

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Abderrahmane MIRA – Bejaia
Faculté de Technologie
Département de Génie Electrique

Projet de fin d'études

En vue de l'obtention du diplôme de Master

Filière : Electronique

Option : Télécommunication

Thème :

**Méthodes Supervisées et Non Supervisées en Classification des
Signaux ECG**

Présenté par :

Mr DEGAMOUS Hamid

Encadré par :

Mr. A. ALLICHE

Mr. A. MEKHMOUKH

Promotion 2016

Remerciements

Avant toute chose, je remercie Dieu, le tout puissant de m'avoir donné la force et le courage de mener à bien ce modeste travail.

Je tiens à exprimer mes remerciements à mon promoteur Mr A. ALLICHE d'avoir proposé ce thème, ainsi que pour son suivi et ses conseils instructifs et précieux.

Je tiens à exprimer aussi toute ma gratitude à mon Co-promoteurs Mr A.MEKHMOUKH pour son orientation, son suivi, ses conseils, sa disponibilité et pour le temps qu'il m'a consacré tout au long de l'élaboration de ce travail.

Mes remerciements s'adressent aussi aux membres de jury d'avoir acceptés d'examiner et d'évaluer ce travail.

Enfin je tiens à remercier toutes les personnes qui ont contribués d'une manière ou d'une autre, que ce soit par un conseil, une idée ou même leur présence, a la réalisation de ce mémoire.

Dédicace

A ceux qui m'ont élevés par une bonne éducation, qui m'ont allumé le chemin du savoir depuis mon enfance jusqu'à ma soutenance pour que je puisse atteindre mon objectif. A ceux qui m'ont procurés le confort, la paix. A ceux qui ont tout fais pour moi

Mes très cher parents que dieu les garde.

A mes sœurs ainsi qu'à mon frère

Lynda, Nassima & Massi.

A toute ma famille qui m'a soutenu et aidé durant tout mon parcours

Mes grands parents, mes tantes, mes oncles, tous mes cousins et cousines.

Je vous remercie infiniment pour vos encouragements.

A ceux avec qui j'ai passé les meilleurs souvenirs de ma vie que ce soit les bons ou les mauvais, mes ami(e)s sans exception:

Nassim, Djidji, Nabil, Sosso, Rafik, Mina, Belhou, Soria, Sisana, Tata,...

Hamid

Table des matières

Introduction générale	1
Chapitre I : Le cœur et l'électrocardiographie	
I.1. Introduction	3
I.2. Système Cardiovasculaire	3
I.2.1. le cœur	3
I.2.2 Fonctionnement électrique du cœur	4
I.3. L'Electrocardiographie :	5
I.3.1. Principe de l'ECG	5
I.3.2. Ondes et intervalles enregistrées sur un ECG	6
I.3.3. le rythme cardiaque	8
I.4. Les pathologies cardiaques	8
I.4.1. Rythme sinusal	9
I.4.2. Arythmies courantes	9
I.4.2.1. Extrasystole	10
I.4.2.2. Bradycardie	10
I.4.2.3. Tachycardie	10
I.4.2.4. Fibrillation Auriculaire	11
I.4.2.5. Flutter Auriculaire	11
I.5. Artefacts visibles sur l'ECG	11
I.5.1. Bruits techniques	11
I.5.2. Artefacts physiques	12
I.6. Base de données MIT-BIH	13
Conclusion	13
Chapitre II : Méthodes de classification	
II.1. Introduction	14
II.2. Prétraitement des signaux ECG	14
II.2.1. Dé-bruitage du signal ECG par la méthode des ondelettes	15
II.2.1.1. Etat de l'art	15
II.2.1.2. Transformée en ondelettes continues	17
II.2.1.3. Transformée en ondelettes discrètes	18
II.2.2. Détection du complexe QRS	18
II.1.3. Détection de pic R	23
II.1.4. Détection de l'intervalle R-R	24

Table des matières

II.2. La classification	25
II.3. Objectif de la classification	25
II.4. Méthode non supervisée	26
II.4.1. Le K-means	28
II.4.2. FCM	30
II.5. Méthodes supervisées	32
II.5.1. Réseaux de neurones	32
II.5.1.1. Définition	32
II.5.1.2. Eléments de base du réseau neuronal	33
II.5.1.3. Architecture des réseaux de neurones	35
II.5.1.4. Apprentissage du réseau	35
II.5.1.5. Avantages et inconvénients des réseaux de neurones	36
Conclusion	36

Chapitre III: Présentation de l'interface graphique.

III.1. Introduction	37
III.2. Fenêtre de présentation générale	37
III.3. Fenêtre de commande principale	38
III.4. Prétraitement de l'ECG	39
III.5. Détection du complexe QRS et de l'intervalle R-R	41
III.6. Classification	45
Conclusion	51
Conclusion et Perspectives.....	52

Liste des Figures et des Tableaux

Liste des figures

Figure 1.1. Anatomie du cœur.....	4
Figure 1.2. Système de conduction électrique du cœur.....	5
Figure 1.3. Etapes successives de dépolarisation/repolarisation du cœur qui se traduisent sur le plan électrocardiographique par différentes ondes P, QRS, T.....	6
Figure 1.4. Ondes liées à la contraction et décontraction du cœur.....	6
Figure I.5. Rythme sinusal.....	9
Figure II.1. Principe de la classification.....	14
Figure II.2. Etape de la classification du signal ECG.....	15
Figure II.3. Principe de dé-bruitage par ondelettes.....	16
Figure II.4. Complexe QRS.....	19
Figure II.5. Diagramme des étapes de l’algorithme de Pan et Tompkins.....	19
Figure II.6. Réponse en fréquence du filtre dérivateur représenté à l’échelle logarithmique.....	21
Figure II.7. (a) : ECG filtré, (b) : sortie du filtre dérivateur, (c) : sortie du filtre intégrateur.....	22
Figure II.8. Réponse en fréquence du filtre proposé par Pan & Tompkins.....	23
Figure II.9. Détection de l’onde R. (a) signal filtré, (b) sélection de la double différence des pics, (c) détection du complexe QRS, (d) détection des ondes R.....	24
Figure II.10. Organigramme de classification.....	26
Figure II.11. Les deux types de clustering hiérarchique/non hiérarchique.....	27
Figure II.12. Partition hiérarchique.....	27
Figure II.13. Schéma d’un neurone biologique.....	33
Figure II.14. Neurone formel.....	34
Figure II.15. Différents types de fonctions d’activation pour le neurone artificiel.....	34
Figure II.16. Réseaux de neurones les plus utilisés.....	35
Figure III.1. Fenêtre de présentation.....	37
Figure III.2. Fenêtre de commande principale.....	38
Figure III.3. Fenêtre de Prétraitement.....	39
Figure III.4. Prétraitement de l’ECG_112.....	40
Figure III.5. Prétraitement de l’ECG_118.....	40
Figure III.6. Prétraitement de l’ECG_230.....	41
Figure III.7. Fenêtre de Détection du complexe QRS et de l’intervalle R-R.....	41
Figure III.8. Détection de R et de l’intervalle R-R du signal ECG_112.....	42

Liste des Figures et des Tableaux

Figure III.9. Détection de R et de l'intervalle R-R du signal ECG_118	42
Figure III.10. Détection de R et de l'intervalle R-R du signal ECG_230	43
Figure III.11. Tracé du complexe QRS du signal ECG_112	44
Figure III.12. Tracé du complexe QRS du signal ECG_118	44
Figure III.13. Tracé du complexe QRS du signal ECG_230	45
Figure III.14. Fréquence cardiaque des signaux ECG_112, 118 & 230 respectivement.	45
Figure III.15. Fenêtre de classification	46
Figure III.16. Classification du signal ECG_112 avec la méthode K-Means	47
Figure III.17. Classification du signal ECG_118 avec la méthode K-Means	47
Figure III.18. Classification du signal ECG_230 avec la méthode K-Means	48
Figure III.19. Classification du signal ECG_112 avec la méthode FCM	58
Figure III.20. Classification du signal ECG_118 avec la méthode FCM	49
Figure III.21. Classification du signal ECG_230 avec la méthode FCM	49
Figure III.22. Réseau de Neurones	50
Figure III.23. Résultats de Détection des PVC pour : (a): ECG_112, (b): ECG_118 & (c): ECG_230	51

Liste des tableaux

Tableau III.1. Intervalles R-R	43
--------------------------------------	----

Liste des abréviations

ECG : Electrocardiogramme.

NAV : Nœud Auriculo Ventriculaire.

NS : Nœud Sinusal.

ESV : Extrasystole Ventriculaire.

MIT-BIH :

TF : Transformée de Fourier.

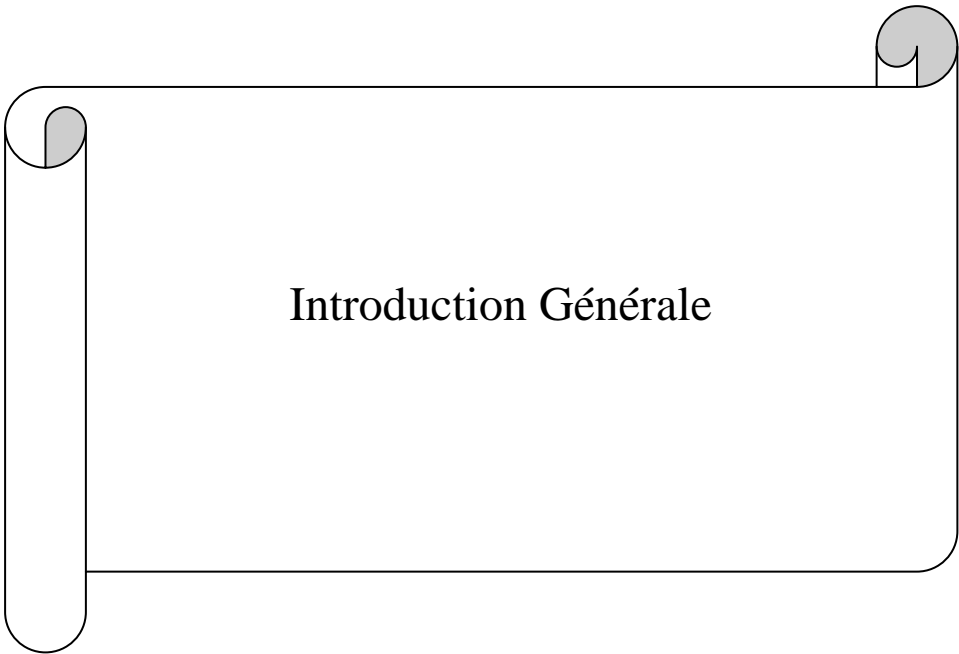
TOD : Transformée en Ondelettes Discrète (DWT).

TOC : Transformée en Ondelettes Continue (CWT).

RII : Réponse Impulsionnelle Infinie.

RN : Réseaux de Neurones.

PVC : Premature Ventricular Contraction.



Introduction Générale

Introduction Générale

Introduction générale :

Les maladies cardiovasculaires constituent un problème majeur de santé publique ; avec un peu plus de 18 millions de personnes atteintes dans le monde. Les facteurs de risques sont multiples : tabac, obésité, hypertension artérielle, diabète...etc.

Le cœur est l'organe central du système cardiovasculaire, il peut être affecté par de nombreuses pathologies qui peuvent être dans certains cas bénignes, comme certaines tachycardies, ou s'avérer dangereuses, comme l'infarctus du myocarde ; c'est pour cela que l'étude et le suivi de son activité présente un intérêt majeur pour chaque être humain afin de détecter et de prévenir une possible dégénérescence en arythmies sévères.

Avec l'évolution des techniques, les médecins disposent aujourd'hui d'outils performants pour observer le fonctionnement du muscle cardiaque et de poser leur diagnostic. Parmi les examens cardiologiques utilisés, l'Électrocardiogramme (ECG).

L'électrocardiogramme (ECG) est un enregistrement de l'activité électrique du cœur en fonction du temps. Le potentiel est recueilli par des électrodes posées à la surface du corps. Il se présente sous forme d'une suite de déflexions (ondes) répétitives représentant chaque phase du fonctionnement du cœur. Chaque déformation visible sur ces ondes peut être attribuée à un dysfonctionnement cardiaque. L'ECG joue un rôle important dans la surveillance et le diagnostic des patients en raison de la rapidité à mettre en place, l'efficacité et la fiabilité pour le diagnostic. Ce pendant, l'ECG n'est performant que s'il est enregistré sur une longue durée, l'analyse d'un tel enregistrement nécessite des méthodes d'extraction des paramètres et de classification. Beaucoup de méthodes de classification ont été proposées ces dernières années. On peut scinder ces méthodes en deux familles qui sont les méthodes « supervisées » et « non supervisées ».

Ce présent travail sera consacré à l'étude des techniques de classification énoncé dans le paragraphe précédent, dont l'organisation est subdivisée en trois (03) chapitres comme suit :

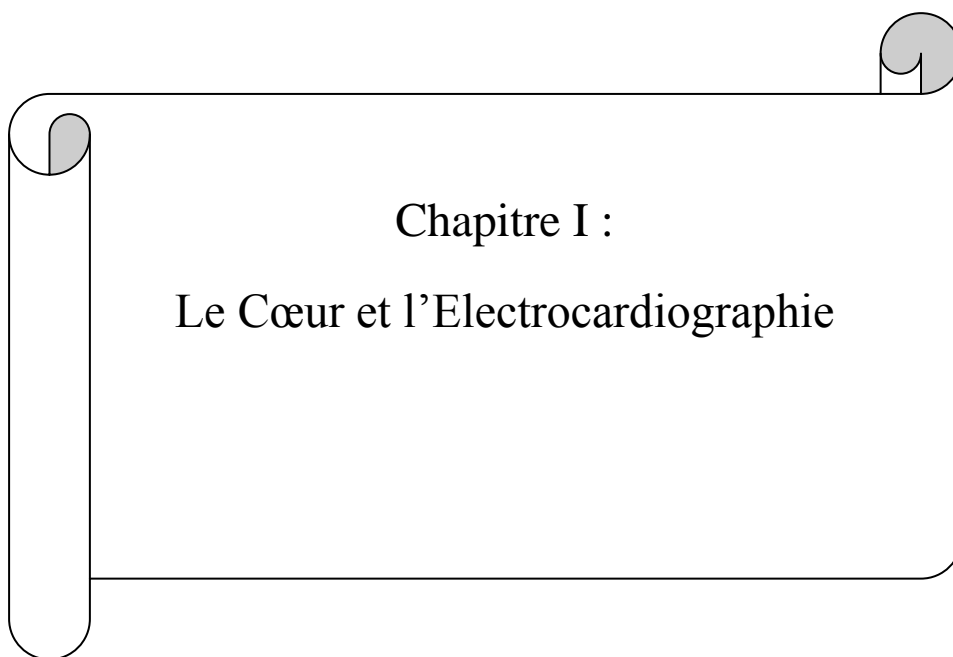
Introduction Générale

Le premier chapitre, présentera des généralités concernant l'aspect physiologique du cœur et son fonctionnement, l'électrocardiogramme (ECG) et les multiples artefacts pouvant affecter l'ECG, ainsi que de quelques pathologies cardiaques typiques, nous présenterons aussi la base de données MIT-BIH.

Le deuxième chapitre sera consacré à la phase de prétraitement des signaux ECG ainsi qu'à la présentation des différentes méthodes de classifications citées précédemment.

Dans le troisième chapitre, nous développerons et présenterons une interface graphique sous MATLAB illustrant les différentes méthodes d'analyse des signaux ECG.

Enfin, nous terminerons ce travail par une conclusion générale et des perspectives de recherche dans ce domaine.



Chapitre I :
Le Cœur et l'Electrocardiographie

1.1. Introduction :

Depuis la découverte de l'existence de la petite circulation (pulmonaire) et de la grande circulation (systémique) par William Harvey en 1615, de nombreux travaux ont enrichis la connaissance de la circulation sanguine.

Dans ce premier chapitre, nous allons présenter le système cardiovasculaire et plus précisément le cœur ainsi que son activité électrique, puis de manière détaillée nous expliquerons le principe de l'Electrocardiographie (ECG), et pour finir, parler des pathologies cardiaques et des moyens de les diagnostiquer. Pour illustration, nous avons considéré les enregistrements recueillis par la base de données MIT-BIH Arrhythmia. L'introduction à la cardiologie est limitée dans ce chapitre aux notions utiles à la compréhension.

1.2. Système Cardiovasculaire :

Le système cardiovasculaire appelé aussi système circulatoire assure la circulation du sang de manière continue, ce qui permet d'alimenter des tissus cellulaires en oxygène et en nutriments indispensables à la vie et d'évacuer les déchets vers les reins et le dioxyde de carbone vers les poumons. Ce système est composé d'un organe pompe qui est le cœur et d'un réseau permettant le transport du sang qui est le système vasculaire (composé de veines et d'artères).

1.2.1. le cœur :

Le cœur est un muscle creux appelé Myocarde situé dans la cage thoracique, dans une zone appelée médiastin antérieur, entre les deux poumons. Il est composé de quatre cavités cardiaques qui sont : l'oreillette droite, l'oreillette gauche, le ventricule droit et le ventricule gauche. Le cœur est séparé en deux parties : la droite (cœur droit) et la gauche (cœur gauche) ; chacun ayant une oreillette et un ventricule. Les deux parties sont aussi séparées par une paroi appelée « septum », comme le montre la figure I.1. [1, 2, 3]

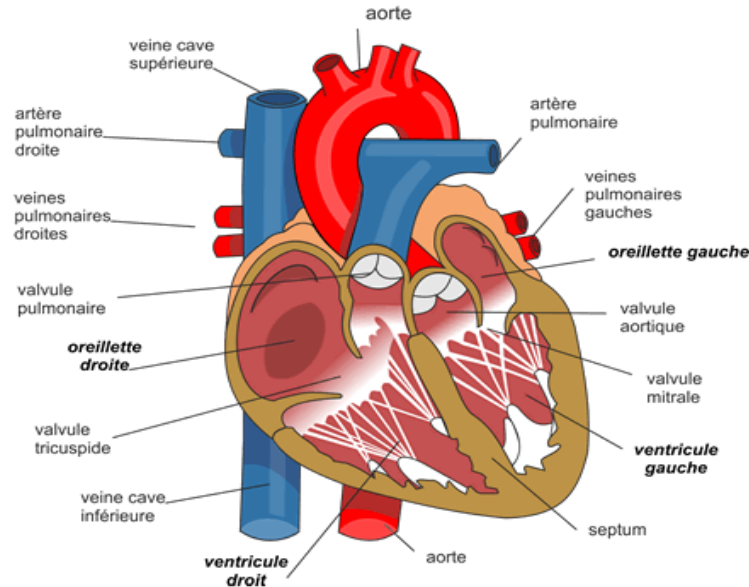


Figure 1.1. Anatomie du cœur.

Le sang pauvre en oxygène arrive au cœur par la veine cave. Il y rentre par l'oreillette droite, et en est chassé par sa contraction appelée systole auriculaire qui le déplace dans le ventricule droit. La contraction ventriculaire (systole ventriculaire) propulse le sang du ventricule droit vers les paumons ou il va se charger en oxygène. De retour au cœur par les veines pulmonaires, le sang s'accumule dans l'oreillette gauche puis, lors de la contraction auriculaire, passe dans le ventricule gauche qui lors de la systole ventriculaire l'envoie vers les organes par l'artère Aorte.

La phase de contraction et la phase de relaxation sont respectivement appelées « Systole » et « Diastole ». Cette suite de systole et de diastole sont coordonnées par une impulsion électrique qui le stimule.

Dans le prochain paragraphe, nous allons aborder le processus de contraction cardiaques du point de vue électrique. [1, 2, 3]

1.2.2 Fonctionnement électrique du cœur :

Le fonctionnement du cœur repose essentiellement sur la systole et la diastole qui est généré par une impulsion électrique qui excite les cellules musculaires dans un ordre bien établi afin que la contraction soit la plus efficace possible.

Le courant électrique est issu d'un point précis du cœur, appelé nœud sinusal, situé au sommet de l'oreillette droite. Le courant qui est de quelques millivolts est généré par un ensemble de cellule. Il se propage en tache d'huile dans le muscle cardiaque, a partir du nœud sinusal provoquant ainsi les

contractions en circulant dans les deux oreillettes du cœur ; de là, le courant électrique est acheminé vers la zone séparant oreillettes et ventricules au niveau du Nœud Auriculo-ventriculaire (NAV), situé en bas de l'oreillette droite.

A partir du NAV, l'impulsion progresse dans les deux ventricules simultanément, empruntant ainsi deux voies conductrices très rapides qui sont le faisceau de His, situé dans la partie haute du septum inter-ventriculaires, et les fibres de Purkinje, arrivant alors au cœur elles provoquent la contraction de celui-ci. Le nombre de contractions observées par minutes correspond au rythme cardiaque ou « pouls ». [4]

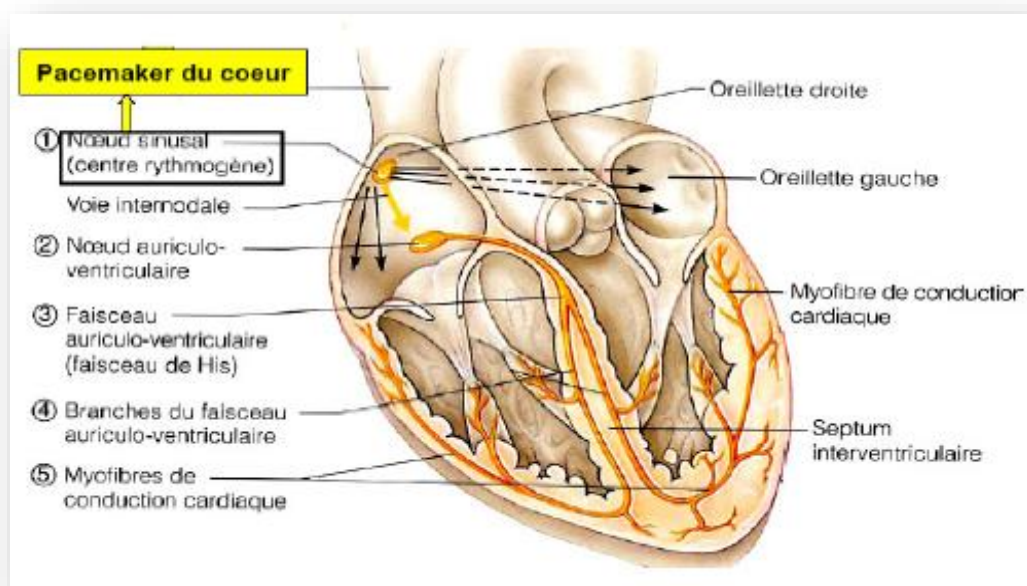


Figure 1.2. Système de conduction électrique du cœur.

1.3. L'Electrocardiographie :

1.3.1. Principe de l'ECG :

L'électrocardiographie (ECG) est un examen qui mesure et enregistre l'activité électrique du cœur. Elle est utilisée pour dépister, soigner ou assurer le suivi d'une pathologie cardiaque. Le corps humain est considéré comme étant électriquement conducteur. Par conséquent, les potentiels d'action générés au niveau des fibres cardiaques lors de l'activité mécanique cardiaque peuvent être réalisés par la pose d'électrodes au niveau du thorax, des poignets et des chevilles ; grâce auxquelles il est possible d'enregistrer l'activité électrique du cœur. [5]

La contraction du cœur est caractérisée par des ondes appelées onde P, QRS, T, U. [6]

I.3.2. Ondes et intervalles enregistrés sur un ECG :

Les informations enregistrées par l'ECG sont présentées sous forme d'ondes électriques, de formes et de durées particulières qui se répètent à chaque cycle cardiaque.

Ces ondes traduisent les différents phénomènes mécaniques relatifs au parcours du potentiel d'action de la stimulation cardiaque et dont les étapes sont illustrées sur la figure suivante :

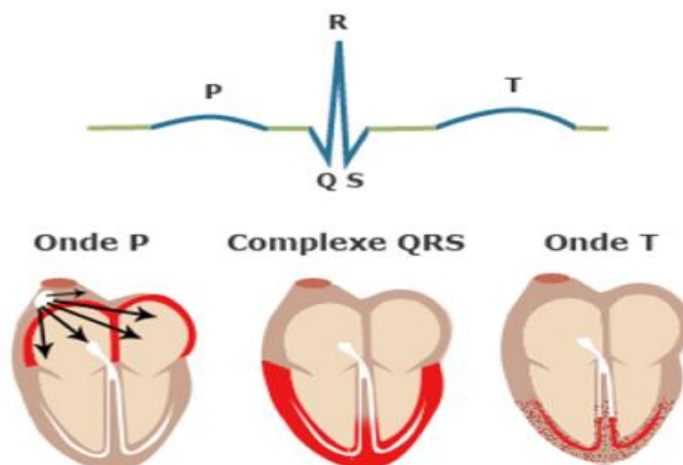


Figure 1.3. Etapes successives de dépolarisation/repolarisation du cœur qui se traduisent sur le plan électrocardiographique par différentes ondes P, QRS, T.

D'après le signal électrocardiographique (ECG), le processus de contraction et de décontraction du myocarde se présente sous forme de séquence de déflexions positives et négative superposées à une ligne du potentiel zéro (ligne de base) qui correspond à l'absence des phénomènes cardiaques comme illustré dans la figure suivante :

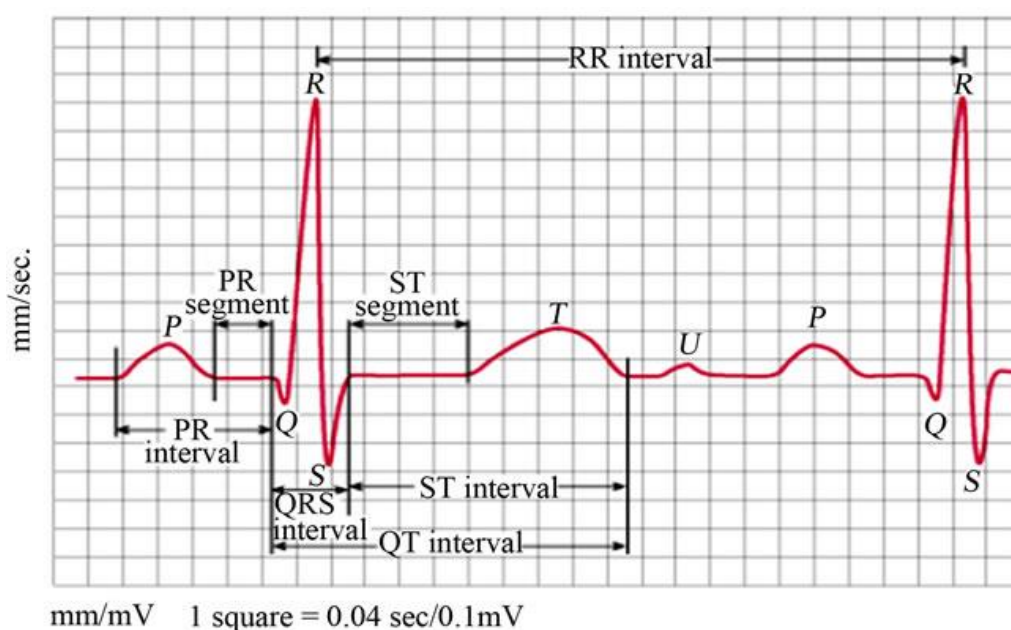


Figure 1.4. Ondes liées à la contraction et décontraction du cœur.

Pour chaque battement cardiaque, l'électrocardiogramme (ECG) enregistre quatre ondes successives (Figure I.4). [6]

➤ **Onde P :**

Première onde enregistrée. Elle correspond à la dépolarisation des oreillettes quand l'impulsion électrique se propage à partir du nœud sinusal. L'onde P, a une amplitude inférieure ou égale à 0.2mV et une durée inférieure à 120ms. C'est l'onde qui précède le complexe QRS. [5]

➤ **Complexe QRS :**

Il est composé de 3 ondes accolées qui suivent l'onde P, et qui correspond à la dépolarisation des ventricules. Pour un adulte sain, son amplitude est comprise entre 5 et 20mV, et sa durée est inférieure à 0.12s. [5] Par définition :

- L'onde Q : première onde négative du complexe.
- L'onde R : première onde positive du complexe.
- L'onde S : première onde négative après l'onde R.

➤ **Onde T :**

Elle correspond au courant de repolarisation des ventricules. Elle est de faible amplitude et ne témoigne d'aucun événement mécanique. Cette onde succède au complexe QRS.

➤ **Onde U :**

Onde positive qui suit l'onde T, dans certaines occasions. Elle est de faible amplitude et elle est visible dans certains cas notamment chez les athlètes. L'onde U est souvent associée aux processus de repolarisation ventriculaire tardive du myocarde.

En plus de différentes ondes citées, il existe un certain nombre d'intervalles et de segments qui portent des informations très utiles. [6]. Les intervalles et les segments les plus importants sont :

➤ **Intervalle RR :**

Il correspond au délai entre 2 dépolarisations ventriculaire. Cet intervalle représente l'espace entre deux ondes R consécutives et dont la durée détermine la fréquence cardiaque instantanée FC (N^{bre} de battements/minute) donnée par la formule suivante :

$$FC = \frac{1}{\text{période}} * 60$$

➤ **Intervalle PR :**

Il correspond au délai entre la fin de la dépolarisation des oreillettes (début de l'onde P), et le début de la dépolarisation ventriculaire (début du QRS). Sa durée reflète la conduction entre le NAV qui est normalement comprise entre 120 et 200ms.

➤ **Intervalle QRS :**

Cet intervalle représente le temps total de dépolarisation ventriculaire. Sa durée est de 70 à 100 ms.

➤ **Intervalle ST :**

Il correspond à la phase où les cellules ventriculaires sont dépolarisées ; il n'y a donc a priori pas de propagation électrique. Le segment ST va de la fin du complexe QRS jusqu'au début de l'onde T. sa durée est comprise entre 50 et 150 ms.

➤ **Intervalle QT :**

Cet intervalle est mesuré depuis le début du complexe QRS jusqu'à la fin de l'onde T. il représente la durée totale de la dépolarisation et de la repolarisation ventriculaire. Sa valeur dépend de la fréquence cardiaque FC. Il est normalement compris entre 350 et 440 ms.

1.3.3. le rythme cardiaque :

La fréquence cardiaque ou le rythme cardiaque peut être mesuré par le pouls. C'est le nombre de pulsation (battement) cardiaques par minute. Il doit, théoriquement, être de 70 pulsations/min chez un homme au repos et doit être plus élevé lors de la pratique d'un sport, pendant la digestion,...etc. S'il apparaît trop élevé, on parle de tachycardie. S'il apparaît trop bas on parle de bradycardie. Le rythme cardiaque est un outil de diagnostic indispensable pour les troubles cardiaques, notamment en cas d'arythmie.

1.4. Les pathologies cardiaques :

Le meilleur outil pour diagnostiquer les arythmies est l'électrocardiogramme. Dans l'analyse de l'ECG, les pathologies sont détectées en fonction de leur différence avec le rythme idéal qui est le rythme sinusal. Chaque déviation visible sur l'ECG par rapport au rythme sinusal peut être attribuée à une anomalie physiologique. [7,8]

1.4.1. Rythme sinusal :

Rythme normal du cœur selon l'électrocardiogramme. Il est avec une synchronisation correcte de la contraction du cœur suivant la séquence nœud sinusal. Il est caractérisé par un rythme cardiaque régulier, compris entre 60 et 80 battements / minute chez l'adulte au repos, avec un intervalle régulier entre chaque battements normaux.

On dit qu'un ECG est sinusal si:

- Chaque QRS, est précédé d'une onde P.

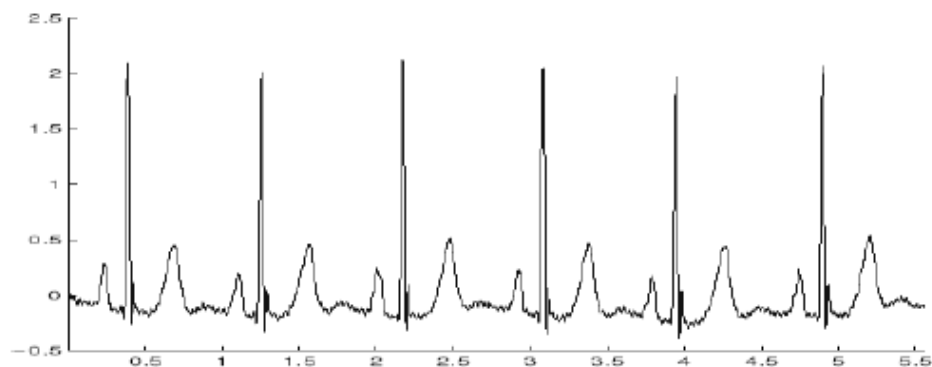


Figure 1.5. Rythme sinusal.

1.4.2. Arythmies courantes :

Les arythmies cardiaques sont une cause majeure de mortalité et constituent une partie très importante de la problématique des maladies cardiovasculaires. L'arythmie survient lorsque le cœur bat de façon irrégulière ou s'il bat à moins de 60 pulsations ou plus de 100 pulsations / minute, sans que cela soit justifié par un effort physique par exemple.

L'arythmie est le trouble cardiaque le plus fréquent. Les impulsions électriques qui contrôlent les battements du cœur se produisent de façon désordonnée ou ne passe pas par les circuits électriques habituels. La contraction cardiaque qui résulte s'écarte du rythme normal. Le battement supplémentaire qu'elle provoque est appelé extrasystole, par opposition à la contraction normale appelée systole. [7]

La durée d'une arythmie varie beaucoup d'un individu à l'autre et dépend aussi du type d'arythmie. [7]

Les arythmies sont classées selon le lieu où elles prennent naissance. Si le trouble se déclare dans l'oreillette ou le NAV, on parlera d'arythmie supra-ventriculaire ; s'il apparaît dans les ventricules, on parlera d'arythmies ventriculaire. Si une arythmie accélère le rythme cardiaque, on parle de tachycardie (tachy=rapide) ; si elle ralentit, on parle alors de bradycardie (brady=lent). [7]

Il existe plusieurs types d'arythmies, voici les principales :

I.4.2.1. Extrasystole :

Les extrasystoles sont des battements du cœur provoqués, non par les impulsions du nœud sinusal, mais par des impulsions déclenchées ailleurs dans les oreillettes (extrasystoles supra-ventriculaires) ou dans les ventricules (extrasystoles ventriculaires). Ces battements perturbent le rythme sinusal et s'accompagnent d'une sensation de palpitation ou de pause cardiaque. La fréquence cardiaque est supérieure à 100bpm. [11]

I.4.2.2. Bradycardie :

C'est un battement lent du rythme cardiaque, sa fréquence est inférieure à 60bpm ; elle est dite d'origine sinusale, jonctionnelle, ou ventriculaire, selon l'origine de l'impulsion électrique au début des battements considérés. [11]

- **Bradycardie sinusale** : le rythme est de l'ordre de 40bpm, elle est caractérisé par la présence d'une P avant le complexe QRS, l'origine de la contraction ventriculaire est la dépolarisation des oreillettes, comme lors des battements normaux.
- **Bradycardie jonctionnelle** : le rythme est de l'ordre de 35bpm. L'impulsion électrique n'est plus transmise des oreillettes aux ventricules. La jonction AV a donc pris le relais, et les ondes d'origine ventriculaire de forme normale ne sont pas précédées d'onde P.
- **Bradycardie ventriculaire** : le rythme est une succession d'ESV à une fréquence très lent, entre 15 et 40bpm, appelé rythme d'échappement ventriculaire. Ce type de pathologie nécessite une pose d'un stimulateur cardiaque pour éviter une possible syncope.

I.4.2.3. Tachycardie :

A l'inverse de la bradycardie, la tachycardie est caractérisée par le fait que la fréquence est supérieure à 100bpm. Elle peut être d'origine sinusale, auriculaire ou ventriculaire. [11]

- **Tachycardie Sinusale** : le rythme est compris entre 100 et 180bpm. On note la présence d'une onde P avant les complexes QRS. Pour s'assurer du caractère sinusal et ne pas confondre avec la tachycardie auriculaire il faudrait comparer la forme de l'onde P observé lors des battements normaux. Elle est considérée comme pathologique si elle est de longue durée.

- **Tachycardie auriculaire** : le rythme est compris entre 120 et 200bpm. Cette accélération du rythme cardiaque ne vient ni des oreillettes ni des ventricules, elle a pour origine un foyer ectopique, une boucle de stimulation (Flutter) ou une voie qui court-circuite la voie AV.
- **Tachycardie ventriculaire** : Elle a pour origine un ou plusieurs foyers ectopiques ventriculaires. Les battements ont donc la forme d'ESV très rapprochées. Ce rythme est dangereux à cause de sa possible évolution en FV qui, elle conduit au décès du patient si elle n'est pas traitée dans les minutes qui suivent son apparition, car les battements sont très rapprochés.

I.4.2.4. Fibrillation Auriculaire :

La FA est la forme d'arythmie la plus répandues. Le rythme cardiaque peut atteindre les 200bpm. Le cœur se contracte de manière anarchique, rapide et irrégulière. On n'observe plus d'activité auriculaire précise (onde P) à cause de la désynchronisation au niveau des oreillettes. Ces mouvements anarchiques transmettent quelques impulsions au NAV qui les transmet aux ventricules et entraîne une systole ventriculaire normale mais irrégulière. L'alimentation des tissus n'étant plus assurée, une intervention est impérative, sinon l'épuisement de ses forces de contraction conduira à l'insuffisance cardiaque. [11]

I.4.2.5. Flutter Auriculaire :

La fréquence de l'onde P peut atteindre 300bpm ; à cette fréquence, le nœud AV n'arrive pas à conduire les impulsions électriques vers les ventricules. Dans ce type de pathologie, il existe un risque majeur pour la santé du patient, dû aux turbulences dans le flux sanguin au niveau des oreillettes. [11]

I.5. Artefacts visibles sur l'ECG :

Sur tout enregistrement électrocardiographique il peut apparaître des événements indésirables pouvant brouiller le tracé et parfois induire en erreur le diagnostic final. Ces bruits sont reconnaissables par l'œil expérimenté qui les identifie avant d'effectuer un diagnostic. [3]

Ces effets indésirables peuvent avoir plusieurs sources : techniques ou physiques.

I.5.1. Bruits techniques :

Lors de l'enregistrement le matériel utilisé peut être source de bruit s'il n'est pas manipulé avec précaution. Les bruits les plus courants sont présentés ci-dessous. [3]

➤ **Bruit dû au secteur :**

Le réseau électrique peut parfois brouiller le signal ECG avec une onde dont l'harmonique principale est à 50 Hz. Ce type de bruit apparaît sur tout l'enregistrement et peut être assez fort mais il s'élimine facilement avec un filtre sélectif.

➤ **Bruits dû aux mouvements des électrodes :**

Des sauts brusques de la ligne de base apparaissent lorsque les électrodes sont connectées incorrectement. L'effet sur le tracé peut aller de la simple diminution d'amplitude à l'apparition de pics lorsque les électrodes sont en contact intermittent avec la peau. Ce type de bruit intermittent a bande spectrale large s'élimine difficilement car son énergie est dans la même gamme de fréquence que le complexe QRS.

➤ **Autres bruits courants :**

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

1.5.2. Artefacts physiques :

Les artefacts physiques sont dus aux activités électriques du corps humains telles que les commandes de contraction des muscles ou la respiration. [7,9]

➤ **Mouvements de la ligne de base :**

L'activité respiratoire peut faire osciller la ligne de base de l'ECG lors de l'enregistrement jusqu'à un rythme régulier. D'autres perturbations peuvent aussi déplacer la ligne de base temporairement, ainsi que le mauvais contact entre la peau et l'électrode. Ces perturbations peuvent être filtrées vu que leur énergie se situe dans une bande de fréquence basse.

➤ **Bruit myoélectrique ou tremblement somatique :**

La contraction d'un muscle est commandée par une dépolarisation des cellules musculaires et bien que les électrocardiographies soient construites pour être surtout sensibles aux fréquences du myocarde, l'ECG enregistre les contractions des muscles squelettiques. L'aspect le plus courant est une oscillation à haute fréquence liée à la tension musculaire d'un sujet qui n'est pas détendu.

1.6. Base de données MIT-BIH :

En 1975, les laboratoires de l'hôpital de Beth Israël à Boston et le MIT ont réalisés une base de données MIT-BIH pour les arythmies cardiaques, qui a commencé à être distribuée en 1980.

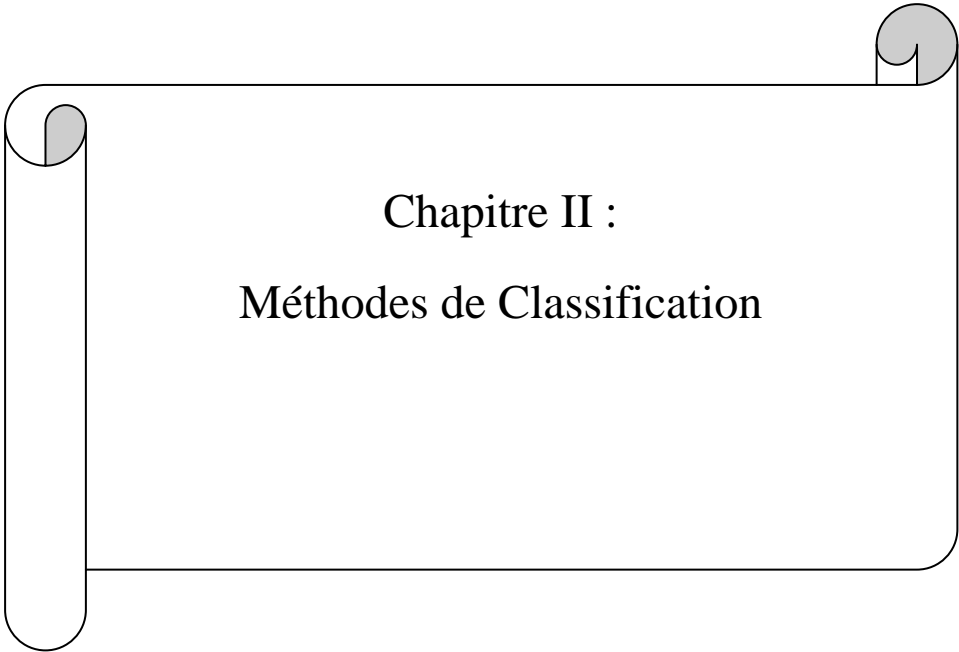
Cette base contient 48 enregistrements extraits d'une demi-heure d'enregistrements ambulatoires à deux voies de l'ECG, obtenus à partir de 47 sujets étudiés par le laboratoire BIH Arrhythmia entre 1975 et 1979. Vingt trois enregistrements ont été choisis au hasard d'un ensemble de 4000 enregistrements ambulatoires de 24 heures d'ECG rassemblées d'une population mélangée de patients hospitalisés (60%) et de patients non hospitalisés (40%) à l'hôpital de Beth Israël à Boston, les 25 enregistrements restants ont été choisis parmi les mêmes enregistrements pour inclure des arythmies moins communes mais cliniquement significatives.

Les enregistrements ont été échantillonnés à une fréquence $f_e = 360 \text{ Hz}$ avec une résolution de 11 bits sur une gamme de 10 mV. Deux cardiologues ou plus annotent indépendamment chaque enregistrement. [10]

Conclusion :

Dans ce chapitre, nous avons introduit les éléments de base de l'électrophysiologie du cœur, de son fonctionnement ainsi que de son activité électrique. Par la suite nous avons parlé de l'électrocardiogramme (ECG), de son principe ainsi que des différentes ondes et intervalles enregistrés. Quelques pathologies cardiaques ont été présentées par la suite. Les différents artefacts qui peuvent affecter le signal ECG enregistré ont été également exposés. Nous avons terminé ce chapitre en parlant de l'intérêt du traitement automatique des signaux ECG et d'une présentation de la base de données MIT-BIH.

Nous allons voir dans le deuxième chapitre les méthodes de traitement des données de l'ECG ainsi que les méthodes de classifications.



Chapitre II :
Méthodes de Classification

II.1. Introduction :

Actuellement le domaine médical exige de nouvelles techniques et technologies, afin d'évaluer l'information d'une manière objective. Cela est dû aux développements récents que connaît le domaine de l'électronique qui a poussé l'informatique à un stade de plus en plus avancé. Ceci a permis d'avoir du matériel plus performant permettant d'exécuter des algorithmes complexes dans le domaine de la classification de données. Ces méthodes de classification permettent de regrouper des objets, des signaux,... dont les éléments ont des propriétés d'appartenance communes, comme le montre la figure suivante :

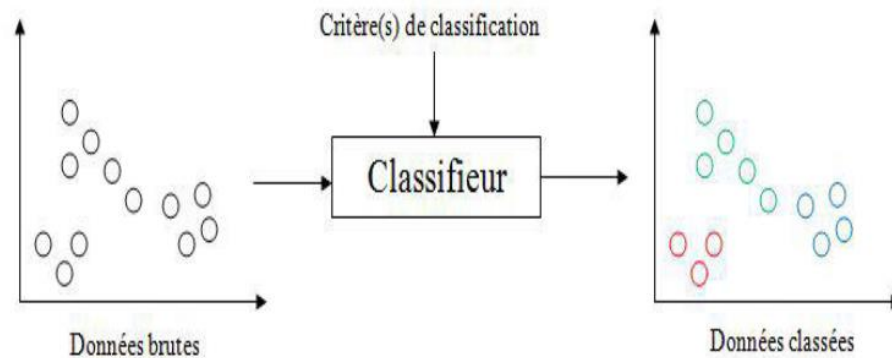


Figure II.1. Principe de la classification.

Dans ce chapitre, nous allons aborder le problème de la classification des battements cardiaques, et nous présenterons la méthode adoptée pour rassembler en une même famille les battements cardiaques dont le tracé ECG est de même nature (sain ou malade) et ainsi les distinguer. Les deux méthodes de classification dont nous parlerons dans ce chapitre sont ; la méthode supervisée et la méthode non supervisée. Mais avant d'entamer l'étape de classification nous devons d'abord effectuer une étape de prétraitement du signal ECG.

II.2. Prétraitement des signaux ECG :

Le prétraitement du signal ECG s'avère nécessaire pour éliminer les diverses perturbations qui dégradent la qualité du signal. Ces perturbations peuvent être des interférences avec le réseau électrique, les oscillations de la ligne de base provoquées par les mouvements ou la respiration du patient, etc. tous ces phénomènes conduisent à la dégradation de la qualité du signal enregistré et

Chapitre II : Méthodes de Classification

rendent difficiles son diagnostic. L'élimination de ces perturbations se fera tout en gardant les caractéristiques des ondes utiles qui composent le signal.

Les premières étapes dans la classification des arythmies cardiaques sont : le dé-bruitage des signaux ECG, la détection des complexes QRS et la mesure de la distance entre deux pics R successifs.

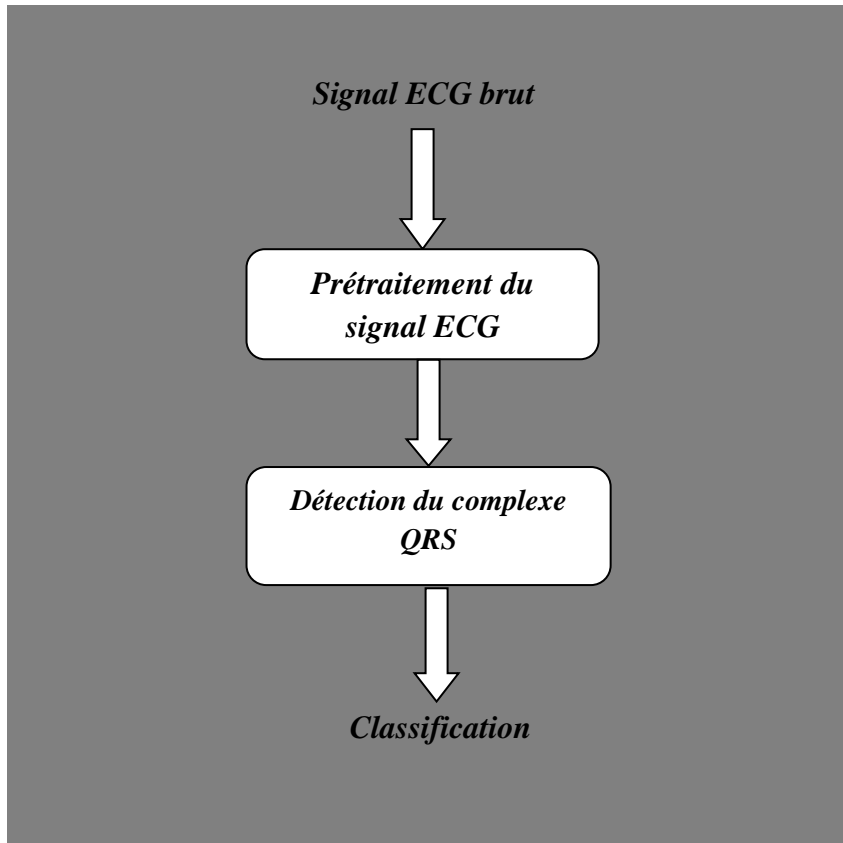


Figure II.2. Etape de la classification du signal ECG.

II.2.1. Dé-bruitage du signal ECG par la méthode des ondelettes:

II.2.1.1. Etat de l'art :

Traditionnellement, les enregistrements ECG sont analysés dans le domaine du temps par des médecins qualifiés. Cependant, les conditions pathologiques du signal ne peuvent pas toujours être évidentes dans le domaine temporel. La transformée de Fourier été l'une des techniques les plus utilisées dans le traitement de signal, qui fournit les informations contenues dans le domaine fréquentiel. L'inconvénient de la TF est qu'elle ne peut pas traiter efficacement les signaux non stationnaires.

Chapitre II : Méthodes de Classification

Les ondelettes sont adaptées au dé-bruitage de signaux contenant des transitions rapides. On utilise un seuillage pour supprimer les coefficients d'ondelettes là où le signal est lisse et les coefficients petits par un filtre passe bas, et on garde les coefficients là où ils sont grands.

Les informations relatives au cœur sont reflétées par son activité électrique en fonction du temps et la fréquence ; de ce fait, les études exigent des méthodes capables de décrire la variation qualitative du signal en temps et en fréquence. La transformée en ondelettes est une méthode puissante qui effectue une analyse des structures locales d'un signal avec un zoom en fonction de l'échelle considérée.

Ces dernières années, des techniques basées sur la transformée en ondelettes sont utilisées dans le dé-bruitage des signaux, car elle fournit à la fois le temps et la fréquence des informations et peut surmonter cette limitation.

La transformée en ondelettes a été appliquée à une vaste gamme d'analyses comme : l'extraction des caractéristiques, fonction de détection de réduction du bruit et des données, la compression. Ici l'analyse par les ondelettes est effectuée dans le but d'enlever les bruits pouvant corrompre le signal ECG. [12]

Le schéma du principe de dé-bruitage basé sur la transformée en ondelettes est illustré dans la figure (II.3).

Généralement, il existe trois étapes successives pour le dé-bruitage :

- Application de la transformée en ondelettes au signal affecté par du bruit.
- Filtrage des coefficients obtenus, en se basant sur un certains nombres de critères.
- Calcul de la transformée inverse, à partir des coefficients issus de l'étape de seuillage précédente.

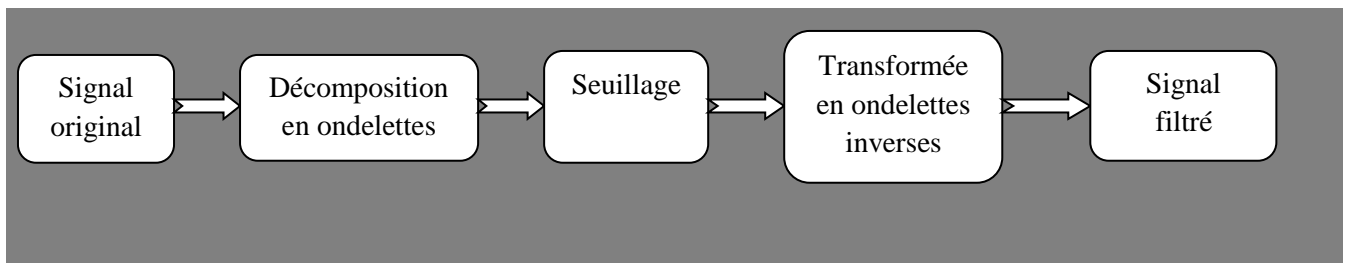


Figure II.3. Principe de dé-bruitage par ondelettes

Le seuillage consiste à comparer les coefficients d'ondelettes à un seuil et de les remettre à zéro si leurs amplitudes sont inférieures à ce seuil. L'efficacité et la fiabilité de l'opération de filtrage sont fortement liées au choix du seuil.

Chapitre II : Méthodes de Classification

Plusieurs techniques de seuillage ont été développées, Donoho a utilisé deux types de fonctions de seuillage, le seuillage dur et le seuillage doux. [12]

- **Seuillage dur :**

Soit d_n^j les coefficients de détail issus du signal $f(t)$ à la résolution j , et T_j la valeur du seuil de sélection correspondant. A l'issue du seuillage, on obtient le vecteur de coefficients seuillés d_{ns}^j tels que :

$$d_{ns}^j = \begin{cases} d_n^j & \text{si } |d_n^j| > T_j \\ 0 & \text{si } |d_n^j| \leq T_j \end{cases} \quad (\text{II.1})$$

- **Seuillage doux :**

Le seuillage doux est similaire au seuillage dur sauf qu'il diminue de T_j la valeur des coefficients d_n^j se trouvant au dessus du seuil :

$$d_{ns}^j = \begin{cases} d_n^j - T & \text{si } d_n^j > T_j \\ d_n^j + T & \text{si } d_n^j < -T_j \\ 0 & \text{si } |d_n^j| \leq T_j \end{cases} \quad (\text{II.2})$$

Il existe deux types de transformées en ondelettes qui sont :

II.2.1.2. Transformée en ondelettes continues :

La valeur de la transformée en ondelette continue (CWT) de la fonction $f(t)$ est donnée par l'équation (II.3), dans laquelle l'ondelette $\Psi_{\tau,a}(t)$ correspond à l'ondelette mère $\Psi(t)$ localisé au temps τ et mise à l'échelle a (équation II.4). [13,14]

$$CWT(\tau, a) = \int_{-\infty}^{\infty} f(t) \cdot \Psi_{\tau,a}(t) dt = \int_{-\infty}^{\infty} f(t) \cdot \frac{1}{\sqrt{a}} \cdot \Psi\left(\frac{t-\tau}{a}\right) dt \quad (\text{II.3})$$

$$\Psi_{\tau,a}(t) = \frac{1}{\sqrt{a}} \cdot \Psi\left(\frac{t-\tau}{a}\right) \quad (\text{II.4})$$

II.2.1.3. Transformée en ondelettes discrètes :

La transformée discrète (DWT) est beaucoup plus répandue, plutôt que de passer le signal par des filtres passe bande de façon parallèle comme pour la CWT, elle utilise des filtres numériques passe bas et passe haut en cascade. [15]

Les équations (II.5) et (II.6) présentent la décomposition à partir du signal discret f_i du niveau i en une partie à haute fréquence d_{i+1} et une partie à basse fréquence f_{i+1} , via les filtres numériques passe bas h et passe haut g .

$$f_{i+1,n} = \sum_l f_{i,l} \cdot h_{(l-2n)} \quad (II.5)$$

$$d_{i+1,n} = \sum_l f_{i,l} \cdot g_{(l-2n)} \quad (II.6)$$

Les filtres numériques ont l'avantage de posséder un support véritablement compact en temps et en fréquence, contrairement aux filtres analogiques qui présentent toujours un certain recouvrement. La transformée discrète peut résulter en une base orthogonale pour obtenir une représentation non redondante du signal, ce que la CWT ne peut fournir.

II.2.2. Détection du complexe QRS :

La détection des complexes QRS a été abordée par plusieurs chercheurs, celle-ci présente plusieurs difficultés vu qu'en pratique, la morphologie des complexes QRS varie dans le temps et elle est différente d'une personne à l'autre.

La détection des complexes constitue une étape très importante dans l'analyse du signal ECG. Généralement leur détection est accomplie par un simple seuillage du signal puisque les ondes R sont plus grandes que les autres ondes en termes d'amplitude. Mais parfois, dans certains cas l'amplitude de l'onde T est comparable à celle de l'onde R, ce qui pourrait induire des erreurs dans le résultat final de détection.

De plus, l'onde R peut avoir parfois une faible amplitude et une morphologie très variable d'un cycle cardiaque à un autre. Ainsi, une bonne détection des complexes QRS est essentielle. Cela nécessite donc un traitement du signal très adéquat dû aux difficultés rencontrées.

La plupart des algorithmes de détection procèdent en deux étapes : une première étape au cours de laquelle le signal passe par un filtre passe bande qui élimine le bruit et les ondes P et T; le signal subit après une transformation non linéaire, par exemple la dérivation pour identifier les fortes pentes

Chapitre II : Méthodes de Classification

autour de l'onde R, et l'élévation au carré pour quantifier l'énergie des QRS. La deuxième étape consiste en une prise de décision selon des critères de seuillage.

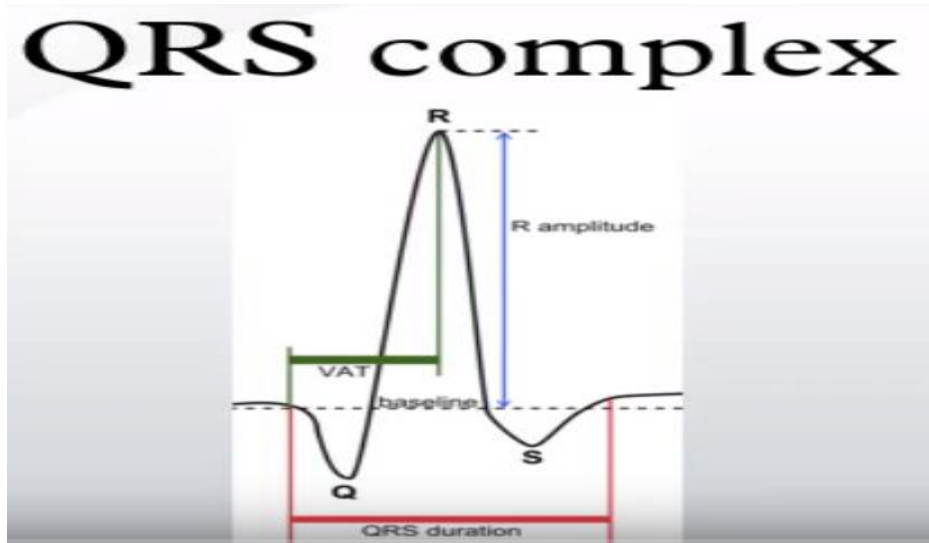


Figure II.4. Complexe QRS.

L'algorithme de détection des complexes QRS que nous utiliserons est celui développé par Pan et Tompkins [17]. Cet algorithme comprend six phases de détection comme le montre la figure suivante :

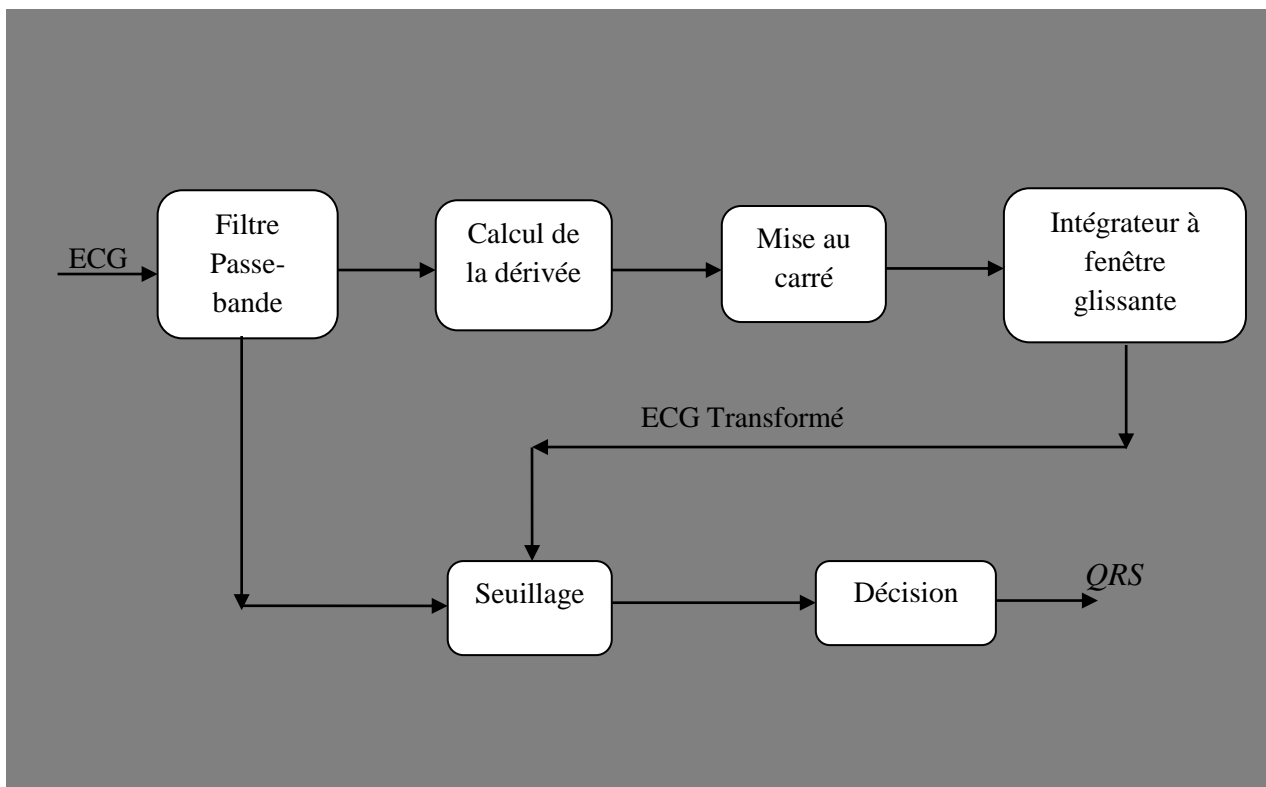


Figure II.5. Diagramme des étapes de l'algorithme de Pan et Tompkins.

Chapitre II : Méthodes de Classification

Le travail de Pan et Tompkins dans leur publication, est focalisé essentiellement sur la détection des complexes QRS. Les auteurs ont alors proposés des filtres qui doivent préserver la densité spectrale de puissance des segments QS. [17]

➤ **Filtrage passe-bande** : d'après les données physiologiques, les ondes R ont une énergie maximale dans la bande 5-15 Hz [16]. Il est donc normal de commencer par un filtrage du signal dans cette bande. Le filtre utilisé est un filtre récursif de bande passante 5-15 Hz. La sortie y_1 du filtre passe bande sera injectée dans un filtre dérivateur.

➤ **Dérivation** : le complexe QRS étant bref (entre 0.02 et 0.2), et de forte amplitude, la dérivée du signal à ce niveau présente des valeurs maximales élevées. On continue donc le traitement du signal par l'application d'un filtre dérivateur numérique y_2 afin de mettre en évidence les fortes pentes qui distinguent les complexes QRS des composantes de l'ECG de basses fréquences telles que les ondes P et T [17]. Un tel filtre est généré par l'équation suivante :

$$y_2(nT) = \left(\frac{1}{8T}\right) [-y(nT - 2T) - 2y(nT - T) + 2y(nT + T) + y(nT + 2T)]. \quad (II.7)$$

La fonction de transfert du filtre est donnée par :

$$H(z) = \left(\frac{1}{8T}\right) (-z^{-2} - 2z^{-1} + 2z^1 + z^2). \quad (II.8)$$

Sa réponse en amplitude est :

$$|H(wT)| = \left(\frac{1}{4T}\right) [\sin(2wT) + 2 \sin(wT)]. \quad (II.9)$$

La figure (II.6) montre que la réponse en fréquence de ce filtre est linéaire entre 0 Hz et 30 Hz. Il approxime donc une dérivée idéale dans cette gamme de fréquence, à laquelle appartiennent les complexes QRS. [17]

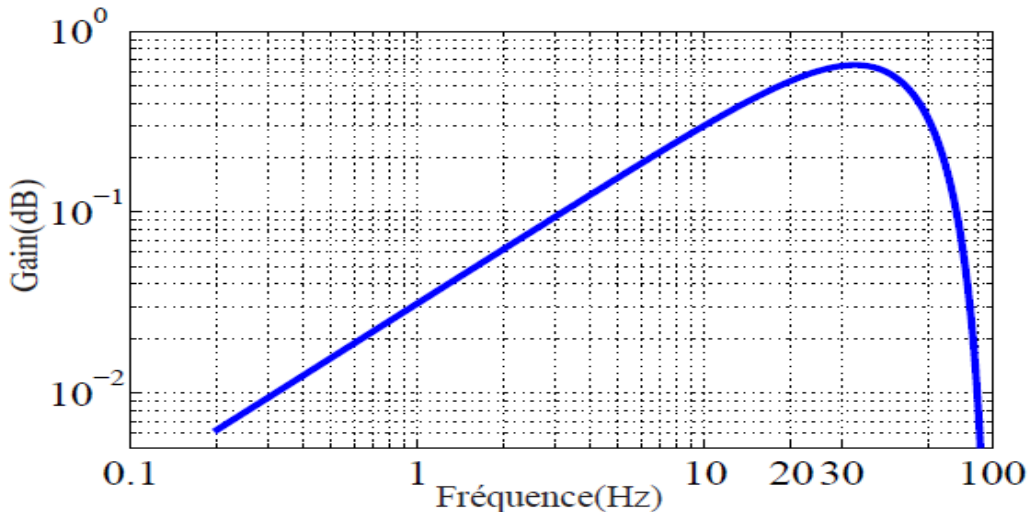


Figure II.6. Réponse en fréquence du filtre dérivateur représenté à l'échelle logarithmique.

➤ **Transformation non linéaire** : afin d'accentuer le contraste mis en évidence par la dérivation, on élève au carré le signal obtenu [17] comme dans l'équation suivante :

$$y_3(nT) = [y_2(nT)]^2 \quad (II.10)$$

Ce qui permet d'éliminer le signe et d'assurer la symétrie de détection entre les ondes R et les ondes QS.

➤ **Intégration** : l'intégration du signal suivi du filtrage passe bas permet d'obtenir un maximum unique pour chaque complexe. 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 ; si, au contraire, elle est trop petite, on obtient plusieurs pics pour une même onde R. La taille choisie est de 0.15 seconde ce qui correspond au double de la largeur moyenne d'un complexe QRS. [17]

L'équation du filtre intégrateur est donnée par :

$$y_4(n) = \frac{1}{N} [y_3(n - (N - 1)) + y_3(n - (N - 2)) + \dots + y_3(n)] \quad (II.11)$$

La longueur N de la fenêtre a été choisie égale à 54 échantillons pour une fréquence d'échantillonnage de 360 Hz. Elle correspond au complexe QRS le plus large possible. La sortie de ce filtre permet en plus de la détection des pics R, la mesure de la durée des ondes QRS.

La figure (II.7) nous montre la sortie du filtre dérivateur et celle du filtre intégrateur.

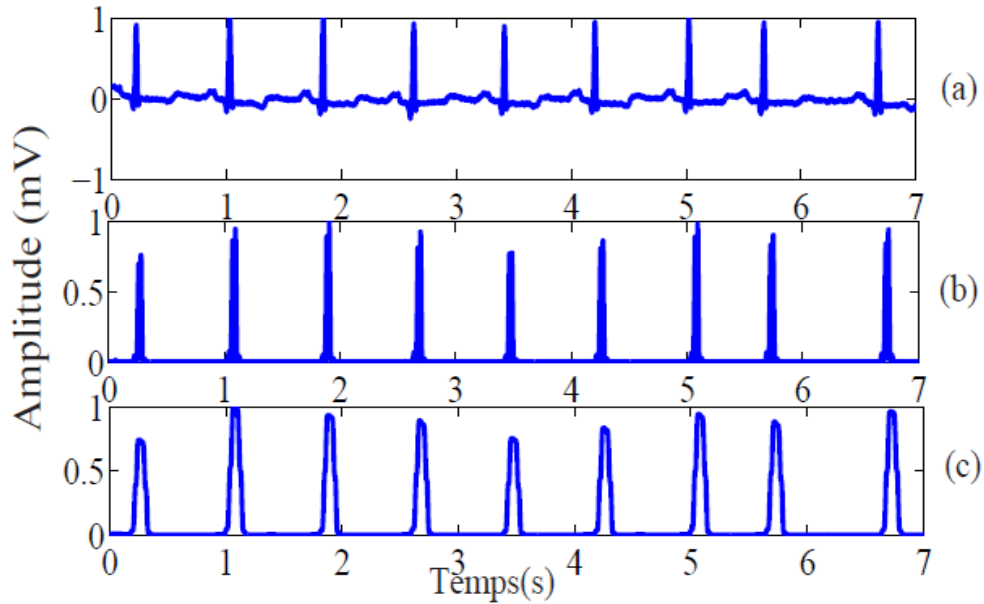


Figure II.7. (a) : ECG filtré, (b) : sortie du filtre dérivateur, (c) : sortie du filtre intégrateur.

➤ **Filtrage Passe Bas :** le signal obtenu passe par filtre passe bas afin d'enlever le bruit haute fréquence résiduel, et obtenir un maximum unique pour chaque complexe. Le filtre est un filtre à Réponse Impulsionnelle Infinie (RII) [17] dont la fonction de transfert du filtre est donnée par :

$$H(z) = \frac{(1-z^{-6})^2}{(1-z^{-1})^2} \quad (II.12)$$

Sa réponse en amplitude est :

$$|H(\omega T)| = \frac{\sin^2(3\omega T)}{\sin^2\left(\frac{\omega T}{2}\right)} \quad (II.13)$$

Où T est la période d'échantillonnage. L'équation du filtre est :

$$y_5(nT) = 2y(nT - T) - y(nT - 2T) + y(nT) - 2y(nT - 6T) + y(nT - 12T) \quad (II.14)$$

La fréquence de coupure et le gain linéaire de ce filtre sont respectivement $f_c = 18 \text{ Hz}$ et $G = 36 \text{ dB}$. La courbe de réponse du filtre est présentée sur la figure suivante :

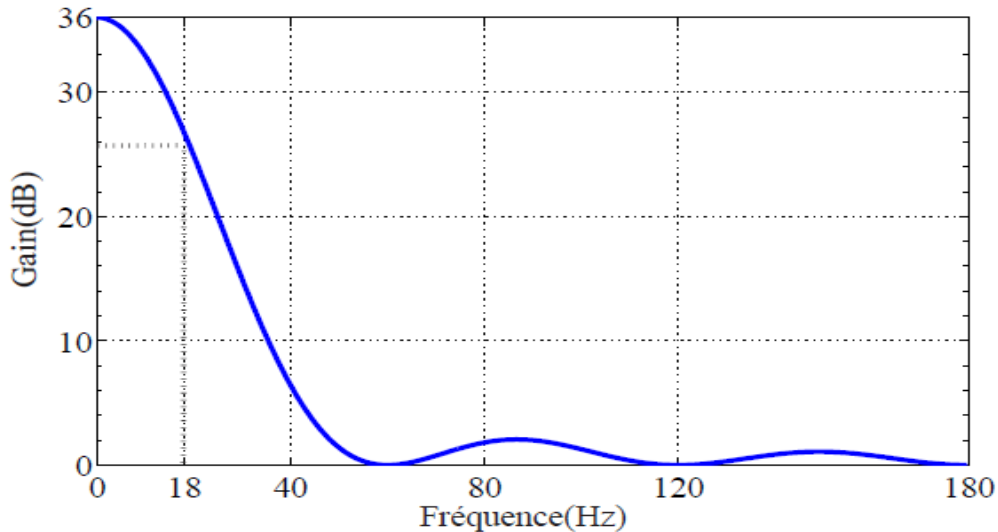


Figure II.8. Réponse en fréquence du filtre proposé par Pan & Tompkins.

➤ **Seuillage adaptatif** : a l'issue du traitement précédent, le signal disponible possède un maximum absolu pour chaque complexe QRS ; il possède en outre d'autres maxima locaux, de plus faible amplitude en générale, qui correspondent soit à du bruit, soit aux ondes T. Cette étape consiste en une recherche de maxima capable de ne pas prendre en considération les maxima trop faibles qui peuvent correspondre au bruit en effectuant un seuillage en amplitude et ceux trop proches qui risquent d'être introduits par les ondes T en effectuant un seuillage temporel. L'amplitude des battements peuvent varier fortement au cours d'un même enregistrement et d'un enregistrement à l'autre, les seuils utilisés ne peuvent pas être fixés a priori : ils sont donc calculés tout au long de l'analyse. A la fin du seuillage on aura y_5 . [17]

➤ **Décision** : le signal filtré et le signal transformé sont ensuite utilisés par un étage de décision pour détecter les candidats. Les règles mises en jeu dans cet étage sont assez complexes. Elles sont basées sur l'estimation du niveau de bruit et du dernier intervalle RR pour mettre à jour des seuils adaptatifs. Cet algorithme est le plus largement utilisé dans la littérature.

II.1.3. Détection de pic R : [26]

Les ondes R sont en général de plus grande amplitude que les autres. Mais ce n'est pas le cas pour tous les patients : parfois, l'onde T est d'amplitude comparable, ce qui pourrait être une sérieuse cause d'erreur.

Les pics R sont les pics positifs des complexes QRS, ceux-ci sont détectés par la comparaison de l'ampleur relative de chacune des régions QRS. Une recherche pour les maximums a été faite sur l'importance relative de chaque fenêtre pour éliminer les erreurs dues à la ligne de base. [23]

- i. Pour chaque complexe QRS détecté et le maximum d'amplitude minimale des valeurs du vecteur de données ECG sont calculées.
- ii. La moyenne des valeurs minimales et maximales est soustraite de tous les points de données de ce complexe pour obtenir les amplitudes relatives.
- iii. La position du maximum des amplitudes relatives est le point de l'emplacement de R du complexe QRS correspondant. La valeur maximale du complexe QRS n'est pas sélectionnée comme étant l'emplacement de R pour éliminer la possibilité de détection de l'onde S.

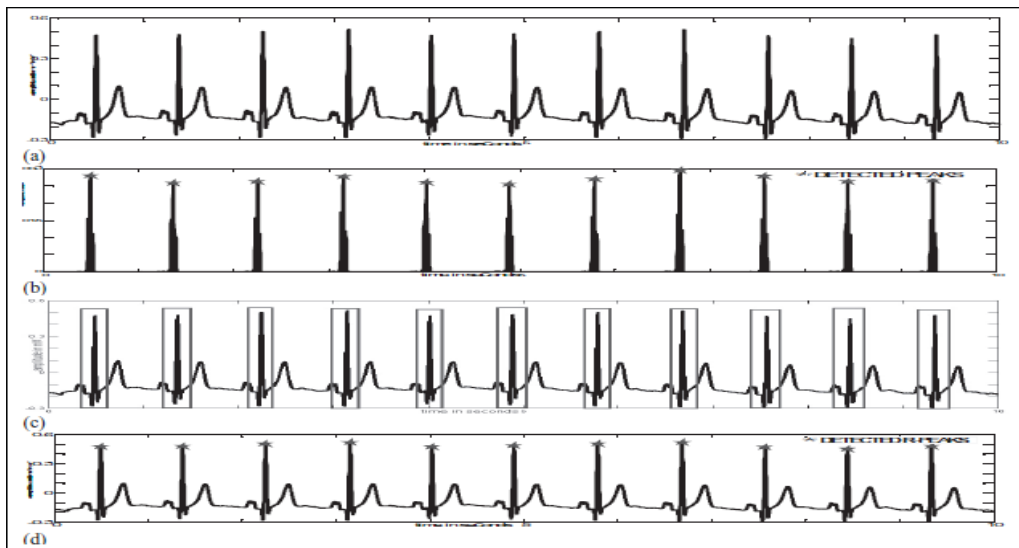


Figure II.9. Détection de l'onde R. (a) signal filtré, (b) sélection de la double différence des pics, (c) détection du complexe QRS, (d) détection des ondes R.

II.1.4. Détection de l'intervalle R-R :

Les ondes R ainsi obtenues précédemment peuvent ne pas être exactes. Il se peut qu'il y est des pics manquants ou de fausses détection. Pour garantir la précision de détection des intervalles RR, ils sont traités en fonction de certains critères. [23]

- i. La différence minimale entre deux pics successifs R est considérée comme pouvant être de 200ms. Les pics détectés à moins de 200ms sont considérés comme étant du bruit et seront éliminés.
- ii. L'intervalle R-R moyen pour 5 pics successifs, est pris en compte par 2 ondes R correspondant au pic le plus élevé et pris comme référence pour le traitement de l'intervalle R-R.
- iii. Tous les intervalles R-R successifs sont traités par comparaison avec l'intervalle R-R moyen calculé. Dans ce cas on a deux possibilités :

Cas 1 : si l'intervalle RR entre deux pics détecté est inférieur a 70 % de l'intervalle RR moyen, le deuxième pic est éliminé.

Cas 2 : si l'intervalle RR entre les deux pics est supérieur a 180% de l'intervalle moyen, une recherche du pic R est déclenché dans cet intervalle avec une diminution de seuil pour le signal de différence.

II.2. La classification :

La classification est la tâche qui consiste à regrouper, un ensemble de données de telle manière que les données d'un même groupe appelé cluster sont plus proches les unes aux autres que celles des autres groupes, au sens du critère de similarité choisi. C'est une technique d'analyse statistique de données très utilisée dans de nombreux domaines, y compris dans le domaine médical ainsi que dans le traitement de signal.

La classification est très importante pour l'exploitation des différentes connaissances acquises et a l'organisation du travail de chacun au sein de l'ensemble. Nous pouvons donc, regrouper les méthodes de classifications en deux grandes familles, en considérant l'intervention ou non d'un « attribut classe » au fur et a mesure du processus de classification, ces deux méthodes sont : « la méthode non supervisée » et « la méthode supervisée ». [24].

II.3. Objectif de la classification :

La classification joue un rôle dans toutes les sciences et techniques qui font appel à la statistique multidimensionnelle. C'est l'une des techniques les plus utilisées en médecine. Tel que les systèmes d'aide au diagnostic constituent aujourd'hui l'un des moyens indispensable dans le domaine médical.

L'objectif de la classification, est la recherche d'une partition où les objets d'une même classe doivent être semblables, et les objets de différentes classes devraient être différents. Ceci est fait avec l'optimisation d'un critère visant à regrouper les signaux dans des classes, chacune le plus homogène possible. [25]

Notre but est de détecter et de classer automatiquement les données de l'ECG, c'est pourquoi le prétraitement du signal ECG est nécessaire avant d'extraire les paramètres du signal avec exactitude. La figure (II.10) nous montre l'organigramme de classification.

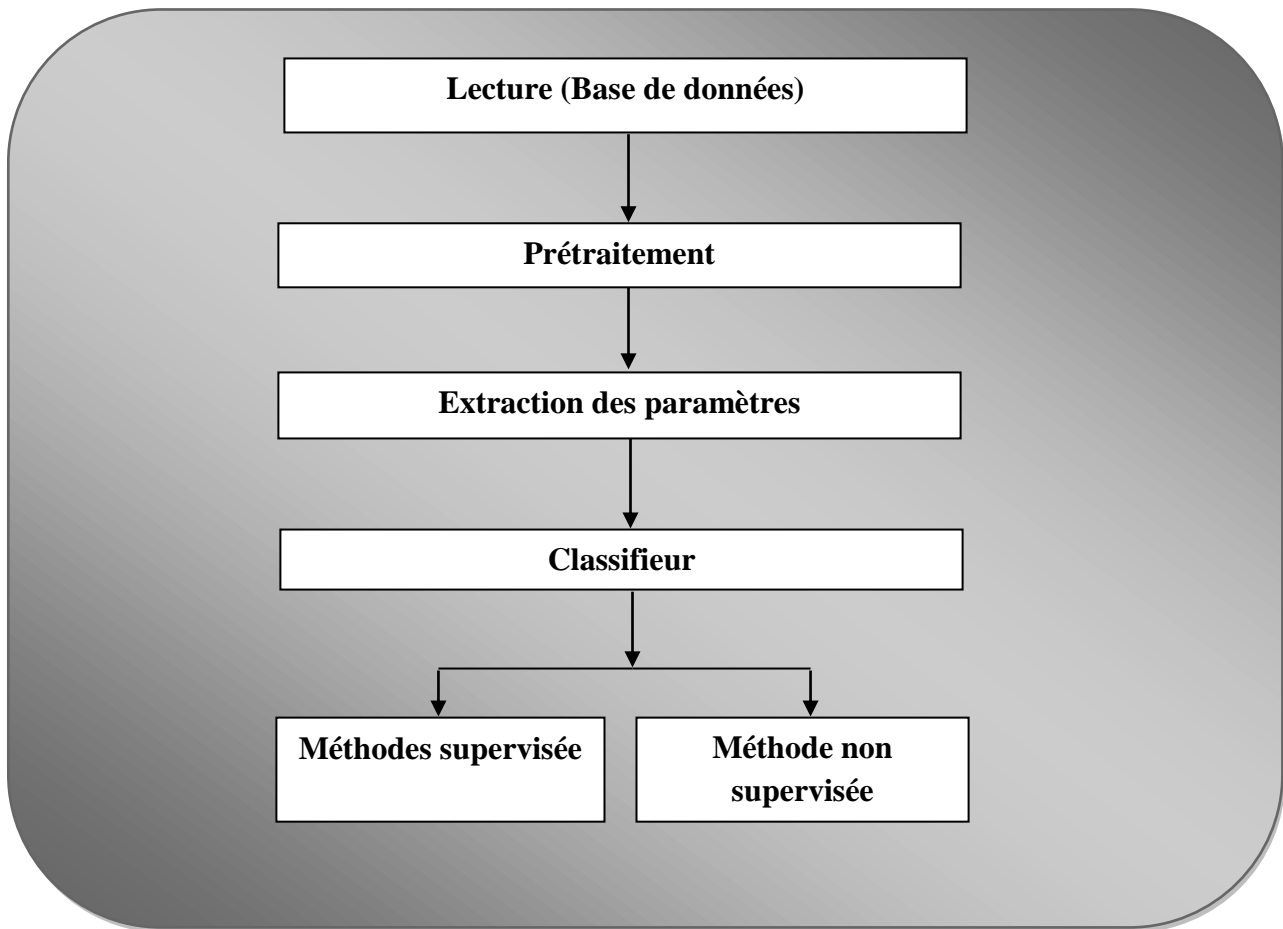


Figure II.10. Organigramme de classification.

II.4. Méthode non supervisée :

L'intérêt des méthodes non supervisées est qu'elles ne nécessitent aucune base d'apprentissage et aucune tâche d'étiquetage manuel préalable. La seule tâche de l'expert se situe à la fin du processus pour identifier les anomalies en calculant les centres des classes et en affectant chaque signal à sa classe.

L'apprentissage non supervisé consiste à inférer des connaissances sur des classes sur la seule base des échantillons d'apprentissage, et sans savoir à quelle classe ils appartiennent.

On distingue aussi les approches de classification non hiérarchiques et les méthodes de classification hiérarchiques.

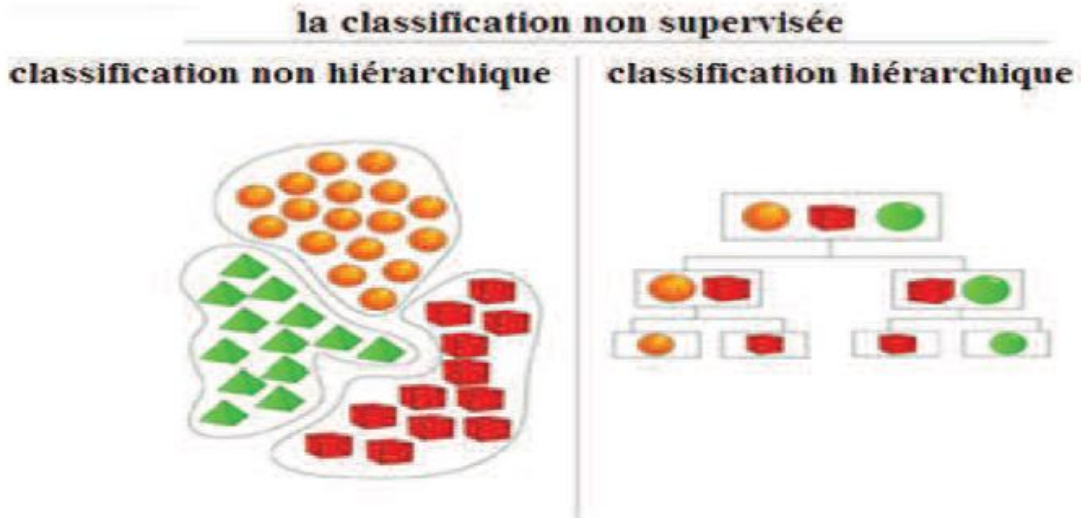


Figure II.11. Les deux types de clustering hiérarchique/non hiérarchique.

a. Méthode non hiérarchique :

Cette méthode produit directement une partition en un nombre fixé de classes.

Regrouper n individus en k classes de telle sorte que les individus d'une même classe soient le plus semblables possible et que les classes soient bien séparées. [18]

b. Méthode hiérarchique :

La classification hiérarchique consiste à effectuer une suite de regroupements en classes de moins en moins fines en agrégeant à chaque étape les objets ou les groupes d'objets les plus proches. Elle fournit ainsi un ensemble de partitions de l'ensemble d'objets [19]. Cette approche utilise la notion de distance, qui permet de refléter l'homogénéité ou l'hétérogénéité des classes. Ainsi, on considère qu'un élément appartient à une classe s'il est plus proche de cette classe que de toutes les autres. La figure suivante est une illustration du principe des méthodes hiérarchiques. [18]

Dans cette figure suivante, on représente la suite de partition d'un ensemble $\{a, b, c, d, e\}$

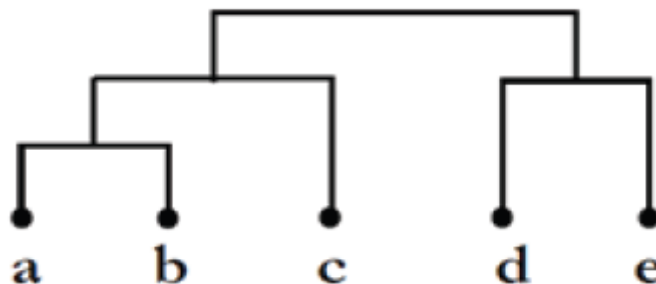


Figure II.12. Partition hiérarchique.

- Dans le cadre de ce travail nous n'utiliserons que deux méthodes de classifications qui sont : la méthode des K-means, et la méthode FCM.

II.4.1. Le K-means :

a. Définition :

L'algorithme K-means est l'un des plus connus et des plus utilisés en termes d'apprentissage non supervisé. Appelé aussi algorithme des centres mobiles, il sélectionne d'abord, au hasard un ensemble de centres pour représenter les classes. Sur la base des caractéristiques choisies, tous les modèles restants sont attribués à leur centre le plus proche. L'algorithme calcule alors les nouveaux centres de gravité en prenant la moyenne de tous les profils appartenant à une même classe. L'opération est répétée jusqu'à ce qu'il n'y ait aucun changement dans les valeurs des centroïdes et des affectations ou la classe pour un groupe particulier. Un algorithme K-means fonctionne bien lorsque les classes sont bien séparées. Cependant, il faut qu'on spécifie à l'avance le nombre de classes pour qu'il sélectionne les centres de gravité. Par exemple, un centroïde peut être piégé dans une valeur optimale locale si un choix inapproprié des centroïdes initiaux a été fait. [20]

Le principe de cette méthode est de minimiser le critère d'erreur (fonction objectif) par rapport aux centres des classes :

$$\mathcal{J}(\Psi, c) = \sum_{i=1}^n \sum_{k=1}^c \|x_i - \Psi_k\|^2 \quad (II.19)$$

Avec :

- Ψ : centre des classes.
- c : nombre de classes.
- x_i : nombre de données.
- $\| \cdot \|$: norme Euclidienne entre chaque donnée x_i et le centre des classes Ψ .

La mise à jour des centres des classes Ψ se fait selon l'équation suivante :

$$\Psi_k = \frac{1}{n} \sum_{i=1}^n x_i \quad (II.20)$$

b. Algorithme K-means :

L'algorithme général qui résume la méthode est le suivant :

Etape 1 : choisir aléatoirement les C centres initiaux $\{\psi_1, \psi_2, \psi_3, \dots, \psi_c\}$ à partir des points $\{x_1, x_2, \dots, x_n\}$

Etape 2 : assigner x_i $i=1 \dots N$ à la classe k ; $k=1 \dots c$ si:

$$\|x_i - \psi_k\| \leq \|x_i - \psi_p\| ; p=1 \dots c; \text{ avec: } i \neq p$$

Etape 3 : mettre à jour les centres des classes par la formule (II.20).

Etape 4 : répéter de l'étape 2 et 3 jusqu'à atteindre la convergence : $\Psi_k^* = \Psi_k$; $k = 1, \dots, c$.

Une fois l'algorithme a convergé, on a donc une estimation des classes (C_1, \dots, C_n) et des centres (Ψ_1, \dots, Ψ_k). On connaît donc l'appartenance de chaque donnée x_i .

c. Avantages et contraintes de l'algorithme K-means :

L'avantage de cet algorithme est qu'il a une :

- Bonne résistance aux données erronées.
- Flexibilité avec tout type de distance.

Par contre il présente des contraintes qui sont :

- Nécessité de spécifier le nombre de clusters k ,
- Complexité à chaque itération,
- Convergence vers un minimal local,
- Algorithme fortement sensible au bruit.

II.4.2. FCM:

a. Définition :

La méthode des C-Moyens Flous (FCM ; Fuzzy C-Means) est une technique de regroupement de données dans laquelle un ensemble de données sont regroupées en n groupes avec chaque point de données dans l'ensemble appartenant à chaque groupe a un certain degré. Par exemple, un certain point de donnée qui se trouve à proximité du centre d'une classe aura un haut degré d'appartenance à cette classe et un autre point de données qui se trouve loin du centre d'une classe aura un faible degré d'appartenance à cette classe. James Bazdek était l'un des premiers à appliquer un raisonnement flou. [21]

L'algorithme FCM commence par une estimation initiale des centres des classes, qui sont destinées à marquer l'emplacement moyen de chaque classe. L'estimation initiale pour ces centres est très probablement incorrecte. Ensuite, l'algorithme attribue toutes les données indiquant un degré d'appartenance pour chaque classe. En mettant à jour itérativement les centres des classes et les qualités d'adhésion pour chaque point de données, il déplace itérativement les centres au bon endroit au sein d'un ensemble de données. Cette itération est basée sur la minimisation d'une fonction objectif qui représente la distance de tout point de données d'un centre de classe pondérée par le degré d'appartenance de ce point. [21] Bazdek a utilisé le concept de la logique floue, où les décisions sont prises par des pondérations analogiques, et l'a appliqué à cette fonction objectif J . Ce faisant J a été défini comme :

$$J_m(\mathbf{u}(\cdot), \mathbf{v}) = \sum_{k=1}^N \sum_{i=1}^c (\mathbf{u}_{ik})^m d_{ik} \quad (II.21)$$

d_{ik} : Distance.

c : Nombre de classes.

m : Indice du flou.

N : Nombre d'observations.

Les centres et les degrés d'appartenance sont, pour une valeur de m donnée, calculés à l'aide des deux formules suivantes :

$$\mathbf{u}_{ik} = \frac{(\mathbf{d}_{ik})^{-2}}{\sum_{j=1}^c (\mathbf{d}_{jk})^{-2}} \quad ; \quad 1 \leq k \leq m ; 1 \leq i \leq c \quad (II.22)$$

$$\mathbf{v}_i = \frac{\sum_{k=1}^N (\mathbf{u}_{ik})^m \mathbf{x}_k}{\sum_{k=1}^N (\mathbf{u}_{ik})^m} \quad ; \quad 1 \leq i \leq c \quad (II.23)$$

Avec :

u_{ik} : Degré d'appartenance.

v_i : Centres des classes.

Il y a deux facteurs dans l'utilisation de la procédure FCM qui nécessitent encore la discussion. L'un de ces points est la sélection optimale du paramètre m , car il n'y a aucun moyen automatisé de le sélectionner. Le second facteur est la manière dont la distance d_{ik} est calculée. D'une manière générale, cette distance peut être calculée grâce à la forme quadratique : [22]

$$d_{ik} = \|x_k - v_i\|^2 \quad (\text{II.24})$$

Le processus itératif d'arrête lorsque la partition devient stable, c'est-à-dire lorsqu'elle n'évolue plus significativement, entre deux itérations successives. Ceci s'exprime de manière générale par la vérification de l'expression (II.25) ou le terme de gauche traduit une norme matricielle et le coefficient ε définit le seuil de convergence :

$$|\bar{J}_m(u(\cdot), v) - J_m(u(\cdot), v)| < \varepsilon \quad (\text{II.25})$$

b. Algorithme FCM :

Etape 1 : Choisir le nombre de classes C .

Etape 2 : Initialiser la matrice de degré d'appartenance U par des valeurs aléatoires dans l'intervalle $[0, 1]$, ainsi que les centres C_k (initialisation aléatoire).

Etape 3 : Calculer le prototype v par la relation (II.23)

Calculer $J_m(u(\cdot), v)$ par la relation (II.21)

Etape 4 : Mettre à jour la matrice de degré d'appartenance par la relation (II.22).

Calculer $\bar{J}_m(u(\cdot), v)$ par la relation (II.21)

Etape 5 : Répéter les étapes 3 à 4 jusqu'à satisfaction du critère d'arrêt qui est :

$$|\bar{J}_m(u(\cdot), v) - J_m(u(\cdot), v)| < \varepsilon$$

- Les algorithmes FCM souffrent du manque de prise en compte d'informations et fournissent des classifications moins efficaces en présence de bruit.

II.5. Méthodes supervisées :

C'est l'ensemble des techniques qui visent à deviner l'appartenance d'un individu à une classe en s'aidant uniquement des valeurs qu'il prend. Cette méthode construit un modèle représentatif d'un certain nombre de données organisées en classes qu'on appelle généralement le corpus d'apprentissage, puis utilise ce modèle pour classer de nouvelles données, c'est-à-dire de prédire leur classe en vu de leur caractéristiques. La construction de ce modèle relève de l'apprentissage automatique supervisé.

L'objectif de la classification supervisée est principalement de définir des règles permettant de classer des objets dans des classes à partir de variables qualitatives ou quantitatives caractérisant ces objets.

La plupart des algorithmes d'apprentissage supervisées tentent de trouver un modèle (une fonction mathématique) qui explique le lien entre les données d'entrée et les classes de sortie.

Il existe de nombreuses méthodes qui traitent de la classification supervisée :

- Arbre de décision.
- Classification naïve bayésienne.
- Réseau de neurones...etc.

Parmi ces méthodes nous utiliserons la méthode de classification par les réseaux de neurones.

II.5.1. Réseaux de neurones :

II.5.1.1. Définition :

Les réseaux de neurones sont des outils très utilisés pour la classification, l'estimation, la prédiction et la segmentation. Ils sont issus de modèles biologiques, et sont constitués d'unités élémentaires (neurones) organisés selon une architecture. [26]

Un réseau de neurones est un modèle de calculs dont sa conception est très schématiquement inspirée du fonctionnement des neurones biologiques. Les réseaux de neurones sont généralement optimisés par des méthodes d'apprentissage de type probabiliste, en particulier bayésien.

II.5.1.2. Eléments de base du réseau neuronal :

a. Modèle neurophysiologique :

Le neurone est une cellule du système nerveux spécialisée dans la communication et le traitement d'informations. Ils sont composés d'un noyau entouré de structures qui partent en étoile, appelées dendrites, et d'un long prolongement mesurant une dizaine de centimètres, appelé axone.

Les neurones moteurs issus du système nerveux, transmettent les ordres via l'influx nerveux qui se propage par l'intermédiaire de ces neurones. La communication entre les neurones s'appelle une synapse et se fait par la libération d'une substance appelée neurotransmetteur qui va être captée par les récepteurs du neurone sous-jacent. [27]

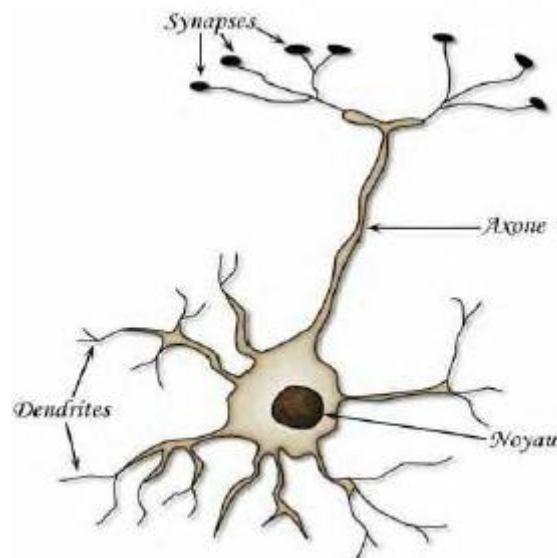


Figure II.13. Schéma d'un neurone biologique.

b. Modèle mathématique :

Par analogie avec le neurone biologique, chaque neurone artificiel est un processeur élémentaire. Il reçoit un nombre variable d'entrée en provenance de neurones en amont. A chaque entrée est associé un poids w , (pour weight en anglais), représentant la force de la connexion. Chaque processeur élémentaire est doté d'une sortie unique qui se ramifie pour alimenter un nombre variable de neurones en aval comme le montre la figure suivante : [27]

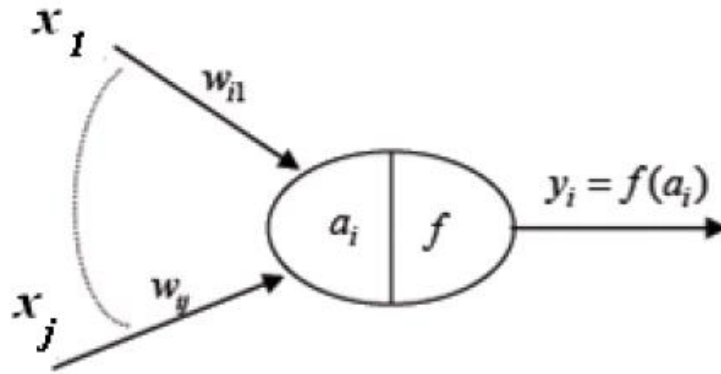


Figure II.14. Neurone formel.

Le comportement du neurone artificiel se compose de deux phases [27]:

La première phase représente le prétraitement des données reçues en calculant la somme pondérée des entrées selon l'expression suivante :

$$a_i = \sum_j w_{ij} x_j \quad (II.26)$$

Dans la deuxième phase, une fonction de transfert calcule la valeur de l'état interne du neurone. Cette valeur sera transmise aux neurones en aval. Plusieurs possibilités existent pour spécifier la fonction d'activation du modèle. Les fonctions mathématiques les plus souvent utilisées sont les fonctions linéaires, sigmoïdes et gaussiennes, qui sont schématisées sur la figure (II.15). Le neurone peut être défini mathématiquement comme étant une fonction algébrique, a priori non linéaire et bornée, avec des entrées x et paramétrée par les poids w_{ij} .

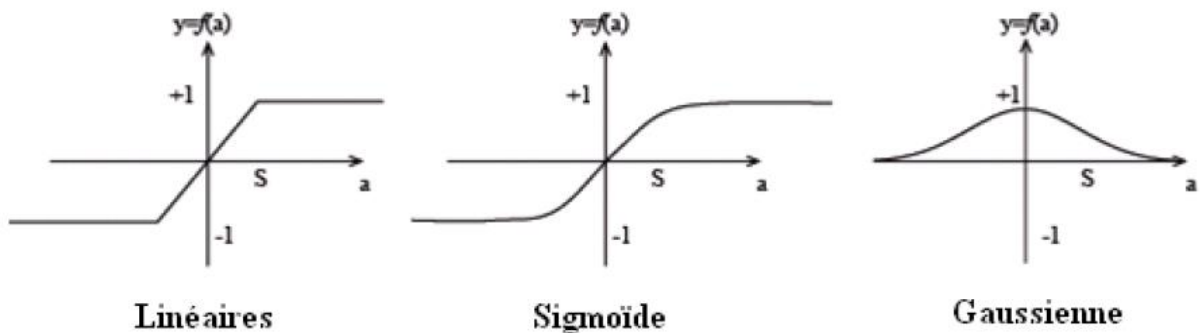


Figure II.15. Différents types de fonctions d'activation pour le neurone artificiel.

II.5.1.3. Architecture des réseaux de neurones :

La structure du réseau de neurones, est le nombre de couches et de nœuds, la façon dont sont interconnectés les différents nœuds et le mécanisme d'ajustement des poids. Le choix de cette structure détermine les résultats qui seront obtenus et constitue le point délicat dans la mise en œuvre d'un réseau de neurones.

On distingue deux familles de réseaux de neurones ; les réseaux bouclés et les réseaux non bouclés. [28]

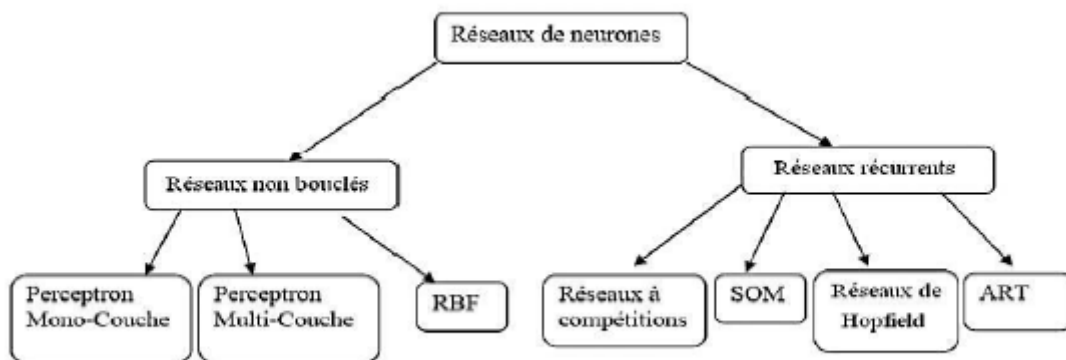


Figure II.16. Réseaux de neurones les plus utilisés.

II.5.1.4. Apprentissage du réseau :

L'apprentissage du réseau est indéniablement la propriété la plus intéressante des réseaux de neurones. Toutes les informations que possède un réseau de neurone se trouvent dans les poids des connexions entre les différents neurones, celui-ci acquiert cette information durant la phase d'apprentissage. Durant cette phase les poids des neurones sont calculés de telle manière que les sorties du réseau soient aussi proches que possible des sorties désirées. La procédure d'apprentissage est divisée en deux grandes catégories : l'apprentissage supervisé et l'apprentissage non supervisé. [29]

a. Apprentissage supervisé :

Cette méthode implique l'existence d'une personne qui a pour rôle d'évaluer le succès (ou l'échec) du réseau. On présente des entrées au réseau de neurones puis on compare les sorties obtenues aux sorties désirées.

b. Apprentissage non supervisé :

Dans cette méthode, les données ne contiennent pas d'informations sur une sortie désirée, et il n'y a pas de superviseur pour évaluer le réseau. L'apprentissage non supervisé implique la fourniture à un réseau autonome d'une quantité d'exemples suffisante, tel que celui-ci en dégage les régularités automatiquement.

II.5.1.5. Avantages et inconvénients des réseaux de neurones :

Comme chaque méthode les réseaux de neurones ont aussi des avantages et des inconvénients qu'on va énumérer dans ce qui va suivre :

a. Avantages :

- Résistance au bruit ou au manque de fiabilité des données.
- Implémentation des systèmes possible par l'utilisation des paramètres descriptifs au lieu des règles prédéfinies sur les maladies.
- Capacité d'analyse en temps réel et d'une façon rapide et fiable.
- Elimine les erreurs dues à la fatigue humaine.
- Capacité de traiter de grandes quantités de données.

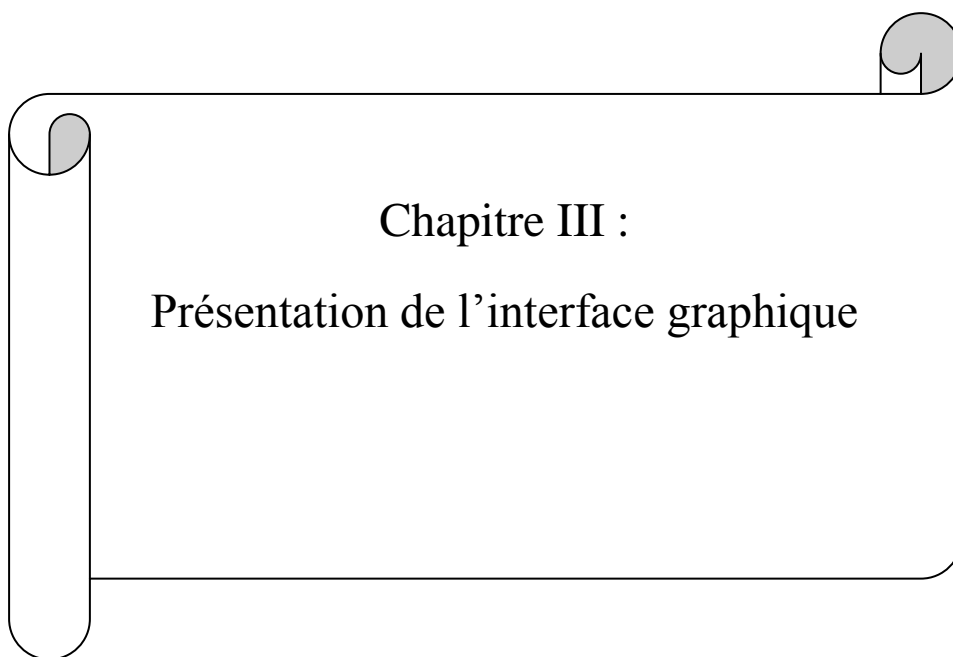
b. Inconvénients :

- Le choix des valeurs initiales des poids du réseau et le réglage du pas d'apprentissage, qui jouent un rôle dans la vitesse de convergence.
- Absence de méthode permettant de définir la meilleure topologie du réseau et le nombre de neurone à placer dans la couche cachée.

Conclusion :

On a vu dans ce chapitre les méthodes de prétraitement des signaux ECG ainsi que les différentes méthodes de classification qui sont les méthodes non supervisées dont le K-means et le FCM, ainsi que les méthodes non supervisées dont les réseaux de neurones.

Dans le prochain chapitre on expliquera les différentes étapes élaborées pour l'illustration de notre interface graphique, et présenter les différentes approches de classification cité précédemment dans ce chapitre.



Chapitre III :
Présentation de l'interface graphique

Chapitre III : Présentation de l'interface graphique

III.1. Introduction :

Dans ce dernier chapitre, après avoir testé les méthodes utilisées précédemment, nous allons illustrer les différents résultats sur une interface graphique qui facilitera la manipulation des données ECG.

Cette interface permettra à l'utilisateur de décrire la procédure utilisée et de choisir le signal à traiter.

L'interface a été réalisée et développée sous MATLAB (version 7.9.0) avec l'utilitaire GUI (Guide User Interface).

III.2. Fenêtre de présentation générale :

Une fois l'application lancée, la fenêtre de présentation générale s'affiche. Cette fenêtre est utilisée pour accéder à la fenêtre de commande principale illustrée sur la figure (III.1).

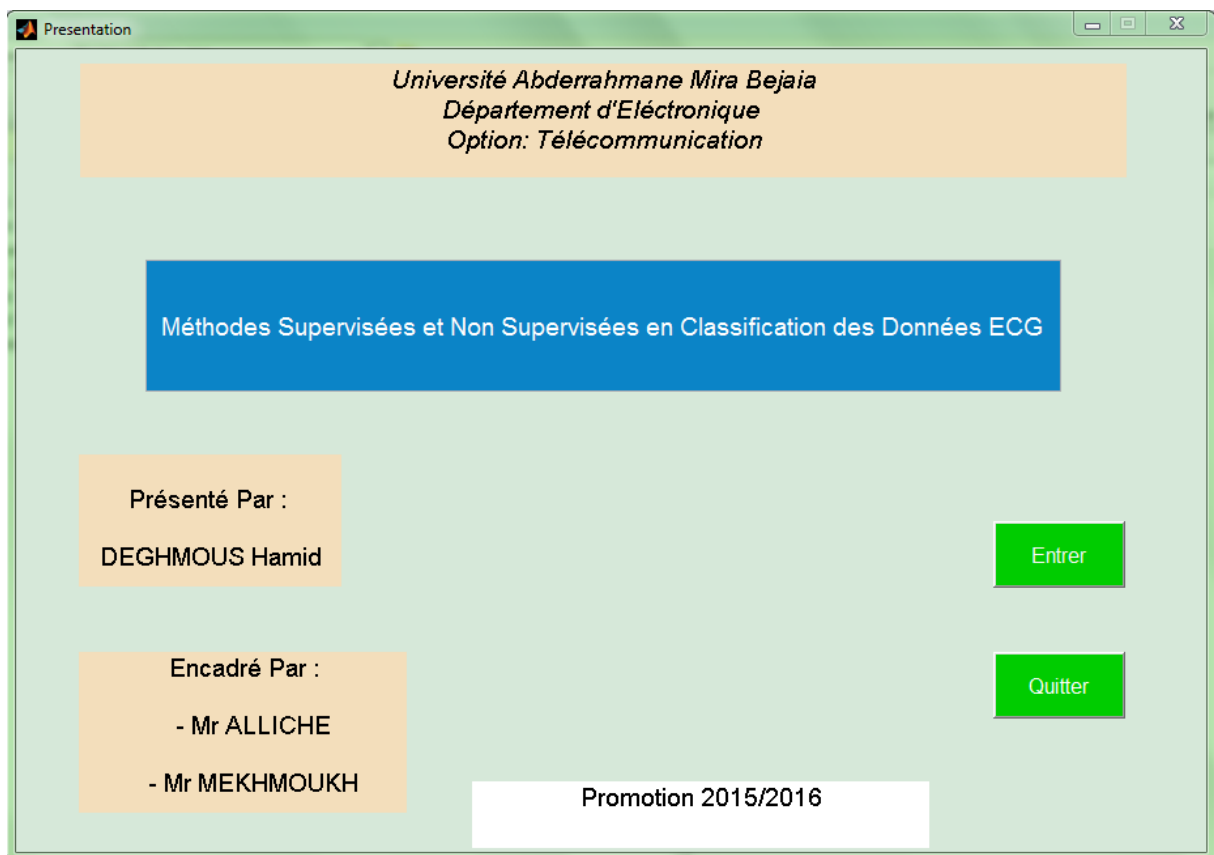


Figure III.1. Fenêtre de présentation.

Chapitre III : Présentation de l'interface graphique

- Le bouton « **Entrer** » permet d'accéder à la fenêtre de commande dans laquelle on peut choisir la commande à tester.
- Le bouton « **Quitter** » permet de quitter complètement l'interface.

III.3. Fenêtre de commande principale :

Après avoir cliqué sur le bouton « Entrer », l'utilisateur aura l'opportunité d'accéder à la fenêtre de commande illustrée sur la figure (III.2).

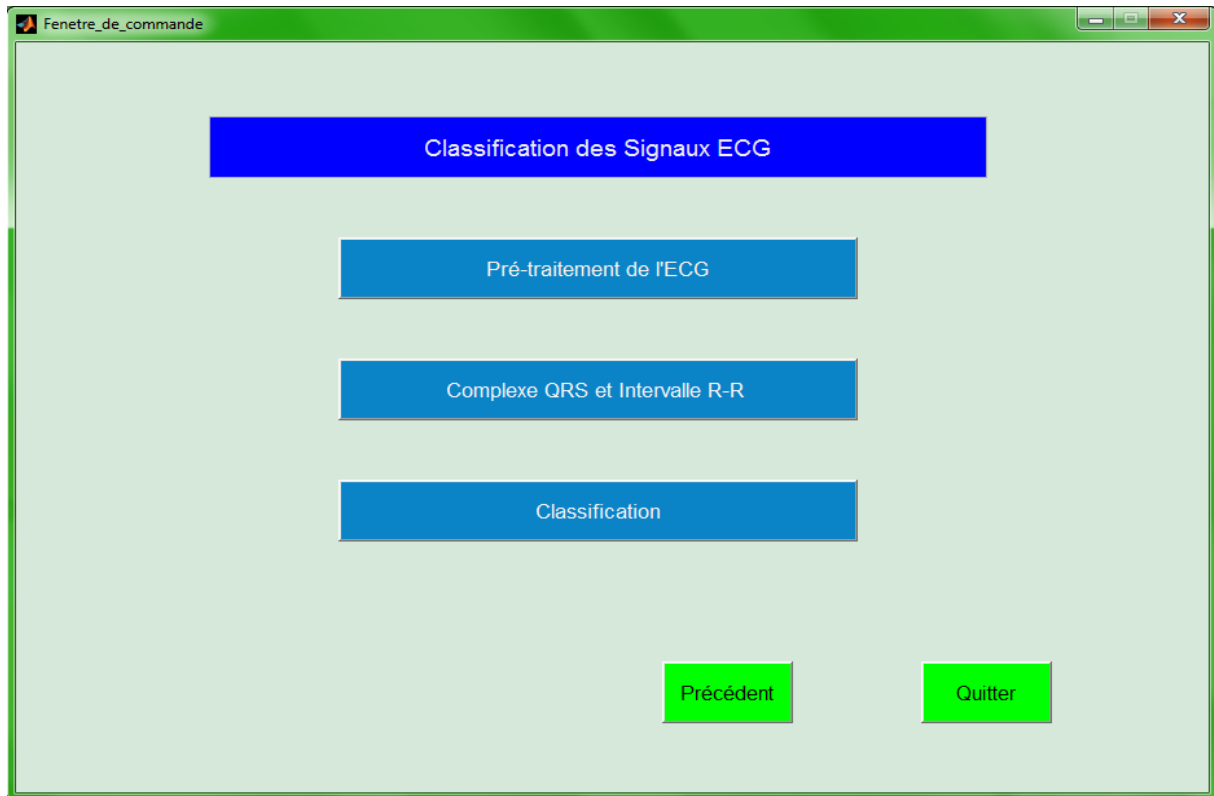


Figure III.2. Fenêtre de commande principale.

Les différents boutons présentés donnent accès à d'autres fenêtres, dont les fonctions sont :

- **Prétraitement de l'ECG** : Cette étape réalise la fonction de filtrage du signal ECG qui élimine les différents bruits présent sur le signal.
- **Complexe QRS et Intervalle R-R** : Cette étape détecte les ondes R, le complexe QRS ainsi que l'intervalle R-R présent sur l'ECG, de plus elle nous donne le nombre de battements par minute.
- **Classification** : Cette étape réalise la classification selon les deux méthodes utilisées à savoir la méthode supervisée et non supervisée.

Chapitre III : Présentation de l'interface graphique

III.4. Prétraitement de l'ECG :

Un simple clique sur ce bouton nous permet d'accéder à la fenêtre graphique suivante :



Figure III.3. Fenêtre de Prétraitement.

Dans tous les exemples qui suivront, nous avons choisi trois (03) signaux de la base **MIT-BIH**, qui sont (112, 118 & 230).

Celle-ci comporte :

- Une zone de lecture du fichier ECG :

En cliquant sur le bouton « **charger le fichier** », nous permettra de choisir le signal ECG à charger, puis une zone de texte éditable pour choisir la « fréquence d'échantillonnage », et enfin en appuyant sur le bouton « Tracer », on aura le tracé temporel du signal ECG correspondant.

- Une zone de bruitage et de dé-bruitage :

Dans cette zone on a aussi une zone éditable qui permet de choisir le taux de bruit ajouté au signal. En cliquant sur le premier bouton « appliquer » on aura le tracé temporel du signal ECG avec un bruit gaussien.

Chapitre III : Présentation de l'interface graphique

Le deuxième bouton « **Dé-bruitage** », nous donnera en appuyant dessus le signal ECG sans bruit.

Les figures suivantes nous illustrent les différentes méthodes citées précédemment.

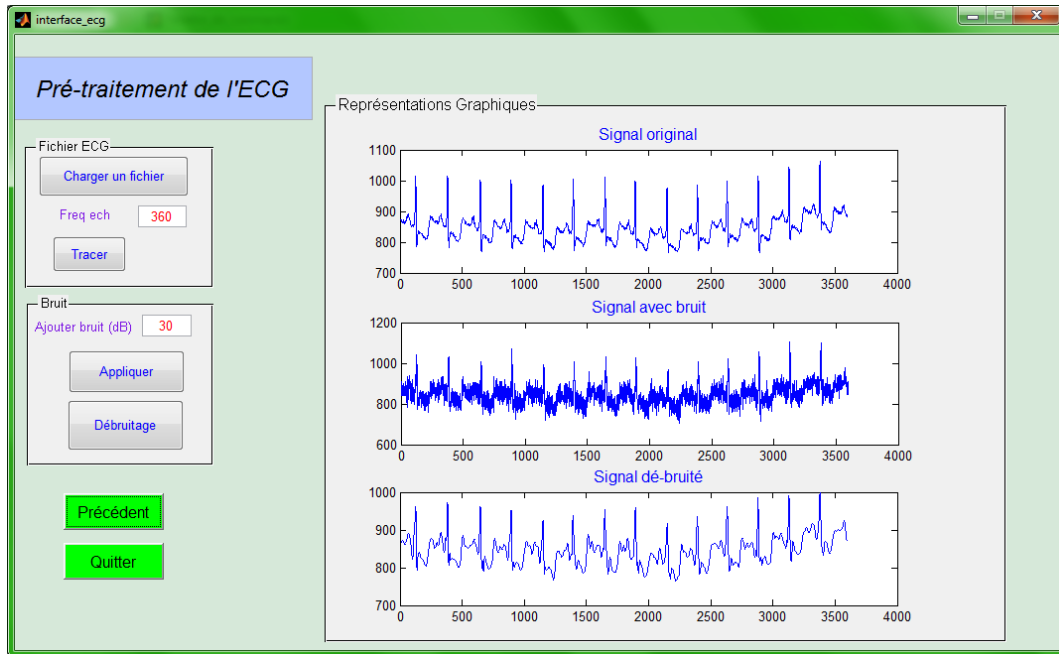


Figure III.4. Prétraitement de l'ECG_112

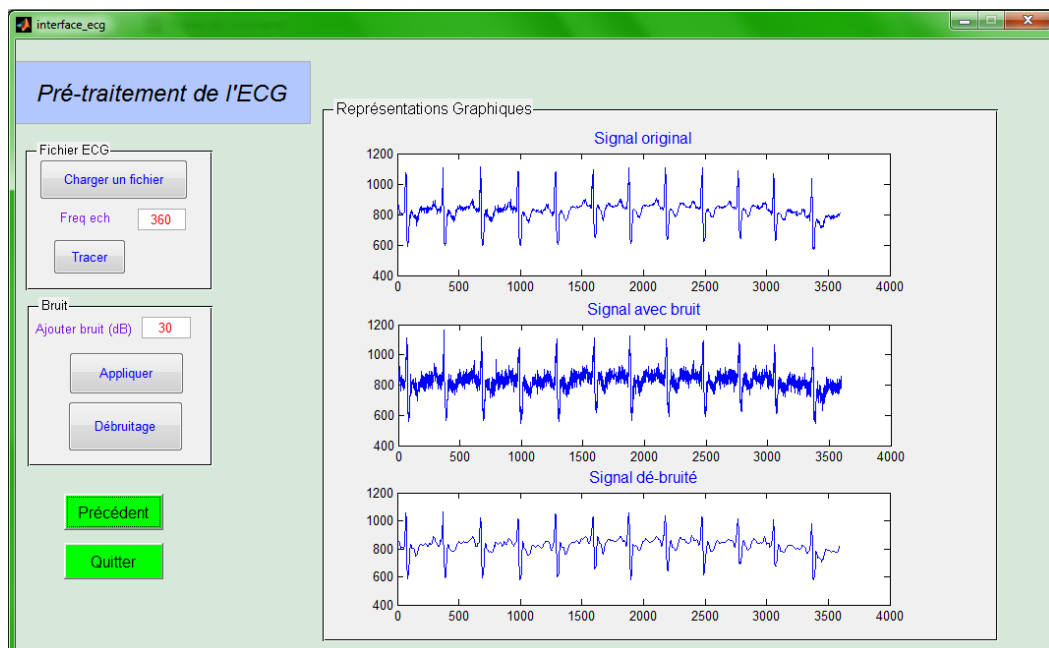


Figure III.5. Prétraitement de l'ECG_118

Chapitre III : Présentation de l'interface graphique

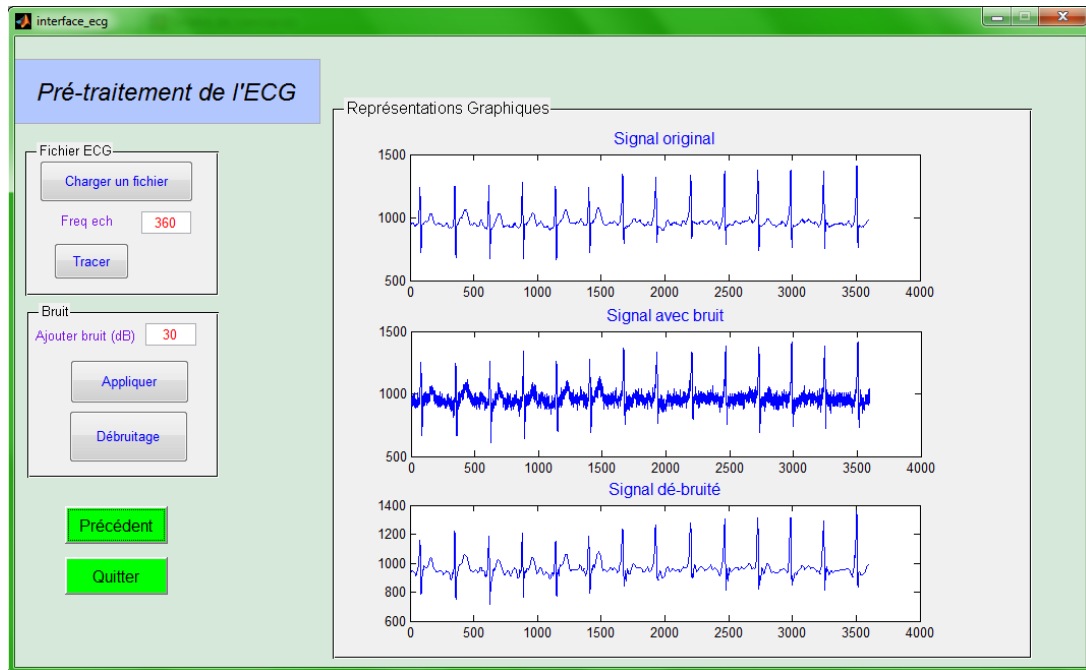


Figure III.6. ECG_230.

III.5. Détection du complexe QRS et de l'intervalle R-R:

En appuyant sur le bouton « **Complexe QRS et intervalle R-R** », on aura la fenêtre suivante qui s'affiche :

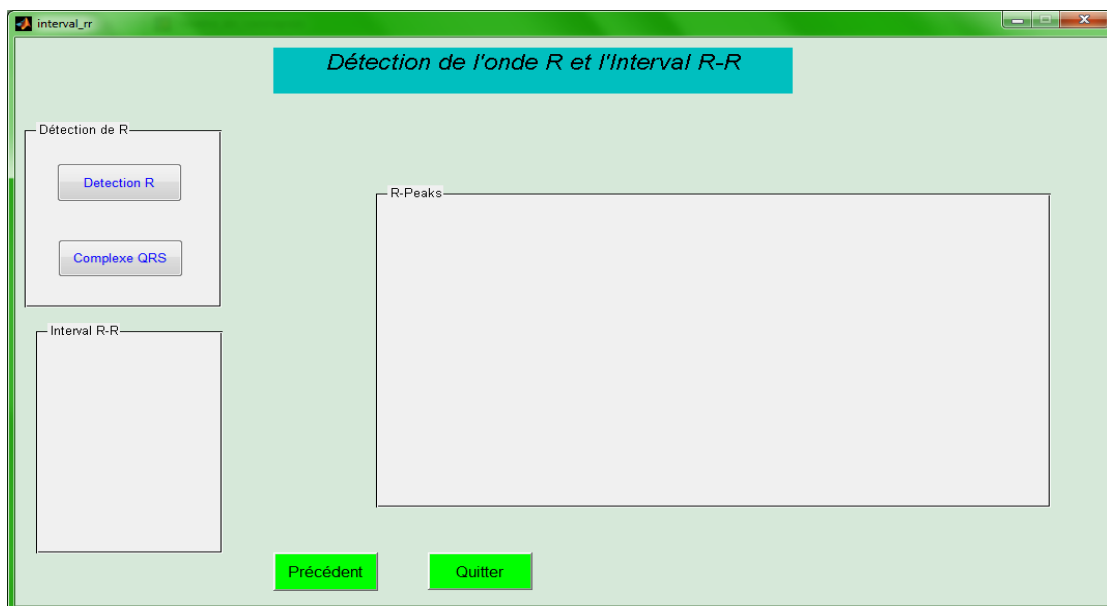


Figure III.7. Fenêtre de Détection du complexe QRS et de l'intervalle R-R

Chapitre III : Présentation de l'interface graphique

Cette fenêtre comporte :

- Un bouton « **Détection R** »:

Ce bouton permet l'affichage des ondes R ainsi que de l'intervalle entre chaque onde détectée.

Comme illustré ci-dessous :

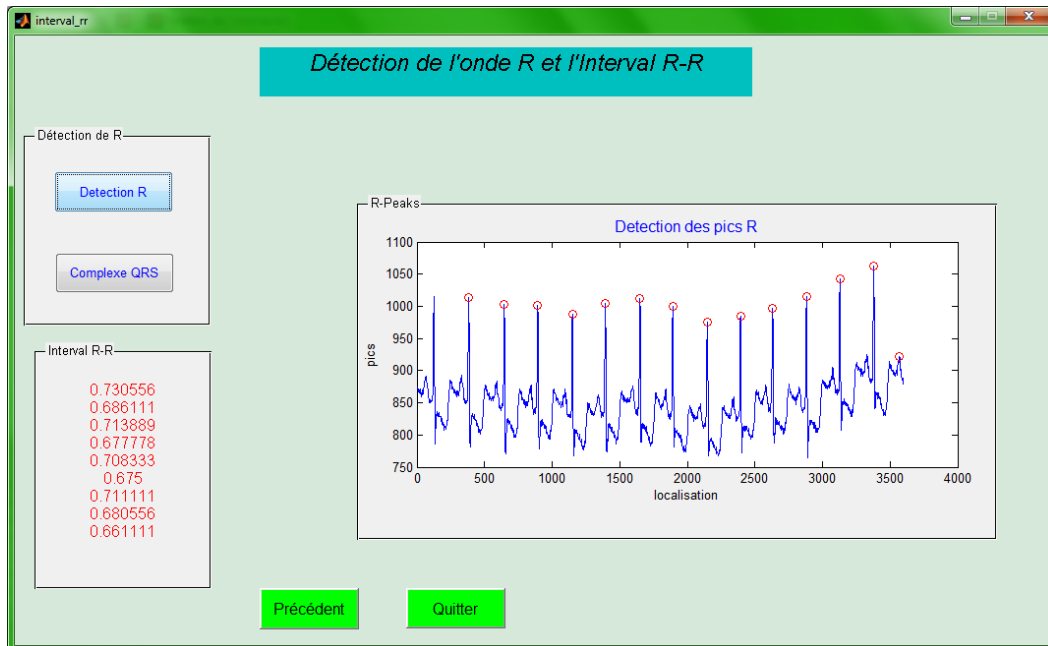


Figure III.8. Détection de R et de l'intervalle R-R du signal ECG_112.

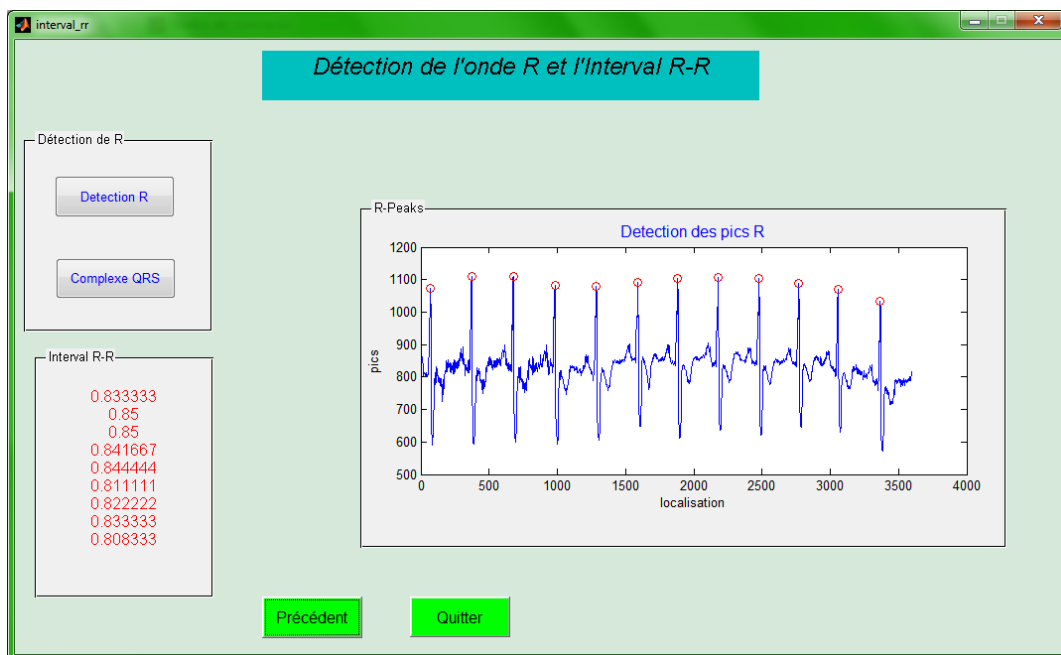


Figure III.9. Détection de R et de l'intervalle R-R du signal ECG_118.

Chapitre III : Présentation de l'interface graphique

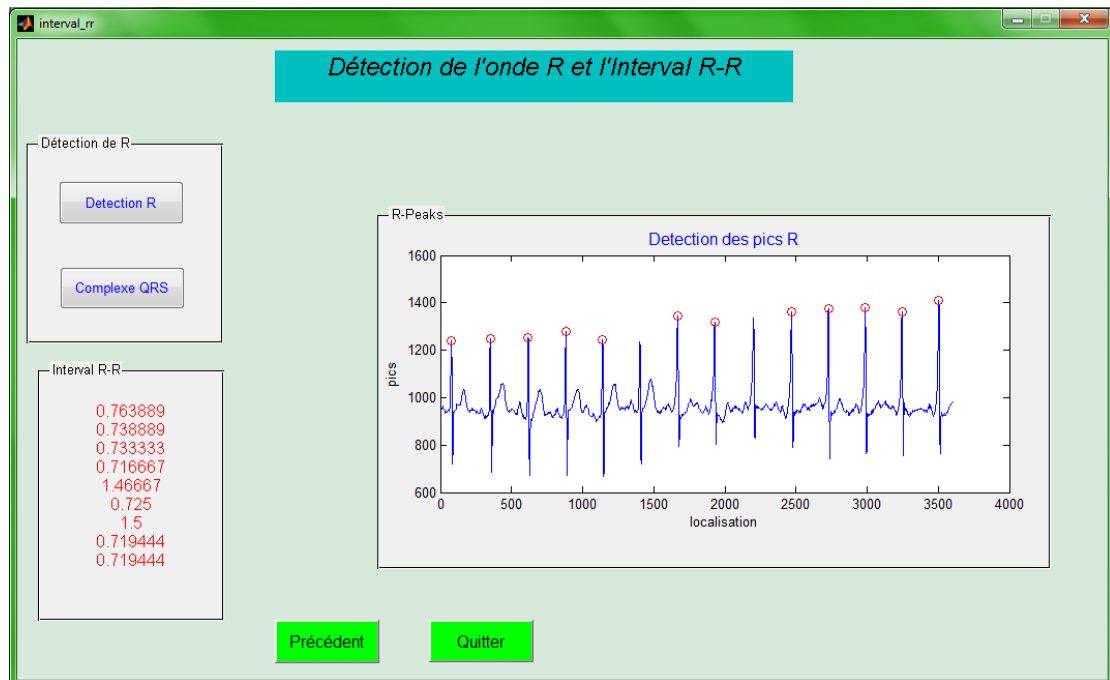


Figure III.10. Détection de R et de l'intervalle R-R du signal ECG_230.

Comme on peut le voir sur les figures (112 & 230), l'algorithme a sauté certaines ondes R, pour l'ECG_112 il a raté une seule onde et pour l'ECG_230 il a raté deux ondes.

Le tableau ci-dessous résume les différents intervalles R-R des signaux précédemment vus :

ECG	Intervalles R-R
ECG_112	0.730556 0.686111 0.713889 0.677778 0.708333 0.675 0.711111 0.680556 0.661111
ECG_118	0.833333 0.85 0.85 0.841667 0.844444 0.811111 0.822222 0.833333 0.808333
ECG_230	0.763889 0.738889 0.733333 0.716667 1.46667 0.725 1.5 0.719444 0.719444

Tableau III.1. Intervalles R-R.

Chapitre III : Présentation de l'interface graphique

Les différentes ondes R détectées sont représentées par un cercle rouge sur les figures précédentes. Les intervalles entre les ondes R-R sont affichés du côté gauche de la fenêtre et regroupés dans le tableau précédent.

- Un bouton « Complexe QRS » :

Ce bouton permet la détection du complexe QRS ainsi que des différentes ondes présentes sur le signal ECG.

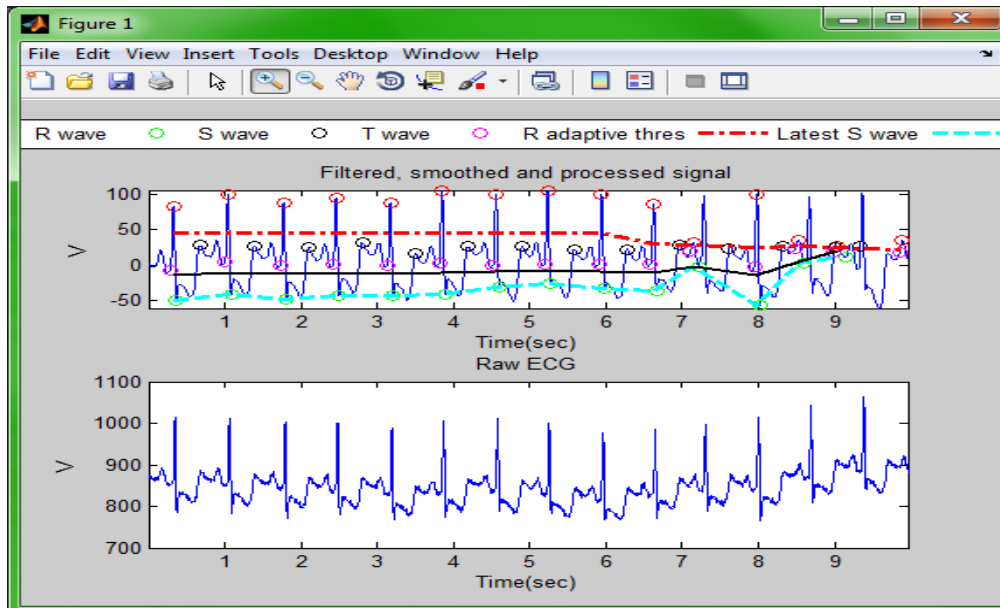


Figure III.11. Tracé du complexe QRS du signal ECG_112.

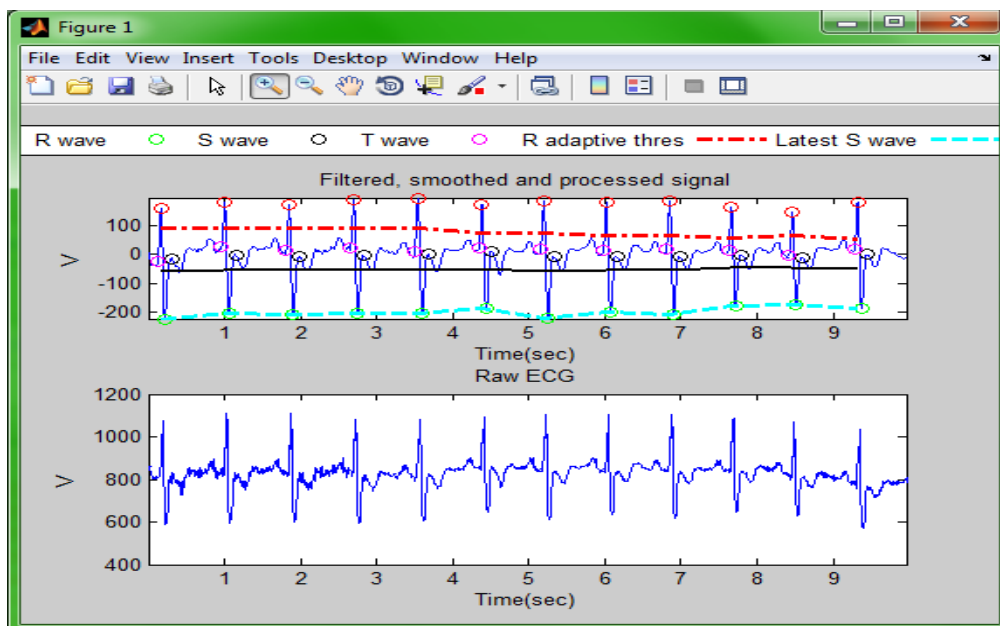


Figure III.12. Tracé du complexe QRS du signal ECG_118.

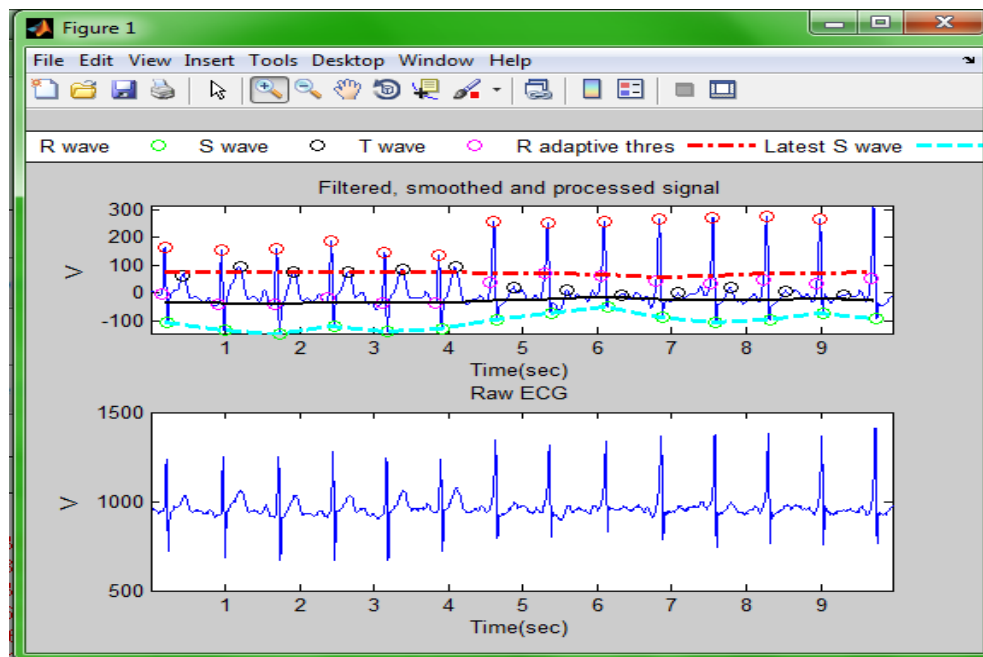


Figure III.13. Tracé du complexe QRS du signal ECG_230.

Les figures précédentes nous ont montrées les différentes ondes enregistrées dont le complexe QRS, sur les signaux ECG_112, 118, 230 respectivement.

En plus du complexe QRS ça nous affiche aussi la fréquence du rythme cardiaque pour les différents signaux ECG pris, qui seront illustrés ci-dessous :

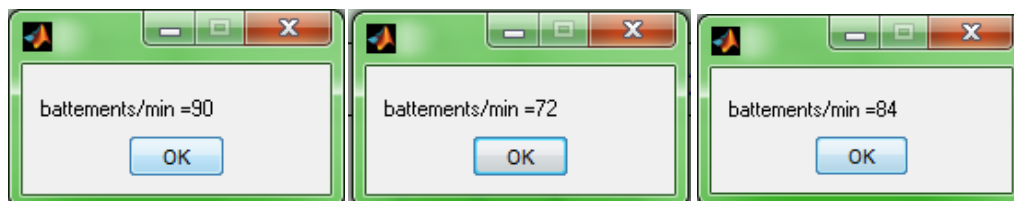


Figure III.14. Fréquence cardiaque des signaux ECG_112, 118 & 230 respectivement.

III.6. Classification :

Ce bouton va nous permettre, en cliquant dessus, de nous diriger vers une autre fenêtre qui est la fenêtre de classification, comme le montre la figure suivante :

Chapitre III : Présentation de l'interface graphique

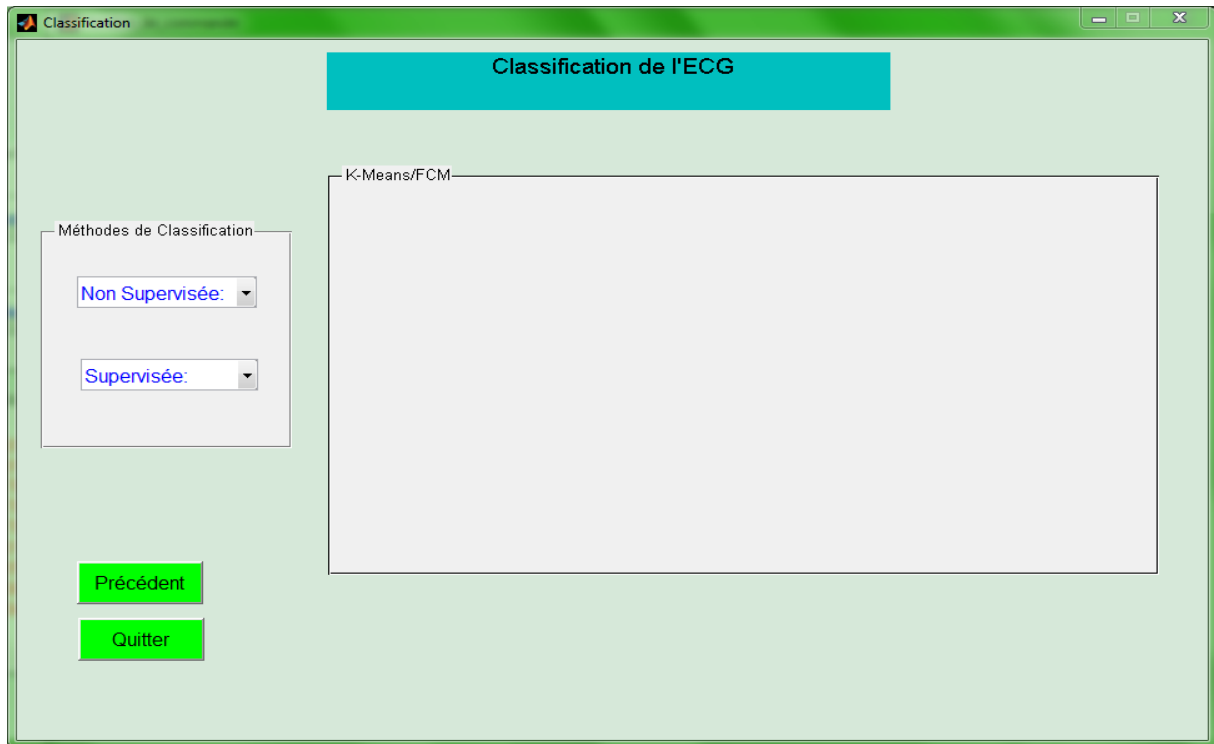


Figure III.15. Fenêtre de classification.

En appliquant la classification, on aura deux type de classe dans notre cas, une classe de type « patient sein » et une autre de type « patient malade ».

Comme on peut le voir sur la figure (III.15), la classification se fait à partir de la zone « Méthodes de Classification », d'où l'on choisi l'une des deux méthodes de classification suivantes : « la méthode supervisée » et « la méthode non supervisée ».

- «Classification Non Supervisée» : en choisissant cette méthode dans le menu déroulant on a deux choix de classification :
 - Choix de la méthode « K-Means », on aura une classification, comme nous le montre les figures ci-dessous pour les signaux précédemment cités :

Chapitre III : Présentation de l'interface graphique

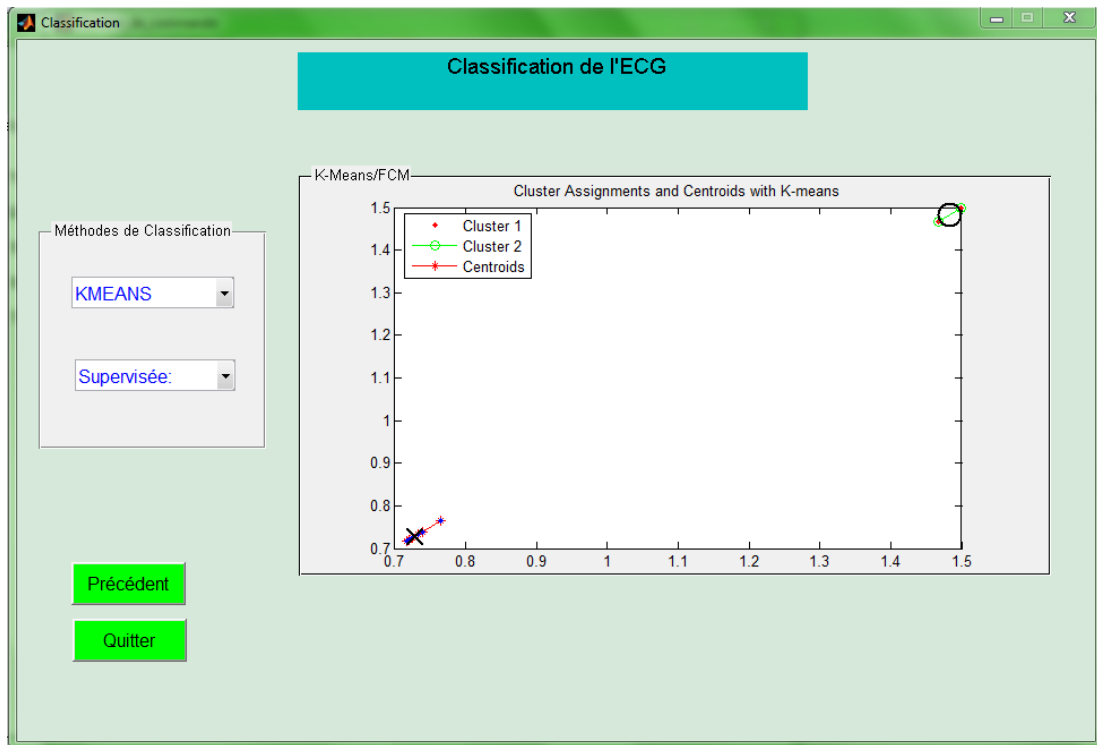


Figure III.16. Classification du signal ECG_112 avec la méthode K-Means.

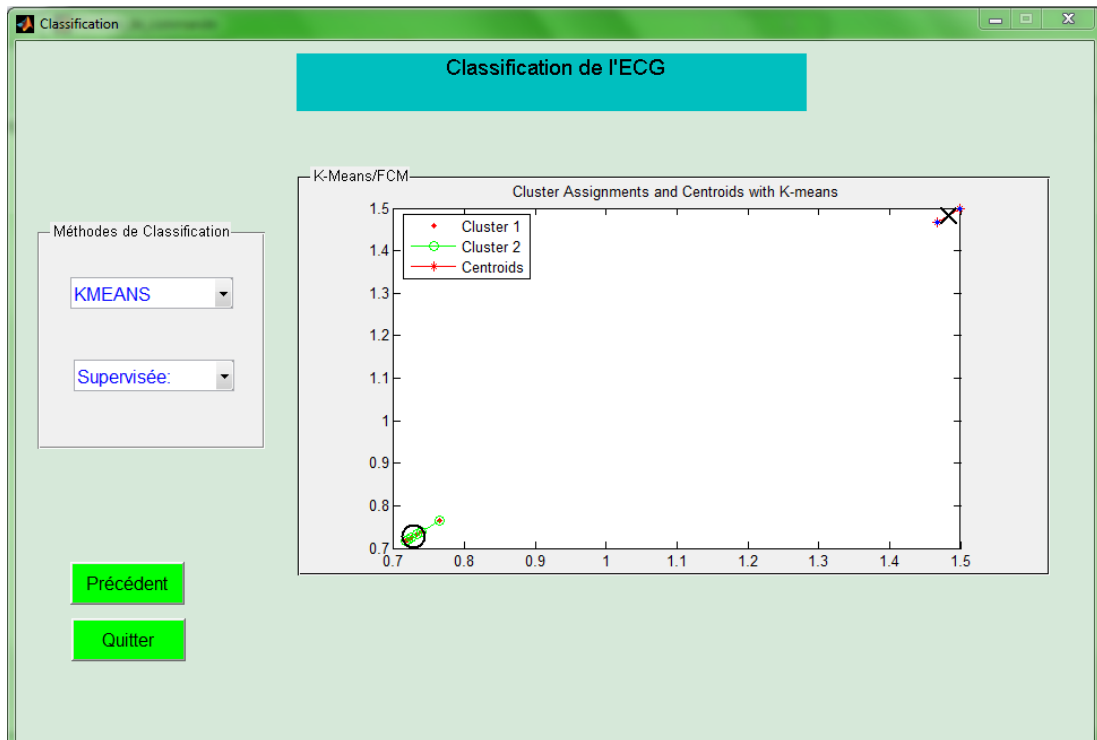


Figure III.17. Classification du signal ECG_118 avec la méthode K-Means.

Chapitre III : Présentation de l'interface graphique

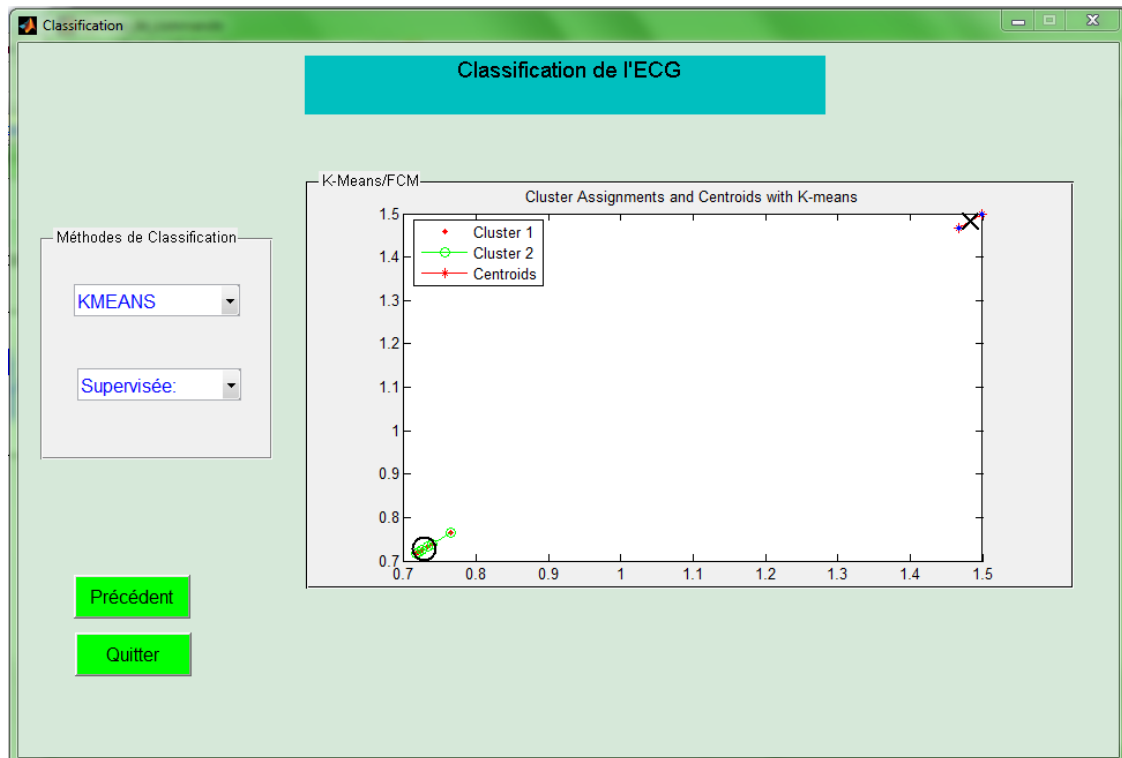


Figure III.18. Classification du signal ECG_230 avec la méthode K-Means.

- En choisissant « FCM », on aura une autre classification des signaux précédents comme seront illustrés sur les figures suivantes :

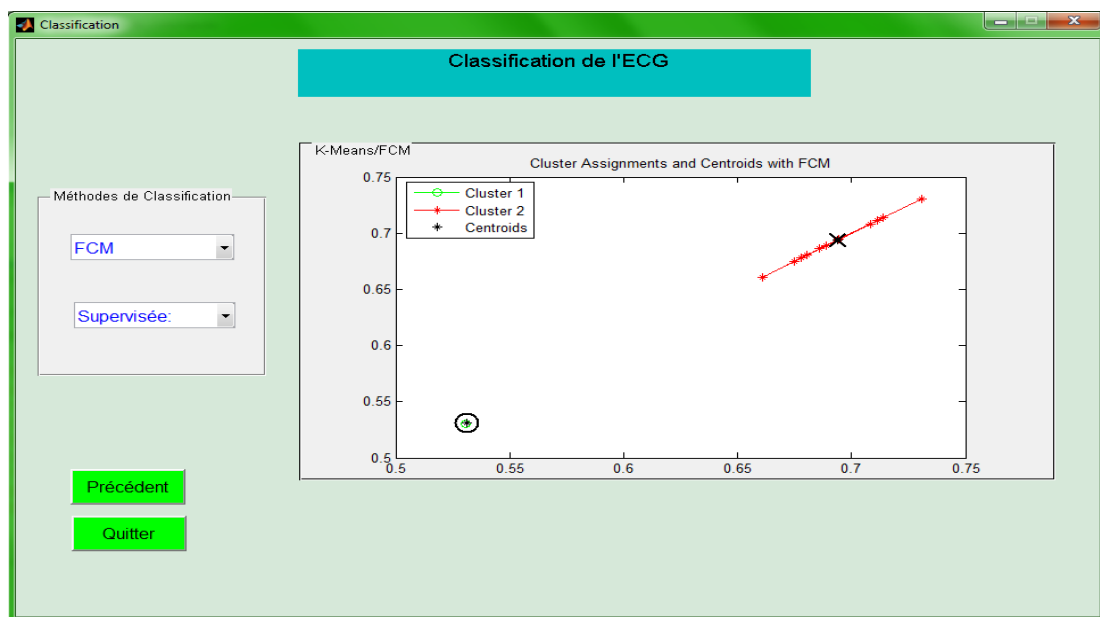


Figure III.19. Classification du signal ECG_112 avec la méthode FCM.

Chapitre III : Présentation de l'interface graphique

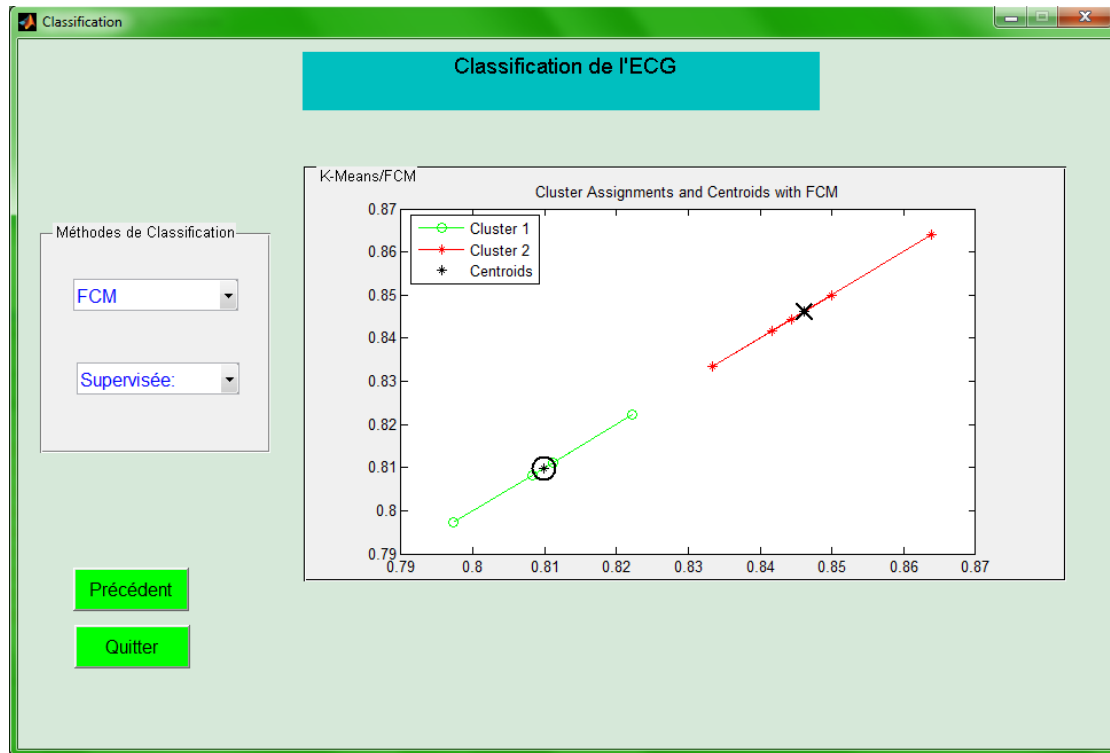


Figure III.20. Classification du signal ECG_118 avec la méthode FCM.

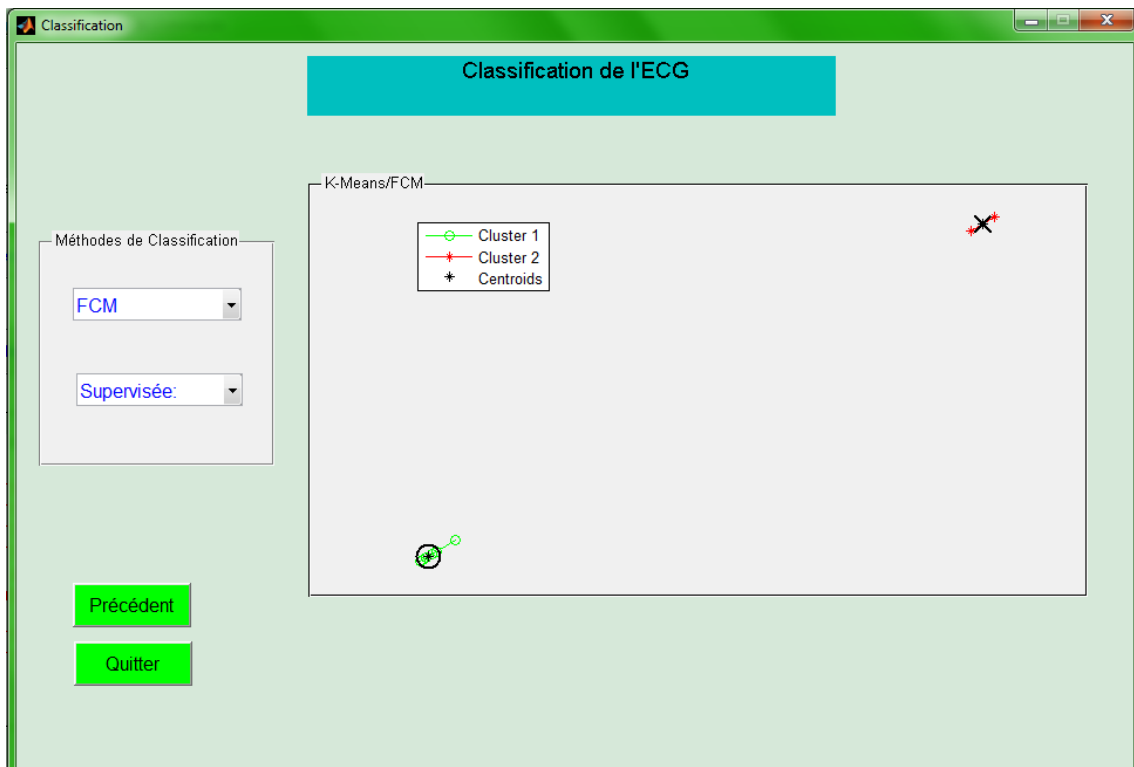


Figure III.21. Classification du signal ECG_230 avec la méthode FCM.

Chapitre III : Présentation de l'interface graphique

- « Classification Supervisées » : avec cette méthode nous avons utilisés “les réseaux de neurones”.

En choisissant cette méthode dans le menu déroulant nous serons dirigés vers une fenêtre externe qui représente le réseau de neurones, comme nous le montre la figure suivante :

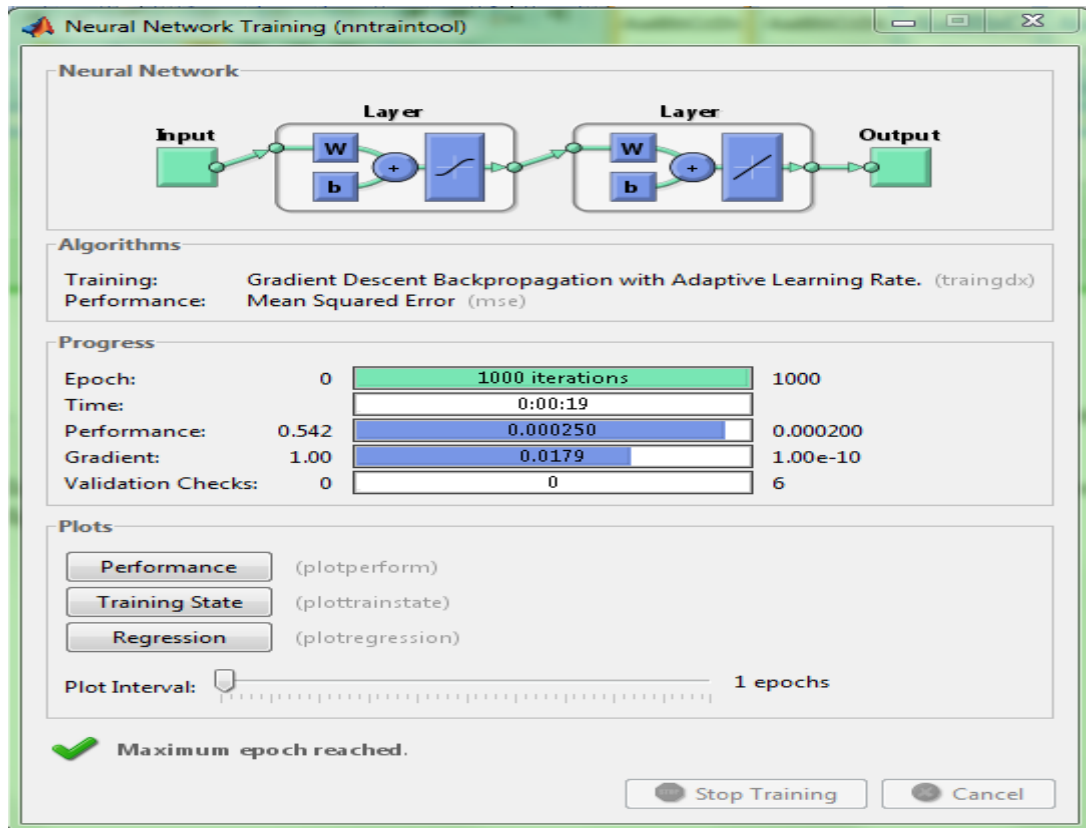


Figure III.22. Réseau de Neurones.

Ce type de réseau fonctionne uniquement pour une certaine catégorie d'anomalies qui sont de type PVC (Premature Ventricular Contractions), et qui sont des battements prématurés lancés dans les ventricules.

Lorsque la fenêtre sera affichée le réseau commence la détection des anomalies en faisant des itérations. On a fixé le nombre d'itération dans ce cas à 500 itérations.

Chapitre III : Présentation de l'interface graphique

Les résultats obtenus sont montrés dans la figure ci-dessous :

PVC :89994	PVC :11027	PVC :35312
PVC :90248	PVC :38489	PVC :126566
PVC :102167	PVC :80838	PVC :144355
PVC :102425	PVC :131196	PVC :202057
PVC :107633	PVC :131478	PVC :236350
PVC :108127	PVC :165955	PVC :249557
PVC :248978	PVC :177888	PVC :313375
PVC :249234	PVC :180270	PVC :313989
PVC :250519	PVC :204421	PVC :330408
PVC :250773	PVC :267814	PVC :339789
PVC :251030	PVC :345862	PVC :395128
PVC :251281	PVC :359279	PVC :414673
PVC :251536	PVC :489189	PVC :421890
PVC :251788	PVC :497897	PVC :526222
PVC :258938	PVC :531799	PVC :531197
PVC :259185	PVC :556792	PVC :553599
nombre PVC detecté :16	PVC :571256	PVC :561346
	PVC :571904	PVC :571196
		PVC :572064
		PVC :595661
		PVC :598611
		PVC :629087
		PVC :629434
		PVC :637334
		PVC :639080
	nombre PVC detecté :18	nombre PVC detecté :25

(a) (b) (c)

Figure III.23. Résultats de Détection des PVC pour :

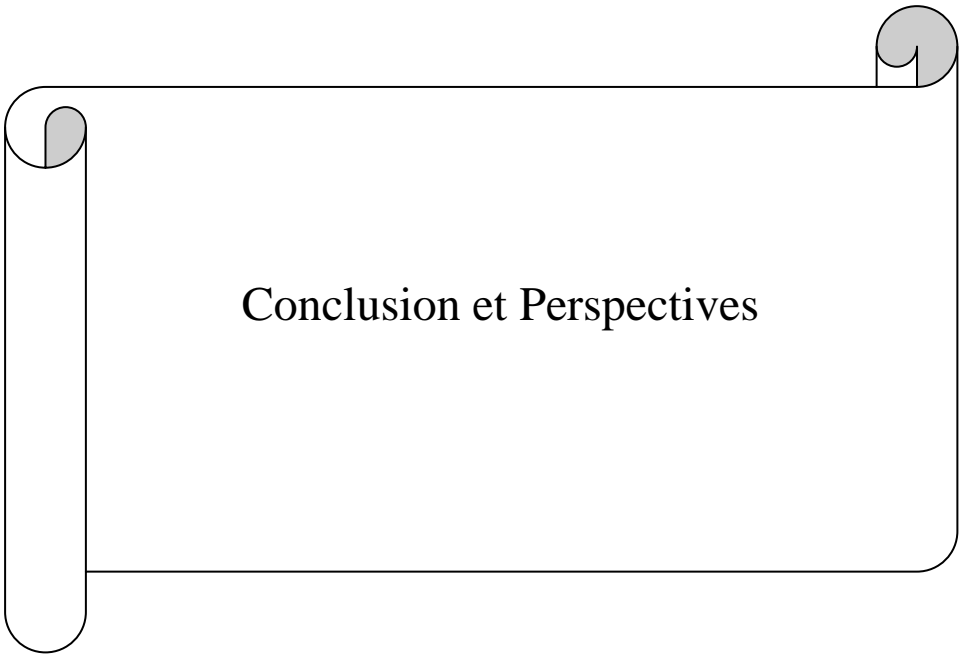
(a): ECG_112, (b): ECG_118 & (c): ECG_230.

La figure précédentes nous montrent l'intervalle du nombre des échantillons lus.

Dans chaque fenêtre il y a la possibilité soit de revenir à la fenêtre de commande principale en appuyant sur « Précédent », soit de quitter complètement l'interface en appuyant sur « Quitter ».

Conclusion :

Ce chapitre nous a permis de présenter notre interface graphique par les quelques tests effectués sur les signaux ECG de la base de donnée MI-BIH par les deux méthodes de classification à savoir la méthode supervisée et la méthode non supervisée.



Conclusion et Perspectives

Conclusion et Perspectives

Conclusion et perspectives :

Les différentes méthodes présentées au cours de ce mémoire ont pour but le prétraitement et la classification des signaux ECG pour donner au corps médical les appuis et les outils d'aide au diagnostic des différentes pathologies cardiaques.

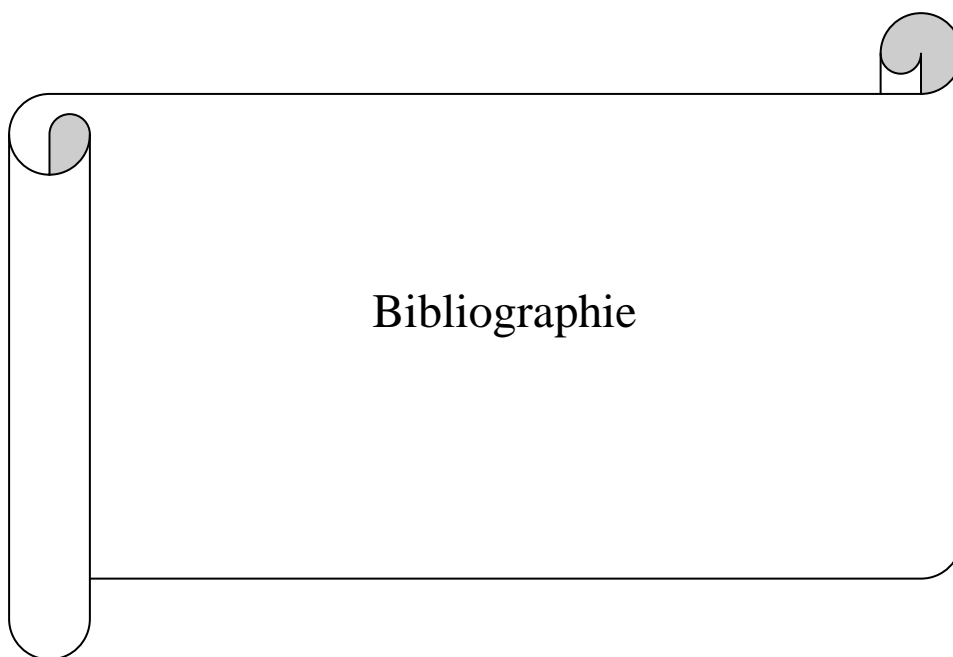
Nous avons abordés le problème du prétraitement des signaux ECG par une approche basée sur la méthode des ondelettes dans le but d'éliminer les différents bruits pouvant affecter le signal ECG, car un enregistrement bruité affecte la détection des complexes QRS et diminue la robustesse de chaque méthode de classification utilisée.

Pour la détection du complexe QRS et de l'intervalle R-R, la méthode que nous avons utilisée est basée sur celle développée par les deux chercheurs Pan & Tompkins. Ces méthodes de prétraitement sont utilisées dans le but d'améliorer la classification.

Nous avons utilisé deux méthodes pour la classification des signaux ECG qui sont « la méthode non supervisée » et « la méthode supervisée ». Dans la première méthode nous avons utilisés « le K-Means & le FCM », dans la seconde méthode c'est « le Réseau de Neurones ».

Le but de ces méthodes de classification est d'analyser, de déterminer la nature de l'ECG et de faire la classification a partir des données extraites dans les signaux.

En perspective nous souhaitons que notre travail soit amélioré au niveau de la classification des signaux ECG par l'introduction de nouvelles méthodes comme les chaînes de Markov Caché afin d'extraire d'autres informations tel que les amplitudes des différentes ondes, ou par l'amélioration des algorithmes existants comme le FCM en augmentant sa robustesse au bruit avec l'intégration de l'information spatiale (FCM_S).



Bibliographie

Bibliographie

- [01] Clifford, G. D., Azuaje, F., & McSharry, P. (2006). *Advanced methods and tools for ECG data analysis*. Artech House, Inc..
- [02] TALBI, M. L. (2011). *Analyse et traitement du signal électrocardiographique (ECG)*.
- [03] Cabasson, A. (2008). *Estimation et analyse des intervalles cardiaques* (Doctoral dissertation, Université Nice Sophia Antipolis).
- [04] Tchiotsop, D. (2007). *Modélisations polynomiales des signaux ECG. Applications à la compression* (Doctoral dissertation, Université de Reims Champagne-Ardenne).
- [05] Dubois, R. (2004). *Application des nouvelles méthodes d'apprentissage à la détection précoce d'anomalies en électrocardiographie* (Doctoral dissertation, UNIVERSITÉ PARIS 6).
- [06] Portet, F. (2005). *Pilotage d'algorithmes pour la reconnaissance en ligne d'arythmies cardiaques* (Doctoral dissertation, Université Rennes 1).
- [07] Mackay, J., Mensah, G. A., Mendis, S., & Greenlund, K. (2004). *The atlas of heart disease and stroke*. World Health Organization.
- [08] S. A. Jones, M.S Ed, MHA, EMT-P, « *ECG Notes Interpretation and Management Guide* », F. A. Davis Company, Philadelphia, USA, 2005.
- (Jones, S. A. (2009). *ECG notes: Interpretation and management guide*. FA Davis.)
- [09] Adamec, J., & Adamec, R. (2000). *ECG Holter: manuel d'interprétation électrocardiographique*. Médecine et hygiène.
- [10] MIT-BIH Arrhythmia Database, <http://www.physionet.org/physiobank/database>.
- [11] Clinique Cardiologique CHU de GRENOBLE, Sémiologies et Pathologies Cardiovasculaires, <http://www-sante.ujf-grenoble.fr>
- [12] Donoho, D. L. De-noising by soft-shresholding. *IEEE Transactions on Information Theory*, 1, 995(4), 1.
- [13] Mallat, S. (1999). *A wavelet tour of signal processing*. Academic press.
- [14] Zerubia, J. (2012). *Analyse en ondelettes M-bandes en arbre dual; application à la restauration d'images* (Doctoral dissertation, Université de Marne-la-Vallée).
- [15] Simard, V. (2005). *Transformée en ondelettes pour un système d'acquisition de signaux corticaux implantable* (Doctoral dissertation, École polytechnique).
- [16] Benitez, D., Gaydecki, P. A., Zaidi, A., & Fitzpatrick, A. P. (2001). The use of the Hilbert transform in ECG signal analysis. *Computers in biology and medicine*, 31(5), 399-406.

Bibliographie

- [17] Pan, J., & TOMPKINS, W. (1985). A real-time QRS detection algorithm. IEEE Transaction on biomedical Engineering, vol. *BME*, 32.
- [18] Laouamer, L. (2006). *Approche exploratoire sur la classification appliquée aux images*. Université du Québec à Trois-Rivières.
- [19] Govaert, G. (1990). Classification binaire et modèles. *Revue de statistique appliquée*, 38(1), 67-81.
- [20] Clifford, G. D., Azuaje, F., & McSharry, P. (2006). *Advanced methods and tools for ECG data analysis*. Artech House, Inc..
- [21] Mathworks, T. (2009). Matlab users guide. *The Mathworks*.
- [22] Tzanakou, E. M. (Ed.). (1999). *Supervised and unsupervised pattern recognition: feature extraction and computational intelligence*. CRC Press.
- [23] Sadhukhan, D., & Mitra, M. (2012). R-Peak detection algorithm for ECG using double difference and RR interval processing. *Procedia Technology*, 4, 873-877.
- [24] Kacimi, M.A & Kaced, M (2011). Classification non supervisée des signaux ECG. Mémoire de fin d'étude en vue de l'obtention du diplôme d'ingénieur d'état en électronique, université A. Mira de Bejaia.
- [25] Messaoud, A., Messaoud, M. B., Kachouri, A., & Sellami, F. Classification des arythmies cardiaques par logique floue à partir de signaux ECG. *SETIT2005*.
- [26] S Tufféry, (2005). « data mining et statistique décisionnelle, l'intelligence dans les bases de données ». Groupe bancaire français, université de Rennes 1 et paris Dauphine.
- [27] McCULLOCH W. S., PITTS W. A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics*, 5, p. 115-133, 1943.
- [28] G. Dreyfus, J. M. Martinez, M. Samuelides, M. B. Gordon, F. Badran, S. Thiria, L. Héroult, RESEAUX DE NEURONES : METHODOLOGIE ET APPLICATIONS, Eyrolles, 2002.
- [29] M.H. Hassoun, FUNDAMENTALS OF ARTIFICIAL NEURAL NETWORKS, The MIT Press, 1995.

Résumé :

Le signal électrocardiogramme (ECG) est l'un des outils les plus utilisés dans le diagnostic et le suivi de l'état cardiaque des patients. Il représente l'activité électrique du cœur en fonction du temps. Il contient aussi des informations qui permettent la distinction des maladies cardiovasculaires.

La classification de ces signaux est une tâche très importante en traitement de signal. La classification est partagée en deux grandes familles qui sont la classification « supervisée » et « non supervisée », et un grand nombre de chercheurs ont développés ces méthodes de classification.

Dans ce travail nous nous sommes intéressés à ces différentes méthodes de classification qui sont, « K-Means et FCM », pour la méthode non supervisée et « les réseaux de neurones », pour la méthode supervisée.

La classification ne peut se faire qu'après avoir traité le signal ECG et l'extraction des paramètres caractérisant ce signal; comme le dé-bruitage, la détection du complexe QRS et la détection de l'intervalle R-R.

Abstract :

The electrocardiogram (ECG) signal is one of the most used tools in the diagnosis and monitoring of cardiac status of patients. It represents the electrical activity of the heart over time. It also contains information that enables the distinction of cardiovascular disease.

The classification of these signals is a very important task in signal processing. Many researchers have developed these classification methods which are divided into two major groups: supervised and unsupervised classification.

In this study we were interested in these different classification methods which are « K -means and FCM » for unsupervised method and « neural networks » for supervised method.

The classification can be done after the ECG signal treatment and extraction of parameters characterizing this signal; such de-noising, detection of QRS complex and detection of the R -R interval.