

Université Mohamed Boudiaf - M'sila

FACULTE DE TECHNOLOGIE

DEPARTEMENT D'ELECTRONIQUE



Numéro de série :

Numéro d'inscription :

Mémoire

Présentée pour l'obtention du diplôme de

Master

Filière : Electronique

Spécialité : Instrumentation

THEME

**Extraction des caractéristiques
et classification des signaux ECG**

Présenté par

LAMANI Meriem & HALITIM Rania

Soutenue le :/09/2020

Devant le jury composé de :

Nom & Prénom

Grade

Qualité

OUALI Mohammed Assam

MCB

Président

Encadreur

LADJAL Mohamed

MCA

Co-encadreur

Examineur

Remerciements

La réalisation de ce mémoire a été possible grâce au concours de plusieurs personnes à qui je voudrais témoigner toute ma gratitude.

*Je voudrais tout d'abord adresser toute ma reconnaissance aux directeurs de ce mémoire **Dr. Ouali mohammed Assam et Dr. LADJAL Mohamed**, pour leurs patiences, leurs disponibilité et surtout leurs judicieux conseils, qui ont contribué à alimenter ma réflexion, je le remercie de m'avoir encadré, orienter et aidé. J'adresse mes sincère remerciements à tous les professeurs de l'université de M'sila, qui m'ont fourni les outils nécessaires à la réussite de mes études universitaires. Je remercie mes très chers parents, qui ont toujours été là pour moi, je remercie ma sœur et mes frères pour leur encouragement. Je voudrais exprimer ma reconnaissance envers les amis et collègues qui m'ont apporté leur soutien moral et intellectuel tout au long de ma démarche. Un grand merci à **RAYHANA ALLOUACHE** pour ces conseils concernant mon style d'écriture, ils ont grandement facilité mon travail.*

Dédicace

*Je dédie ce modeste travail, en première lieu, aux êtres qui me sont les plus chers ;
Mes parents qui ont toujours aimé me voir réussir.*

*A mon père pour son soutien moral et financier, leurs énormes sacrifices et leurs
Encouragements qu'ils m'ont apportés tout au long de toutes les années d'études.*

*A ma mère ange de l'âme, source de l'espoir et de la tendresse, pour ton amour, pour
tous ce que tu as sacrifié pour moi, Tu es maman idéale*

*A mes chères frères « Taha », « Azzedine » et « Yahya », et ma chère sœur « HIBA » Pour
tous les bons moments passés avec vous.*

« Je vous souhaite une vie joyeuse pleine de succès dans tous les domaines »

*A tous mes amies et surtout :
« Cheyma », « Rania », « Lilla »*

Tous mes collègues de l'étude sans exception.

A toute la famille « LAMANI »

LAMANI Meryem

Dédicace

Je dédie ce modeste travail, en première lieu, aux êtres qui me sont les plus chers

Mes parents, pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leurs prières tout au long de mes études,

A mes sœurs Assia et Samah pour leurs encouragements permanents, et leur soutien moral,

A mes frères, Hamza, Oussama et Mouhamed , pour leur appui et leur encouragement,

A toute ma famille HALITIM pour leur soutien tout au long de mon parcours universitaire, et surtout Nouari ,Rabie , Dhaifullah et Samir

Encouragements qu'ils m'ont apportés tout au long de toutes les années d'études.

A mes amies :

« Meriem », «Mounira », «Noor Huda»

A toute la famille « HALITIM » , « OUALI» et «LADJAL»

Tous mes collègues de l'étude sans exception.

Merci d'être toujours là pour moi.

HALITIM Rania

Liste des abréviations

ECG : Électrocardiogramme ;
OD : Oreillette droite ;
OG : Oreillette Gauche ;
VCS : Veine cave supérieur ;
VCI : Veine cave inférieur ;
AD : auricule droite ;
VD : ventricule gauche ;
VA : ventricule droite ;
AO : L'aorte ascendante ;
AP : Artère pulmonaire ;
ESV : Extrasystole ventriculaire ;
TV : Tachycardie ventriculaire ;
AAMI : Association for the Advanced of Medical Instrumentation ;
NFLS : Nonlinear Filtering Scheme ;
SVM : Machine à vecteur de support ;
SV : Vecteur de support ;
RNA : Réseaux de neurones artificiels ;
BBO : Biogéographie-Based Optimization ;
HSI : Habitat Suitability Index ;
SIV : Suitability Index Variables ;

LISTE DES TABLEAUX

Tableau 1. 1 Annotations de la base de données MIT-BIH PhysioBank.....	15
Tableau 3.1 Différentes fonctions d'activations utilisées dans les RNA.....	35
Tableau 4.1 Enregistrements ECG et nombres de battements utilisés dans ce travail.....	47
Tableau 4. 2 Terminologie de la BBO.....	52
Tableau 4.3 Paramètres de la <i>BBO</i>	53

Liste des Figures

Figure 1.1	L'anatomie du cœur.....	4
Figure 1.2	Vue droite du cœur.....	5
Figure 1.3	Vue gauche du cœur.....	5
Figure 1.4	La physiologie du cœur.....	6
Figure 1.5	Fonctionnement électrique.....	7
Figure 1.6	L'onde P.....	8
Figure 1.7	Le complexe QRS.....	9
Figure 1.8	L'onde T.....	9
Figure 1.9	Extrasystole ventriculaire (ESV).....	10
Figure 1.10	Tachycardie ventriculaire	11
Figure 1.11	La fibrillation atriale.....	12
Figure 1.12	Différents fichiers de la base de données MIT-BIH PhysioBank.....	13
Figure 1.13	L'enregistrement 233.dat de la base de données MIT-BIH PhysioBank.....	14
Figure 1.14	Fichier :hea de l'enregistrement 100.dat de la base de données MIT-BIH PhysioBank.....	14
Figure 2.1	Localisation des transitions d'un signal échantillonnés.....	21
Figure 2.2	Caractéristiques d'un battement cardiaque.....	23
Figure 2.3	Les étapes de la méthode Pan and Tompkins.....	23
Figure 3.1	Schéma synoptique d'un classificateur flou supervisé.....	28
Figure 3.2	Exemple d'un hyperplan séparateur.....	29
Figure 3.3	Exemple de vecteurs de support.....	29
Figure 3.4	Exemple de marge maximal.....	30
Figure 3.5	Classification par SVM.....	30
Figure 3.6	Diagramme du Réseau de neurones.....	32
Figure 3.7	Neurone biologique.....	33
Figure 3.8	Neurone formel.....	34

Figure 3.9 Les réseaux de neurones non bouclés.....	37
Figure 3.10 Schéma d'un réseau de neurones monocouche.....	37
Figure 3.11 Schéma d'un réseau de neurones Perceptron multicouches.....	38
Figure 3.12 Schéma d'un réseau de neurones à connexions locales.....	38
Figure 3.13 Schéma de réseau de neurones bouclé.....	39
Figure 3.14 Forme canonique d'un réseau de neurones bouclé.....	40
Figure 3.15 Apprentissage non supervisé.....	41
Figure 3.16 Apprentissage supervisé.....	41
Figure 4.1 Organigramme de la procédure adoptée	46
Figure 4.2 Résultats de simulation de la méthode Pan and Tompkins.....	48
Figure 4.3 Paramètres du réseau de neurones à optimiser.....	49
Figure 4.4 Migration des espèces.....	50
Figure 4.5 Organigramme de l'algorithme <i>BBO</i>	53
Figure 4.6 Illustration de deux solutions candidates S ₁ et S ₂	56
Figure 4.7 Organigramme de l'approche utilisée.....	57

Table des matières

Remerciements	I
Dédicace	II
LISTE DES TABLEAUX	V
LISTE DES FIGURES	VI
INTRODUCTION GENERAL	1

CHAPITRE I

1. Introduction	4
2. L'anatomie du cœur humain	4
3. La physiologie cardiaque	6
4. Etude d'un cycle cardiaque	6
4.1 L'activité mécanique du cœur	6
4.2 L'activité électrique du cœur	7
5. Electrocardiogramme (ECG)	7
5.1 Les caractéristiques de l'ECG.....	8
5.2.1 Onde P.....	8
5.2.2 Le complexe QRS	9
5.2.3 L'onde T	9
5.2.4 L'intervalle PR	10
6. Les maladies cardiaques	10
6.1 Extrasystole ventriculaire(ESV).....	10
6.2 la tachycardie Ventriculaire.....	10
6.3 la fibrillation atrial (la fibrillation auriculaire).....	11
7. Le standard AAMI	12
8. La base de données MIT-BIH PhysioBank	12
8.1 Le fichier de donnée(.dat)	12
8.2 Le fichier (.hea).....	13
8.3 Le fichier (.atr).....	14
9. Conclusion	15

CHAPITRE II

Notions et techniques d'extraction des caractéristiques

1. Introduction	16
2. Définition et intérêt	17
3. Méthodes d'extraction des caractéristiques	18
3.1 Méthodes basées sur du filtrage dérivatif	18
3.2 Extraction par Analyse en composante principale (ACP).....	18
3.2.1 Principe	19
3.3 Extraction de caractéristiques par les filtres numériques	20
3.4 L'extraction des caractéristiques par la méthode NFLS.....	21
3.4.1 Principe	21
3.5 L'extraction par la méthode Pan and Tompkins	22
3.5.1 Les étapes de traitement de la détection du complexe QRS ..	23
4. Conclusion	25

CHAPITRE III

Notions et techniques de classification

1. Introduction	26
2. La classification	27
3. Les techniques de classification	27
3.1 La classification par la logique floue.....	27
3.1.1 Représentation d'un ensemble flou.....	27
3.2 La Classification par la machine à vecteur de support (SVM).....	28
3.2.1 Définition	28

3.2.2	Principe et fonctionnement générale de SVM.....	29
3.3	La Classification par les réseaux de neurones artificiels (<i>RNA</i>).....	31
3.3.1	<i>Définition</i>	31
3.3.2	<i>Neurone biologique</i>	32
3.3.3	<i>Neurone formel</i>	33
3.3.4	<i>Fonction d'activation</i>	34
3.3.5	<i>Architecture des réseaux de neurons</i>	36
3.3.5.1	Les réseaux de neurones non bouclés.....	36
a.	Réseaux de neurones monocouches	
b.	Réseaux de neurones multicouches	
c.	Réseaux de neurones à connexions locales	
3.3.5.2	Les réseaux de neurones bouclés.....	39
3.3.6	Apprentissage	40
3.3.6.1	Apprentissage non supervise.....	40
3.3.6.2	Apprentissage supervisé	41
3.3.6.3	Règles d'apprentissage	42
3.3.7	Applications des réseaux de neurones artificiels.....	42
3.3.7.1	Classification.....	42
3.3.7.2	Prédiction	42
3.3.7.3	Reconnaissance des formes (pattern recognition).....	43
3.3.7.4	Optimisation	43
4.	Conclusion	43

CHAPITRE IV

Extraction des caractéristiques et classification des battements cardiaques : Méthode & résultats

1.	Introduction	44
2.	Méthode et résultats	45
2.1	Préparation des bases de données.....	46
2.2	Extraction des caractéristiques.....	47
2.3	Sélection des caractéristiques par la méthode mRMR.....	48

2.4 Classification.....	49
2.4.1 Paramètres de classificateur à ajuster.....	49
2.4.2 L'algorithme d'optimisation à base de biogéographie (BBO).....	49
2.4.2.1 Biogéographie.....	50
2.4.2.2 Principe de l'algorithme BBO	51
2.4.2.3 Optimisation et biogeography.....	51
2.4.3 Étapes de l'algorithme d'optimisation basée sur la biogéographie	52
A. Initialisation	
B. Évaluation de la fonction HSI	
C. Sélection	
D. Opérateur de migration	
E. Opérateur de mutation	
2.4.4 L'approche de classification à base de réseau de neurones artificiels et L'algorithme d'optimisation à base de biogéographie (BBO).....	57
2.5 Résultat.....	57
3. Conclusion	57
Conclusion générale.....	59
Bibliographie	60

Introduction générale

Les maladies cardiovasculaires constituent un problème majeur de santé publique, elles sont les premières causes de mortalité dans le monde. Les facteurs de risques sont multiples : tabac, sédentarité, obésité, hypertension artérielle, diabète, facteurs génétiques aussi parfois.

Le cœur est l'organe central du système cardiovasculaire, il peut être affecté de nombreuses pathologies qui peuvent être bénignes, comme certaines tachycardies par exemple, soit s'avère très sérieuses, comme l'infarctus du myocarde. Avec l'évolution des techniques, les médecins disposent aujourd'hui d'outils performants pour observer le fonctionnement du muscle cardiaque et dresser ainsi leur diagnostic. Parmi les examens cardiologiques possibles, l'électrocardiogramme (ECG) est l'examen le plus couramment effectué.

L'électrocardiogramme (ECG) est la mesure de l'activité électrique durant la contraction du cœur, il est caractérisé par des ondes appelées P, Q, R, S et T. Les ondes Q, R et S forment le complexe QRS. L'invention de l'ECG a valu à Willem Einthoven le prix de Nobel de médecine en 1924. De nos jours, l'ECG est largement utilisé et présente jusqu'à douze paires d'électrodes d'enregistrement. La différence de potentiel enregistrée entre chaque paire d'électrodes constitue le signal qui représente l'activité électrique du cœur depuis plusieurs points de vue. La fréquence d'échantillonnage de tels ECG modernes se situe entre 250 et 500 Hz. Néanmoins, ce type de procédures est uniquement utilisé pour des mesures très courtes. Malheureusement, une mesure de l'ECG de courte durée peut ne permettre le diagnostic de certaines pathologies comme Les arythmies cardiaques des épisodes ischémiques transitoires ou des ischémies silencieuses du myocarde. Dans ce genre de situations, les médecins ont recours à un enregistrement ECG de long terme en utilisant moins d'électrodes (seulement deux ou trois) : le Holter.

L'objectif de ce mémoire de fin d'étude est la conception d'un système expert pour le diagnostic des pathologies ou des arythmies cardiaques en se basant sur le signal électrocardiogramme. La détection de ces arythmies est un sérieux problème sur le plan clinique et technologique (moyens et approches). Elles sont variées et les manières de les aborder sont également variées, souvent cela dépend de quel point de vue où on se place. Il est possible de définir les arythmies cardiaques par l'irrégularité des battements cardiaques ou comme un groupe de battements irréguliers. Elles forment deux groupes : le premier comprend la fibrillation et la tachycardie ventriculaires qui sont dangereuses et nécessitent une thérapie immédiate avec un défibrillateur ; le second concerne les arythmies non dangereuses mais qui exigent une thérapie préventive d'autres problèmes. Ces troubles cardiaques sont aussi classés suivant la fréquence et le rythme décrit par le nombre de cycles ou contractions ventriculaires (dépolariation des ventricules) par minute, repérés sur l'ECG par le nombre de complexe QRS.

Dans ce travail tout un système expert pour le diagnostic des maladies cardiaques sera présenté en détail, ce système est présenté dans la figure ci-dessous. Pour la première étape qui est la détection des battements cardiaques, la base de données MIT-BIH PhysioBank (MIT-BIH Arrhythmia Database) sera utilisée.

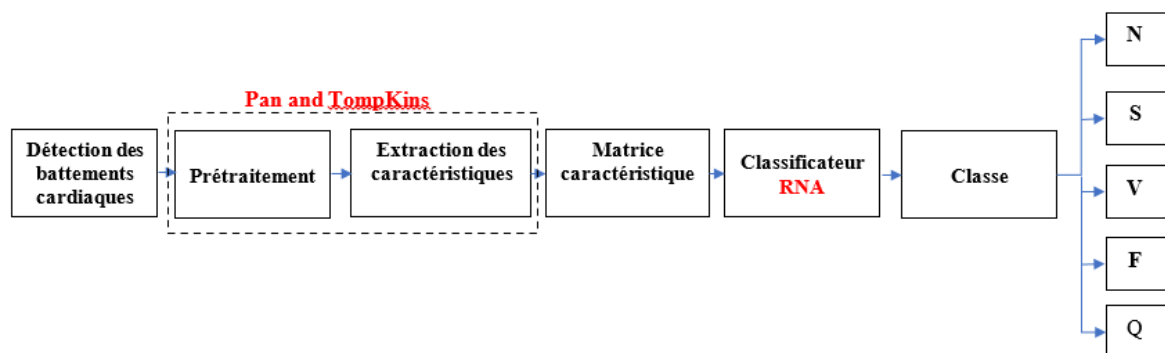


Figure : Les étapes de classification du signal ECG.

Après la détection des battements cardiaques, une étape de pré-traitement et d'extraction des caractéristiques sera nécessaire. Dans notre travail la méthode de Pan and TompKins sera adoptée.

Après l'étape de pré-traitement et d'extraction des caractéristiques la tâche suivante sera la classification. Pour pouvoir classifier les battements cardiaques, il faut commencer par définir des classes. Dans notre cas, c'est la classification recommandée par l'AAMI « Association for the Advancement of Medical Instrumentation », qui sera utilisée, les classes sont au nombre de cinq à savoir : les battements normaux (N), les battements anormaux (S, V et F), et les battements non connus (Q). Les réseaux de neurones artificiels seront utilisés comme

classificateur avec la particularité d'utiliser un algorithme d'optimisation pour ajuster les paramètres de ce classificateur au lieu des techniques basées sur le gradient et cela pour éviter le problème de minimum local.

Ce mémoire s'articule en quatre chapitres qui nous permettront de présenter les différents aspects de notre travail.

- ✚ Le premier chapitre présente brièvement le cœur et son fonctionnement, ainsi que le signal électrocardiogramme et ses caractéristiques.
- ✚ Le deuxième chapitre présente les notions de bases relatives à l'extraction des caractéristiques ainsi que quelques techniques d'extraction des caractéristiques à savoir : l'analyse en composants principales (ACP), les méthodes basées sur le filtrage dérivatif, extraction de caractéristiques par les filtres numériques, L'extraction de caractéristiques par la méthode NFLS, afin d'expliquer la technique dont nous adaptons qui est Pan and Tompkins.
- ✚ Le troisième chapitre introduit les notions de base relatives à la classification et consiste les techniques les plus populaires de la classification des signaux ECG à savoir : La classification par la logique floue, classification par la machine à vecteur de support (SVM) et la Classification par les réseaux de neurones artificiels (RNA)
- ✚ Dans le dernier chapitre, nous exposons la structure proposée pour le diagnostic des maladies cardiaques et les résultats trouvés.

On termine ce mémoire par une conclusion générale mettant en relief les résultats obtenus.

CHAPITRE I

Anatomie du cœur et signal électrocardiogramme

Le travail proposé dans ce mémoire de fin d'étude touche le signal électrocardiogramme ECG de différents angles tels que son dé-bruitage et son diagnostic. Nous exposons dans ce premier chapitre l'anatomie du cœur à fin de comprendre la genèse de ces signaux (ECG). En outre, la morphologie du signal électrocardiogramme et ses caractéristiques.

Sommaire

10. Introduction
 11. L'anatomie du cœur humain
 12. La physiologie cardiaque
 13. Etude d'un cycle cardiaque
 14. Electrocardiogramme (ECG)
 15. Les maladies cardiaques
 16. Le standard AAMI
 17. Les maladies cardiaques
 18. La base de données MIT-BIH PhysioBank
 19. Conclusion
-

1. Introduction

Le cœur est un muscle qui se contracte sans cesse, dès les premières heures de notre vie jusqu'à notre mort. Dans une vie entière, il battra environ de trois milliards de fois. Il a pour fonction de pomper et propulser le sang vers tous les organes de notre corps. Il pompe environ cinq litres de sang à la minute.

Le travail proposé dans ce mémoire se déroule autour du signal électrocardiogramme et vis à développer un système expert pour le diagnostic des maladies cardiovasculaires. Nous exposons dans ce premier chapitre de présenter l'anatomie du cœur, sa physiologie, son fonctionnement électrique, les caractéristiques du signal électrocardiogramme et finalement nous exposons le standard AAMI.

2. L'anatomie du cœur humain

Le cœur est un organe creux et musculaire comparable à une pompe, qui assure la circulation du sang dans les veines et les artères. Sa forme est similaire à un cône inversé (sa base vers le haut et à droite et son apex en bas et à gauche).

Le cœur se situe dans le médiastin, c'est la partie médiane de la cage thoracique délimitée par les deux poumons, le sternum et la colonne vertébrale. Il se trouve un peu à gauche du centre du thorax. Le cœur pèse environ 300 grammes chez l'homme adulte, 250 grammes chez la femme. Il est capable de propulser, au repos, 4 à 5 litres de sang par minute.

Le cœur est considéré comme une pompe musculaire subdivisée en quatre cavités : deux oreillettes et deux ventricules (Figure. 1), permettant de propulser le sang vers toutes les cellules du corps humain. Les paires oreillette ventricule droite et gauche constituent respectivement les cœurs droit et gauche[1].

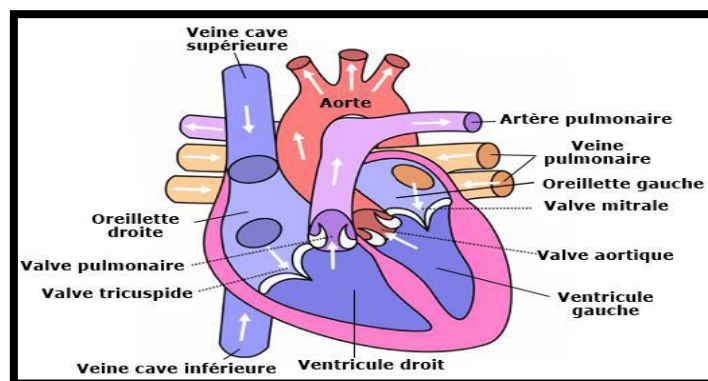


Figure. 1: L'anatomie du cœur.

La partie droite du cœur est constituée principalement par l'oreillette droite (OD) ou atrium droit qui reçoit en haut la veine cave supérieure (VCS) et en bas la veine cave inférieure (VCI) et se prolonge en haut et en avant par l'auricule droite (AD). En arrière de l'oreillette

droite, apparaît le bord postéro-droit de l'oreillette gauche (OG) recevant les deux veines pulmonaires droites, supérieure et inférieure. Au-dessus de l'oreillette gauche et derrière la veine cave supérieure, sous la veine grande azygos, émerge la branche droite de l'artère pulmonaire. En avant, le ventricule droit (VD) est peu visible. Il donne naissance à l'artère pulmonaire. Entre l'artère pulmonaire et la veine cave inférieure, apparaît la face droite de l'aorte ascendante (Ao)[2].

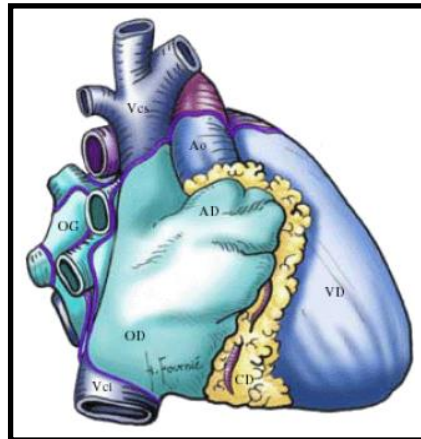


Figure. 2 : vue droite du cœur.

La partie gauche du cœur est formée surtout par le ventricule gauche (VG). En avant de lui, apparaît l'infundibulum du ventricule droit (VD) d'où naît l'artère pulmonaire (AP) qui se bifurque en branches droite et gauche, au-dessous de la crosse aortique (Ao) à laquelle elle est reliée par le ligament artériel. En arrière du ventricule gauche, apparaît la face gauche de l'oreillette gauche (OG) ou atrium gauche qui reçoit les deux veines pulmonaires gauches et se prolonge vers le haut et l'avant par l'auricule gauche (AG). Dans le sillon inter-ventriculaire antérieur, chemine l'artère de même nom (IVA) accompagnée par la grande veine coronaire qui passe ensuite dans le sillon auriculo-ventriculaire gauche et se termine dans le sinus veineux coronaire[2].

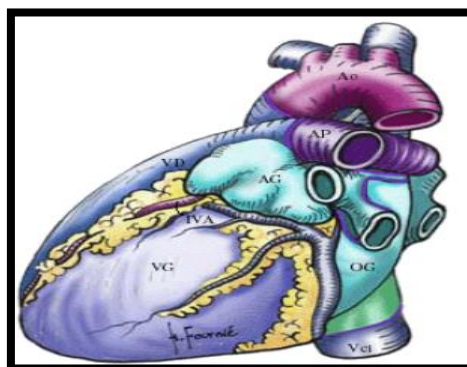


Figure. 3 : vue gauche du cœur.

3. La physiologie cardiaque

Le septum évite le passage direct du sang. Les valves assurent le passage unidirectionnel coordonné du sang depuis les oreillettes vers les ventricules. Le cœur droit est dit veineux et le cœur gauche est dit artériel. Les parois des ventricules sont plus épaisses, et leur contraction est plus importante pour la distribution du sang contre la résistance artérielle. Du sang appauvri en oxygène par son passage dans le corps entre dans l'atrium droit par trois veines, la veine cave supérieure, la veine cave inférieure et le sinus coronaire. Le sang passe ensuite vers le ventricule droit. Celui-ci le pompe vers les poumons par l'artère pulmonaire. Après avoir perdu son dioxyde de carbone dans les poumons et s'y être pourvu de dioxygène, le sang passe par les veines pulmonaires vers l'oreillette gauche, le sang oxygéné entre dans le ventricule gauche, celui-ci est la chambre pompante principale, ayant pour but d'envoyer le sang par l'aorte vers toutes les parties du corps [2].

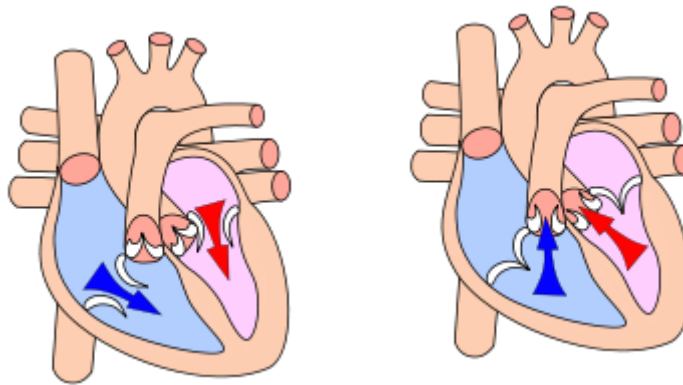


Figure. 4 : La physiologie du cœur.

4. Etude d'un cycle cardiaque

Chaque battement du cœur entraîne une séquence d'événements mécaniques et électriques collectivement appelés cycle cardiaque [5].

4.1. L'activité mécanique du cœur

L'activité mécanique du cœur est engendrée par une activité électrique dont le foyer est le nœud sinusal situé dans l'oreillette gauche, et qui se propage ensuite vers l'oreillette droite et les ventricules. L'essentiel de cette activité mécanique est la contraction simultanée des deux oreillettes puis la contraction puis la contraction simultanée des deux ventricules. La phase du cycle pendant laquelle le myocarde se contracte et appelée systole, il se relâche pendant la phase de la diastole. Les activités auriculaires et ventriculaires ne sont pas simultanées [3].

4.2. L'activité électrique du cœur

Le cœur comprend un système de conduction électrique automatique qui assure chacun de ses battements. La contraction du muscle cardiaque est régie par une impulsion électrique qui prend sa naissance dans le nœud sinusal (ou de Keith et Flack) situé au niveau de la veine cave supérieure, dans l'oreillette droite comme illustré dans la figure. 4. Ce nœud est constitué d'un ensemble de cellules auto-excitables qui génère un courant de dépolarisation 60 à 100 fois par minute. Ce nœud est considéré comme le "pacemaker" du cycle cardiaque.

Cette excitation électrique se propage dans les deux oreillettes induisant la systole auriculaire et arrive au nœud auriculo-ventriculaire (ou d'Achoff-Tawara) seul point de passage entre les oreillettes et les ventricules situé dans le septum inter-ventriculaire (Figure. 4).

A ce niveau, l'excitation électrique marque une courte pause avant de passer au faisceau de His et par suite au réseau de Purkinje induisant la systole ventriculaire. Cette pause est très importante pour permettre une stimulation retardée par rapport aux oreillettes ce qui permet aux ventricules d'être complètement remplis au moment de la contraction auriculaire [1].

Ce système électrique explique la régularité du rythme cardiaque et assure la coordination des contractions auriculo-ventriculaires [1].

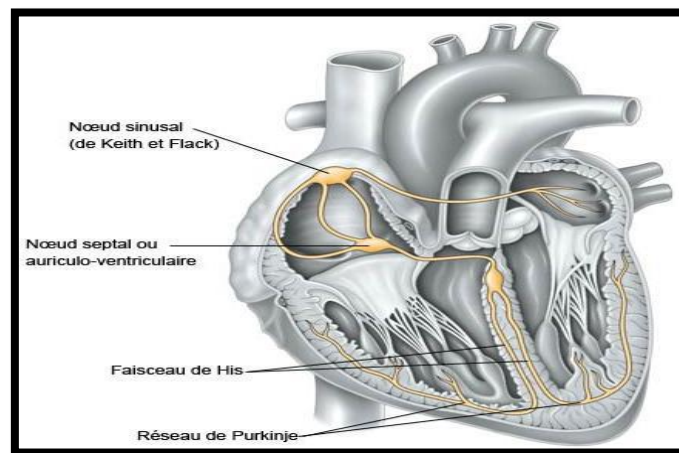


Figure. 5 : Fonctionnement électrique.

5. Electrocardiogramme (ECG)

Le terme électrocardiogramme (ECG) désigne en général les techniques permettant de visualiser les différences de potentiel qui résultent de l'excitation du cœur. Ces potentiels naissent à la limite entre les zones excitées et celles non excitées du myocarde et sont mesurés entre deux points de la surface du corps. Une fibre cardiaque en cours de dépolarisation peut être assimilée à un dipôle électrique. A un instant donné le front de l'onde d'activation formé par l'ensemble des dipôles élémentaires crée un champ électrique. L'enregistrement de

l'évolution temporelle du champ électrique résultant, effectué au moyen d'électrodes cutanées, est nommé l'électrocardiogramme de surface et appelé simplement ECG [6].

5.2. Les caractéristiques de l'ECG

Le processus de dépolarisation et de repolarisation des structures myocardiques se présente dans l'ECG comme une séquence de déflexions ou ondes superposées à une ligne de potentiel zéro, appelée ligne isoélectrique ou ligne de base. Ces déflexions sont dites positives si elles sont situées au-dessus de la ligne isoélectrique sinon elles sont dites négatives. Pour chaque battement cardiaque l'ECG enregistre principalement trois ondes successives comme montré sur la figure. 6.

5.2.1. Onde P

C'est la première onde détectable. Elle apparaît quand l'impulsion électrique se propage à partir du nœud sinusal pour dépolariser les oreillettes (voir Figure. 6). Sa masse musculaire relativement faible entraîne une variation de potentiel faible (moins de 0.25 mv). La progression de l'onde de dépolarisation dans les oreillettes est beaucoup plus lente que dans les ventricules. Par conséquent, la région des oreillettes autour du nœud sinusal est dépolarisée très en avance par rapport aux régions plus éloignées. Puis, le front de repolarisation prend le même sens que celui de la dépolarisation et le vecteur résultant instantané est orienté vers le nœud sinusal. Cela produit une onde de repolarisation à l'inverse de l'onde de dépolarisation P. Normalement, l'onde de repolarisation des oreillettes apparaît au moment où le complexe QRS est produit. Comme ce signal est beaucoup plus intense que le premier, l'onde de repolarisation est cachée [10].

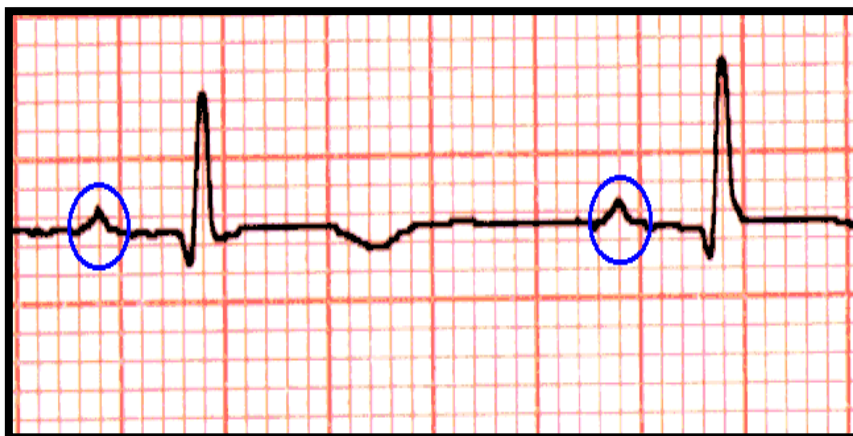


Figure. 6: l'onde P [4].

5.2.2. Le complexe QRS

C'est un ensemble de déflexions positives et négatives qui correspondent à la contraction des ventricules. Pour un cas normal, il a une durée inférieure à 0.12 seconde et son amplitude variable est comprise entre 5 et 20 mV.

Il est constitué de trois ondes :

- ✓ **L'onde Q** : première déflexion négative ;
- ✓ **L'onde R** : première déflexion positive ;
- ✓ **L'onde S** : déflexion négative qui suit l'onde R.

Sa forme est variable selon les dérivations utilisées (emplacement des électrodes sur le corps). La morphologie des complexes QRS selon que le battement cardiaque est normal ou arythmique.



Figure. 7: Le complexe QRS[4].

5.2.3. L'onde T

Correspond au courant de repolarisation des ventricules. Cette onde succède le complexe QRS. L'onde T normale a une amplitude plus faible que le complexe QRS [4].

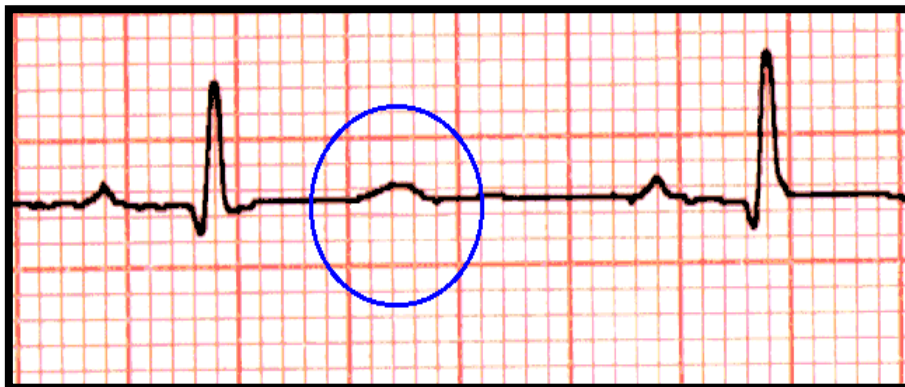


Figure. 8: l'onde T.

5.2.4. L'intervalle PR

Cet intervalle mesure la durée entre le début de l'onde P et le début de Q ou de R. Sa durée est comprise entre 120 et 180 ms. La phase de repolarisation ST-T est beaucoup plus longue (300-400 ms) que la phase de dépolarisation ventriculaire (85-95 ms) [4].

6. Les maladies cardiaques

6.1. Extrasystole ventriculaire (ESV)

L'extrasystole ventriculaire (ESV) est un battement anormal suffisamment répandu pour être présenté en introduction. Les ESV s'observent sur quasiment tous les enregistrements, principalement en période de récupération après un effort. Bien que leur présence n'indique aucune pathologie particulière, si, de façon récurrente, leur nombre par minute est supérieur à 6, elles peuvent être un signe précurseur d'une tachycardie ventriculaire, qui elle, constitue une pathologie majeure. Contrairement aux battements normaux qui ont pour origine la dépolarisation des cellules sinusales, l'ESV naît de la dépolarisation spontanée d'un petit groupe de cellules ventriculaires, appelé alors foyer ectopique ventriculaire. L'impulsion électrique créée n'emprunte pas la voie normale de conduction (faisceau de His), et se propage donc plus lentement dans les ventricules. La contraction ventriculaire ainsi étalée dans le temps perd de son efficacité. Le tracé d'un battement ESV est caractérisé par deux propriétés : l'onde R n'est pas précédée d'une onde P, puisqu'il n'y a pas eu d'activité auriculaire préalable, et la durée du complexe est supérieure à la durée d'un complexe QRS normal (Figure. 9)[7].

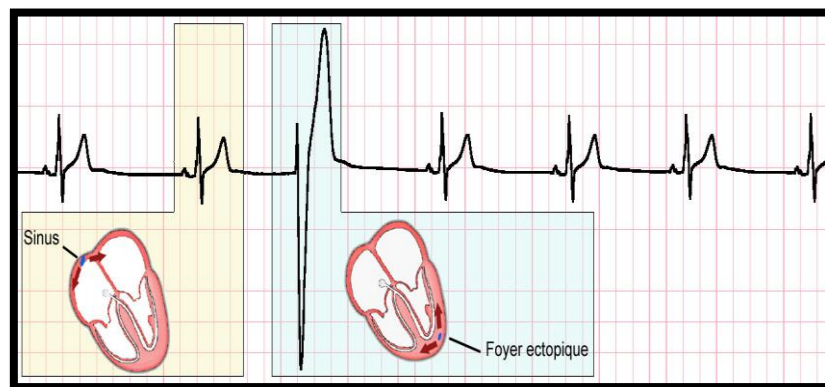


Figure. 9: Extrasystole ventriculaire (ESV)[7].

6.2. La tachycardie ventriculaire

Une tachycardie ventriculaire est une tachycardie dont l'origine est située en dessous de la bifurcation du faisceau de His. L'influx chemine donc entièrement dans le myocarde ventriculaire, en aval du faisceau de His. La tachycardie ventriculaire (TV) représente le type même de l'urgence cardiologique. Sa gravité provient de sa fréquente mauvaise tolérance

hémodynamique, et du risque toujours possible de transformation en fibrillation ventriculaire et de mort subite (Figure. 10)[8].

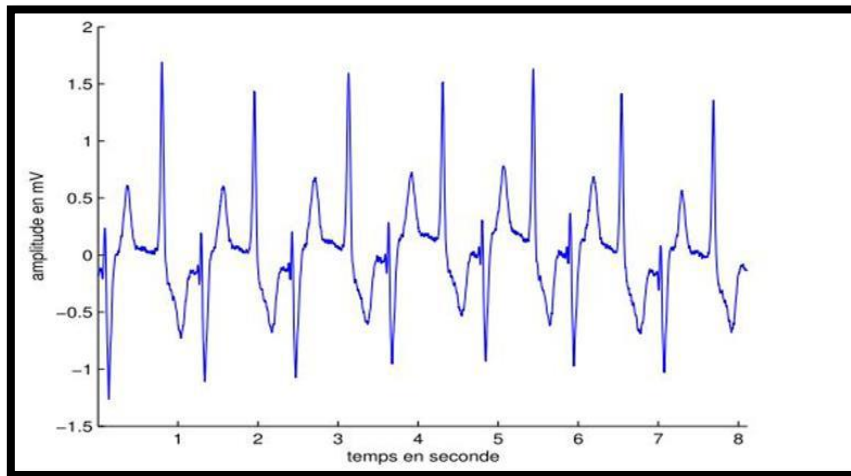


Figure. 10: Tachycardie ventriculaire[8].

6.3. La fibrillation atriale (Fibrillation auriculaire)

La fibrillation atriale est une arythmie supra ventriculaire sans aucune organisation. La dépolarisation est fractionnée en une multitude de fronts de directions et d'amplitudes différentes, réalisant une activité électrique au niveau des oreillettes totalement désordonnée. Cette activité, le plus souvent soutenue, ne laisse au myocarde auriculaire aucun « repos » électrique. Elle se traduit sur l'ECG par la disparition des ondes auriculaires organisées au profit d'une activité continue ressemblant à une sorte de sinusoïde irrégulière. La fréquence de dépolarisation est variable d'un point à l'autre au sein des oreillettes, mais de toutes façons le plus souvent très élevée. De multiples fronts de dépolarisation se présentent ainsi au niveau du nœud auriculo ventriculaire qui remplit son rôle de « filtre » en ne laissant passer de façon aléatoire que quelques fronts, la fréquence des ventricules devenant complètement irrégulière, généralement entre 90 et 140/min en fonction de l'état de perméabilité du nœud.

L'absence des systoles auriculaires n'est pas grave en soi, mais elle est cependant responsable d'une baisse significative de l'efficacité cardiaque à deux niveaux : d'une part, parce que le cœur ne bénéficie pas de la systole auriculaire qui assure une partie du remplissage sanguin des ventricules, et, d'autre part, parce que le rythme moyen est généralement supérieur à 100 bat/min et peut atteindre 200 bat/min. Le risque majeur lié à cette pathologie est celui du flutter, c'est-à-dire de l'envoi possible, dans la circulation, d'embolies formés au niveau des oreillettes (particulièrement préjudiciables s'il s'agit de l'oreillette gauche, puisque le ventricule gauche envoie le sang dans la circulation générale, et en priorité vers le cœur et le cerveau) [1].

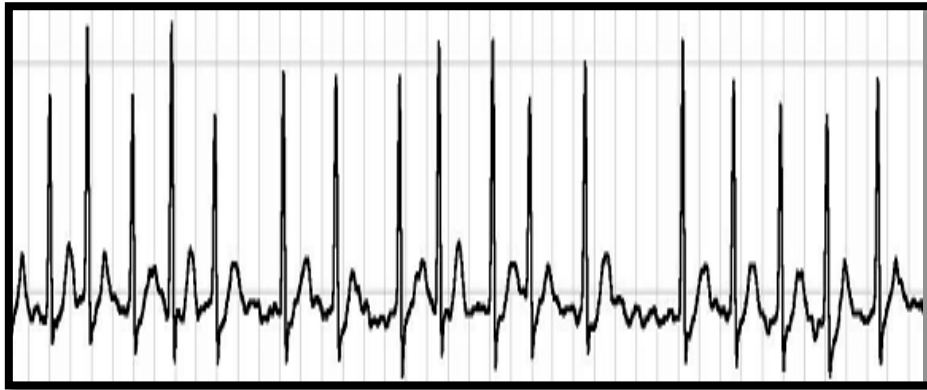


Figure. 11 : La fibrillation atriale.

6.4. Tachycardie ventriculaire

C'est un rythme cardiaque rapide, régulier ou non, et qui prend son origine dans les ventricules (au-delà de 100 battements à la minute). Cette forme d'arythmie ne représente que la minorité des cas d'arythmie et reflète en général un mauvais fonctionnement des ventricules. Une personne atteinte de tachycardie ventriculaire se sentira mal, sera pâle, aura la peau moite et des palpitations. Si la fréquence cardiaque est très rapide, elle peut même souffrir d'étourdissements et d'évanouissement [6].

7. Le standard AAMI

Pour pouvoir classifier les battements cardiaques, il faut commencer par définir des classes. De nombreuses classifications peuvent être trouvées dans la littérature. Dans ce mémoire de fin d'étude, c'est la classification recommandée par l'AAMI (*Association for the Advancement of Medical Instrumentation*) qui a été choisie :

- ✓ **Classes N** : battement normal ;
- ✓ **Classe S** : battement ectopique supra ventriculaire ;
- ✓ **Classe V** : battement ectopique ventriculaire ;
- ✓ **Classe F** : battement de fusion ;
- ✓ **Classe Q** : rythme inconnu.

8. La base de données MIT-BIH Physio-Bank

Pour la validation de ce travail nous avons choisi d'utiliser la base de données des signaux électrocardiogrammes MIT-BIH Physio-Bank. Cette base de données contient 48 extraits d'une demi-heure d'enregistrements ECG ambulatoires à deux canaux (DII et V5). Vingt-trois enregistrements ont été choisis au hasard parmi une série de 4000 enregistrements d'ECG ambulatoires 24 heures sur 24 recueillis auprès d'une population mixte de patients

hospitalisés (environ 60%) et de patients externes (environ 40%) à l'hôpital Beth Israel de Boston; les 25 enregistrements restants ont été sélectionnés dans le même groupe pour inclure des arythmies moins communes mais cliniquement significatives qui ne seraient pas bien représentées dans un petit échantillon aléatoire. Chaque enregistrement est échantillonné à une fréquence de 360 Hz par canal avec une résolution de 11 bits sur une plage de 10 mV. L'avantage majeur de cette base de données est qu'elle contient un grand nombre de pathologies cardiaques, ce qui permet de valider les algorithmes sur un grand nombre de signaux ECG. Les enregistrements sont numérotés de 100 à 124 pour le premier groupe qui comporte une variété de formes d'ondes et de 200 à 234 pour le deuxième qui comporte une variété de cas pathologiques. Chaque enregistrement a été annoté indépendamment par plusieurs cardiologues ce qui permet d'avoir des études plus fiables. L'annotation correspond à l'instant d'apparition du pic R et au type d'arythmie. Pour chaque enregistrement de la base de données, il existe trois fichiers différents avec les extensions suivantes : `.dat`, `.hea` et `.atr` (Figure. 12).

Reference annotations	Signals	Header
100 atr	100.dat	100.hea
101 atr	101.dat	101.hea
102 atr	102.dat	102.hea
103 atr	103.dat	103.hea
104 atr	104.dat	104.hea
105 atr	105.dat	105.hea
106 atr	106.dat	106.hea
107 atr	107.dat	107.hea
108 atr	108.dat	108.hea
109 atr	109.dat	109.hea
111 atr	111.dat	111.hea
112 atr	112.dat	112.hea
113 atr	113.dat	113.hea
114 atr	114.dat	114.hea
115 atr	115.dat	115.hea
116 atr	116.dat	116.hea

Figure. 12 : différents fichiers de la base de données MIT-BIH PhysioBank.

8.1. Le fichier de données (.dat)

Contient les données numérisées du signal ECG, deux différents signaux correspondants aux deux dérivations (MLII et V5) sont stockés dans le même fichier (Ex : 233.dat) (Figure. 13).

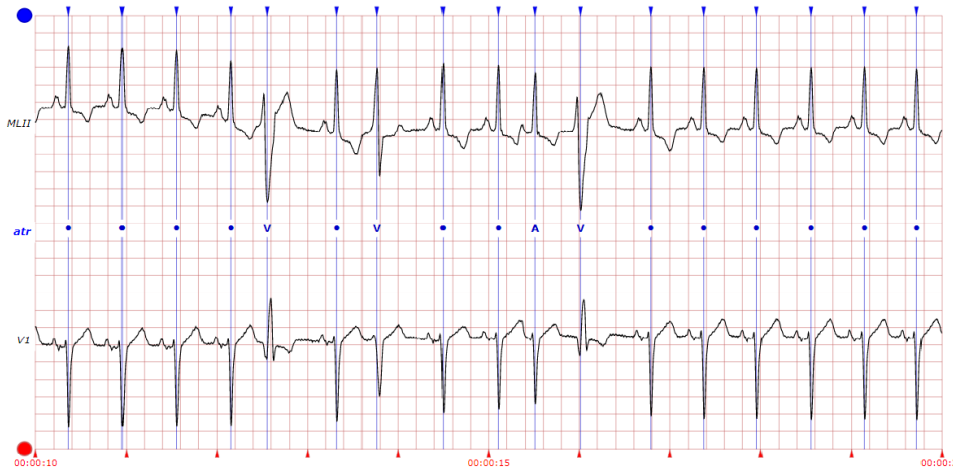


Figure. 13 : L'enregistrement 233.dat de la base de données MIT-BIH PhysioBank.

8.2.Le fichier (.hea)

Contient les paramètres d'interprétation du fichier de données correspondantes tels que : la fréquence d'échantillonnage, gain d'amplification, dérivations utilisées, nombre d'échantillons, pathologie,,etc (Figure. 14).

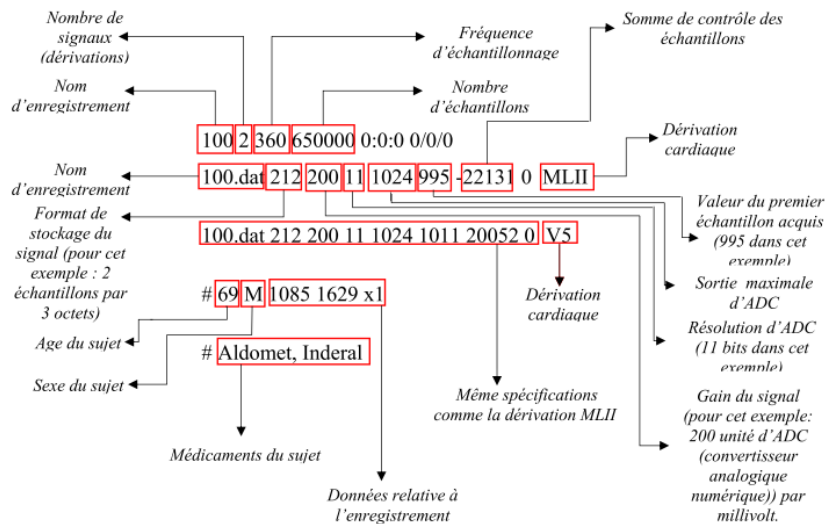


Figure. 14 : Fichier :hea de l'enregistrement 100.dat de la base de données MIT-BIH PhysioBank.

8.3.Le fichier (.atr)

Comporte les positions ou les moments d'apparition des pics R du signal ECG, ces emplacements ont été marqués manuellement par plusieurs cardiologues, on donne également

dans ce fichier un nombre indiquant l'ordre du pic R ainsi qu'une marque correspondant au type d'arythmie (Figure. 13,Tableau.1).

Le Tableau.1représente les différents codes et leurs descriptions des annotations de la base de données MIT-BIH PhysioBank :

Table. 1 : Annotations de la base de données MIT-BIH PhysioBank.

Code	Description
N	Battement Normal (<i>Normal beat</i>)
L	Battement de bloc de branche gauche (<i>Left bundle branch block beat</i>)
R	Battement de bloc de branche droit (<i>Right bundle branch block beat</i>)
B	Battement de bloc de branchement (non spécifié) (<i>Bundle branch block beat (unspecified)</i>)
A	Battement prématuré auriculaire (<i>Atrial premature beat</i>)
a	Battement prématuré auriculaire aberré (<i>Aberrated atrial premature beat</i>)
J	Battement prématuré nodal (jonctionnel) (<i>Nodal (junctional) premature beat</i>)
S	Battement prématuré ou ectopique supraventriculaire (auriculaire ou nodal) (<i>Supraventricular premature or ectopic beat (atrial or nodal)</i>)
V	Contraction ventriculaire prématurée (<i>Premature ventricular contraction</i>)
r	R-on-T contraction ventriculaire prématurée (<i>R-on-T premature ventricular contraction</i>)
F	Fusion du battement ventriculaire et normal (<i>Fusion of ventricular and normal beat</i>)
e	Battement d'échappement atriale (<i>Atrial escape beat</i>)
j	Battement d'échappement nodal (jonctionnel) (<i>Nodal (junctional) escape beat</i>)
n	Battement d'échappement supraventriculaire (atrial ou nodal) (<i>Supraventricular escape beat (atrial or nodal)</i>)
E	Battement d'échappement ventriculaire (<i>Ventricular escape beat</i>)
f	Fusion du battement rythmé et normal (<i>Fusion of paced and normal beat</i>)
Q	Battement inclassable (<i>Unclassifiable beat</i>)

9. Conclusion

Plusieurs activités électriques peuvent être détectées au niveau du corps humain (ECG,EMG, EEG). Le signal électrocardiogramme (ECG) est considéré dans notre cas. L'électrocardiogramme représente l'activité électrique du cœur. Ce signal peut être utiliser dans le domaine médical pour détecter les anomalies de fonctionnement du cœur (diagnostic). L'électrophysiologie cardiaque a été expliquée en détail dans ce chapitre où nous avons essayé d'expliquer et de vulgariser toutes les notions liées au signal électrocardiogramme, ainsi que la base de données MIT-BIH PhysioBank d'où nous avons pris les signaux sur lesquels nous avons validé nos expérimentations.

CHAPITRE II

Notions et techniques d'extraction des caractéristiques

Le diagnostic des maladies du cœur est la dernière étape de toute une chaîne d'acquisition qui commence par les capteurs, amplification, conversion analogique numérique, filtrage et en fin le diagnostic par un système expert contenant une base de données. Cette opération se compose de deux étapes clés qui sont l'extraction des caractéristiques et la classification. Dans ce chapitre nous allons présenter les notions de base relatives à la classification.

Sommaire

- 20. Introduction
 - 21. Définition et intérêt
 - 22. Méthodes d'extraction des caractéristiques
 - 23. Conclusion
-

1. Introduction

L'ECG est un signal bioélectrique facilement accessible qui offre des informations importantes concernant les anomalies cardiaques, ce signal permet de représenter la contraction cyclique et la relaxation des muscles cardiaque humains. L'activité du muscle cardiaque est contrôlée par des impulsions électriques qui sont transmises à travers un réseau nerveux, ces impulsions électriques sont suffisamment fortes peut être détectées par des électrodes placées sur la peau humaine, Le complexe QRS est la partie la plus significative du signal ECG, il donne les informations utiles sur le comportement cardiaque. Par conséquent, la détection fiable du complexe QRS demeure un domaine de recherche très important, cette détection est difficile, et cela est dû principalement à la morphologie du complexe qui change de manière significative d'un état sain à un état malade. Les chercheurs ont développé plusieurs méthodes de l'extraction de caractéristiques de l'ECG. Chaque méthode possède ses avantages et ses inconvénients en termes de performance et complexité. Mais avec le développement des outils informatiques (matériel/logiciel), le poids de la complexité a diminué par rapport à la performance.

2. Définition et intérêt

L'étape d'extraction des caractéristiques est la clé du succès de la classification des battements cardiaques à l'aide du signal ECG. Toute information extraite de la pulsation utilisée pour discriminer son type peut être comme une caractéristique. Les caractéristiques peuvent être extraites sous différentes formes directement à partir de la morphologie du signal ECG dans le domaine temporel et/ ou dans le domaine fréquentiel.

L'extraction des caractéristiques est un processus de réduction de dimensionnalité par lequel un ensemble initial de données brutes est réduit à des groupes plus gérables pour le traitement, qui va souvent de pair avec la classification. Cette étape réduisant efficacement la quantité de données à traiter, tout en décrivant toujours avec précision et complètement l'ensemble de donnée d'origine. L'étape de l'extraction est utile lorsque vous devez réduire le nombre de ressources au traitement sans perdre d'informations importantes ou pertinentes.

Une caractéristique est une valeur décrivant une propriété des signaux ECG, les caractéristiques sont généralement regroupées dans un vecteur qu'on appelle 'vecteur de caractéristiques [9]. Après les opérations de prétraitement (filtrage...etc.), il est nécessaire d'extraire les caractéristiques des signaux ECG afin de les utiliser dans la phase de classification. Les approches existantes ont été regrouper en deux catégories principales : fiduciaire et non fiduciaire [10].

- **Approches fiduciaires** : sont les méthodes les plus utilisées pour faire l'extraction des attributs d'un signal ECG (les pics, les intervalles, les pentes ou d'autres mesures d'une onde) qui sont simplement basées sur l'information temporelle et morphologique lesquelles apparaissent au fur et à mesure d'un rythme cardiaque ECG.
- **Approches non fiduciaires** : les caractéristiques naturelles d'un signal ECG ont fait un grand effet sur la localisation d'une information invariante dans le temps, celle-ci peut être considérée un challenge dans le traitement des signaux non stationnaires, l'information fréquentielle a montré un atout dans plusieurs domaines, la transformée de Fourier, l'ondelette...etc. sont des outils utilisés pour extraire ces informations [10].

Le but du processus d'extraction de caractéristiques est de sélectionner et de conserver les informations pertinentes à partir du signal d'origine.

3. Méthodes d'extraction des caractéristiques

3.1. Méthodes basées sur du filtrage dérivatif

Il est généralement supposé que les débuts et les fins d'ondes de l'ECG soient liés à des changements brusques dans les variations du signal ECG. De ce fait, plusieurs méthodes pour la segmentation de l'ECG basées sur la différentiation numérique ont été proposées afin d'exploiter les changements de pente du signal. Ce type de méthodes a été largement utilisé pour la délimitation des ondes cardiaques. Pour ces méthodes, les hypothèses sur le modèle d'observations sont :

- ✚ Le signal est partiellement connu ;
- ✚ On a cependant quelques informations a priori sur les caractéristiques du signal, notamment l'intervalle de temps où l'onde d'intérêt se trouve ;
- ✚ Le signal peut subir des changements de forme et de morphologie,
- ✚ Les signaux parasites n'existent pas.

Le signal ECG est traité par un filtre dérivatif $G_1(z)$ puis par un filtre de lissage $G_2(z)$ tels que ;

$$G_1(Z) = 1 - Z^{-6}. \quad (1)$$

$$G_2 = \frac{1-Z^{-8}}{1-Z^{-1}} \quad (2)$$

Cette méthode permet de délimiter les complexes QRS et les ondes T [11].

3.2. Extraction par Analyse en composante principale (ACP)

L'ACP est une méthode mathématique linéaire qui appartient de la famille d'analyse de donnée qui consiste à transformer des variables liées entre elles qu'on appelle des variables corrélées. L'ACP est une technique bien établie pour l'extraction des caractéristiques pertinentes de l'ensemble de données ECG, son principe est de rechercher les directions de l'espace (les axes) qui maximise la variance des données et minimise l'écart au carré par rapport aux axes [12].

La première composante principale présente la variabilité la plus élevée, et la deuxième composante est orthogonale à la première composante principale et ainsi de suite. Le pourcentage de la variabilité totale des données est fixé comme un seuil pour sélectionner le nombre de composantes principales.

3.2.1. Principe

Soit le signal électrocardiogramme donné par le vecteur suivant :

$$y(k) = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(M) \end{bmatrix} \quad (3)$$

Où M est le nombre d'échantillons du signal ECG. L'ensemble des battements cardiaques est représenté par la matrice :

$$Y = [y_1, y_2, \dots, y_N] \quad (4)$$

L'ACP comprend les étapes suivantes :

1. Calcule du vecteur moyen, le vecteur moyen de chaque battement du cœur est calculé par l'équation suivante :

$$\bar{y} = \frac{1}{M} \sum_{i=1}^M y_i \quad (5)$$

2. Calcule des données ajustées moyennes comme suit :

$$y_{adj_i} = y_i - \bar{y} \quad (6)$$

$$Y_{adj} = [y_{adj_1} y_{adj_2} \dots y_{adj_N}] \quad (7)$$

3. Calcule de la matrice de covariance :

$$C = \frac{1}{M-1} \sum_{i=1}^M (y_i - \bar{y})^T (y_i - \bar{y}) \quad (8)$$

4. Calcule des vecteurs propres e_i et les valeurs propres λ_i de la matrice de covariance :

$$C.e_i = \lambda_i . e_i, \quad i=1, \dots, N \quad (9)$$

5. Choisir des composants et former un vecteur caractéristique. Le vecteur propre avec la valeur la plus élevée est le composant principale. Ensuite, les vecteurs propres sont classés par valeurs propres du plus élevé au plus bas, ce qui renvoie les

composants par ordre d'importance. Ainsi le pourcentage de variance r_K de chaque valeur propre est obtenu comme suit :

$$r_K = \frac{\sum_{i=1}^K \lambda_i}{\sum_{i=1}^N \lambda_i} . \quad (10)$$

6. Dérivation du nouvel ensemble de données par la formule suivante :

$$\text{Données finales} = \text{vecteur d'entité de ligne} * \text{ajustement des données de ligne} \quad (11)$$

Où le vecteur d'entité de ligne est la matrice avec des vecteurs propres dans les colonnes transposées. Une hypothèse faite pour l'extraction de caractéristiques par PCA est que la plupart des informations des vecteurs d'observation sont contenues dans le sous-espace couvert par les m premiers axes principaux, ou $m < p$ pour un espace de données à p dimensions. Par conséquent, chaque vecteur de données d'origine peut être représenté par son vecteur composant principale de dimensionnalité m .

3.3. Extraction de caractéristiques par les filtres numériques

Cette méthode consiste à appliquer un filtre différentiateur au signal ECG dont l'équation aux différences est décrite comme suit :

$$Y_0(n) = X(n) - X(n-4), \quad 4 < n < 8991. \quad (12)$$

L'étape suivante consiste à appliquer un deuxième filtre passe bas sur le signal $Y_0(n)$ décrit par l'équation aux différences :

$$Y_1(n) = Y_0(n) + 4 * Y_0(n-1) + 6 * Y_0(n-2) + 4 * Y_0(n-3) + Y_0(n-4); \quad (13)$$

Le signal filtré $Y_1(n)$ est scruté à la recherche du premier point i dont l'amplitude est supérieure au seuil $\alpha_1 = 21$, ce point constitue alors le début d'une zone de recherche de $160ms$. Chaque point est classé comme appartenant à la ligne isoélectrique, ou à la zone QRS, ou comme étant un bruit selon un nombre des dépassement alternés des seuils α_1 et d'un deuxième seuil de magnitude négative $\alpha_2 = -21$. Si aucun dépassement des seuils n'est enregistré alors le point i sera classé appartenant à la ligne isoélectrique.

Par la suite on introduit une méthode qui consiste à lisser le signal ECG en utilisant un filtre à moyenne glissante définie par l'équation aux différences suivantes :

$$Y_0(n) = \frac{|x(n-1) + 2x(n) + x(n+1)|}{4}, \quad 1 < n < 8190 \quad (14)$$

L'étape suivante consiste à appliquer au signal $Y_0(n)$ résultat un filtre passe bas dont l'équation aux différences est décrite comme suit :

$$YI(n) = [1/(2m+1)] * \sum_{k=n-m}^{n+m} Y_0(k), \quad m < n < 8191 - m \quad (15)$$

La différence entre l'entrée et la sortie du filtre passe bas est calculée et élevée au carré conformément à l'équation :

$$Y_2(n) = (Y_0(n) - Y_1(n))^2, m < n < 8191 - m \quad (16)$$

Le résultat est à son tour filtré par l'équation :

$$Y_3(n) = Y_2(n) * (\sum_{k=n-m}^{n+m} Y_2(k))^2, m < n < 8191 - m \quad (17)$$

Un quatrième vecteur $Y_4(n)$ est calculé comme suit :

$$\begin{cases} Y_4(n) = Y_3(n) & \text{si } [Y_0(n) - Y_0(n - m)][Y_0(n) - Y_0(n - n + m)] > 0 \\ Y_4(n) = 0 & \text{sinon} \end{cases} \quad (18)$$

Finalement un QRS candidat est détecté pour tout point de vecteur $Y_4(n)$ qui dépasse le seuil décrit par l'équation suivante [3] :

$$\text{Seuil} = 0.125 * \max [Y_4(n)] \quad (19)$$

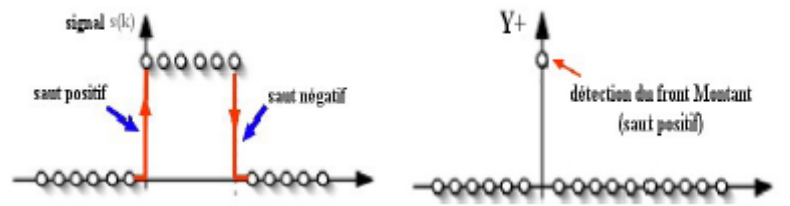
3.4. L'extraction des caractéristiques par la méthode NFLS

L'approche NFLS (*Nonlinear Filtering Scheme*) réalise un filtrage non linéaire, on utilise cette approche pour localiser le complexe QRS et la délimitation des ondes P et T en localisant leur début et fin, mais aussi pour le dé-bruitage de ces signaux.

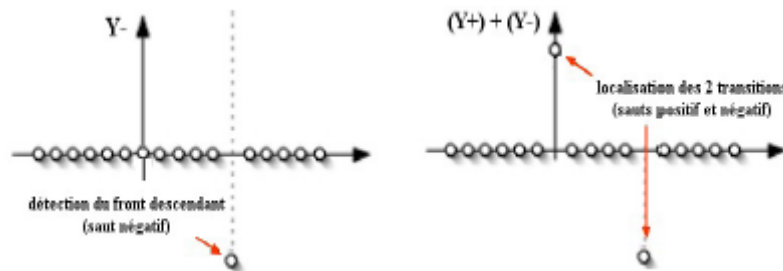
3.4.1. Principe

Le principe de l'approche NFLS est basé sur la détection des transitions d'un signal échantillonné $S(k)$. Le signal Y_+ (Figure. 1) représente les points montants avec une localisation choisie sur le point haut de la transition (saut positif), et le signal Y_- contient les détections des points descendants montants avec une localisation choisie sur le point bas de la transition (saut négatif). La somme ($Y_+ + Y_-$) permet de localiser toutes les transitions du signal. Deux filtres de détection $F_+(z)$ et $F_-(z)$ sont proposées [5] :

$$\begin{cases} F_+(z) = 1 - z^{-1} \\ F_-(z) = F(z - 1) \end{cases} \quad (20)$$



(a) exemple de signal avec un saut positif et un saut négatif (b) localisation de la transition montante (saut positif) dans le signal Y_+



(c) Localisation de la transition descendante (saut négatif) dans le signal Y_- signal initial par la somme: $Y_+ + Y_-$

Figure. 1 : Localisation des transitions d'un signal échantillonné.

La réponse de ces deux filtres est sélectionnée à l'aide d'un opérateur de seuillage et permet donc d'obtenir deux signaux $Y_+ + Y_-$ contenant les sauts positifs et négatifs du signal. Les expressions de ces composantes décrivent comme suit :

$$\begin{cases} Y_+ = T(F_+(z)S(z)) \\ Y_- = -T(F_-(z)S(z)) \end{cases} ; \quad (21)$$

L'opérateur de seuillage est défini par l'expression suivante :

$$T(x) = \begin{cases} x & \text{si } x \geq 0 \\ 0 & \text{si non} \end{cases} \quad (22)$$

Ces étapes concernent la localisation des pics R, il reste que la localisation de complexe, on applique l'approche NFLS sur le signal ECG dont il extrait la composante qui contient les transitions positives correspondant aux segments QR du complexe QRS et la composante qui contient les transitions négatives correspondant aux segments RS du complexe QRS [5].

3.5. L'extraction par la méthode Pan and Tompkins

Les travaux de *Pan and Tompkins* influencent grandement la détection du complexe QRS par rapport à d'autre. Cette technique est l'un des algorithmes les plus importants pour la détection des complexes QRS. Partout où la précision de l'extraction de la forme d'onde de l'électrocardiogramme (ECG) joue un rôle essentiel dans l'amélioration du diagnostic de toutes maladies cardiaques. L'ECG normal devrait être composé de

plusieurs parties, notamment l'onde P, le complexe QRS et l'onde T, ces ondes reflètent l'activité du cœur, telle que l'onde P produite par la contraction musculaire des oreillettes et sa durée indique l'élargissement des oreillettes. L'onde Q donne la première valeur négative et est généralement censée être inférieure de 25 % à la valeur de l'onde R.

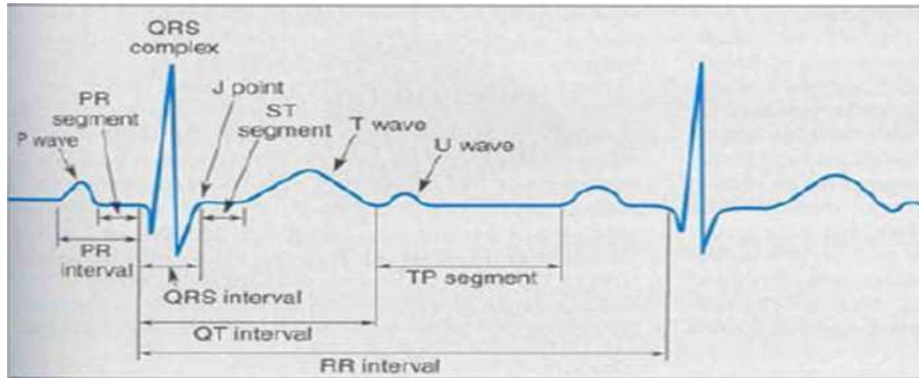


Figure. 2: Caractéristiques d'un battement cardiaque[13].

L'algorithme de Pan and Tompkins est basé sur des analyses numériques des pentes, de l'amplitude et de la largeur. La méthodologie suivie est que l'ECG passe à travers un filtre passe-bande numérique spéciale, Il peut réduire les fausses détections causées par les différents types d'interférences présentes dans le signal ECG. Ensuite le signal filtré passe par les phases de dérivation, de quadrature. Enfin un seuil adaptatif est appliqué.

L'algorithme ajuste automatiquement les seuils et les paramètres périodique pour adapter les changements de la morphologie du QRS.

3.5.1. Les étapes de traitement de la détection du complexe QRS

Les différentes étapes de l'algorithme sont illustrées dans la figure suivante :

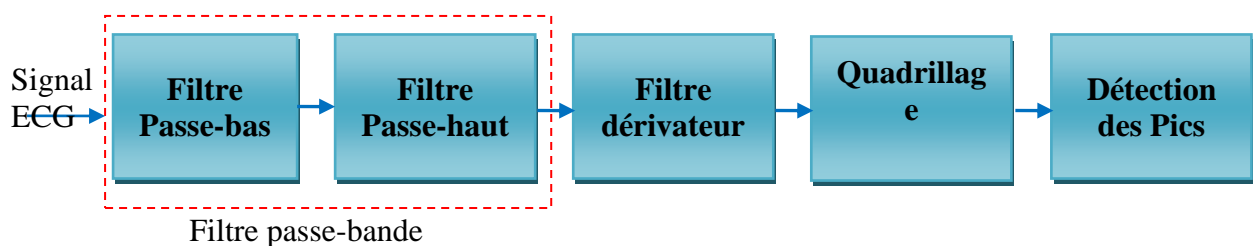


Figure. 3 : Les étapes de la méthode Pan and Tompkins.

Le signal passe par des phases de filtrage, de dérivation, d'élévation au carré avant que les pics soient détectés.

❖ *Le filtre passe-bande*

Dans la première étape, le signal passe à travers un filtre passe-bande. Le filtre passe bande de réduit le bruit dans le signal ECG en s'adapter au spectre du complexe QRS. Ce atténue le bruit dû aux influence musculaire, aux interférences des lignes électriques et à la

déviations de la ligne de base. Le filtre mis en œuvre dans cet algorithme est composé de filtre passe haut et passe bas en cascade.

- ✓ La fonction de transfert de filtre passe bas est décrite comme suit :

$$H(z) = \frac{Y(z)}{X(z)} \quad (23)$$

$$H(z) = (1 - Z^{-6})^2 / (1 - Z^{-1})^2 \quad (24)$$

- ✓ La fonction de transfert de filtre passe haut est décrite comme suit :

$$H(z) = (-1/32 + Z^{-16} - Z^{-17} + Z^{-32}/32) / (1 - Z^{-1}) \quad (25)$$

Où $X(Z)$ et $Y(Z)$ sont respectivement la transformée en Z des signaux d'entrée $x(n)$ et de la sortie $y(n)$.

❖ *Filtre dérivateur*

L'étape de traitement suivante se fait par l'application d'un filtre dérivateur, qui est une technique standard pour trouver les pentes élevées qui distinguent normalement les complexes QRS des autres ondes ECG. La procédure dérivée supprime les composantes basse fréquence des ondes P et T, et fournit un gain important aux composantes haute fréquence provenant des pentes élevées du complexe QRS.

- ✓ La fonction de transfert adopté du filtre dérivateur est donnée par :

$$H(z) = 0.1 * (-2 * Z^{-2} - Z^{-1} + Z^1 + 2 * Z^2) \quad (26)$$

❖ *Quadrillage*

L'opération de quadrillage rend le résultat positif et met en évidence les grandes différences résultant des complexes QRS, les petites différences résultant des ondes P et T sont supprimées. Les composantes haute fréquence du signal liées au complexe QRS sont encore améliorées. Il s'agit d'une transformation non linéaire qui consiste à élever au carré point par point les échantillons du signal.

- ✓ L'équation de cette étape est :

$$y(n) = y(n)^2 \quad (27)$$

Finalement la forme d'onde carrée passe à travers un intégrateur de fenêtre mobile. Ce dernier a pour objectif d'obtenir sous forme d'onde des informations supplémentaires en plus de la pente de l'onde R.

L'étape suivante est l'intégration du signal. Cette opération produit un signal contenant les informations concernant la pente et la largeur du complexe QRS.

La taille de la fenêtre d'intégration doit être adaptée à la largeur moyenne d'un complexe QRS. Si elle est trop grande, le maximum est décalé en temps par rapport à la position du R, au contraire, elle est petite on obtient plusieurs pics pour une même onde R. après l'intégration de signal, le signal disponible possède un maximum absolu pour chaque complexe QRS [8].

Finalement le complexe QRS sera détecté, les pics locaux du signal intégré sont trouvés. Un pic est défini comme le point dont le signal change de direction (de direction croissante à une direction décroissante), il est important de suivre l'évolution des paramètres cycle par cycle afin d'obtenir le vecteur caractéristique qui contient quelques paramètres qui sont : les amplitudes des pics, l'intervalle RR puisque son évolution permet de suivre au cours du temps le rythme cardiaque, et les différents intervalles entre différents pics. Après la détection de ces différents instants caractéristiques de chacune des ondes de l'ECG, nous pouvons maintenant classifie les battements cardiaques à travers le vecteur caractéristique.

4. Conclusion

Dans ce chapitre, nous avons présenté les notions de base relatives à l'étape d'extraction des caractéristiques ainsi quelques méthodes d'extraction. Parmi ces méthodes on trouve la méthode de Pan and Tompkins qu'on va l'utiliser dans le quatrième chapitre pour extraire les caractéristiques des battements. Dans le chapitre suivant nous allons présenter les notions de base relatives à la deuxième étape de l'opération du diagnostic des maladies cardiaques qui est l'étape de classification.

CHAPITRE III

Notions et techniques de classification

Le diagnostic des maladies du cœur est la dernière étape de toute une chaîne d'acquisition qui commence par les capteurs, amplification, conversion analogique numérique, filtrage et en fin le diagnostic par un système expert contenant une base de données. Cette opération se compose de deux étapes clés qui sont l'extraction des caractéristiques et la classification. Dans ce chapitre nous allons présenter les notions de base relatives à la classification.

Sommaire

- 24. Introduction
 - 25. La classification
 - 26. Les techniques de classification
 - 27. Conclusion
-

1. Introduction

La reconnaissance et la classification précise de divers types d'arythmies sont essentielles pour le traitement correct du patient cardiaque. Par conséquent, la détection automatique d'arythmie à partir d'un enregistrement électrocardiographique a été un sujet très important. Plusieurs techniques de classification ont été développées dans la littérature, dans ce chapitre on va présenter les notions de base relatives à la classification ainsi que quelques techniques de classification.

2. La classification

Après la phase d'extraction des caractéristiques il reste donc que la phase de la classification à l'aide de vecteur de caractéristiques qui est trouvé. Le but principal de la phase de classification est de classer les battements cardiaques automatiquement selon le type d'arythmie. Les méthodes de classification peuvent être supervisées ou non supervisées. Les méthodes supervisées demandent à l'utilisateur une description des classes tandis que celles non supervisées sont indépendantes de l'utilisateur. Elles sont plutôt des méthodes de groupement statistique qui trient les objets selon leurs propriétés et forment des ensembles ayant des caractéristiques similaires [14].

3. Les techniques de classification

3.1. La classification par la logique floue

La logique floue fut développée par Lotfi A. Zadeh en 1965 à partir de sa théorie des sous-ensembles flous, la logique floue appelée aussi « logique linguistique » car ses valeurs de vérité sont des mots de langage courant : 'plutôt vrai, presque faux, loin, si loin, pré de, grand, petit...etc.', conduit à ce qu'on appelle la logique floue. En fournissons une base pour un raisonnement approximatif, c'est-à-dire un mode de raisonnement qui n'est ni exacte ni très inexacte, l'approche de la logique floue permet l'introduction du concept de degré d'appartenance, qui détermine les « forces » avec lesquelles un individu appartient aux différentes classes. Cela repose sur le fait que le concept flou ne cherche pas un point de rupture qui décide de l'appartenance d'un individu à une classe, mais qu'elle raisonne plutôt sur la base d'un intervalle de valeurs.

3.1.1. Représentation d'un ensemble flou

L'approche par la logique floue permet l'introduction du concept de degré d'appartenance, qui détermine les « forces » avec lesquelles un individu appartient aux différentes classes. Cela repose sur le fait que le concept flou ne cherche pas un point de

rupture qui décide de l'appartenance d'un individu à une classe, mais qu'elle raisonne plutôt sur la base d'un intervalle de valeurs [14].

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 : $x = [p_1 ; p_2 ; \dots ; p_Q] \in \mathcal{R}^Q$. 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 tel que l'intensité du niveau de gris, l'aspect circulaire d'une cellule, des électrocardiogrammes, etc.

Un classificateur flou est toute application :

$$D : \mathcal{R}^Q \rightarrow [0,1]^M \quad (1)$$

Le résultat de la classification est alors donné par :

$$D(x) = [\mu_1(x) ; \dots ; \mu_M(x)] \quad (2)$$

où $\mu_i(x)$ représente le degré d'appartenance de x à la classe C_i (Figure 1).

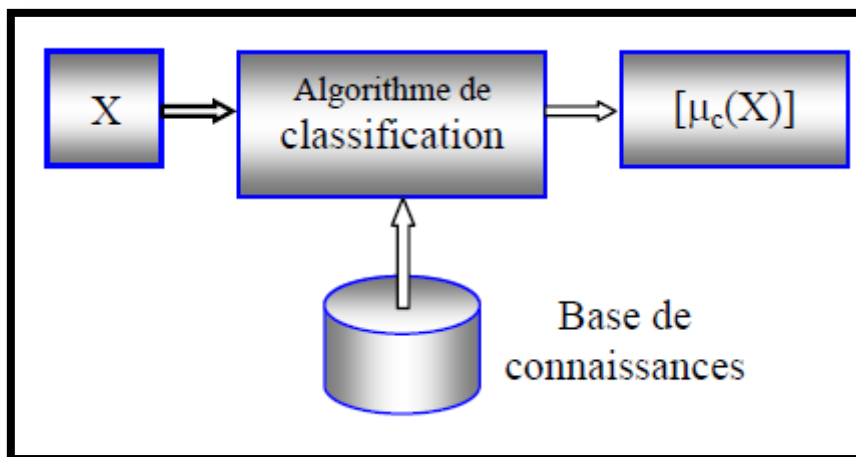


Figure. 1 : Schéma synoptique d'un classificateur flou supervisé.

3.2. La Classification par la machine à vecteur de support (SVM)

3.2.1. Définition

La machine à vecteur de support (SVM) est un algorithme d'apprentissage supervisé capable d'effectuer une classification et une régression. Le SVM est généralement utilisé par les tâches de classification, il est utilisé pour trouver un hyperplan de séparation optimal, qui génère une marge maximale entre deux catégories de données. Pour construire un OSH, SVM

mappe des données dans un espace d'entités de dimension supérieure. SVM effectue ce mappage non linéaire à l'aide d'une fonction de noyau. Ensuite, SVM construit un OSH linéaire entre deux catégories de données dans l'espace d'entité supérieure. Les vecteurs de données qui sont les plus proches de la SST dans l'espace de fonctionnalité supérieure sont appelés vecteur de support (SV) et contiennent toutes les informations requises pour la classification.

3.2.2. Principe et fonctionnement général de SVM

Pour deux classes d'exemples donnés, le but de SVM est de trouver un classificateur qui va séparer les données et maximiser la distance entre ces deux classes. Avec SVM, ce classificateur est un classificateur linéaire appelé hyperplan [15].

La figure suivante représente un exemple d'un hyperplan qui sépare deux ensembles de point.

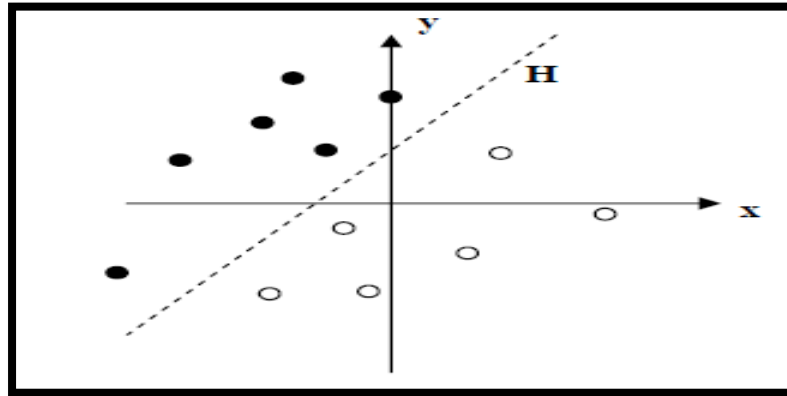


Figure. 2 : Exemple d'un hyperplan séparateur [15].

Les points les plus proches, qui seuls sont utilisés pour la détermination de l'hyperplan, sont appelés vecteurs de support.

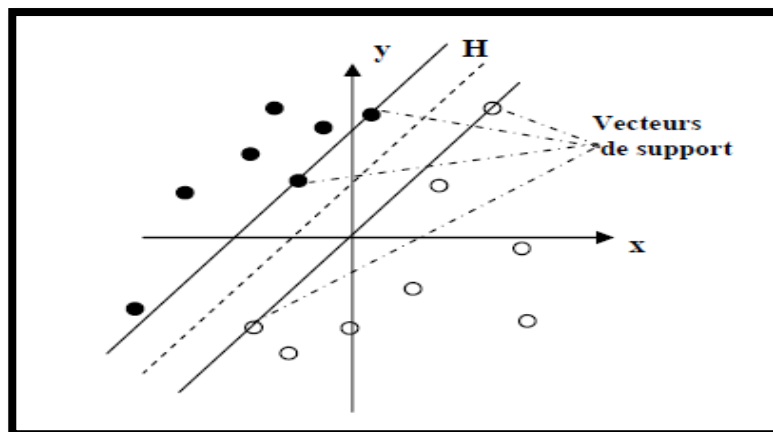


Figure 3 : Exemple de vecteurs de support [15].

Il est évident qu'il existe une multitude d'hyperplan valide mais la propriété remarquable des SVM est que cet hyperplan doit être optimal. Nous allons donc en plus chercher parmi les hyperplans valides, celui qui passe « au milieu » des points des deux classes d'exemples. Intuitivement, cela revient à chercher l'hyperplan le « plus sûr » [16].

En effet, supposons qu'un exemple n'ait pas été décrit parfaitement, une petite variation ne modifiera pas sa classification si sa distance à l'hyperplan est grande. Formellement, cela revient à chercher un hyperplan dont la distance minimale aux exemples d'apprentissage est maximale [16]. On appelle cette distance « marge » entre l'hyperplan et les exemples. L'hyperplan séparateur optimal est celui qui maximise la marge. Comme on cherche à maximiser cette marge, on parlera de séparateurs à vaste marge [16].

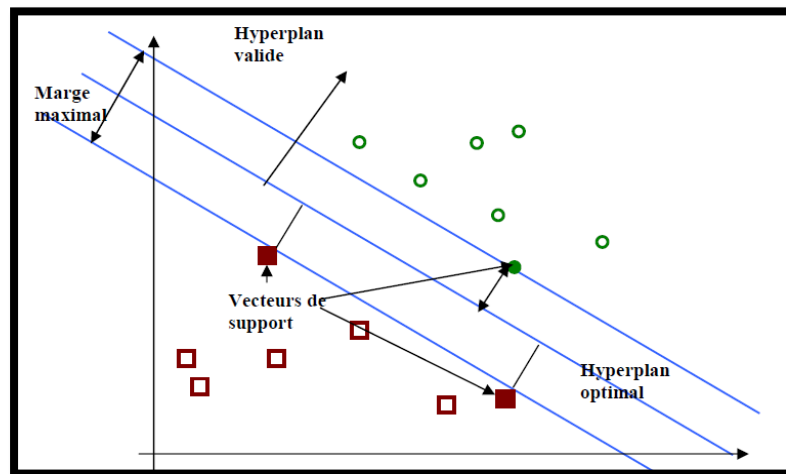


Figure. 4 : Exemple de marge maximale [15].

Les bases des machines à vecteurs de support et leur fonctionnement sont mieux compris avec un exemple simple : Soit un nuage de points de natures différentes (points rouges, points bleus), l'objectif recherché est de trouver une frontière de décision (hyperplan séparateur), qui puisse séparer le nuage de points en deux régions en commettant un minimum d'erreurs, c'est-à-dire (trouver l'hyperplan optimal), montre qu'il existe en effet plusieurs hyperplans séparateurs dont les performances en apprentissage sont identiques (le risque empirique est le même), mais dont les performances en généralisation peuvent être très différentes pour résoudre ce problème [12].

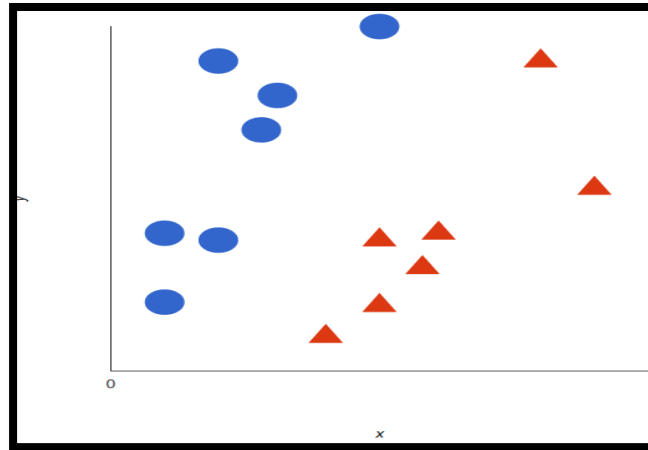


Figure. 5 : classification par SVM.

Une machine à vecteurs de support prend ces points de données et sort l'hyperplan (qui en deux dimensions c'est simplement une ligne) qui sépare le mieux les balises. Cette ligne est la frontière de décision ça veut dire que tout ce qui tombe d'un côté, nous classerons en bleu, et tout ce qui tombe de l'autre en rouge.

3.3. La Classification par les réseaux de neurones artificiels (RNA)

3.3.1. Définition

Un réseau de neurones peut être considéré comme un modèle mathématique de traitement réparti, composé de plusieurs éléments de calcul non linéaire (neurones), opérant en parallèle et connectés entre eux par des poids.

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

Les neurones artificiels sont souvent utilisés sous forme de réseaux qui diffèrent selon le type de connexions entre les neurones, une cinquantaine de types peuvent être dénombrée. En guise d'exemples nous citons : le perceptron de rosemblat, les réseaux de Hopfield, etc. Ces derniers sont les plus utilisés dans le domaine de la modélisation et de la commande des procédés. Ils sont constitués d'un nombre fini de neurones qui sont arrangés sous forme de couches. Les neurones de deux couches adjacentes sont interconnectés par des poids. L'information dans le réseau se propage d'une couche à l'autre, on dit qu'ils sont de type « feed-forward ». Nous distinguons trois types de couches [17] :

- **Couche d'entrée:** les neurones de cette couche reçoivent les valeurs d'entrée du réseau et les transmettent aux neurones cachés. Chaque neurone reçoit une valeur, il ne fait pas donc de sommation.

- **Couches cachées** : chaque neurone de cette couche reçoit l'information de plusieurs couches précédentes, effectuent la sommation pondérée par les poids, puis la transforme selon sa fonction d'activation qui est en général une fonction sigmoïde. Par la suite, il envoie cette réponse aux neurones de la couche suivante.
- **Couche de sortie**: elle joue le même rôle que les couches cachées, la seule différence entre ces deux types de couches est que la sortie des neurones de la couche de sortie n'est liée à aucun autre neurone[18].

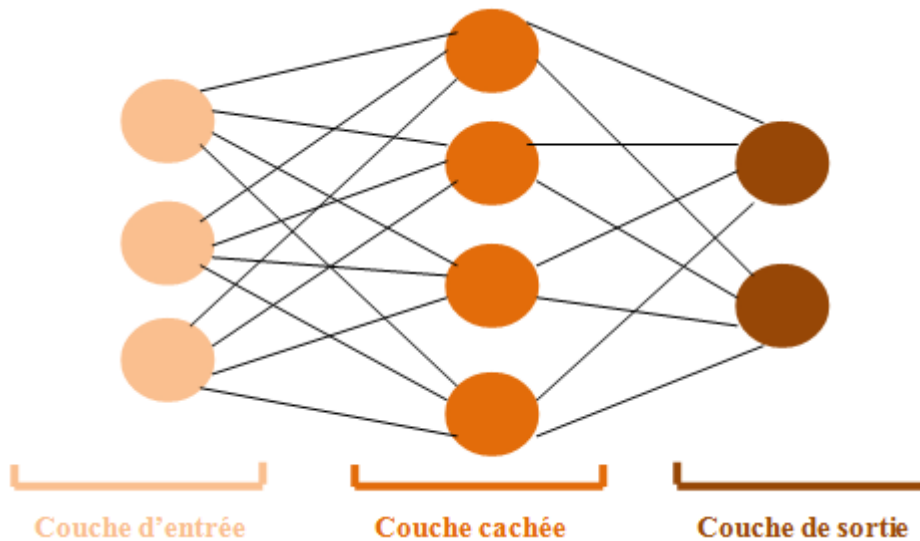


Figure. 6 : Diagramme du Réseau de neurones.

3.3.2. Neurone biologique

Le cerveau humain possède deux hémisphères latéraux reliés par le corps calleux et d'autres ponts axonaux, il pèse moins de deux kilogrammes et contient mille milliards de cellules, dont 100 milliards sont des neurones constitués en réseaux.

Le neurone biologique (figure. 7) est une cellule vivante spécialisée dans le traitement des signaux électriques. Les neurones sont reliés entre eux par des liaisons appelées axones. Ces axones vont eux-mêmes jouer un rôle important dans le comportement logique de l'ensemble.

Ils conduisent les signaux électriques de la sortie d'un neurone vers l'entrée (synapse) d'un autre neurone. Les neurones font une sommation des signaux reçus en entrée et en fonction du résultat obtenu vont fournir un courant en sortie [19].

Les neurones sont des cellules nerveuses décomposables en 4 parties principales (figure.7). Les dendrites, sur lesquelles les autres cellules entrent en contact synaptique : c'est par les dendrites que se fait la réception des signaux. Le corps de la cellule ou noyau, c'est l'unité de

traitement. L'axone est la partie où passent les messages accumulés dans le corps de la cellule. Enfin, à la sortie du neurone on trouve les synapses, par lesquelles la cellule communique avec d'autres neurones, ce sont des points de connexion par où passent les signaux de la cellule.

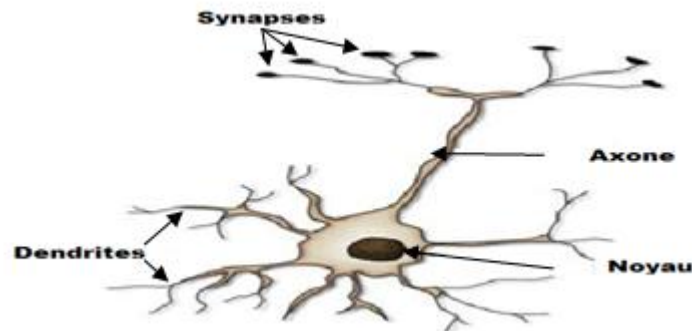


Figure. 7 : Neurone biologique [19].

Un neurone stimulé envoie des impulsions électriques ou potentielles d'action, à d'autres neurones. Ces impulsions se propagent le long de l'axone unique de la cellule. Au point de contact entre neurones, les synapses, ces impulsions sont converties en signaux chimiques. Quand l'accumulation des excitations atteint un certain seuil d'activation, le neurone engendre un potentiel d'action, pendant une durée de 1ms. Le neurone émettant le signal est appelé le neurone pré-synaptique et celui recevant ce signal, neurone post-synaptique.

3.3.3. *Neurone formel*

Un RNA est un assemblage d'éléments de structure identique appelés cellules (ou neurones) interconnectées à l'instar des cellules du système nerveux des vertébrés. Chaque point de connexion (appelé coefficient ou poids) entre deux cellules joue le rôle d'une synapse, l'élément principal d'interaction entre les neurones. Ces connexions ou poids synaptiques ont un rôle primordial dans le fonctionnement parallèle et adaptatif des neurones d'où la notion de réseaux connexionnistes. La représentation mathématique du neurone est introduite par McCulloch et Pitts (1943). Le neurone formel est considéré comme étant la pierre angulaire de la structure des RNA et comprennent cinq éléments importants :

- **Les entrées** : les valeurs, considérées comme entrées du réseau, sont généralement normalisées.

- **Les poids et le biais** : Les poids $W_{k0}, W_{k1} \dots W_{km}$ sont des valeurs reliant les neurones de deux couches différentes ou ceux de la même couche (selon le mode de connectivité). Le biais est représenté par x_0 .
- **La fonction de sommation (combinaison)**:

$$Y = \sum_{i=0}^m x_i w_{ki} + x_0. \quad (3)$$
- **La fonction de transfert (d'activation)**
- **La sortie** : le résultat obtenu Y [19].

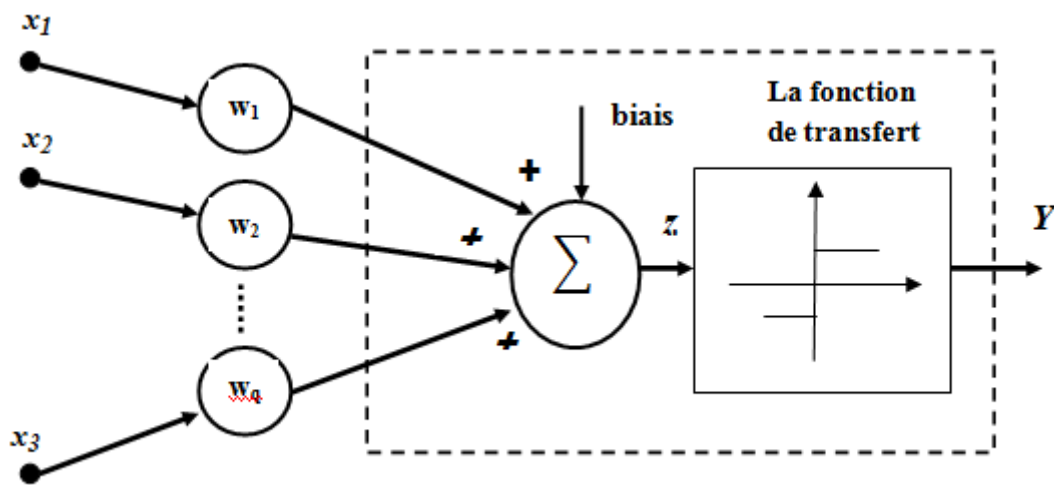


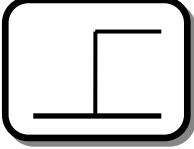
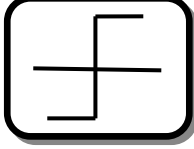
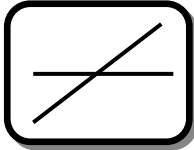


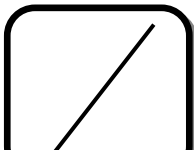
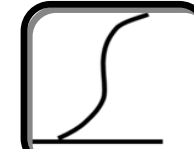
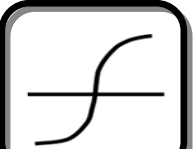
Figure. 8 : Neurone formel.

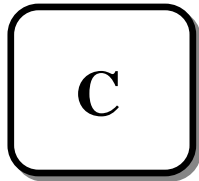
3.3.4. Fonction d'activation

La fonction d'activation (ou fonction de seuillage, ou encore fonction de transfert) sert à introduire une non-linéarité dans le fonctionnement du neurone. Les fonctions de seuillage présentent généralement trois intervalles :

- ✓ En dessous du seuil, le neurone est non actif (souvent dans ce cas, sa sortie vaut 0 ou -1).
- ✓ Aux alentours du seuil, une phase de transition.
- ✓ Au-dessus du seuil, le neurone est actif (souvent dans ce cas, sa sortie vaut 1) dans sa première version, le neurone formel était implémenté avec une fonction à seuil, mais de nombreuses versions existent[20]. Ainsi, le neurone de McCulloch et Pitts a été généralisé de différentes manières, en choisissant d'autres fonctions d'activations, comme les fonctions énumérées dans le (tableau .1).

Tableau. 1 : Différentes fonctions d'activations utilisées dans les RNA .

<i>Nom de la fonction</i>	<i>Relation entrée/sortie</i>	<i>graphe</i>	<i>Nom MATALB</i>
<i>Seuil</i>	$y=0$ si $s < 0$ $y=1$ si $s \geq 0$		Hardlim
<i>Seuil symétrique</i>	$y=-1$ si $s < 0$ $y=1$ si $s \geq 0$		Hardlims
<i>Linéaire</i>	$y=s$		Purelin
<i>Linéaire saturée</i>	$y=0$ si $s \leq 0$ $y=s$ si $0 \leq s \leq 1$ $y=1$ si $s \geq 1$		Satlin
<i>Linéaire saturée symétrique</i>	$y=-1$ si $s < -1$ $y=s$ si $-1 \leq s \leq 1$ $y=1$ si $s > 1$		Satlins
<i>Linéaire positive</i>	$y=0$ si $s \leq 0$ $y=s$ si $s \geq 0$		Poslin
<i>Sigmoïde</i>	$y = \frac{1}{1+\exp^{-s}}$		Logsig
<i>Tangente hyperbolique</i>	$y = \frac{e^s - e^{-s}}{e^s + e^{-s}}$		Tansig

<i>Compétitive</i>	$y=1$ si s maximum $y=0$ autrement		Compet
--------------------	---	--	--------

3.3.5. Architecture des réseaux de neurones

L'architecture d'un réseau de neurones est l'organisation des neurones entre eux au sein d'un même réseau. Autrement dit, il s'agit de la façon dont ils ordonnaient et connectés. La majorité des réseaux de neurones utilise le même type de neurones. Quelques architectures plus rares se basent sur des neurones dédiés. L'architecture d'un réseau de neurones dépend de la tâche à apprendre. Un réseau de neurones est en général composé de plusieurs couches de neurones, des entrées jusqu'aux sorties. On distingue deux grands types d'architectures de réseaux de neurones :

- ✚ Les réseaux de neurones non bouclés.
- ✚ Les réseaux de neurones bouclés [21].

3.3.5.1. Les réseaux de neurones non bouclés

Un réseau de neurones non bouclé réalise une (ou plusieurs) fonctions algébriques de ses entrées, par composition des fonctions réalisées par chacun de ses neurones. Un réseau de neurones non bouclé est représenté graphiquement par un ensemble de neurones "connectés" entre eux, l'information circulante des entrées vers les sorties sans « retour en arrière »; si l'on représente le réseau comme un graphe dont les nœuds sont les neurones et les arêtes des « connexions » entre ceux-ci, le graphe d'un réseau non bouclé est acyclique. Le terme de "connexions" est une métaphore : dans la très grande majorité des applications, les réseaux de neurones sont des formules algébriques dont les valeurs numériques sont calculées par des programmes d'ordinateurs, non des objets physiques (circuits électroniques spécialisés) ; néanmoins, le terme de connexion, issu des origines biologiques des réseaux de neurones, est passé dans l'usage, car il est commode quoique trompeur. Il a même donné naissance au terme du connexionnisme [22].

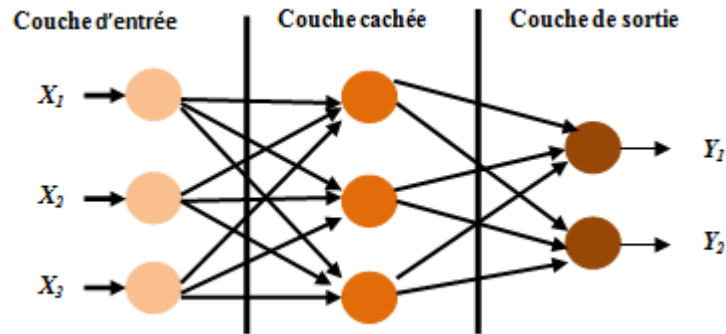


Figure. 9 : Les réseaux de neurones non bouclés.

a. Réseaux de neurones monocouches

La structure d'un réseau monocouche est telle que des neurones organisés en entrée soient entièrement connectés à d'autres neurones organisés en sortie par une couche modifiable de poids ([figure. 10](#))[23].

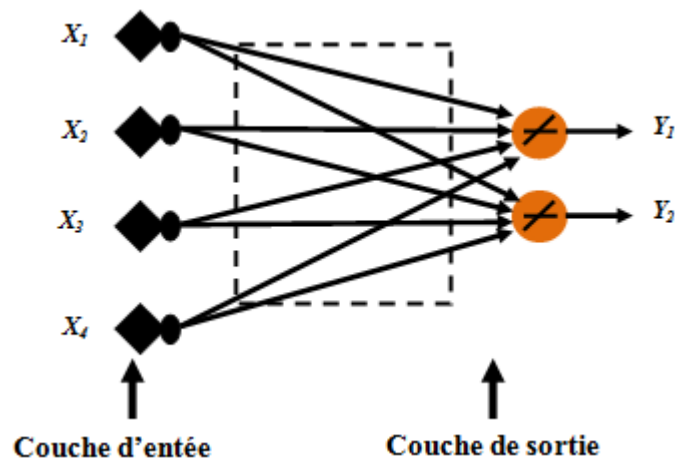


Figure. 10 : Schéma d'un réseau de neurones monocouche.

b. Réseaux de neurones multicouches

Les neurones sont arrangés par couches. Il n'y a pas de connexion entre neurones d'une même couche, et les connexions ne se font qu'avec les neurones de couches avalent. Habituellement, chaque neurone d'une couche est connecté à tous les neurones de la couche suivante et celle-ci seulement. Ceci nous permet d'introduire la notion de sens de parcours de l'information (de l'activation) au sein d'un réseau et donc définir les concepts de neurones d'entrée, neurone de sortie. Par extension, on appelle couche d'entrée l'ensemble des neurones

d'entrée, couche de sortie l'ensemble des neurones de sortie. Les couches intermédiaires n'ayant aucun contact avec l'extérieur sont appelées couches cachées. La figure. 11 représente un réseau de neurones non bouclé qui a une structure particulière, très fréquemment utilisée : il comprend des entrées, deux couches de neurones cachés et des neurones de sortie. Les neurones de la couche cachée ne sont pas connectés entre eux. Cette structure est appelée Perceptron multicouche.

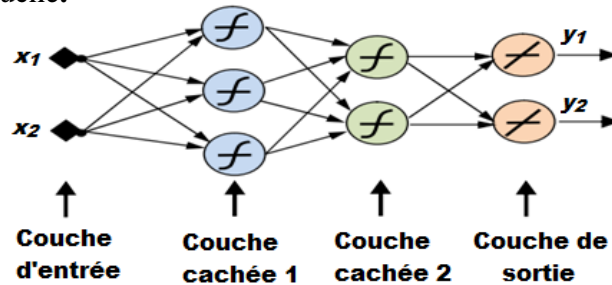


Figure. 11 : Schéma d'un réseau de neurones Perceptron multicouches.

On note aussi que les réseaux multicouches sont beaucoup plus puissants que les réseaux simples à une seule couche. En utilisant deux couches (une couche cachée et une couche de sortie), à condition d'employer une fonction d'activation sigmoïde sur la couche cachée, on peut entraîner un réseau à produire une approximation de la plupart des fonctions, avec une précision arbitraire (cela peut cependant requérir un grand nombre de neurones sur la couche cachée). Sauf dans des rares cas, les réseaux de neurones artificiels exploitent deux ou trois couches [24].

c. Réseaux de neurones à connexions locales

Il s'agit d'une structure multicouche, mais qui à l'image de la rétine conserve une certaine topologie. Chaque neurone entretient des relations avec un nombre réduit et localisé de neurones de la couche avalent. Les connexions sont donc moins nombreuses que dans le cas d'un réseau multicouche classique (figure.12).

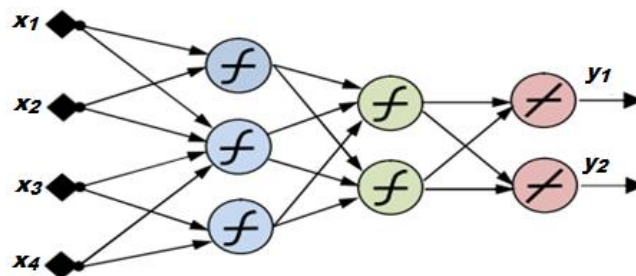


Figure. 12 : Schéma d'un réseau de neurones à connexions locales.

3.3.5.2. Les réseaux de neurones bouclés

Contrairement aux réseaux de neurones non bouclés dont le graphe de connexions est acyclique, les réseaux de neurones bouclés peuvent avoir une topologie de connexions quelconque, comprenant notamment des boucles qui ramènent aux entrées la valeur d'une ou plusieurs sorties. Pour qu'un tel système soit causal, il faut évidemment qu'à toute boucle soit associé un retard : un réseau de neurones bouclé est donc un système dynamique, régi par des équations différentielles ; comme l'immense majorité des applications sont réalisées par des programmes d'ordinateurs, on se place dans le cadre des systèmes à temps discret, où les équations différentielles sont remplacées par des équations aux différences. Il s'agit donc des réseaux de neurones avec retour en arrière (feedback network or récurrent network), (**Figure. 13**).

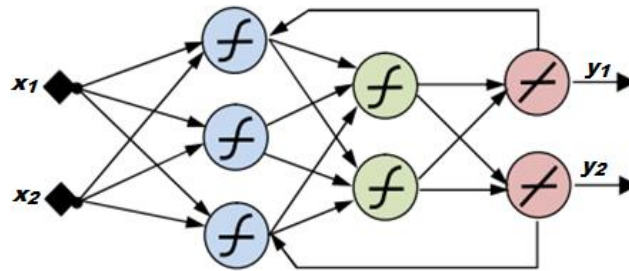


Figure. 13 : Schéma de réseau de neurones bouclé.

Un réseau de neurones bouclé à temps discret est donc régi par une (ou plusieurs) équations aux différences non linéaires, résultant de la composition des fonctions réalisées par chacun des neurones et des retards associés à chacune des connexions. La forme la plus générale des équations régissant un réseau de neurones bouclé est appelée **forme canonique** :

$$x(k+1) = \varphi[x(k), u(k)] \quad (4)$$

$$y(k) = \psi[x(k), u(k)] \quad (5)$$

Où φ et ψ sont des fonctions non linéaires réalisées par un réseau de neurones non bouclé (mais pas obligatoirement, un perceptron multicouche), et où k désigne le temps (discret). La forme canonique est représentée sur la figure 14. Tout réseau de neurones, aussi compliqué soit-il, peut être mis sous cette forme canonique, de manière complètement automatique [25].

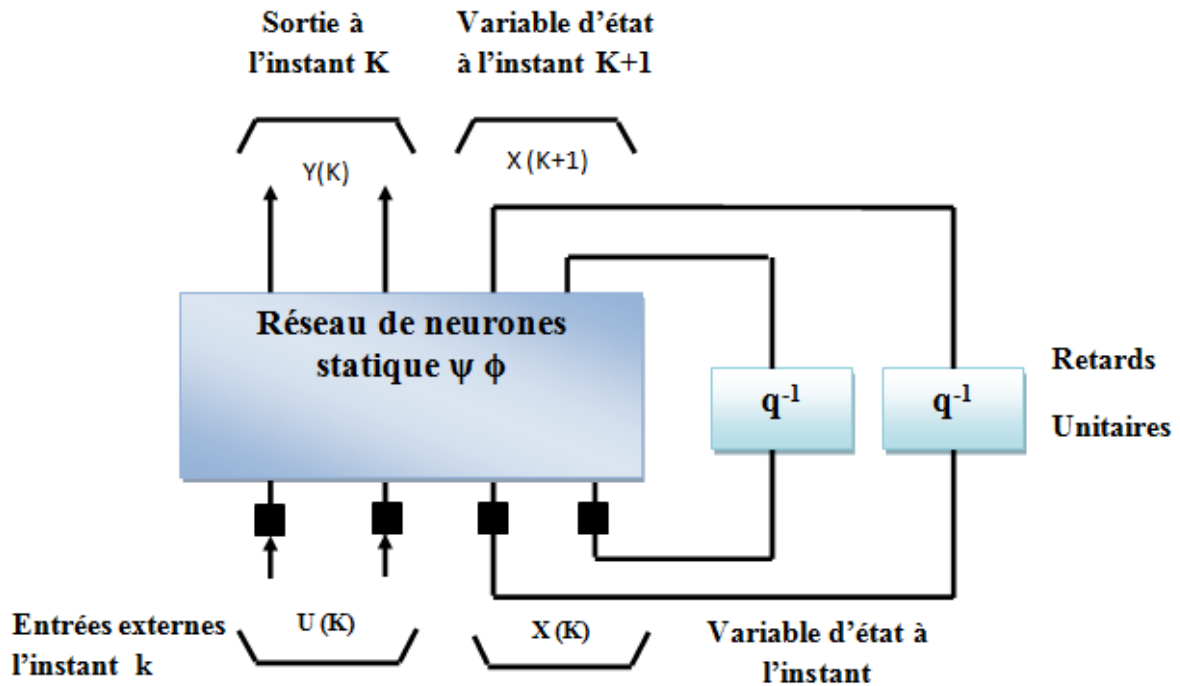


Figure. 14 : Forme canonique d'un réseau de neurones bouclé.

3.3.6. Apprentissage

Si le neurone est considéré comme étant la pierre angulaire du réseau de neurones artificiels, l'apprentissage peut être défini comme le processus le plus important pour la construction d'une structure performante du réseau de neurones en se déroulant sur l'adaptation des poids. Pour cela, il existe deux méthodes d'apprentissage : l'apprentissage supervisé et non supervisé. Les deux méthodes précédentes, fournissent au réseau neuronal une base d'apprentissage qui représente les données sur lesquelles le réseau peut apprendre et réaliser la phase d'apprentissage[26].

3.3.6.1. Apprentissage non supervisé

Que signifie le fait de former un réseau de neurones sans supervision ? Comme indiqué précédemment, le réseau de neurones est muni d'une base d'apprentissage, qui est une collection de valeurs d'entrée bien définies. Le réseau de neurones non supervisé n'est pas muni de données attendues ou sorties anticipées. L'apprentissage non supervisé est généralement utilisé pour entraîner les réseaux de neurones utilisés pour la classification et le data mining[26].

La figure suivante résume la méthode d'apprentissage non supervisé.

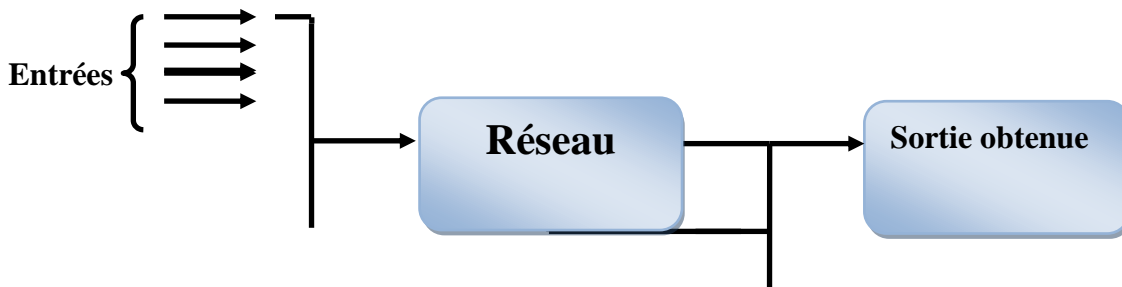


Figure. 15 : Apprentissage non supervisé.

3.3.6.2. Apprentissage supervisé

La méthode d'apprentissage supervisée est similaire à la méthode d'apprentissage non supervisé, dans laquelle les ensembles d'apprentissage sont fournis. Tout comme avec l'apprentissage non supervisé, les ensembles d'apprentissage spécifient les signaux d'entrée pour le réseau neuronal. La principale différence est que dans l'apprentissage supervisé, les résultats attendus sont fournis.

L'apprentissage supervisé est basé sur la comparaison directe entre l'entrée actuelle du RNA et la sortie désirée correcte, aussi connue comme la sortie cible. Il est souvent formulé comme la minimisation d'une fonction d'erreur comme l'erreur quadratique moyenne totale entre l'entrée actuelle et la sortie désirée additionnée à toutes les données disponibles. Une descente de gradient, basée sur l'algorithme d'optimisation telle que la rétro-propagation, peut alors être utilisée pour ajuster itérativement les connexions des poids dans le RNA afin de minimiser l'erreur [29].

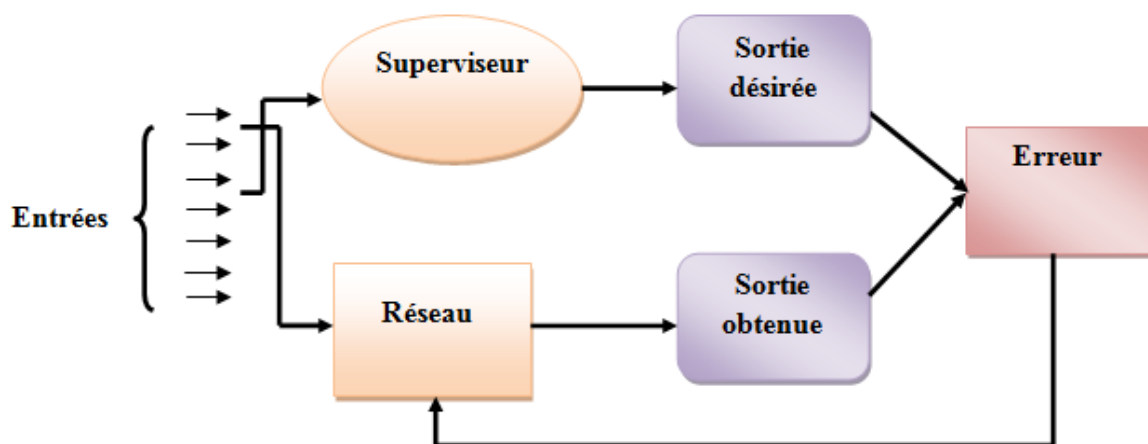


Figure.16 : Apprentissage supervisé.

3.3.6.3. Règles d'apprentissage

L'apprentissage est une procédure adaptative par laquelle les connexions (ou synapses) des neurones sont ajustés face à une source d'information. Il existe au moins 23 types de règles d'apprentissage qui peuvent être regroupés en trois catégories : les règles d'apprentissage supervisé, non supervisé, sont renforcées. Mais l'objectif fondamental de l'apprentissage reste le même : soit la classification, l'approximation de fonction ou encore la prévision. Dans tous les cas, la procédure adaptative de mise à jour des poids obéit essentiellement à un objectif : la recherche analytique du minimum de la fonction de coût (ou fonction d'erreur) dans un espace multidimensionnel. Il existe deux techniques de recherche du minimum de la fonction d'erreur : la première est une technique de recherche du minimum local basé sur la méthode du gradient décroissant ; la seconde est une recherche du minimum global fondée sur des approches statistiques ou génétiques [26].

3.3.7. Applications des réseaux de neurones artificiels

Les RNA sont considérés comme une nouvelle approche de traitement de l'information par apprentissage de cette information et la rend disponible à l'utilisation afin de résoudre un tel problème. Il existe de nombreux problèmes qui peuvent être résolus par un réseau de neurones. Toutefois, les réseaux de neurones sont généralement utilisés pour traiter des types particuliers de problèmes. Les quatre types de problèmes qui sont souvent résolus par les réseaux de neurones sont les suivants :

3.3.7.1. Classification

La classification est le processus de classement des entrées en groupes. Par exemple, une compagnie d'assurance peut vouloir classer les demandes d'assurance dans les différentes catégories de risques, ou une organisation en ligne peut vouloir de son système de messagerie de classer le courrier entrant dans des groupes de messages spam et non-spam [27].

3.3.7.2. Prédiction

La prédiction est une autre application des réseaux neuraux. Étant donné une série temporelle de données d'entrée, un réseau de neurones peut prédire les valeurs futures. La précision de la prédiction dépend de nombreux facteurs, tels que la quantité et la pertinence des données d'entrée. Par exemple, les réseaux de neurones sont généralement appliqués à des problèmes de prédiction de l'évolution des marchés financiers [27].

3.3.7.3.Reconnaissance des formes (pattern recognition)

Reconnaissance des formes est l'une des utilisations les plus courantes des réseaux neuraux. Pattern recognition est une forme de classification et est tout simplement la capacité de reconnaître un motif. Le modèle doit être reconnu même s'il n'est pas clair. Exemple : la reconnaissance des visages [28].

3.3.7.4.Optimisation

Une autre application des réseaux de neurones est l'optimisation qui peut être appliquée à de nombreux problèmes pour lesquels une solution est recherchée. Le réseau de neurones peut ne pas toujours trouver la solution optimale mais il cherche à trouver une solution acceptable. Les RNA sont utilisés dans le domaine des sciences cognitives où l'on cherche à développer des modèles capables de manifester des capacités d'apprentissage et d'adaptation à leur environnement (Blayo et Verleysen 1996). Les deux concepts de base à l'origine des différents types de modèles neuronaux sont l'architecture et l'apprentissage [29].

4. Conclusion

Ce chapitre é été consacré à la présentation des notions de base relatives à la classification ainsi quelques techniques de classification à savoir : la logique floue, l'SVM et les réseaux de neurones artificiels. Dans le chapitre suivant on essayer de développer un système expert pour le diagnostic des maladies cardiaques à base de la technique d'extraction des caractéristiques nommée Pan and Tompkins et les réseaux de neurones artificiels.

CHAPITRE IV

Extraction des caractéristiques et classification des battements cardiaques : Méthode & résultats

Dans ce chapitre nous allons présenter une structure pour le diagnostic des maladies cardiaques. Cette structure est composée des deux étapes : Extraction des caractéristiques en utilisant la technique Pan and Tompkins, et classification à base de réseau de neurones artificiels, la particularité ici et l'ajout d'une technique d'optimisation baptisée BBO pour ajuster les paramètres de réseau de neurones artificiels..

Sommaire

- 28. Introduction
 - 29. Méthode et résultats
 - 30. Conclusion
-

1. Introduction

Tout ce qui affecte le cœur et l'empêcher de fonctionner normalement fait partie des maladies du cœur. On groupe souvent ces maladies selon qu'elles sont congénitales (c-à-d présentes à la naissance) ou acquises (c-à-d attribuable à une autre condition). Parmi les maladies du cœur, on compte :

- **La cardiomyopathie** : une maladie du muscle cardiaque,
- **L'insuffisance cardiaque congestive** : fonctionnement insuffisant de la pompe cardiaque qui entraîne l'accumulation de liquide dans les poumons et ailleurs dans le corps ;
- **Les troubles des valves cardiaques** : d'origine congénitale (p.ex. maladie d'Ebstein) ou acquis (Ex : péricardite infectieuse);
- **La maladie du péricarde** : une infection du péricarde, la membrane qui enveloppe le cœur tel un sac.

L'électrocardiographie est un procédé rapide, simple et indolore, qui consiste à amplifier et enregistrer soit sur un papier soit sous forme de données les impulsions électriques traversant le cœur (électrocardiogramme). Cet enregistrement, l'électrocardiogramme (ECG), donne des informations concernant la partie du cœur qui initie chaque battement cardiaque (le stimulateur cardiaque, appelé nœud sino-auriculaire ou sinusal), les voies de conduction nerveuse, ainsi que la vitesse et le rythme cardiaque. Ces informations peuvent être extraites soit par un expert en utilisant l'enregistrement sur un papier, soit automatiquement en utilisant un système expert contenant une base de données. Dans ce mémoire de fin d'étude on a adopté la deuxième méthode. Généralement, ces systèmes se composent de deux parties, la première partie consiste à l'extraction des caractéristiques des différents battements constituant le signal électrocardiogramme, la deuxième étape sera le classement de ces battements selon le type de pathologie atteinte. Dans allons essayer de développer un système expert pour le diagnostic des maladies cardiaques à base de la technique Pan and Tompkins et les réseaux de neurones artificiels. Généralement on utilise on ajuste les paramètres des réseaux de neurones artificiels en utilisant des techniques à base du gradient ce qui nous pose le problème de minimum local, pour surmonter ce problème on a opté à utiliser un algorithme d'optimisation pour ajuster ces paramètres. L'algorithme qui sera utilisé est baptisé BBO (Biogeography-based optimization).

2. Méthode et résultats

Dans ce chapitre, nous allons présenter une structure pour la classification des arythmies cardiaques. Dans ce travail, cinq types de battements d'arythmie recommandés par AAMI (Association for the Advancement of Medical Instrumentation) seront considérés : battement normal (N), battement anormal d'origine supra-ventriculaire (S), battement anormal d'origine ventriculaire (V), battement résultant de la fusion d'un battement V et d'un battement N (F), et battement inconnu (Q) (voir chapitre 1). Pour la première étape qu'est l'extraction des caractéristiques. La technique Pan and Tompkins discutée précédemment dans le chapitre. 2 sera utilisée. Afin d'extraire les caractéristiques des battements nous allons avoir un vecteur caractéristique qui sera utilisé comme une entrée d'un classificateur à base de réseau de neurones. Les paramètres de classificateur (RNA) seront optimisés dans la phase d'apprentissage en utilisant un algorithme d'optimisation à savoir la BBO pour éviter le problème de minimum local. La Figure.1 représente l'organigramme de la procédure adoptée.

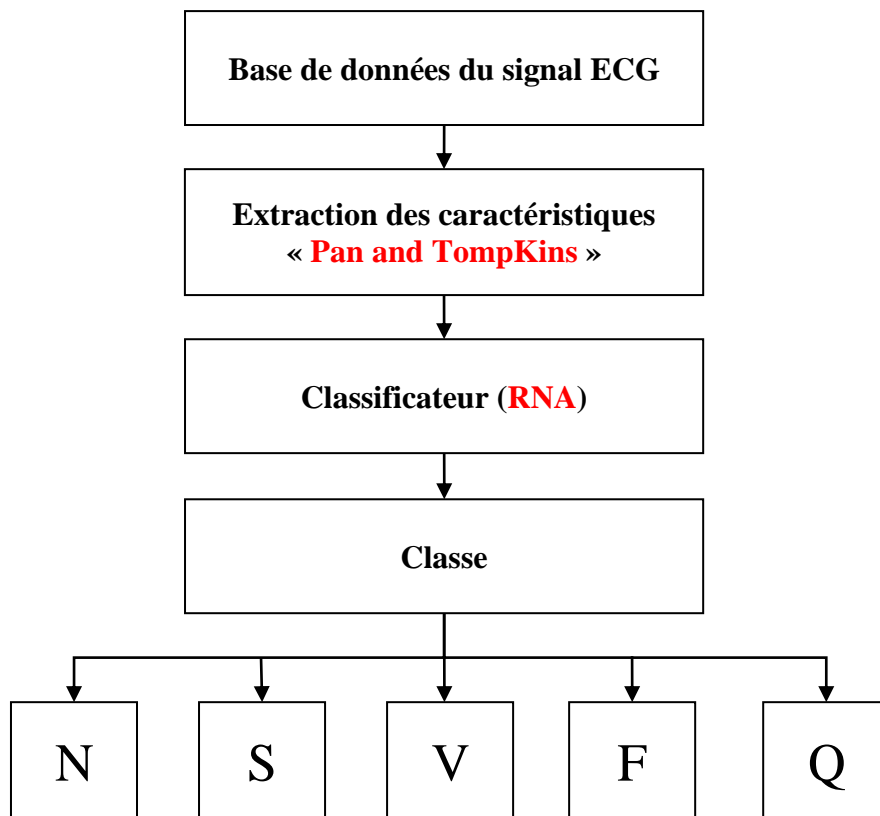


Figure. 1 : Organigramme de la procédure adoptée.

2.1. Préparation des bases de données

La base de données MIT-BIH PhysioBank ([Chapitre. 1](#)) nous a permis de créer deux sous bases de données : une pour l'apprentissage et l'autre pour les tests qui

serviront l'entraînement et l'évaluation des performances du classificateur. Les battements sélectionnés sont regroupés selon les cinq classes recommandées par le standard AAMI (Chapitre. 1) sont données dans le Tableau.1.

Table. 1 : Enregistrements ECG et nombres de battements utilisés dans ce travail.

Classe	ECG Record (Sous classe)	Base d'apprentissage	Base de test
N	100, 103, 105, 113, 115, 119, 121	200	100
S	100, 103, 113, 121, 202, 205, 209, 210	200	100
V	100, 105, 119, 121, 123, 202	200	100
F	202, 205, 210	200	100
Q	105, 203, 208	200	100
Total		1000	500

2.2.Extraction des caractéristiques

Pour la détection du complexe QRS nous avons utilisé la méthode de Pan et Tompkins, car elle est de complexité raisonnable et peu excessive en temps de calcul. Le but final de la méthode Pan et Tompkins est la localisation du pic R pour le calcul de la fréquence cardiaque, pour détecter les ondes Q et S, et pour mesurer la durée du complexe QRS. Les principales étapes de cette méthode ont été discutées dans le chapitre 2 :

Les résultats de simulations de la méthode Pan and TompKins sont donnés par la figure. 2. Après avoir détecter le complexe QRS, vecteur caractéristique sera composé des paramètres suivants (pour chaque complexe QRS) :

- ✓ *L'amplitude de l'onde Q ;*
- ✓ *L'amplitude de l'onde R ;*
- ✓ *L'amplitude de l'onde S ;*
- ✓ *L'intervalle SQ ;*
- ✓ *L'intervalle RQ ;*
- ✓ *L'intervalle SR ;*
- ✓ *L'intervalle RR.*

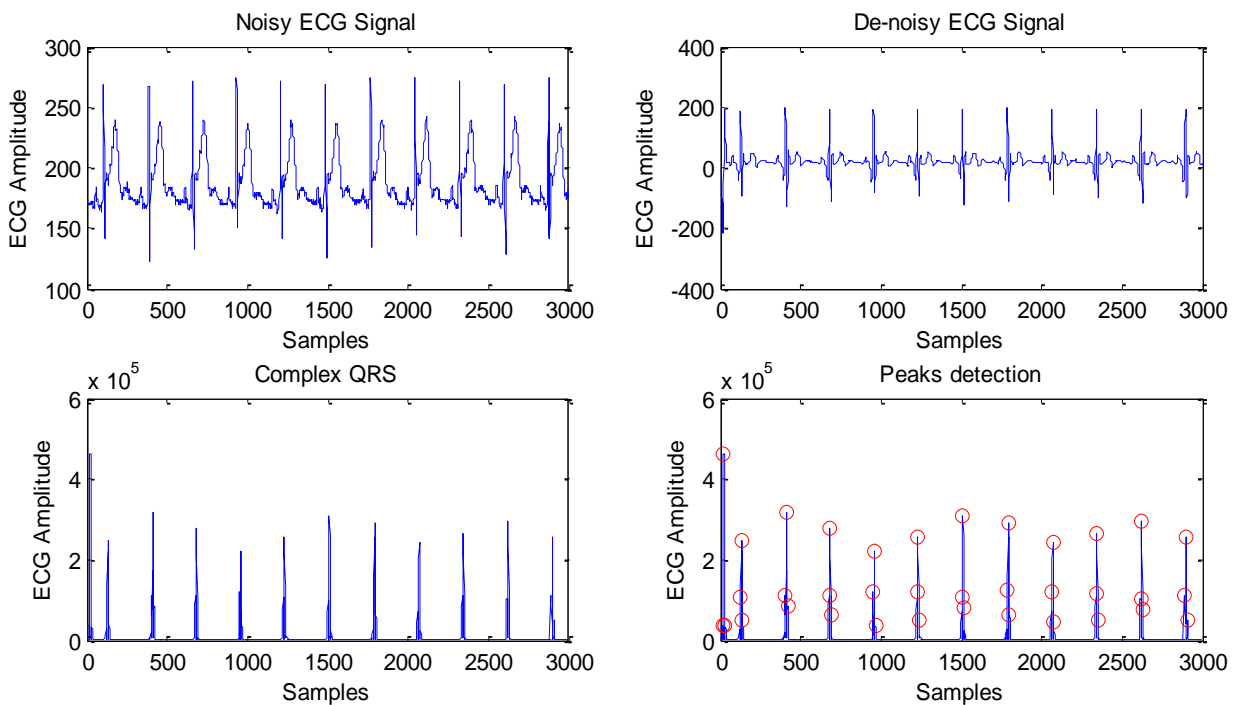


Figure.2 : Résultats de simulation de la méthode Pan and Tompkins.

2.3. Sélection des caractéristiques par la méthode mRMR

Dans les problèmes de classification, et sur tous si la taille de la matrice caractéristique est énorme, nous cherchons toujours à trouver le meilleur sous ensemble de caractéristiques pour choisir le modèle approprié et minimiser les erreurs de classification et aussi le temps de calcul. Plusieurs techniques ont été développées pour résoudre ce problème de sélection à savoir : l'analyse en composantes principales (PCA), l'analyse en composantes indépendantes (ICA), la décomposition en valeurs singulières (SVD) et autres. Dans de mémoire de fin d'étude une méthode baptisée mRMR (*minimum-redundancy maximum relevancy*) est adoptée. L'mRMR est une méthode de filtrage pour la sélection des caractéristiques proposée par Peng. Cette méthode est basée sur des mesures statistiques classiques comme l'information mutuelle, la corrélation...etc. L'idée de base est de profiter de ces mesures pour essayer de minimiser la redondance (mR) entre les variables et de maximiser la pertinence (MR). Peng et al ont utilisé l'information mutuelle pour calculer les deux facteurs mR et MR. Le calcul de la redondance et de la pertinence d'une variable est donnée par l'équation suivante :

$$\text{Redondance } (i) = \frac{1}{|F|^2} \sum_{i,j} I(i,j) \quad (1)$$

$$\text{Pertinence } (i) = \frac{1}{|F|^2} \sum_{i,j} I(i,Y) \quad (2)$$

Avec :

$|F|$: représente la taille de l'ensemble de variables ;

$I(i, j)$: est l'information mutuelle entre la i^{ieme} et la j^{ieme} variable ;

$I(i, Y)$: est l'information mutuelle entre la i^{ieme} variable et l'ensemble des étiquettes de la classe Y .

Le score d'une variable est la combinaison de ces deux facteurs tel que :

$$\text{Score}(i) = \text{Pertinence}(i) - \text{Redondance}(i) \quad (3)$$

2.4. Classification

Le but est de classifier automatiquement les battements cardiaques selon le type d'arythmie. Le choix du classificateur est très important, parce qu'il constitue l'élément de décision dans le système de reconnaissance de formes. L'efficacité du classificateur reste aussi subordonnée à plusieurs autres facteurs à savoir : le choix et la taille de la base d'apprentissage et la méthode d'extraction des caractéristiques. Un classificateur à base de réseau de neurones artificiels sera utilisé dans ce mémoire de fin d'étude dans laquelle nous avons ajouté un étage d'optimisation pour ajuster les paramètres de classificateur.

2.4.1. Paramètres de classificateur à ajuster

La conception de classificateur à base de réseau de neurones artificiels comprend la détermination des paramètres inconnus qui sont les poids de différentes couches du la structure du réseau de neurones artificiels utilisée (Figure. 3).

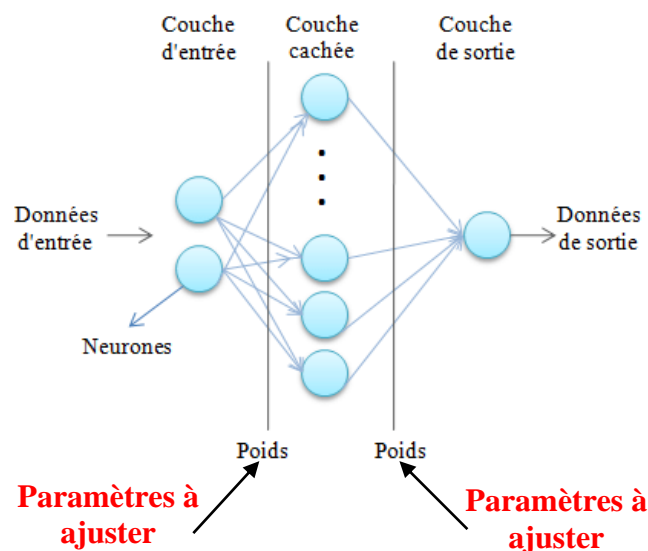


Figure. 3 : Paramètres de réseau de neurones à optimiser.

2.4.2. L'algorithme d'optimisation à base de biogéographie (BBO)

L'algorithme à base de biogéographie (*BBO* : *Biogeography-Based Optimization*), a été développé par Dan Simon en 2008. Inspiré par des études sur la répartition spatiale des

espèces de plantes et d'animaux ainsi que les causes de leur répartition et de leur extinction. Elle traite de façon dont la richesse en espèces (nombres d'espèces) est maintenue dans un système d'île qui sont sujettes à l'immigration et sur lesquelles des espèces s'éteignent, quand une île ne peut pas facilement supporter la population d'une espèce, les membres migrent vers de nouvelles îles et subissent une spéciation. L'algorithme BBO manipule une population d'individus appelés îles (ou habitats). Chaque île représente une solution possible au problème à résoudre. La « fitness » de chaque île est déterminée par son HSI (*Habitat Suitability Index*), qui considère les caractéristiques environnementales telles que les précipitations, la température et la végétation qui constituent une mesure de la qualité d'une solution candidate. Ces caractéristiques peuvent être représentées quantitativement et sont appelées variables de l'indice d'aptitude (*Suitability Index Variables* ou SIV). Une bonne solution au problème d'optimisation est une île avec un grand nombre d'espèces, ce qui correspond à une île avec un faible HSI. Dans l'algorithme BBO, chaque habitat a ses propres taux d'immigration et d'émigration représentant les espèces qui viennent et sortent de l'île. Ces paramètres sont influencés par le nombre d'espèces (S) sur l'île.

2.4.2.1. Biogéographie

La Biogéographie est l'étude de la distribution de la biodiversité dans l'espace et dans le temps, ce qui permet aux nombreuses espèces animales de migrer vers différents habitats ou îles pour leur survie et une meilleure vie. La biogéographie a été étudiée dès le 19^{me} siècle par Alfred Wallace et Charles Darwin.

Dans la science de la biogéographie, une île est définie comme la zone écologique habitée par des plantes et animales et géographiquement isolée d'autres habitats. Chaque île a ses caractéristiques telles que la disponibilité alimentaire, les précipitations, la température, la diversité des espèces, la sécurité, etc. (Figure.4). La qualité d'une île est mesurée par son indice d'adéquation (*Suitability Index*). Les îles avec un indice élevé ont plus adaptées à la vie et ont donc une grande population.

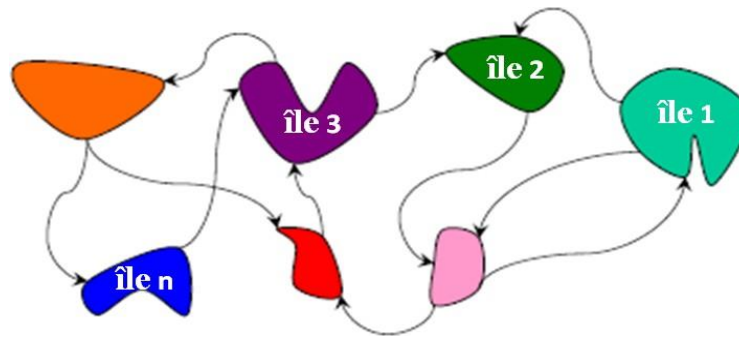


Figure .4 : Migration des espèces.

2.4.2.2.Principe de l’algorithme BBO

BBO, est un algorithme basé sur une population d'individus appelés île (ou habitats). Chaque île représente une solution possible au problème à résoudre. La « fitness » de chaque île est déterminée par son HSI (*Habitat Suitability Index*) qui est une mesure de la qualité d'une solution candidate, chaque île est représentée par des variables d'indice de qualité SIVs (*Suitability Index Variables*). Un HSI élevé d'une île signifie une bonne performance sur le problème d'optimisation, et un faible HSI signifie une mauvaise performance pour un problème d'optimisation.

Le fonctionnement de *BBO* est basé sur la migration et la mutation. La population initiale représente l'espace de recherche, elle est générée aléatoirement. L'évaluation de la population initiale engendre la migration de certains individus et les descendants vont être mutés. La migration permet de créer un nouvel ensemble d'individus et la mutation permet de fixer la proportion de la population qui sera renouvelée à chaque génération. Les meilleurs individus trouvés sont conservés par l'élitisme (sélection). Les nouveaux descendants remplacent les parents afin de former une nouvelle population.

2.4.2.3.Optimisation et biogéographie

L'application de la biogéographie dans l'optimisation est un exemple d'un modèle de processus naturel pour résoudre les problèmes d'optimisation. Ceci est similaire à ce qui s'est passé dans les dernières décennies avec les algorithmes génétiques, réseaux de neurones, les colonies de fourmis et d'autres domaines de l'intelligence informatique. L'optimisation à base de biogéographie (*BBO*) utilise un vocabulaire similaire à celui de la biogéographie ([Tableau. 2](#)) où chaque habitat est analogue à une solution du problème. Les caractéristiques des solutions (variables de décision) sont appelées variables d'indice d'adéquation (SIV). L'indice d'adéquation de l'habitat (HSI) est analogue au fitness qui permet de mesurer l'adéquation de

l'individu. Les habitats avec un HSI élevé ont tendance à avoir un grand nombre d'espèces, un taux d'immigration faible et un taux d'émigration élevé. Inversement, les habitats avec un HSI faible ont un faible nombre d'espèces, un taux d'immigration élevé et un taux d'émigration faible. L'algorithme BBO nécessite l'emploi de certains termes dont il est utile de préciser la définition.

Table. 2 : Terminologie de la BBO.

La Biogéographie	Algorithme BBO
<i>Habitat ou île</i>	Solution du problème
<i>HSI</i>	Qualité de la solution (fonctioncoût)
<i>SIV</i>	Les Variables du problème d'optimisation

✓ **Variable d'indice d'adéquation**

Une variable d'indice d'adéquation (*SIV:Suitability Index Variable*) est une variable entière, réelle où booléenne qui caractérise l'habitabilité d'une île.

✓ **Habitat**

Un habitat H est une solution du problème représentée généralement par un vecteur initialisé aléatoirement par des variables d'indice d'adéquation. Il est noté :

$$H = [SIV_1, SIV_2, \dots, SIV_k]. \quad (4)$$

✓ **Population** : une population est un ensemble de n habitats.

$$Pop = [H_1, H_2, \dots, H_n]. \quad (5)$$

✓ **Indice d'adéquation de l'habitat**

Un indice d'adéquation de l'habitat (*HSI*) équivalant au fitness, associe une valeur pour chaque individu. Cette valeur a pour but d'évaluer le degré d'adaptation d'un individu à son environnement.

$$HSI = f(H) = f([SIV_1, SIV_2, \dots, SIV_k]) \quad (6)$$

✓ **Taux d'immigration**

Le taux d'immigration $\lambda(H_i)$ est le taux d'entrée des variables (*SIV*) à un habitat.

✓ **Taux d'émigration**

Le taux d'émigration $\mu(H_i)$ est le taux de sortie des variables (*SIV*) d'un habitat.

2.4.3. Étapes de l'algorithme d'optimisation basée sur la biogéographie

L'organigramme fonctionnel présenté dans la Figure .5, illustre l'organigramme générale de l'algorithme *BBO*. L'algorithme commence par initialiser les paramètres et la population initiale, il modifie cette population par des opérateurs spécifiques en construisant de nouvelle population jusqu'à atteindre une qualité (*HSI*) meilleure qu'un seuil préfixé ou un nombre maximal de générations g_{max} .

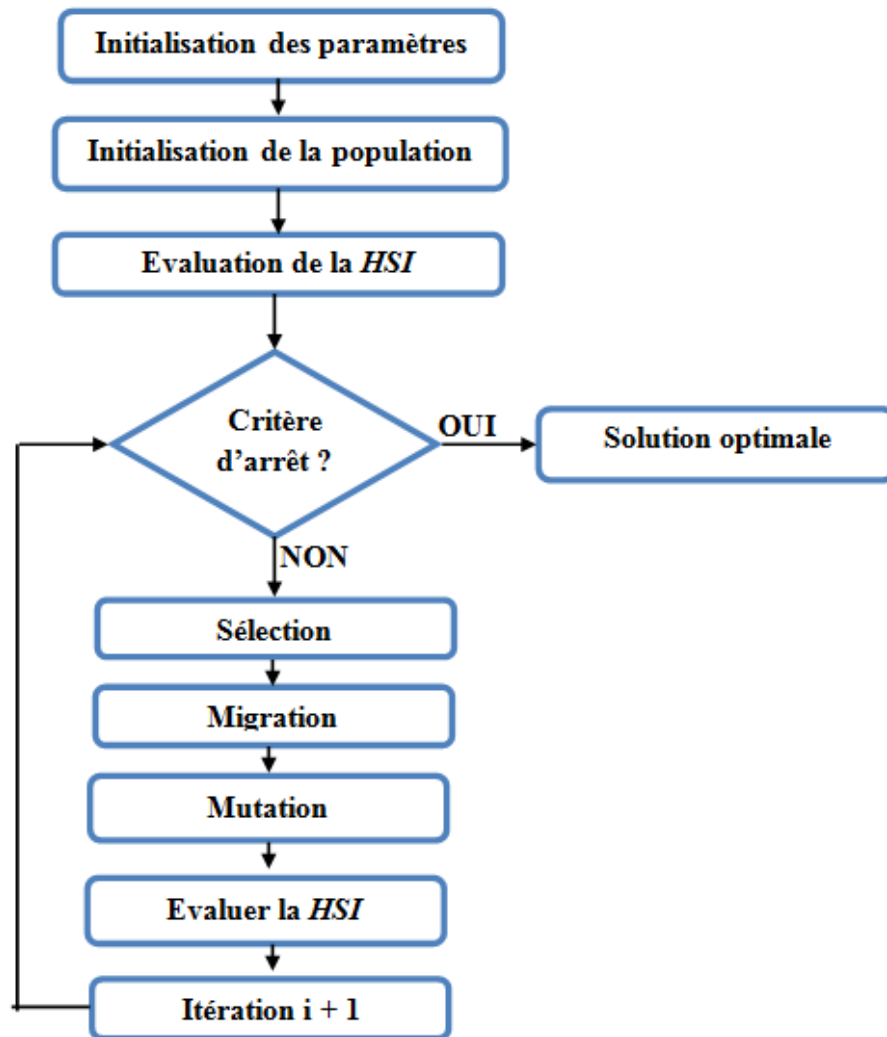


Figure. 5 : Organigramme de l'algorithme *BBO*.

A. Initialisation

Dans cette étape, nous définissons les paramètres de contrôle (Tableau. 3) et nous générons aléatoirement une population initiale de n habitats. La taille de cette population reste constante tout au long de l'algorithme.

Table. 3 : Paramètres de la *BBO*.

<i>Paramètres</i>	<i>Notation</i>
<i>Taille de la population</i>	N_p
<i>Probabilité de mutation</i>	P_m
<i>Taille de la mémoire élite</i>	n_{elit}
<i>Taux d'immigration maximal</i>	I
<i>Taux d'émigration maximal</i>	E
<i>Nombre maximum de générations</i>	G_{max}

B. Évaluation de la fonction *HSI*

Pour évaluer la pertinence d'une solution par rapport à une autre, nous calculons la valeur de la *HSI* correspondante à chaque solution candidate.

C. Sélection

Pour appliquer les opérateurs de l'algorithme BBO comme n'importe quel autre algorithme d'optimisation, nous devons sélectionner les habitats ou individus candidats à cet opérateur. La sélection est fondée sur la qualité des habitats, estimée à l'aide de fonction d'adaptation. Il existe plusieurs méthodes de sélection.

➤ *Sélection par roulette*

La population est représentée comme une roue de roulette, où chaque habitat est représenté par une portion qui correspond proportionnellement à sa valeur de *HSI* (fitness). La sélection d'un individu se fait en tournant la roue. L'un des inconvénients de ce type de sélection est de choisir presque toujours le même habitat s'il en existe un bien meilleur que tous les autres, ce qui cause une perte de diversité dans la population.

➤ *Sélection parrang*

La sélection par rang trie d'abord la population par *HSI*. Ensuite, chaque habitat se voit associé un rang en fonction de sa position. Ainsi le plus mauvais habitat aura le rang 1, le suivant 2, et ainsi de suite jusqu'au meilleur habitat qui aura le rang n . La sélection par rang d'un habitat est la même que par roulette, mais les proportions sont en relation avec le rang plutôt qu'avec la valeur de l'évaluation. Avec cette méthode de sélection, tous les habitats ont une chance d'être sélectionnés. Cependant, elle conduit à une convergence plus lente vers la bonne solution. Ceci est dû au fait que les meilleurs habitats ne diffèrent pas énormément des

plus mauvais.

➤ *Sélection par tournoi*

La sélection par tournoi est l'une des sélections les plus utilisées dans les algorithmes évolutionnaires. Le principe consiste à choisir un sous-ensemble d'individus (S individus) aléatoirement dans la population, puis à sélectionner le meilleur individu dans ce groupe en fonction de son *HSI*. Ce processus est répété jusqu'à l'obtention du nombre d'individus requis. Le nombre de participants à un tournoi, appelé la taille du tournoi, est utilisé pour faire varier la pression de cette sélection. Si ce nombre est grand, alors la pression sera forte et les faibles individus auront une petite chance d'être choisis. En général, un seul gagnant est choisi parmi les participants à un tournoi.

➤ *Élitisme*

A la création d'une nouvelle population, il y a de grandes chances que les meilleurs habitats soient perdus après les opérations de migration et de mutation. Pour éviter cela, nous utilisons la méthode d'élitisme. Elle consiste à copier un ou plusieurs des meilleurs habitats dans la nouvelle génération. Ensuite, nous générons le reste de la population selon le mécanisme de reproduction usuel. L'opérateur de sélection est utilisé pour sélectionner des habitats pour appliquer les deux opérateurs d'évolution : la migration et la mutation.

D. Opérateur de migration

La migration est un opérateur probabiliste utilisé pour modifier chaque solution H_i en partageant des caractéristiques parmi les différentes solutions. L'idée de l'opérateur de migration est basée sur la migration en biogéographie, qui représente le mouvement des espèces entre les différents habitats. La probabilité qu'une solution est sélectionnée pour immigrer ou émigrer dépend de son taux d'immigration λ_i ou le taux d'émigration μ_j . Le processus de migration est défini par la relation :

$$H_i(SIV_k) \leftarrow H_j(SIV_k). \quad (7)$$

L'équation (7) représente comment une caractéristique ou *SIV* d'une solution H_i est remplacée par une caractéristique ou *SIV* d'une solution H_j par l'opération de migration. Dans la *BBO*, chaque solution H_i a son propre taux d'immigration λ_i et taux d'émigration μ_i .

Ces deux taux sont calculés par l'équation (8) et l'équation (9) respectivement.

$$\lambda_i = I * \left(1 - \frac{K_i}{n}\right) \quad (8)$$

$$\mu_i = E * \left(\frac{K_i}{n}\right) \quad (9)$$

I et E représentent les taux maximaux possibles d'immigration et d'émigration respectivement. k_i représente le rang du i habitat après le tri de tous les habitats en fonction de leur HSI . n représente la taille de la population. La Figure. 6 illustre deux solutions candidates S_1 et S_2 à un problème en utilisant des courbes d'immigration et d'émigration symétriques ($E = I$). S_1 représente une mauvaise solution et S_2 représente une meilleure solution. La probabilité d'immigration pour S_1 sera donc plus élevée que celle de S_2 tandis que la probabilité d'émigration pour S_1 sera inférieure à la probabilité d'émigration pour S_2 .

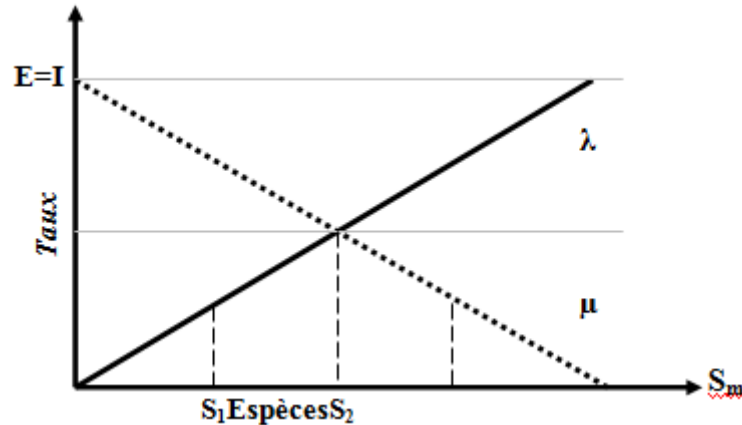


Figure .6 : Illustration de deux solutions candidates S_1 et S_2 .

E. Opérateur de mutation

En plus de l'opérateur de migration, nous avons l'opérateur de mutation. La mutation est un opérateur probabiliste utilisé pour modifier un ou plusieurs SIV d'une solution sélectionnée aléatoirement en se basant sur sa probabilité d'existence P_{ig} pour la génération g . Cet opérateur aide à introduire de nouvelles caractéristiques et augmente la diversité dans la population. La probabilité de mutation m_i est calculée en fonction de la probabilité de la solution exprimée dans l'équation (10).

$$m_i = m_{max} \left(1 - \frac{P_{ig}}{P_{max}} \right) \quad (10)$$

Où :

m_i : Le taux de mutation pour l'habitat i .

m_{max} : Le taux maximum de mutation.

P_{max} : La probabilité maximale d'existence. Avant d'appliquer l'opérateur de la mutation, nous devons calculer P_{ig} , la probabilité d'existence de l'habitat i dans la génération actuelle g . Ce paramètre est calculé à partir de sa valeur dans la génération précédente par l'équation (11).

$$P_{ig} = P_{ig-1} + P_{ig} \quad (11)$$

Les valeurs de P_i avant le début de l'algorithme sont initialisées à $1/n$ pour chaque habitat. P_i est le changement de probabilité d'existence de l'habitat i , il est donné par l'équation

(11) :

$$P_i = \begin{cases} -(\lambda_i + \mu_i) P_i + \mu_{i+j} P_{i+1} & i = 1 \\ -(\lambda_i + \mu_i) P_i + \lambda_{i-1} P_{i-1} + \mu_{i+j} P_{i+1} & 2 \leq i \leq n \\ -(\lambda_i + \mu_i) P_i + \lambda_{i+j} P_{i-1} & i = n \end{cases} \quad (12)$$

2.4.4. L'approche de classification à base de réseau de neurones artificiels et l'algorithme d'optimisation à base de biogéographie (BBO)

L'organigramme de la figure. 7 représente les sept étapes qu'ont été utilisées pour former un classificateur à base de réseau de neurones artificiels et l'algorithme BBO.

2.5. Résultat

La structure utilisée a classé efficacement les différents battements et fournit de résultat substantiel en termes de précision (95.7%).

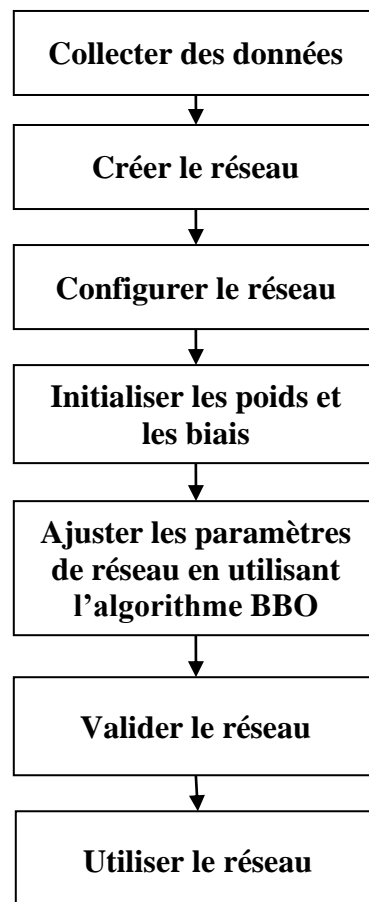


Figure. 7 : Organigramme de l'approche utilisée.

3. Conclusion

Dans ce chapitre nous avons présenté une méthode pour classifier le signalé lectrocardiogramme en cinq classes selon la recommandation d'AAMI. Cette méthode se compose de deux blocs principaux à savoir : le bloc d'extraction des caractéristiques et le bloc de classification. Pour le premier bloc nous avons utilisé une technique pour l'extraction des caractéristiques morphologiques et temporelles connue sous le nom de Pan and Tompkins. Après avoir extrait les vecteurs caractéristiques, ils seront utilisés comme l'entrée d'un classificateur qui représente le deuxième bloc. Ce classificateur est la base de réseau de neurones artificiels avec la particularité d'ajouter l'algorithme d'optimisation à base de biogéographie (BBO) pour ajuster les paramètres de ce classificateur. Finalement Un résultat excellent en termes de précision (95.7%)

CONCLUSION GÉNÉRALE

Les travaux principaux présentés dans ce mémoire visent à créer un système automatique permettant de diagnostiquer les maladies cardiaques. Généralement, ces systèmes se composent de deux phases :

- ✚ La première phase consiste à l'extraction des caractéristiques des différents battements constituant le signal électrocardiogramme
- ✚ La deuxième phase est le classement de ces battements selon le type de pathologie atteinte.

Dans ce mémoire de fin d'étude nous avons traité l'étape de diagnostic des maladies cardiovasculaires. Dans cette étape nous avons présenté l'extraction des caractéristiques et la classification des battements cardiaques selon la pathologie atteinte. Pour l'extraction des caractéristiques nous avons utilisé une méthode simple s'appelle Pan and Tompkins qui consiste à extraire les caractéristiques temporelles et morphologiques du signal électrocardiogramme autour de son complexe QRS. Finalement un vecteur caractéristique sera composé. Ce vecteur a été utilisé comme l'entrée du classificateur qui représente la deuxième phase du système expert de diagnostic des maladies cardiaques. Le classificateur utilisé dans ce mémoire est basé sur les réseaux de neurones artificiels. Nous avons utilisé un algorithme d'optimisation baptisé BBO pour ajuster les paramètres de classificateur, cet algorithme nous a permis de surmonter le problème de minimum local rencontré lors de l'utilisation des techniques d'optimisation à base de gradient. Le résultat obtenu à un taux de reconnaissance de classificateur (95.7%) ce qui montre l'efficacité de la méthode étudiée pour classer les différents battements cardiaques selon le type de pathologie.

BIBLIOGRAPHIE

- [1] BENALI Radhwan. Analyse du signal ECG par réseau adaptif d'ondelettes en vue de la reconnaissance de pathologies cardiaques. Doctorat 2013.
- [2] Christian Cabrol Raphael.Vialle Henry. Guérin-Surville. Anatomie du cœur humain. Niveau PCEM2 - EIA cardiologie .2002 – 2003.
- [3] NAIT-HAMOUD Mohamed Cherif. Segmentation et Classification du Signal ECG Par les SVM Flous Multi-classes. Magistère. Tébessa.2010
- [4] OUISSI Fatima Zohra née BILLAMI. Classification des arythmies cardiaques par les arbres de décision flous. Master . Tlemcen. 2010-2011
- [5] AMARA Houaria et HAMZAOUI Nour El Imane. Débruitage. de Signal Electrocardiogramme (ECG) par la Transformée de Walsh-Hadamard et la Transformée des Ondelettes : Applications dans le Domaine de la Diagnostic des Maladies Cardiaques. Master. Saida. 2015-2016
- [6] Nadia KRICHANE et Samira TAZEBOUDJT. Classification des signaux ECG par les réseaux de neurone probabiliste . MASTER .TIZI-OUZOU. 2015.
- [7] Rémi DUBOIS. Application des nouvelles méthodes d'apprentissage à la détection précoce d'anomalies en électrocardiographie. L'UNIVERSITÉ PARIS 6.
- [8] TRIQUI Bohra. Classification des arythmies cardiaques par la carte TSOM (Temporal Self Organizing Map), Doctorat, Oran, 2017-2018.
- [9] M elle . ADJLOUT Zahia, M elle. BELA Fariza. Extraction du rythme cardiaque foetal (RTF) à partir de l'ECG composite de la mère, Master, TIZI-OUZOU,2014.
- [10] LAKEHAL Soumia, Identification d'un individu par L'ECG (L'électrocardiogramme) , Master, M'sila, 2018
- [11] Nadia KRICHANE, Détection du complexe QRS d'un signal, ingénieur d'état, Tizi-Ouzou,2012.
- [12] M.Fedias, D.Saigaa, l'apport de la couleur à l'authentification de visage, Biskra, 2007.
- [13] Zid-el-khir CHALABI, Abderhim BOUDJEMAOUI, Larbi SAADIA et Nasreddine BERRACHED, Détection et Classification Automatiques d'Arythmies Cardiaques , d'Oran,2009.
- [14] Ali MESSAOUD, Mohamed BEN MESSAOUD, Abdennaceur KACHOURI et Faïçal SELLAMI, Classification des arythmies cardiaques par logique floue à partir de signaux ECG, Tunisie,2005
- [15] Zahia Zidelmal, Reconnaissance d'arythmies cardiaques par support Vector Machine(SVMs) ,Doctorat, TIZIOUZOU,2012.
- [16] M. ZAIZ Faouzi, Les Supports Vecteurs Machines (SVM) pour la reconnaissance des caractères manuscrits arabes, magister, BISKRA ,2010.
- [17] Y. Djeriri, Les Réseaux de Neurones Artificiels.
- [18] Melle. Nadia KRICHANE et Melle. Samira TAZEBOUDJT. Classification des signaux ECG par les réseaux de neurone probabiliste . MASTER .TIZI-OUZOU. 2015
- [19] David PANZOLI, Proposition de l'architecture « Cortexionist » pour l'intelligence comportementale de créature artificielles, Doctorat, France,2008.

- [20] DEGHMOUS Hamid, méthodes supervisé et non supervisées en classification des signaux ECG , Master, Bejaia,2016
- [21] A. Cornuéjols , Une nouvelle méthode d'apprentissage : « Les SVM. Séparateurs a vaste marge »France, Juin 2002.
- [22] Naima IFREK, Lahna BOUSSAID, Etude et application du réseau ELM (Extreme Learning Machine) pour la classification de donnes, master, Tizi-Ouzou,2017.
- [23] Mr.belgacem.Amar, Classification des signaux ECG avec un système-multi-agent neuronale, Magister ,TELMCEN,2012.
- [25] Lotfi Baghli---- Contribution à la commande de la machine asynchrone, utilisation de la logique floue, des réseaux de neurones et des algorithmes génétiques,Doctorat,1999.
- [27]G.DREYFUS---LES R2SEAUX DE NEURONNES, Paris, 1998.
- [28] G. Dreyfus et al.,Eyrolles, Réseaux de neurones : méthodologie et application, 2002.
- [29]Ramla Halimi, Yamina Hammouya, classification d'un signal ECG par RNA(RBF),master,Ourgla,2019

ملخص

يركز العمل المقدم في أطروحة الماستر هذا على تحقيق نظام تشخيص تلقائي للأمراض القلب دون تدخل بشري ، لذلك نقوم بتكبير إشارة مخطط كهربية القلب (ECG) . لأنه يقيس النشاط الكهربائي للقلب ويستخدم على نطاق واسع للكشف عن مرض القلب بسبب بساطته ، ولتحقيق هدفنا من الضروري تمرير إشارة مخطط القلب من خلال خطوتين ، الأولى هي استخراج الخصائص باستخدام طريقة بان وتومبكينز ، وهذه الطريقة هي في الأساس مجموعة من المرشحات من أجل الحصول على متجه الميزات ، والذي سيتم استخدامه كمدخل للمصنف التصنيفات التي تم إجراؤها باستخدام الشبكات العصبية ، نستخدم هذا النظام لتصنيف أنواع ضربات ECG وفقاً لمعيار AAMI ، وأخيراً لتقييم نظامنا ، نستخدم-MIT "BIH Physio Bank".

الكلمات المفتاحية: ECG ، التصنيف ، الاستخراج ، ميزات ، الشبكة العصبية، (م ك ق)، عدم انتظام ضربات القلب

Résumé

Le travail présenté dans ce mémoire de master porte essentiellement sur la réalisation d'un système expert pour diagnostiquer automatiquement les maladie cardiaques sans l'intervention humain, et pour cela nous adaptons sur le signal électrocardiogramme (ECG), car il mesure l'activité électrique du cœur et a été largement utilisé pour détecter les maladie cardiaques en raison de sa simplicité et de sa nature non invasive , et pour atteindre à notre objectif il faut passer le signal électrocardiogramme par deux étape parce essentielles la premier est l'extraction des caractéristique en utilisant la méthode Pan and Tompkins, ce dernier est essentiellement un ensemble de filtres afin d'obtenir le vecteur des caractéristiques, qui sera utilisé comme un entrée pour le classificateur. La classification est effectuée en utilisant les réseaux de neurones, nous utilisons ce système pour classer les signaux ECG en cinq type de battement selon le standard AAMI, et pour évaluer notre système, nous utilisons la base de données « MIT-BIH PhysioBank ».

Les mots clés: ECG classification, extraction, vecteur de caractéristiques, réseau de neurones ,QRS, arythmie,

Abstract

The work presented in this master's thesis focuses on the realization of a a system automatically diagnose heart disease without human intervention, for this we adapt the electrocardiogram (ECG) signal, because it measures the electrical activity of the heart and has been widely used to detect heart disease due to its simplicity, to achieve our target it is necessary to pass the electrocardiogram signal through two steps, the first is the extraction of the characteristics using the Pan and Tompkins method, this method is basically a set of filters in order to get the feature vector, which will be used as an input for the classifier. The classifications done using neural networks, we use this system to classify ECG beat types according to AAMI standard, finally for evaluate our system we use 'MIT-BIH PhysioBank'.

Keywords: ECG , classification, extraction ,features vector, neural network, QRS, Ahythmia