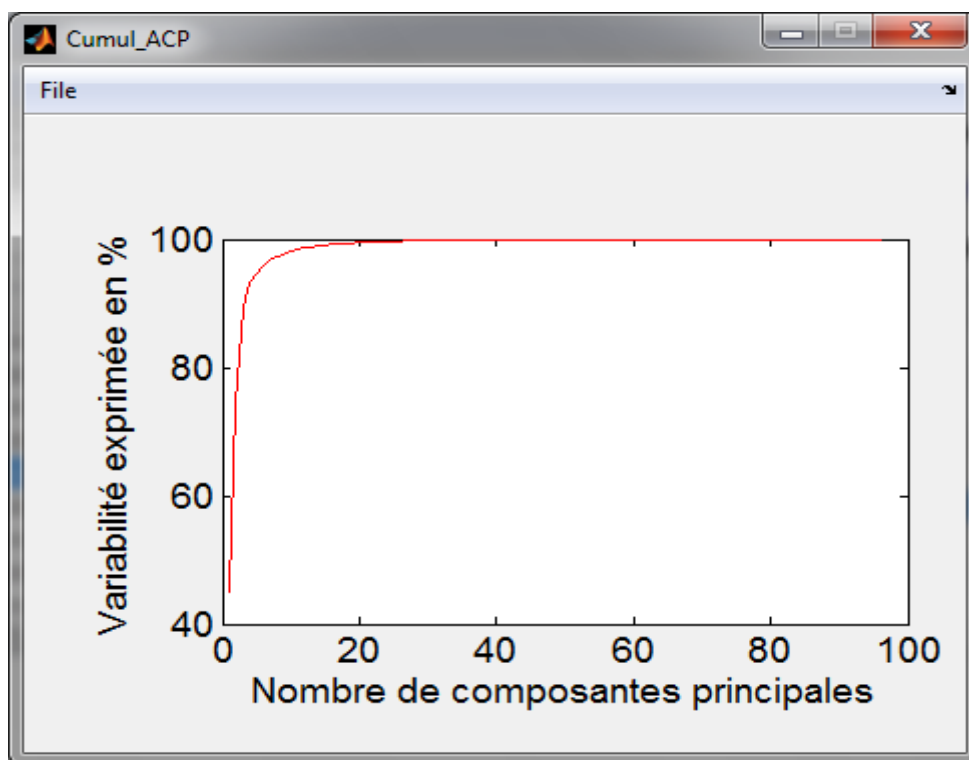


Figure IV.3: Variabilité des données de la base d'apprentissage (l'ACP).

Tandis que pour l'enregistrement 209, la première composante principale n'exprime pas plus de 44% et les 08 premières composantes représentent presque 100% de la variabilité.

Donc, nous remarquons clairement que pour chaque enregistrement, sa morphologie peut être représentée par uniquement une dizaine de paramètres représentatifs. Cette réduction de données permet d'avoir une architecture optimisée des classifieurs d'arythmies.

Le tableau IV.2 récapitule le nombre de composantes principales présentant la variabilité des données, pour chaque enregistrement, toujours en respectant, la contrainte de moins 01% de perte de données.

Échantillon testé	Nbre de composantes principales Retenues
Base d'apprentissage	15
ECG 100	16
ECG 101	09
ECG 103	08
ECG 105	13
ECG 112	10
ECG 116	08
ECG 122	16
ECG 205	08
ECG 209	19
ECG 210	18
ECG 213	08
ECG 220	08
ECG 230	09
	Max→19

Tableau IV.2: Nombre de composantes principales retenues pour chaque ECG.

IV.1.2 Résultats expérimentaux des classifications

Dans ce qui suit, on présente de façon résumée les résultats des deux classifieurs obtenus. Les performances sont représentées respectivement par le taux de classification T_{class} , la sensibilité S_e et la spécificité S_p déjà définis dans le chapitre des modèles de classification.

Le tableau suivant (Tab IV.3) récapitule les résultats des classifications faites sur une base d'apprentissages, c'est une base contenant les descripteurs morphologiques (des vecteurs propres représentant la base initiale suivant le degré de perte de données).

Enregistrement	T_{class}	S_e	S_p
100	91.24	95.10	92.66
101	93.15	97.21	99.12
103	97.17	93.59	95.25
105	91.20	86.33	90.34
112	85.3	92.30	94.41
116	79.58	91.05	93.76
203	93.41	92.19	95.05

205	99.40	96.22	89.14
209	93.06	96.29	91.90
210	93.42	89.45	93.12
213	96.64	88.17	75.67
215	87.17	94.55	89.33
230	90.81	97.73	95.72
Moyenne	91.72	92.70	91.64

Tableau IV.3: Performances du classifieur neuronal à données morphologiques réduites.

IV.3. Architecture d'un classifieur à base des descripteurs morphologiques

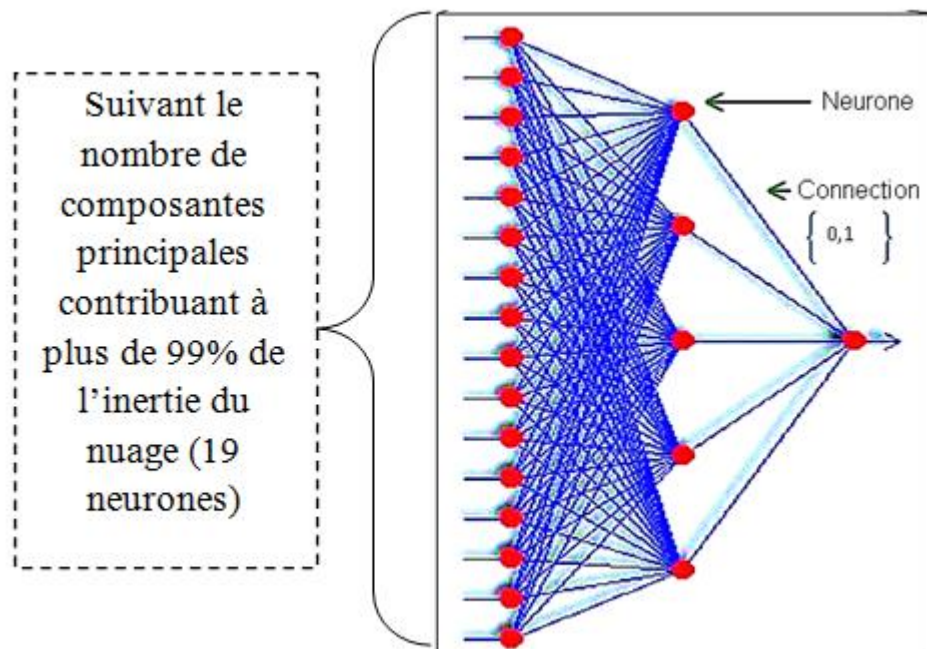


Figure IV.4: Architecture du classifieur à base des descripteurs morphologiques.

La figure IV.4, illustre l'architecture minimale du classifieur à base des caractéristiques morphologiques, il reçoit comme entrée un vecteur (19 entrées) contenant les composantes principales représentant 99% de l'inertie totale pour chaque enregistrement. Il fait son apprentissage neuronal supervisé, qui donne comme résultat le type de classe (N qui vaut ZERO pour normal ou A qui vaut UN pour autres) à laquelle il appartient cet ECG.

IV.4. Présentation de l'application

Après la compilation du programme, nous devons choisir un signal parmi les 13 cas étudiés pour visualiser l'ECG de ce dernier :

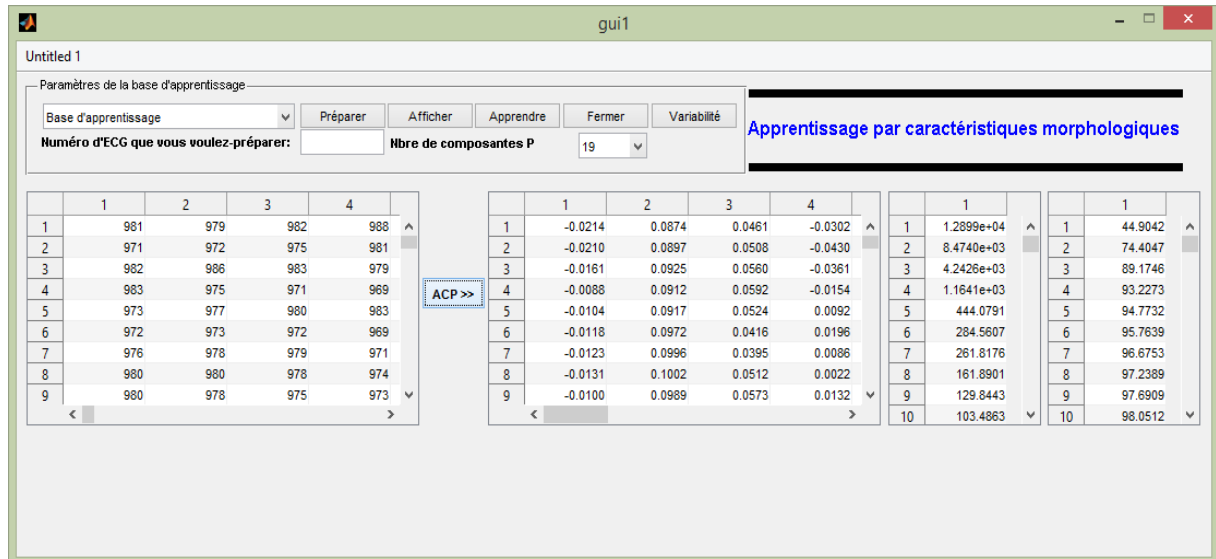


Figure IV.5: Interface principale du classifieur morphologique.

Lorsqu'on prend ce signal, on doit préciser leurs paramètres en appuyant sur la touche (Apprendre) :

- Le nom d'apprentissage.
- Nbr de couches cachées.
- Nbr de neurones de sortie.
- Fonction de transfert.
- Fonction d'apprentissage.
- Nbr d'itérations.
- Erreurs.

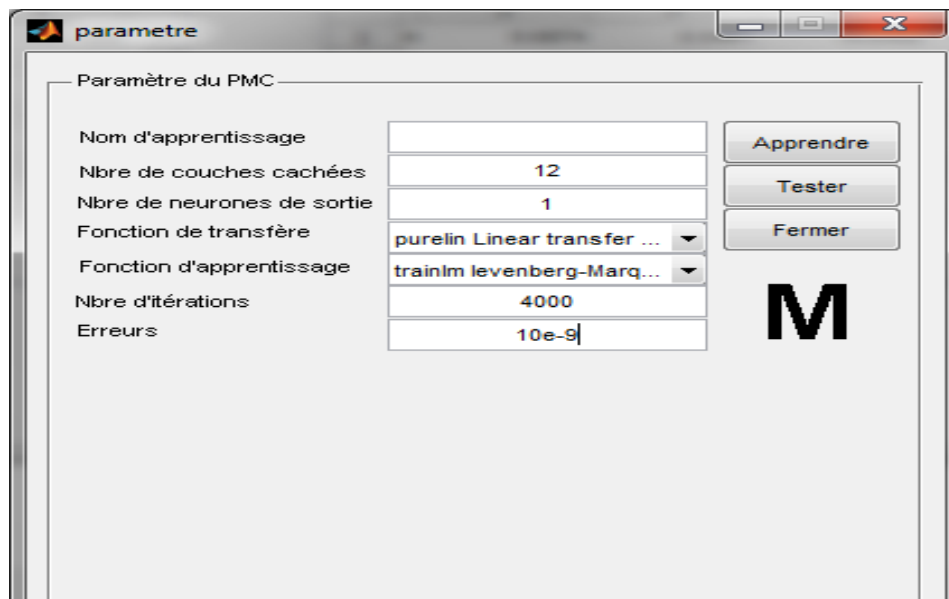


Figure IV.6: Paramétrage du réseau de neurone.

En appuyant sur la touche (Apprendre), on obtient la fonction au réseau neuronal (Neuronal Network Training).

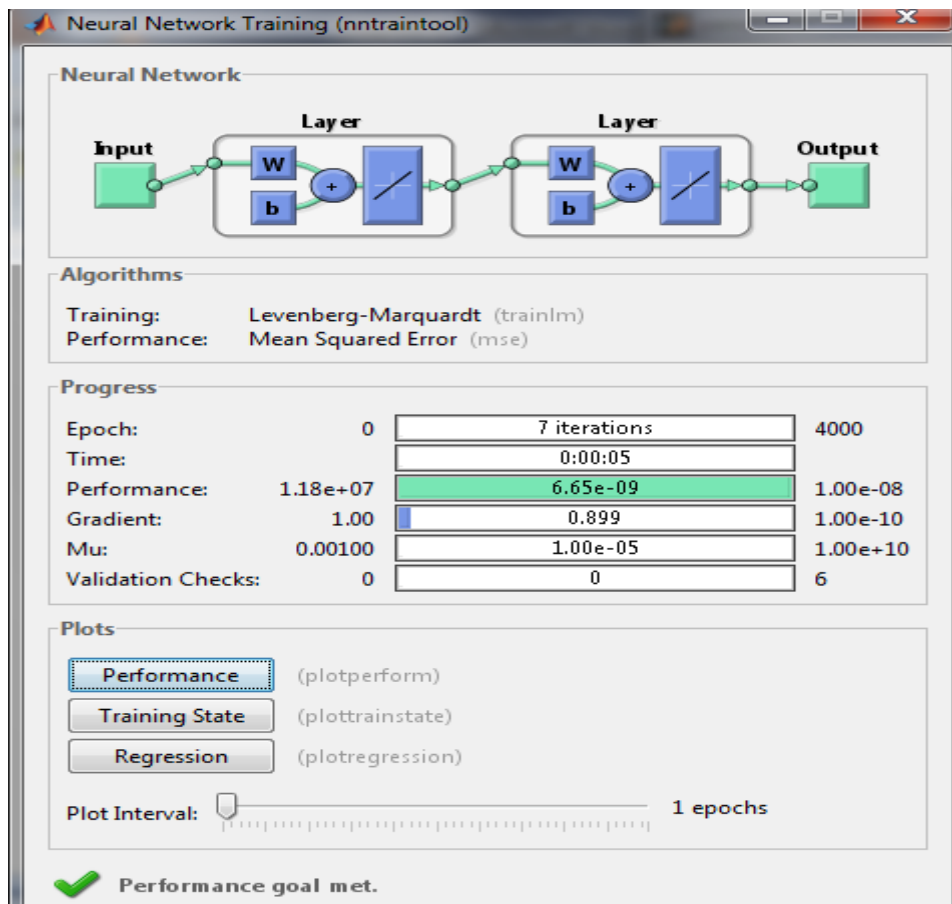


Figure IV.7: Interface d'apprentissage du réseau de neurone.

Une fois que l'opération s'est terminée, si on appui sur la touche (Performance), on obtient la figure(9), on remarque une convergence de la performance d'apprentissage du réseau pour ce signal vers l'erreur précisée du paramétrage en fonction du nombre d'itérations fixé.

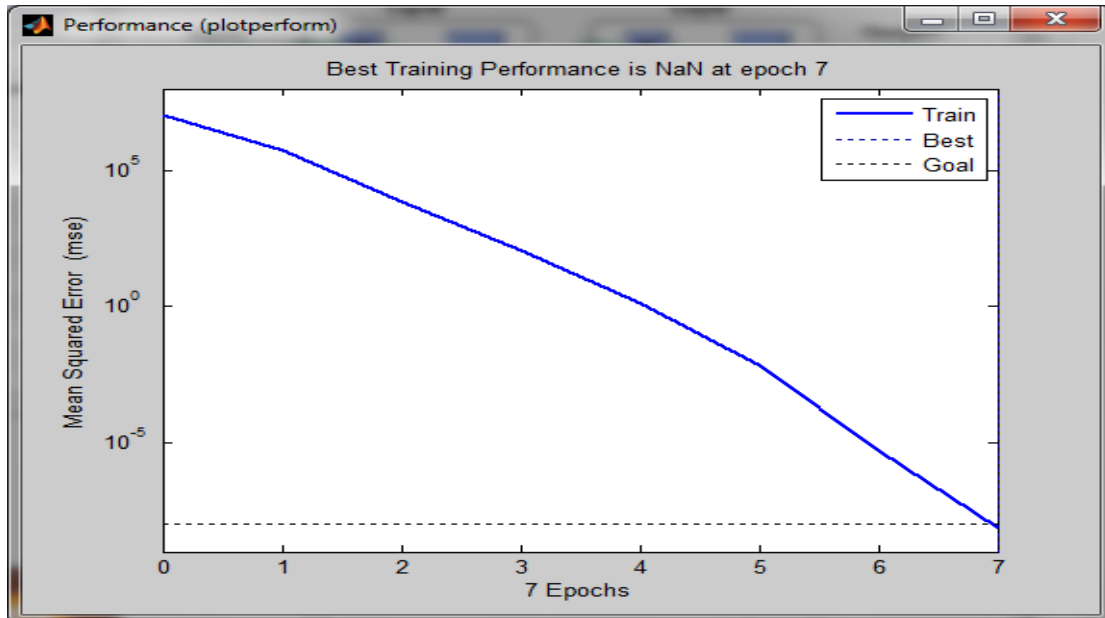


Figure IV.8: Graphe de performance.

Si on appui sur la touche (Training State), on obtient la figure (10), elle représente les graphes du gradient et la moyenne d'erreurs et la validation du test par rapport au nombre d'itérations.

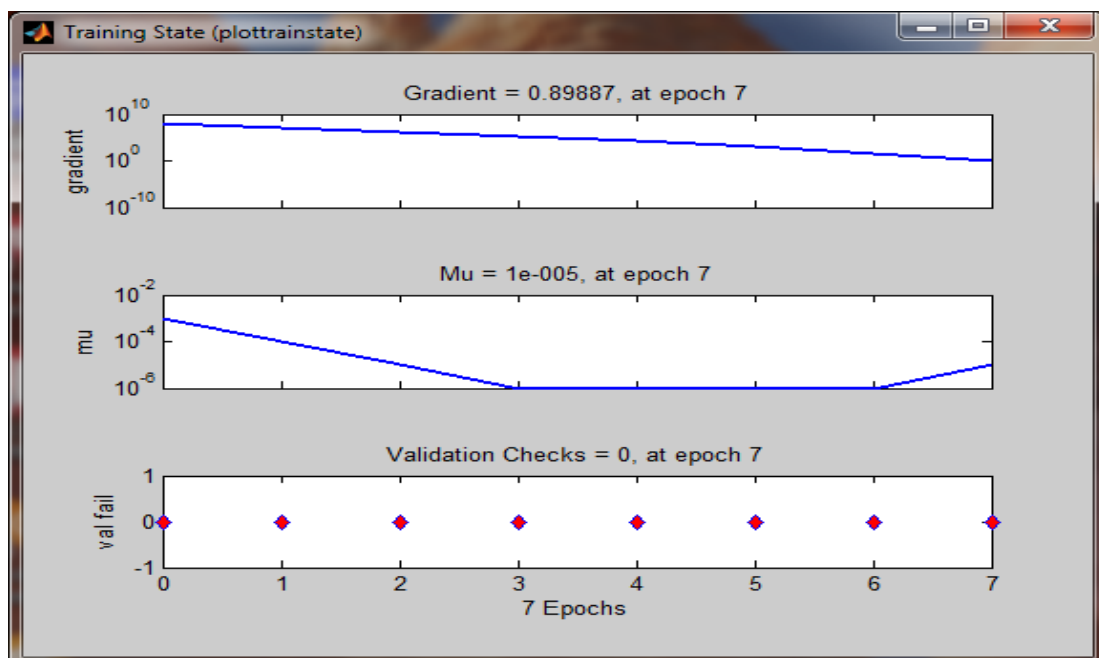


Figure IV.9: Les graphes du gradient, moyenne d'erreurs et validation du test.

La touche (Afficher) dans l'interface classifieur morphologique, permet à ouvrir un signal au choix avec un nombre d'échantillons bien précis pour mesurer les paramètres de l'ECG.

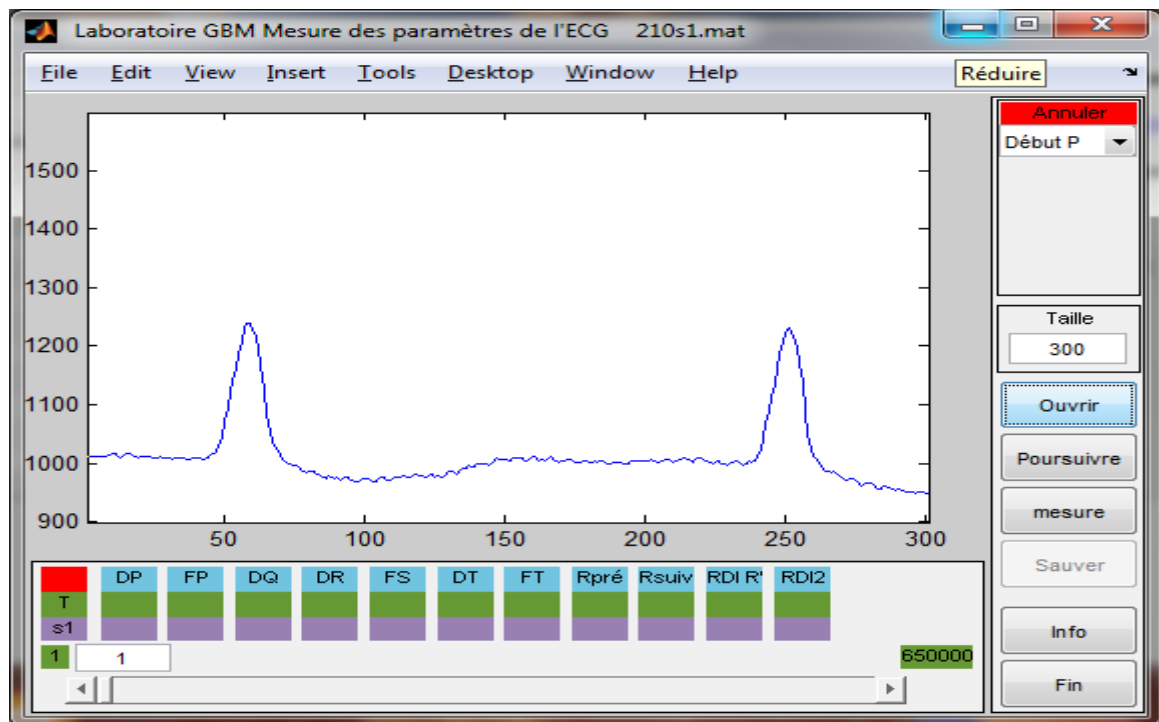


Figure IV.10: Mesure des paramètres de l'ECG du signal 210s1.

Et la touche (Variabilité), permet à afficher la variabilité des données de la base d'apprentissage (l'ACP) comme se présente dans la figure (3).

IV.5. Conclusion

L'objectif de ce chapitre c'était de montrer une solution de la problématique posée « Approche Neuronale pour la Classification Morphologique des Signaux ECG ». C'est une technique pour donner un bon diagnostic médical et dans les meilleurs délais, afin de protéger la santé des humains. Alors, nous avons proposé une méthode de classification à base des descripteurs morphologiques extraites depuis le signal ECG. Cette dernière nécessite un traitement de plus, à savoir la réduction de la taille de l'échantillon étudiée par l'une des méthodes d'analyse de données (ACP).

Cette classification est faite par une méthode d'intelligence artificielle, à l'occurrence les réseaux de neurones, qui ont montré leurs performances dans ce domaine, et surtout où se trouvent des données de type numérique et de grande taille (base de données géantes), ils ont la capacité de donner des bons résultats et dans les meilleurs délais, car le temps joue un rôle primordiale dans le diagnostic médical et l'aide de décision dans ce domaine.

Conclusion Générale

Conclusion générale

Le travail réalisé dans ce mémoire concerne l'application d'une approche neuronale en utilisant des outils de l'IA pour la classification des battements cardiaques, à l'occurrence, les réseaux de neurones artificiels qui ont prouvé leurs performances dans ce domaine (la classification des arythmies cardiaques).

Nous nous sommes intéressés à la classification des battements cardiaques, donc nous avons proposé de créer un modèle de classification, dont ce dernier reçoit des valeurs descriptives morphologiques de l'ECG en question après avoir réduit la taille de l'inertie en entrée par l'intermédiaire de la méthode d'analyse en composantes principales, dont on choisit seulement les composantes principales offrant la moindre perte de données (01%).

Cette classification nous donne une seule sortie qui est soit « N » pour normale, soit « A » pour autres.

Cette approche nous a permis de profiter des avantages des réseaux de neurones artificiels pour faire cette classification.

Cette petite recherche sera, dans le proche futur, améliorée afin de donner des solutions efficaces, opérationnelles et moins coûteuses permettant d'identifier les patients à risque de mort subite rythmique et de prendre des mesures préventives selon le cas rencontré et donc des millions de décès pourraient être évités.

Annexes

Annexe A

Analyse en Composantes Principales (ACP)

a. Introduction

L'analyse en composantes principales (ACP) est une technique descriptive permettant d'étudier les relations qui existent entre les variables, sans tenir en compte, a priori, d'une quelconque structure [38],[39]; son utilisation pour l'exploitation des données remonte au début du siècle dernier. Elle est principalement issue des travaux de psychologues américains [40],[41],[42]. Le but de l'ACP est d'identifier la structure de dépendance entre des observations multivariées, afin d'obtenir une description ou une représentation compacte de ces dernières. L'idée de base de l'ACP est de réduire la dimension de la matrice des données, en retenant le plus possible les variables présentes dans le jeu des données de départ. Cette réduction ne sera possible que si les variables initiales ne sont pas indépendantes et ont des coefficients de corrélation entre elles non nuls. Ces variables initiales sont transformées en de nouvelles variables, appelées composantes principales. Elles sont obtenues par combinaisons linéaires des précédentes et sont ordonnées et non corrélées entre elles. L'analyse en composantes principales cherche à identifier les vecteurs propres orthonormaux et les valeurs propres de dispersion des variables originelles. Les vecteurs propres orthonormaux sont utilisés pour construire les composantes principales et les valeurs propres sont les variances des composantes principales correspondantes [38].

b. Principe de l'analyse en composantes principales

Dans le cadre cette thèse, nous aborderons l'ACP comme une technique de réduction des échantillons de l'ECG dans le cas des descripteurs morphologiques. La théorie sous-jacente à l'analyse en composantes principales est vaste, nous ne passerons donc en revue que les points importants. Du point de vue géométrique l'ACP consiste à effectuer une certaine rotation du repère des variables autour de leurs valeurs moyennes. Cette rotation transforme les n variables non corrélées. Notons que ce sont justement ces variables transformées que l'on a nommé les composantes principales. Supposons qu'on a un vecteur de données aléatoires

$$\mathbf{X}_k = [x_1, \dots, x_m]^t \in \mathbf{R}_m \dots\dots\dots i$$

de moyenne nulle

$$\in \{x(k)\} = 0 \dots\dots\dots ii$$

et de matrice de covariance ou d'auto-corrélation

$$\Sigma = \in \{XX^t\} \in \mathbf{R}^{m.m} \dots\dots\dots iii$$

En analyse en composantes principales, un vecteur caractéristique $t \in \mathbf{R}^l$ est associé à chaque vecteur de données dont il optimise la représentation au sens de la minimisation de l'erreur

d'estimation de \mathbf{x} ou la maximisation de la variance de \mathbf{t} . Les vecteurs \mathbf{t} et \mathbf{x} sont liés par une transformation linéaire

$$t = P^t X \dots\dots\dots \text{iv}$$

où la matrice de transformation $P \in R^{m,l}$ vérifie la condition d'orthogonalité

$$P^t P = I_l \dots\dots\dots \text{v}$$

Les colonnes de la matrice \mathbf{P} forment les vecteurs de base orthonormés d'un sous-espace \mathbf{R}^l de représentation réduite des données. La transformation linéaire s'apparente ainsi à une projection de l'espace des données de dimension m vers un sous espace orthogonal de dimension l .

Les composantes t_j , avec $j = 1, \dots, l$, du vecteur caractéristique \mathbf{t} représentent les composantes projetées du vecteur de données \mathbf{x} dans ce sous-espace.

Au sens de l'ACP, la projection \mathbf{P} est optimale si l'erreur quadratique d'estimation des vecteurs de données \mathbf{x} est minimale. Ce problème d'optimisation s'exprime par :

$$P_{opt} = \arg \min J_e(P) \dots\dots\dots \text{vi}$$

Où J_e représente le critère d'erreur d'estimation de l'ACP. Sous la contrainte d'orthogonalité de la matrice de projection

$$P^t P = I_l \dots\dots\dots \text{vi}$$

ce critère aura la forme :

$$\begin{aligned} J_e(P) &= \varepsilon \left\{ \|X - \hat{X}\|^2 \right\} \\ &= \varepsilon \{ \|X - P P^t X\|^2 \} \\ &= \varepsilon \{ (X - P t)^t (X - P t) \} \\ &= \varepsilon \{ (X^t X - 2 t^t t + t^t t + t^t P P^t t) \} \\ &= \varepsilon \{ \text{trace}(X X^t) - t^t t \} \\ &= \text{trace}(\Sigma) - \varepsilon \{ t^t t \} \dots\dots\dots \text{vii} \end{aligned}$$

Notons que la trace d'une matrice carrée est définie par la somme de ces éléments diagonaux. Du fait que la matrice covariance Σ est indépendante de la matrice des paramètres \mathbf{P} , minimiser J_e revient à maximiser le second terme J_v de son expression :

$$\begin{aligned} J_v(l) &= \varepsilon \{ t^t t \} \\ &= \sum_{j=1}^l \varepsilon \{ t_j^2 \} \dots\dots\dots \text{viii} \end{aligned}$$

Ainsi, la minimisation de l'erreur quadratique d'estimation de \mathbf{x} est équivalente à la maximisation de la variance des projections t_j des données.

En conséquence :

$$P_{opt} = \arg \min J_e(P)$$

$$= \arg \max J_v(P) \dots\dots\dots \text{ix}$$

Le problème de l'ACP, considéré sous l'angle de la maximisation de la variance de projection des données, est celui de la détermination des vecteurs propres de la matrice de covariance Σ .

c. Identification du modèle ACP :

Soit $X_k = [X_1, \dots, X_n]^t$ un vecteur d'observations de n variables aléatoires. On notera \bar{X} le vecteur correspondant à la moyenne des observations de \mathbf{x} et Σ la matrice de covariance de ces m observations.

L'estimation des paramètres du modèle ACP se résume en une estimation des valeurs et vecteurs propres de la matrice de covariance Σ . Puisque la matrice de covariance Σ est carrée de dimension $n \times n$ et symétrique, on peut, selon l'algèbre linéaire, vérifier une relation de type:

$$P^t \Sigma P = L \dots\dots\dots \text{x}$$

Où L est une matrice diagonale et P est une matrice orthonormale. Les éléments diagonaux de $L = \{l_1, \dots, l_n\}$ sont les valeurs propres de la matrice de covariance Σ alors que les vecteurs colonnes de P sont les vecteurs propres de Σ .

Les valeurs propres l_i peuvent être obtenues par la solution de l'équation caractéristique :

$$\det(\Sigma - l_i I) = 0 \dots\dots\dots \text{xi}$$

pour les n valeurs propres. Connaissant les valeurs propres l_i , on peut déduire les vecteurs propres de la matrice de covariance par la solution de

$$(\Sigma - l_i I)A_i = 0 \dots\dots\dots \text{xii}$$

et puis

$$u_i = \frac{A_i}{\sqrt{(A_i^t A_i)}} \dots\dots\dots \text{xiii}$$

pour les n vecteurs propres. Notons que l'équation précédente est aussi appelée l'équation de normalisation. Il est intéressant de constater qu'ici les éléments des vecteurs propres u_i sont les cosinus directeurs des axes générés par une rotation dans l'espace des observations. Nous pouvons donc calculer tout un ensemble de valeurs à partir des ces nouveaux axes. En d'autres mots, on peut transformer les variables $X = [x_1, \dots, x_n]^t$ en un nouvel ensemble de $X = [x_1, \dots, x_n]^t$ par l'application de :

$$Y_i = u_i^t (X - \bar{X}) \dots\dots\dots \text{xiv}$$

Y_i est la $i^{\text{ème}}$ composante principale de l'ensemble des observations. Les nouvelles variables Y_i ont évidemment une moyenne nulle et elles possèdent une variance donnée par l_i (valeur propre de Σ). Les

vecteurs de colonnes \mathbf{u}_i de \mathbf{P} sont orthonormaux car ils sont orthogonaux et de modules unitaires. Donc, si l'on utilise ces vecteurs pour obtenir les composantes principales, tel que montré à l'équation précédente, on obtiendra des composantes principales qui ne seront pas corrélés et de variances égales aux valeurs propres de la matrice de covariance $\mathbf{\Sigma}$. En appliquant une transformation sur les vecteurs colonnes \mathbf{u}_i , on peut obtenir de nouvelles propriétés intéressantes. L'une de ces transformations consiste à effectuer une mise à l'échelle de \mathbf{u}_i par les valeurs propres de la matrice de covariance $\mathbf{\Sigma}$:

$$V_i = u_i \sqrt{l_i} \dots\dots\dots \text{xv}$$

Les composantes obtenues par l'utilisation de $V_i = [v_i, \dots, v_n]^t$ seront de même unité que les variables originales. Nous retrouvons également une transformation dans laquelle les vecteurs \mathbf{u}_i sont divisés par les valeurs propres de la matrice de covariance \mathbf{S} :

$$W_i = \frac{u_i}{\sqrt{l_i}} \dots\dots\dots \text{xvi}$$

Dans ce cas, les composantes obtenues seront de variance unitaire.

Enfin, nous soulignons que l'on peut avoir une mesure de variabilité expliquée par chacune des composantes principales en utilisant les vecteurs transformés de l'équation avant précédente selon

$$Var = V_i V_i^t \dots\dots\dots \text{xvii}$$

avec : $i = 1, \dots, n$

Les valeurs obtenues par cette dernière équation peuvent être utilisées pour indiquer le pourcentage de la variabilité des caractéristiques représentées par chacune des composantes principales.

d. Détermination de la structure du modèle ACP

L'analyse en composantes principales recherche une approximation de la matrice initiale des données X par une matrice rang inférieur issue d'une décomposition en valeurs singulières. La question qui se pose alors, et qui a été largement débattue dans la littérature, concerne le choix du nombre de composantes principales qui doit être retenu. De nombreuses règles sont proposées pour déterminer le nombre de composantes à retenir, [43],[44],[45],[46]. La plupart de ces méthodes sont heuristiques et donnent un nombre de composants subjectif. Toutefois, dans le cadre de l'application de l'ACP à la réduction des échantillons de l'ECG, le nombre de composantes à un impact significatif sur la phase finale de classification. Si peu de composantes sont utilisées, on risque de perdre des informations contenues dans les données de départ en projetant certains échantillons dans le sous-espace des résidus et donc avoir des erreurs de modélisation, ce qui provoque une mauvaise caractérisation.

Si par contre beaucoup de composantes sont utilisées, il y a risque d'avoir des composantes retenues (les composantes correspondantes aux valeurs propres les plus faibles parmi celles retenues dans le modèle) qui sont porteuses de redondance, ce qui indésirable.

Dans leurs travaux de recherche, Qin et al [47] ont proposé une technique basée sur la variance de l'erreur de reconstruction des mesures ; ce critère permet de prendre en compte la notion de redondance entre les variables (échantillons de l'ECG dans notre cas).

On trouve dans la littérature plusieurs critères utilisant différentes techniques pour la détermination du nombre de composantes principales, nous allons présenter deux critères heuristiques utilisés pour la réduction de dimension de la taille de l'ECG.

❖ **Pourcentage cumulé de la variance totale (PVC) :** L'idée de base de cette approche est que chaque composante principale est représentative d'une portion de la variance des mesures de l'information étudiée. Les valeurs propres de la matrice de corrélation sont les mesures de cette variance et peuvent donc être utilisées dans la sélection du nombre de composantes principales.

Pour le choix de l , il faut choisir le pourcentage de la variance totale qu'on veut conserver. Le nombre de composantes est alors le plus petit nombre pris de telle sorte que ce pourcentage soit atteint ou dépassé ; les composantes sont choisies successivement dans l'ordre des variances décroissantes. Le pourcentage de variance expliquée par les l premières composantes est donné par :

$$PVC(l) = 100 \left(\frac{\sum_{j=1}^l \lambda_j}{\sum_{j=1}^m \lambda_j} \right) \% \dots\dots\dots \text{xviii}$$

La variance du bruit étant inconnue a priori, la décision est basée seulement sur le pourcentage de la variance expliquée est un peu arbitraire. Sa capacité à fournir le nombre correct de composantes principales dépendra fortement du rapport signal sur bruit.

❖ **Moyenne des valeurs propres :** Cette règle consiste à ne prendre en considération que les composantes pour lesquelles la valeur propre est supérieure à la moyenne arithmétique de toutes les valeurs propres. En particulier, si on travaille sur les données centrées réduites, cela revient à négliger les composantes dont la variance est inférieure à l'unité

$$\left(\frac{1}{m} \text{trace}(\Sigma) = 1 \right) \dots\dots\dots \text{xix}$$

Dans le cas du modèle ACP calculé à partir de la matrice de covariance Σ , la moyenne arithmétique des valeurs propres est donnée par :

$$\frac{1}{m} \text{trace}(\Sigma) \dots\dots\dots \text{xx}$$

Annexe B

Base de données MIT-BIH

Depuis 1975, les laboratoires de l'hôpital de Beth Israël à Boston et MIT ont réalisé une base de données MIT-BIH qui a été commencée à être distribuée en 1980. Cette base de données contient 48 enregistrements extraits d'une demi-heure des enregistrements ambulatoires à deux voies d'ECG, obtenus à partir de 47 sujets étudiés par le laboratoire d'arythmie de BIH entre 1975 et 1979. 23 enregistrements ont été choisis au hasard d'un ensemble de 4000 enregistrements ambulatoires de 24 heures d'ECG rassemblés d'une population mélangée des patients hospitalisés (60%) et des patients non hospitalisés (40%) à l'hôpital de Beth Israël à Boston (les séries '100'), les 25 enregistrements restants ont été choisis parmi les mêmes enregistrements mais qui en considération des arythmies rarement observés qui ont une signification clinique (les séries '200').

Les enregistrements ont été échantillonnés à une fréquence $f_e = 360\text{Hz}$ avec une résolution de 11 bits sur une gamme de 10 mV. Deux cardiologues ou plus ont indépendamment annoté chaque enregistrement, environ 110,000 annotations ont été incluses avec la base de données. Le tableau B.1 présente les 48 enregistrements de la base de données MIT-BIH.

N° d'enregistrement arythmies	N	A	V	F	P	Q
100	2239	33	1	-	-	-
101	1860	3	-	-	-	2
102	99	-	4	-	2028	-
103	2082	2	-	-	-	-
104	163	-	2	-	1380	18
105	2526	-	41	-	-	5
106	1507	-	520	-	-	-
107	-	-	59	-	2078	-
108	1739	4	17	2	-	-
109	-	-	38	2	-	-
111	-	-	1	-	-	-
112	2537	2	-	-	-	-
113	1789	-	-	-	-	-
114	1820	10	43	4	-	-
115	1953	-	-	-	-	-
116	2302	1	109	-	-	-

117	1534	1	-	-	-	-
118	-	96	16	-	-	-
119	1543	-	444	-	-	-
121	1861	1	1	-	-	-
122	2476	-	-	-	-	-
123	1515	-	3	-	-	-
124	-	2	47	5	-	-
200	1743	30	826	2	-	-
201	1625	30	198	2	-	-
202	2061	36	19	1	-	-
203	2529	-	444	1	-	4
205	2571	3	71	11	-	-
207	-	107	105	-	-	-
208	1586	-	992	373	-	2
209	2621	383	1	-	-	-
210	2423	-	194	10	-	-
212	923	-	-	-	-	-
213	2641	25	220	362	-	-
214	-	-	256	1	-	2
215	3195	3	164	1	-	-
217	244	-	162	-	1542	-
219	2082	7	64	1	-	-
220	1954	94	-	-	-	-
221	2031	-	396	-	-	-
222	2062	208	-	-	-	-
223	2029	72	478	14	-	-
228	1688	3	362	-	-	-
230	2255	-	1	-	-	-
231	314	1	2	-	-	-
232	-	1382	-	-	-	-
233	2230	7	831	11	-	-
234	2700	-	3	-	-	-

Tableau B.1: les 48 enregistrements de la base de données MIT-BIH.

N : Battement normal (Normal beat).

A : Battement auriculaire prématuré (Atrial premature beat).

V : Contraction ventriculaire prématurée (Premature ventricular contraction).

F : Mélange entre les battements ventriculaires et les battements normaux (Fusion of ventricular and normal beat).

P : Pause.

Q : Battement non classifiable (Unclassifiable beat).

Références

Bibliographiques

Références bibliographiques

- [1] Ahmet Yardimci, “Soft computing in medicine,” Akdeniz University, TBMYO, departement of Industrial Automation, Antalya, 07059, Turkey, 2010.
- [2] Handel Mounia, “Classification des arythmies cardiaques par les réseaux de neurones artificiels,” Université des sciences et de technologie d’Oran, département d’informatique, laboratoire SIMPA, El-M’naouar BP 1505, Oran Algérie, 2009.
- [3] E.-P. d’Alché, *Comprendre la physiologie cardiovasculaire*. Flammarion médecine-sciences, 2003.
- [4] R. Dubois, “Application des nouvelles méthodes d’apprentissage à la détection précoce d’anomalies en électrocardiographie,” UNIVERSITÉ PARIS 6, 2004.
- [5] T.Chou, “Normal electrocardiogram in: Electrocardiology in clinical practice,” New-York, San Francisco, London, pp. 3–16, 1979.
- [6] A. Cabasson, “Estimation et analyse des intervalles cardiaques,” Université Nice Sophia Antipolis, 2008.
- [7] *Encyclopedia Universalis*, vol. 8. pp. 141–147, 1996.
- [8] *Alpha*, vol. 8. Grange Batelière, p. 2157.
- [9] R. H. Wasserburger and W. J. Alt, “The normal RS-T segment elevation variant,” *Am. J. Cardiol.*, vol. 8, no. 2, pp. 184–192, 1961.
- [10] Jean Sende, *Guide pratique de l’ECG*, Estem. 2003.
- [11] J. Mackay, G. A. Mensah, S. Mendis, and K. Greenlund, *The atlas of heart disease and stroke*. World Health Organization, 2004.
- [12] P. Szolovits P, “Artificial Intelligence in Medicine,” vol. 11, no. 1, 1997.
- [13] A. Belaïd and Y. Belaïd, *Reconnaissance des formes: méthodes et applications*. Paris: InterEditions, 1992.
- [14] M. A. Chikh Mohamed Amine, “Analyse du signal ECG par les réseaux de neurones et la logique floue: Application à la reconnaissance des battements ventriculaires prématurés,” Thèse de doctorat, Thèse présentée devant l’Université de Tlemcen, 2005.
- [15] C. Leschi C, “Aspects fondamentaux et évolutifs de la reconnaissance de formes. Approche pluridisciplinaire. Notion d’approximation.” thèse de doctorat en informatique et automatique appliquée, PhD thesis, Institut Nationale des sciences Appliquées de Lyon, 1991.
- [16] M. Milgram M, *Reconnaissance des formes: méthodes numériques et connexionnistes*. Paris: A. Colin, 1993.

- [17] I. Guyon, J. Makhoul, R. Schwartz, and V. Vapnik, "What size test set gives good error rate estimates?," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 1, pp. 52–64, 1998.
- [18] G. Mohamed Goismi, "Approche multi-agents pour le monitoring médical," Université Ibn Khaldoun de Tiaret, 2011.
- [19] A. Messaoud, M. B. Messaoud, A. Kachouri, and F. Sellami, "Classification des arythmies cardiaques par logique floue à partir de signaux ECG," *SETIT2005*.
- [20] L. I. Kuncheva and F. Steimann, *Fuzzy diagnosis. Artificial Intelligence in medicine*, vol. 16, 2 vols. Elsevier, 1999.
- [21] T. Q. CHU Thanh-Quang, "Application des méthodes de l'intelligence artificielle à l'analyse des données concernant les canalisations de gaz," Institut de la francophonie pour l'informatique, Rapport de stage, 2006.
- [22] B. Schölkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.
- [23] V. Vapnik, "Pattern recognition using generalized portrait method," *Autom. Remote Control*, vol. 24, pp. 774–780, 1963.
- [24] O. D. Richard, E. H. Peter, and G. S. David, "Pattern classification," *Wiley-Intersci.*, pp. 373–378, 2001.
- [25] L. G. Valiant, "A theory of the learnable," *Commun. ACM*, vol. 27, no. 11, pp. 1134–1142, 1984.
- [26] J. Mercer, "Functions of positive and negative type, and their connection with the theory of integral equations," *Philos. Trans. R. Soc. Lond. Ser. Contain. Pap. Math. Phys. Character*, vol. 209, pp. 415–446, 1909.
- [27] M. A. Aiserman, E. M. Braverman, and L. I. Rozonoer, "Theoretical foundations of the potential function method in pattern recognition," *Avtomat Telemekh*, vol. 25, pp. 917–936, 1964.
- [28] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, pp. 144–152.
- [29] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013.
- [30] D. Maatar, Z. Lachiri, and N. Ellouze, "Classification Automatique d'Arythmies par HMM utilisant les Paramètres Morphologiques dans l'ECG," *TAIMA Tunis.*, 2009.
- [31] Haykin S, *Neural Networks: A Comprehensive Foundation*. 1994.
- [32] P. J. Braspenning, F. Thuijsman, and A. J. M. M. Weijters, *Artificial neural networks: an introduction to ANN theory and practice*, vol. 931. Springer Science & Business Media, 1995.

- [33] G. Dreyfus, J. M. Martinez, M. Samuelides, M. B. Gordon, F. Badran, and S. Thiria, “Apprentissage statistique: Réseaux de neurones: Cartes topologiques Machines à vecteurs supports, eyrolles,” 2004.
- [34] A. Cornuéjols and L. Miclet, *Apprentissage artificiel: concepts et algorithmes*. Editions Eyrolles, 2011.
- [35] C. M. Bishop, *Neural networks for pattern recognition*. Oxford university press, 1995.
- [36] J. C. Nash, *Compact numerical methods for computers: linear algebra and function minimisation*. CRC press, 1990.
- [37] K. Hornik, “Approximation capabilities of multilayer feedforward networks,” *Neural Netw.*, vol. 4, no. 2, pp. 251–257, 1991.
- [38] I. T. Jolliffe, “Principal Component Analysis. I Springer-Verlag,” N. Y., 1986.
- [39] K. I. Diamantaras and S. Y. Kung, *Principal component neural networks: theory and applications*. John Wiley & Sons, Inc., 1996.
- [40] K. Person, *On Lines and Planes of Closest Fit to System of Points in Space. Philosophical Magazine*, 2, 559-572. 1901.
- [41] C. Spearman, “‘General Intelligence,’ Objectively Determined and Measured,” *Am. J. Psychol.*, vol. 15, no. 2, pp. 201–292, 1904.
- [42] H. Hotelling, “Analysis of a complex of statistical variables into principal components,” *J. Educ. Psychol.*, vol. 24, no. 6, p. 417, 1933.
- [43] P. Besse and L. Ferre, “Sur l’usage de la validation croisée en analyse en composantes principales,” *Rev. Stat. Appliquée*, vol. 41, no. 1, pp. 71–76, 1993.
- [44] D. M. Himes, R. H. Storer, and C. Georgakis, “Determination of the number of principal components for disturbance detection and isolation,” in *American Control Conference, 1994*, 1994, vol. 2, pp. 1279–1283.
- [45] L. Ferré, “Selection of components in principal component analysis: a comparison of methods,” *Comput. Stat. Data Anal.*, vol. 19, no. 6, pp. 669–682, 1995.
- [46] S. Valle, W. Li, and S. J. Qin, “Selection of the number of principal components: the variance of the reconstruction error criterion with a comparison to other methods,” *Ind. Eng. Chem. Res.*, vol. 38, no. 11, pp. 4389–4401, 1999.
- [47] S. J. Qin and R. Dunia, “Determining the Number of Principal Components for Best Reconstruction,” *IFAC Proc. Vol.*, vol. 31, no. 11, pp. 357–362, 1998.

Résumé :

Les maladies cardiovasculaires constituent un problème majeur de santé publique. Elles sont responsables de la plupart des décès dans le monde.

Le cœur est l'organe central du système cardiovasculaire, il peut être affecté par de nombreuses pathologies.

L'électrocardiogramme est une présentation graphique du potentiel électrique qui commande l'activité musculaire du cœur.

Dans ce mémoire nous avons appliqué une approche neuronale pour la classification morphologique du signal ECG où nous avons proposé de mettre au point un système de classification des battements cardiaques, ce système est basé principalement sur l'utilisation des réseaux de neurones, dont, le classifieur neuronal reçoit en entrée 19 paramètres morphologiques qui sont les composantes principales retenues de l'inertie en entrée, après l'avoir compressé par la méthode d'analyse en composantes principales ACP, caractérisant un signal ECG. Notre système a l'objectif de faire la classification neuronale en se basant sur les données morphologiques des ECG selon deux classes, N «normales» et A «anormales» qui a été validé sur des signaux de la base de données MIT-BIH. Les résultats qualitatifs et quantitatifs obtenus démontrent l'efficacité de cette approche.

Mots clés :

ECG, les maladies cardio-vasculaires, arythmies cardiaques, classification, réseaux de neurones, apprentissage artificiel, descripteur morphologique, ACP.

المخلص:

أمراض القلب و الأوعية الدموية تشكل خطرا كبيرا على الصحة العمومية، هي المسؤولة على معظم حالات الوفيات في العالم. القلب هو العضو المركزي لنظام القلب و الأوعية الدموية، قد يتأثر بكثير من الأمراض. مخطط نبضات القلب هو تمثيل بياني للجهد الكهربائي الذي يتحكم في النشاط العضلي للقلب. في هذه المذكرة قمنا بتطبيق نظام مورفولوجي لمعاينة نبضات القلب و يعتمد هذا الأخير على الشبكة العصبية حيث يتلقى 19 تعليمة مورفولوجية لتمييز نبضة قلبية واحدة و تصنيفها إلى فئتين «عادية» و «غير عادية» معتمدا على قاعدة بيانات MIT-BIH و تبين النتائج الكمية و النوعية فعالية هذا النظام.

الكلمات الأساسية:

مخطط نبضات القلب، أمراض القلب و الأوعية الدموية، عدم انتظام نبضات القلب، التصنيف، الشبكات العصبية، التعلم الاصطناعي، الواصف المورفولوجي، التحليل إلى مكونات أساسية.