



جامعة أبو بكر بلقايد - تلمسان

Université Abou Bakr Belkaïd de Tlemcen

Faculté de Technologie

Département de Génie Biomédical

Laboratoire de Recherche de Génie Biomédical

MEMOIRE DE PROJET DE FIN D'ETUDES

Pour l'obtention du Diplôme de

MASTER en GENIE BIOMEDICAL

Spécialité : Informatique Biomédicale

Présenté par : DAHMANI Ahmed Azeddine

**UNE APPROCHE DE REGRESSION LOGISTIQUE
MULTINOMIALE POUR LA DETECTION DE
L'ARYTHMIE.**

Soutenu le 30 juin 2019 devant le Jury

M.	MESSADI Mohamed	<i>MCA</i>	Université de Tlemcen	Président
M.	BEHADADA Omar	<i>MCB</i>	Université de Tlemcen	Encadreur
M.	M.BENAZOUZ Mourtada	<i>MCA</i>	Université de Tlemcen	Examineur
Mm.	BENCHAIB Yasmine	<i>MCB</i>	Université de Tlemcen	Co-Encadreur

Année universitaire 2018-2019

Table des matières

1	Introduction générale	1
1	Electrocardiographie et pathologies cardiaques	4
1	Introduction	4
2	Anatomie et physiologie du cœur	4
2.1	Anatomie du cœur	4
2.2	Physiologie cardiovasculaire	8
3	Electrocardiographie	10
3.1	Electrocardiogramme	11
3.2	Les dérivations	13
4	La commande rythmique du cœur	15
4.1	Le système nerveux autonome	15
5	Rythme cardiaque physiologique.	15
6	Les troubles de rythme ou les arythmies cardiaques (rythme cardiaque à l'état pathologique)	16
6.1	Qu'est-ce qu'une arythmie cardiaque?	16
6.2	Les troubles du rythme supraventriculaire	18
6.3	Les troubles du rythme ventriculaire	19
7	Conclusion	21
2	État de l'art	22
1	Introduction	22
2	Traitement de signal	22
3	la base de données	23
4	Travaux reliées	25
5	Travail proposé	28
6	Conclusion	28
3	régression logistique	29
1	Introduction	29

2	Généralité sur la classification	29
2.1	La classification	29
2.2	Approche Paramétrique versus non-paramétrique	30
3	Régression logistique	31
3.1	Définition mathématique du modèle de régression lo- gistique	32
3.2	Intérêts de la régression logistique	32
3.3	Les types de régression logistique	33
4	Modèle de régression logistique multinomiale proposé	35
5	Conclusion	36
4	Résultats et discussion	37
1	Introduction	37
2	Préparation de données	37
3	Classification	38
3.1	phase d'apprentissage	39
3.2	Phase de test	39
3.3	La distribution des descripteurs en boites à moustaches	39
3.4	Étude de corrélation	45
4	Résultats et évaluation	48
5	Études comparatives avec d'autres techniques de classification	50
5.1	Les mesures de performances	50
6	Conclusion	51

Table des figures

1.1	Anatomie et système de conduction cardiaque	5
1.2	Configuration interne et connexions vasculaires	6
1.3	valves cardiaques en diastole et systole	7
1.4	Tissu nodal et de conduction	8
1.5	Variations du potentiel transmembranaire en fonction du temps	9
1.6	Cycle cardiaque (mouvements des parois et des valves (flèches))	10
1.7	Étapes successives de dépolarisation/repolarisation du cœur qui se traduisent sur le plan	11
1.8	L'ECG d'un battement de cœur	12
1.9	triangle d'Einthoven. Le corps est supposé avoir une configura- tion	13
1.10	positions d'électrodes recommandées pour l'enregistrement du système de dérivations unipolaires précordiales (V1-V6).	14
1.11	ECG D'un rythme cardiaque normal	16
1.12	système de conduction cardiaque	17
1.13	ECG d'une Fibrillation auriculaire	18
1.14	Tracé ECG cas d'un Flutter auriculaire	19
1.15	ECG d'un trouble du rythme jonctionnels	19
1.16	ECG d'une Tachycardie ventriculaire	20
1.17	ECG Fibrillation ventriculaire	20
2.1	battement ECG standard	23
4.1	plan de travail	38
4.2	la distribution de la fonctionnalité durée P dans différentes classes	39
4.3	la distribution de la fonctionnalité intervalle PR dans diffé- rentes classes	40
4.4	la distribution de la fonctionnalité complexe QRS dans diffé- rentes classes	40
4.5	la distribution de la fonctionnalité durée T dans différentes classes	41

4.6	la distribution de la fonctionnalité segment ST dans différentes classes	41
4.7	la distribution de la fonctionnalité l'intervalle QT dans différentes classes	42
4.8	la distribution de la fonctionnalité RRp dans différentes classes	43
4.9	la distribution de la fonctionnalité RRS dans différentes classes	43
4.10	la distribution de la fonctionnalité RDI battement dans différentes classes	44
4.11	la distribution de la fonctionnalité durée battement dans différentes classes	44
4.12	la distribution de la fonctionnalité RRs/RRp dans différentes classes	45
4.13	histogramme de mesures de performances de la MLR.	49

Liste des tableaux

1.1	concentrations intra et extra-cellulaires (données en mmoles) des principaux ions impliqués dans les phénomènes électrophysiologies cardiaques et valeurs des potentiels électrochimiques d'équilibre correspondant.	9
2.1	Données d'évaluation extraites de la base de données	24
2.2	Détails du jeu de données de la base de données	24
2.3	les paramètres ou descripteurs de la base de données	25
4.1	Détails du jeu de données, de la base de données après l'élimination de la deuxième classe	38
4.2	les types d'arythmie	38
4.3	Coefficients de corrélation	47
4.4	matrice de confusion	48
4.5	Résultats de Classification de la MLR pour les différents types de battements cardiaques	49
4.6	comparaison avec les techniques de classification de l'état de l'art.	50

ABBREVIATIONS

MLR : Multinomial Logistic Regression

ECG : Electrocardiogramme

PVC : Les contractions ventriculaires prématurées (Premature Ventricular Contraction)

N : Battements Normaux

LBBB : Les blocs de branche gauche (Left Bundle Branch Block)

RBBB : Les blocs de branche droite (Right Bundle Branch Block)

QRS : durée du complexe QRS

RR_p : distance entre le pic R actuel et le pic R précédant

RR_s : distance entre le pic R actuel et le pic R suivant

T_c : taux de classification

Se : sensibilité

Sp : spécificité

Dedicaces

Je dédie ce travail à :
A mon frère AICI Said,
A ma mère,
A mon père,
A mes frères et sœurs,
A toute ma famille,
A DRIS Younes, MERAH Mohamed, DERG Fouad, KARI Mohamed Zakaria
et CHRIAR Amine.
A tous mes enseignants et mes collègues a l'université de Tlemcen,
A tous ceux que j'aime,
Qu'ils trouvent ici l'expression de toute ma reconnaissance.

Remerciement

*Je tiens d'abord à remercier mes encadreurs et mes professeurs **M.BEHADAD Omar** et **Mme.BENCHAIB Yasmine** pour avoir bien voulu encadrer ce travail ainsi que pour leurs riche contribution et ses précieux conseils.*

*J'adresse mes très sincères remerciements à Monsieur **HADJAHMED Ismail**, doctorant à l'université de Tlemcen.*

Je tiens aussi à remercier les membres du jury qui m'ont fait l'honneur de participer à l'examen de ce travail :

- **M.MESSADI Mohamed**, Maitre de conférence, Université de Tlemcen.*
- **M.BENAZOUZ Mourtada**, Professeur d'informatique, Université de Tlemcen.*

Je tiens aussi à remercier l'ensemble du corps enseignant de Département GBM.

Les encouragements de nos amis et de nos collègues étaient la bouffée d'oxygène qui nous ressourçait dans les moments pénibles, de solitude et de souffrance, où l'on a terriblement besoin d'un petit mot, d'un petit geste, aussi soit-il de soutien moral.

Et nous en venons à nos famille, à nos parents, à nos adorables parents, à ceux qui sont toujours présents et continuent de l'être pour faire notre bonheur.

1 Introduction générale

Des changements dans le rythme normal du rythme cardiaque humain peuvent entraîner différentes arythmies cardiaques, pouvant être immédiatement fatales ou causer des lésions irréparables au cœur et durablement endurées. La capacité d'identifier automatiquement les arythmies à partir d'enregistrements d'électrocardiogramme (ECG) est importante pour le diagnostic et le traitement [1].

Les électrocardiogrammes sont très largement utilisés comme moyens peu coûteux et non invasifs d'observer la physiologie cardiaque. Le signal ECG ou électrocardiogramme mesure l'activité électrique des cellules cardiaques spécialisées qui génèrent un potentiel d'action auto-induit répétitif. Chaque potentiel d'action généré entraîne une contraction du muscle cardiaque et donc du rythme cardiaque. Le signal ECG contient une quantité importante d'informations pouvant être exploitées de différentes manières. Le signal ECG permet d'analyser les aspects anatomiques et physiologiques de l'ensemble du muscle cardiaque [2].

Un signal ECG normal consiste en une onde P, QRS, T et U. Les ECG servent à diagnostiquer les maladies cardiaques, à identifier les rythmes cardiaques irréguliers (arythmies), à évaluer les effets des médicaments et à surveiller les procédures chirurgicales. La magnitude, la conduction et la durée de ces potentiels sont détectées en plaçant des électrodes sur la peau du patient. À partir du tracé ECG, les informations suivantes peuvent être déterminées : fréquence cardiaque ; rythme cardiaque, anomalies de la conduction (anomalies dans la façon dont l'impulsion électrique se propage à travers le cœur), maladie coronarienne, anomalie du muscle cardiaque, etc. [3].

Selon le Rapport de l'organisation mondiale de la santé en 2018 [4]. Les maladies cardiovasculaires sont l'un des problèmes de santé les plus préoccupants et la principale cause de mortalité dans le monde, Ainsi, une évaluation cardiaque de faible coût et de haute qualité constitue un défi de taille. En outre, la disponibilité d'une quantité considérable d'informations générée par le développement continu des méthodes et des techniques du Big Data offre de nouveaux défis ainsi que de nouvelles opportunités dans ce domaine.

La détection de l'arythmie cardiaque est un domaine très intéressant, car la contraction ventriculaire (PVC) est un facteur de prédiction efficace de la mort subite. Plusieurs études, au cours de la dernière décennie, se sont concentrées sur des méthodes et des algorithmes de détection et de signification des arythmies cardiaques, dans le but d'atteindre un bon taux de classification.

Plus spécifiquement, de nombreuses méthodes de classification ont été utilisées sur le terrain, telles que les classificateurs de Bayes, les arbres de décision, les apprenants neuronaux et basés sur des règles [5]. Cependant, les méthodes de classification présentant des bons taux de classification ont généralement un faible degré d'interprétabilité, ce qui empêche l'utilisateur (comme un cardiologue) de tirer pleinement parti de cette méthode.

L'interprétabilité de tout système basé sur la connaissance est cruciale, en particulier lorsqu'il s'agit d'applications Big Data [6]. En fait, cela garantit un progrès décisionnel efficace en produisant des connaissances interprétables, qui peuvent facilement être maintenues et évaluées. L'acquisition de connaissances spécialisées est une tâche complexe mais essentielle en raison de son exactitude inhérente, si elle est réalisée par une intervention humaine. Cependant, cela a tendance à être inefficace lorsqu'il s'agit de grands ensembles de données et, de plus, une telle connaissance contient un composant inconscient, difficile à formaliser [7]. Alternativement, les connaissances peuvent également être définies en analysant les informations extraites des données expérimentales, ce qui est susceptible de fournir un aperçu précis des différents paramètres. Il existe en particulier une multitude d'algorithmes et des techniques d'apprentissage automatique pour l'identification des modèles, basés sur les propriétés des indices de précision, qui peuvent être appliqués au processus d'induction de connaissances [8].

Dans notre recherche, nous appliquons l'algorithme MLR (régression logistique multinomiale) pour la classification de différentes arythmies cardiaques avec un meilleur taux de classification et gagner le temps pour les maladies cardiaques pour éviter l'arrêt cardiaque .

Ce mémoire de fin d'études de Master IBM est constitué de 4 chapitres qui sont représentés comme suit :

1. **chapitre 01** : électrocardiographie et pathologies cardiaques, englobe des définitions sur le système cardiovasculaire, les arythmies cardiaques et l'électrocardiographie.
2. **chapitre 02** : état de l'art, expose les travaux du domaine de la classification des arythmies cardiaques et la description de la base de données.

3. **chapitre 03** : régression logistique, des généralités sur la classification, la régression logistique et particulièrement régression logistique multinomiale.
 4. **chapitre 04** : Résultats et discussion, dévoile l'étude expérimentale.
- En dernier lieu, une conclusion générale qui résume les points essentiels de ce travail.

Chapitre 1

Electrocardiographie et pathologies cardiaques

1 Introduction

le système cardio-vasculaire est composé du cœur et des vaisseaux sanguins (artères et veines)[9], le cœur est considéré comme une pompe qui chasse le sang vers les différents organes, à travers ces canaux de façon continue et pulsatile en assurant les besoins métaboliques (l'O₂, et besoins énergétiques : glucose, protéines, lipides etc ...) et ramenant les déchets vers les organes d'épuration, Ex : poumons (CO₂), les reins (produits de dégradation des métabolites) et au foie aussi, le tout dans un état hémodynamique stable de TA (tension artérielle), FC, de volémie efficace.

Dans ce chapitre, nous allons discuter :

- l'anatomie du cœur,
- la physiologie cardiovasculaire,
- l'Electrocardiographie,
- le Rythme cardiaque physiologique,
- les troubles de rythme ou les arythmies cardiaques (rythme cardiaque à l'état pathologique).

2 Anatomie et physiologie du cœur

2.1 Anatomie du cœur

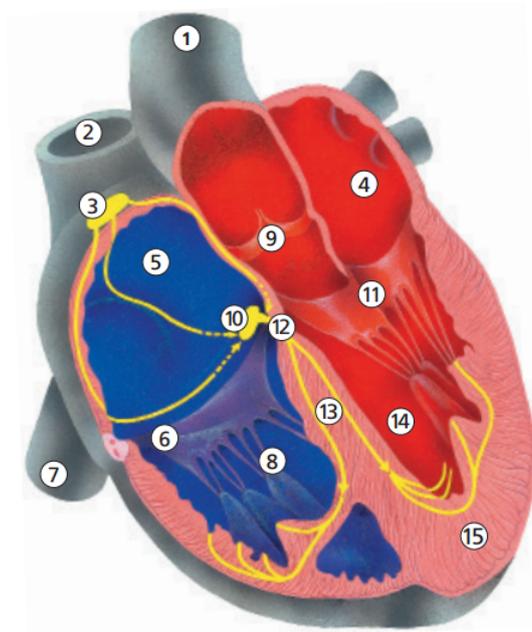


FIGURE 1.1 – Anatomie et système de conduction cardiaque

source : [10]

1. Aorte
2. Veine cave supérieure
3. Nœud sinusal
4. Oreillette gauche
5. Oreillette droite (atrium)
6. Valve tricuspide
7. Veine cave inférieure
8. Ventricule droit
9. Valve aortique
10. Nœud atrio-ventriculaire (AV)
11. Valve mitrale
12. Faisceau de His (atrium)
13. Branches droite et gauche
14. Ventricule gauche
15. Muscle cardiaque (myocarde)

Position

Le cœur est situé au niveau du thorax dans le médiastin antérieur[11], entre les deux poumons en arrière du sternum et en avant de l'œsophage, reposant sur le diaphragme gauche.

Description extérieur

Le cœur est un organe creux, qui pèse environ 300 g chez l'adulte [12] : Il a la forme d'une pyramide dont le sommet est en bas, la base regarde en haut, son grand axe est oblique en bas, en avant et à gauche, Il comprend trois faces antérieure, inférieure et latérale.

Description intérieur

Le cœur est un organe creux formé du cœur droit et du cœur gauche qui sont totalement séparés[13]. Les oreillettes sont séparées par le septum inter auriculaire et les ventricules par le septum inter ventriculaire. Le cœur droit est formé de l'oreillette (OD) et du ventricule droit (VD) qui se communique entre eux par le biais des valves tricuspidiennes. Le cœur gauche est formé de l'oreillette (OG) et du ventricule gauche (VG) qui se communique entre eux par le biais des valves mitrales représenté dans la Figure 1.2.

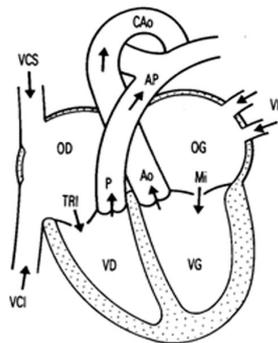


FIGURE 1.2 – Configuration interne et connexions vasculaires

source : [14]

Connexions vasculaires

L'oreillette droite reçoit les veines caves inférieure (VCI) et supérieure (VCS) par lesquelles le sang veineux revient au cœur illustré[9].

Le ventricule droit se vide via l'orifice pulmonaire (P) dans l'artère pulmonaire (AP).

L'oreillette gauche reçoit 4 veines pulmonaires (VP) qui drainent le sang oxygéné des poumons vers l'oreillette gauche.

Le ventricule gauche se vide via l'orifice aortique dans l'aorte (Ao puis Cao = crosse de l'aorte)(Figure1.2).

Appareils valvulaires

Les orifices des grandes artères pulmonaires et aortiques, sont en forme de valves dont chacune comporte un anneau fibreux et trois valvules nommées sigmoïdes[13].

Entre l'étage auriculaire et l'étage ventriculaire il existe 2 appareils valvulaires auriculo-ventriculaires formés d'un anneau fibreux et d'un système valvulaire constitué de valvules, de cordages et de piliers. L'orifice tricuspide qui représente à la figure 1.3 contient : 3 valvules, et le mitral : 2 (grande et petite valves).

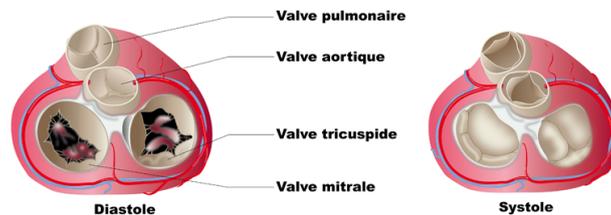


FIGURE 1.3 – valves cardiaques en diastole et systole

source : [16]

Le tissu Nodal et de conduction

Ce tissu électrique donne l'origine aux impulsions électriques puis les conduit vers les cellules myocardiques, engendrant leurs dépolarisations et par la suite la contraction cardiaque.

Il est formé du nœud sinusal(1) siégeant dans la paroi de l'oreillette droite(2) près de l'abouchement de la veine cave supérieure, du nœud auriculo-ventriculaire(3)

qui siège juste au niveau de la jonction de l'oreillette droite avec le ventricule droit. Puis le faisceau de His(4) qui se divise ensuite en deux branches droite(d) et gauche(g) puis se distribue grâce au réseau de Purkinje(6) jusqu'au contact des cellules myocardiques illustré dans la figure 1.4[13] .

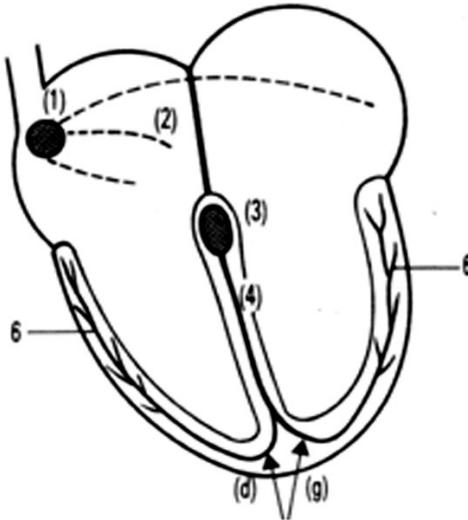


FIGURE 1.4 – Tissu nodal et de conduction

source : [17]

2.2 Physiologie cardiovasculaire

Potentiels de repos et d'action

Au repos, les cellules myocardiques comme toutes les cellules de l'organisme sont « polarisées » chargées positivement à l'extérieur et négativement à l'intérieur. Si deux microélectrodes reliées à un galvanomètre sont placées de part et d'autre de la membrane cellulaire, une différence de potentiel s'inscrit, de l'ordre de -90 mV pour une cellule ventriculaire : c'est le potentiel de repos transmembranaire.

Si la cellule myocardique est stimulée un potentiel d'action apparaît, qui traduit les variations du potentiel transmembranaire en fonction du temps, qui est les résultats des variations des concentrations ioniques (Na^+ , K^+ , Cl^-) de part et d'autre de la membrane cellulaire à travers les canaux voltages ioniques[18].

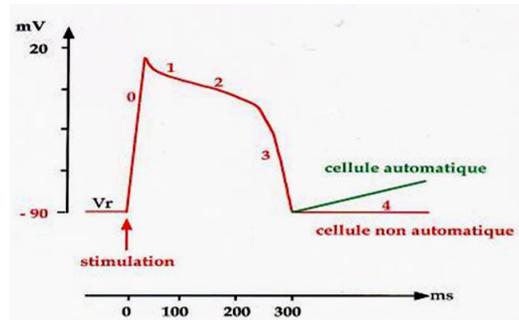


FIGURE 1.5 – Variations du potentiel transmembranaire en fonction du temps

source : [19]

Concentrations ioniques		
	Intra-cellulaire	Extra-cellulaire
Cl	30 mmoles	140 mmoles
Na+	10 mmoles	140 mmoles
Ca++	100 mmoles	2 mmoles
K+	140 mmoles	4 mmoles

TABLE 1.1 – concentrations intra et extra-cellulaires (données en mmoles) des principaux ions impliqués dans les phénomènes électrophysiologiques cardiaques et valeurs des potentiels électrochimiques d'équilibre correspondant.

source : [22]

Le cycle cardiaque

L'activité électrique du cœur est à l'origine de son activité contractile avec un petit décalage dans le temps de l'ordre de quelques dizaines de millisecondes, la contraction cardiaque autrement dite le cycle cardiaque est représenté en gros par la contraction (dite systole ventriculaire) puis la relaxation des ventricules (dite Diastole ventriculaire) essentiellement (la contraction des oreillettes n'a qu'une incidence limitée sur la fonction cardiaque normale). L'alternance d'une systole ventriculaire et d'une diastole ventriculaire forme un cycle cardiaque. Les chiffres (figure 1.6) indiquent en mm de mercure les pressions régnant dans les cavités.

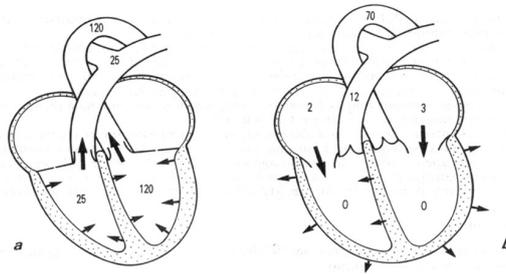


FIGURE 1.6 – Cycle cardiaque (mouvements des parois et des valves (flèches))

source : [20]

3 Electrocardiographie

Une fibre cardiaque en cours de dépolarisation peut être assimilée à un dipôle de courant. A un instant donné, le front de l'onde d'activation formé par l'ensemble des dipôles élémentaires crée un champ électrique qui est fonction des moments dipolaires. L'enregistrement de l'évolution temporelle du champ électrique résultant, effectué au moyen d'électrodes cutanées, se nomme l'électrocardiogramme de surface.

L'apparition de l'électrocardiographie, il y a une centaine d'années, coïncide avec la création du premier système d'enregistrement suffisamment sensible pour mesurer les potentiels électriques cardiaques à partir de la surface du corps. Ce système fut réalisé en 1903 par Willem Einthoven, physiologiste néerlandais de Leyde (1860-1927), considéré comme le père de l'électrocardiographie. Il décrit la succession des ondes P, Q, R, S et T dans le signal électro cardiologique[21].

Il fut également le premier à découvrir certaines anomalies électrocardiographiques telles que les tachycardies et les bradycardies ventriculaires. Il reçut en 1924 le prix Nobel pour l'ensemble de son travail sur l'électrocardiographie. Le système d'enregistrement d'Einthoven, produit et commercialisé pour la première fois en 1908 par la «Cambridge Scientific Instrument Company of London», obtint un succès important et bien d'autres modèles furent développés par la suite. D'autres scientifiques ont activement contribué à l'évolution de l'électrocardiographie. Citons notamment Samojloff, Lepeschkin, Lewis et Wilson[22].

L'électrocardiographie est une procédure simple, rapide et sans douleur, dans laquelle les impulsions électriques du cœur sont amplifiées et enregistrées. Cet enregistrement, l'électrocardiogramme (également appelé électrocardiogramme), fournit des informations sur la partie du cœur qui déclenche chaque battement de cœur (le stimulateur cardiaque, appelé nœud sino-auriculaire

ou sinusal), les voies de conduction nerveuse du cœur, ainsi que la vitesse et le rythme. Du cœur. Parfois, l'ECG peut montrer que le cœur est élargi (généralement en raison d'une pression artérielle élevée) ou que le cœur ne reçoit pas suffisamment d'oxygène en raison d'un blocage dans l'un des vaisseaux sanguins qui alimentent le cœur (les artères coronaires)[22].

3.1 Electrocardiogramme

L'ECG est un enregistrement de l'activité électrique du cœur sur une période donnée. La mesure est faite par le placement stratégique d'électrodes sur la peau du patient qui mesurent les changements de polarisation, se produisant lorsque le cœur se contracte et se détend. La contraction provoque une dépolarisation et la relaxation qui en résulte provoque une repolarisation du tissu. Ces changements de polarisation sont ce que l'on peut voir comme des ondes sur l'ECG [23]

Les différentes parties de l'ECG

Le signal ECG se présente comme une série d'ondes électriques, aux formes et durées particulières qui se répètent à chaque cycle cardiaque. En réalité ces ondes traduisent les différents phénomènes mécaniques relatifs au parcours du potentiel d'action de la stimulation cardiaque et dont les étapes sont successives comme illustré sur la figure 1.7.

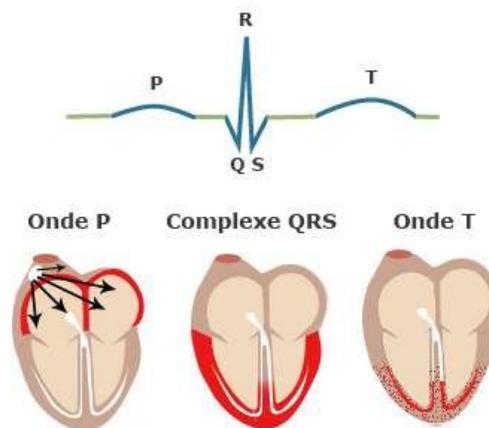


FIGURE 1.7 – Étapes successives de dépolarisation/repolarisation du cœur qui se traduisent sur le plan

source : [24]

Dans un signal électrocardiographique ECG que le processus de contraction et de décontraction du myocarde se présentent comme une séquence de déflexions positives et négatives superposées à une ligne de potentiel zéro (ligne de base) qui correspond à l'absence des phénomènes cardiaques comme illustré dans la figure 1.8 Par convention, on attribue aux ondes principales de l'ECG les lettres P, Q, R, S, T :

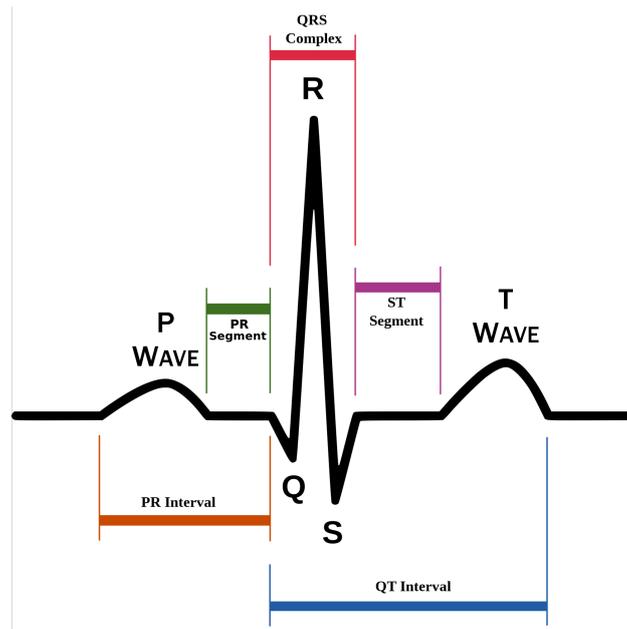


FIGURE 1.8 – L'ECG d'un battement de cœur

source : [25]

L'onde P : elle représente la dépolarisation auriculaire. Sa durée est de l'ordre de 90 ms. C'est une onde positive dont l'amplitude est normalement inférieure ou égale à 0.2 mV [22].

-Le complexe QRS : il correspond à la dépolarisation ventriculaire précédant l'effet mécanique de contraction. Sa durée normale est comprise entre 85 et 95 ms.

- L'onde T : elle correspond à la repolarisation des ventricules. L'onde T normale a une amplitude plus faible que le complexe QRS.

-L'intervalle PR 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).

La phase de repolarisation auriculaire n'apparaît pas sur l'ECG car elle est noyée dans le complexe QRS.

3.2 Les dérivations

Les dérivations périphériques bipolaires standards

Les dérivations I, II et III sont des dérivations bipolaires qui explorent l'activité cardiaque dans le plan frontal [23]. Ce système de référence est schématisé par un triangle équilatéral « triangle d'Einthoven », dont les sommets représentent les localisations des électrodes illustré sur la figure 1.9. Les trois électrodes sont placées respectivement : au bras droit pour R, au bras gauche pour L et à la jambe gauche pour F. On considère que tous les vecteurs résultants instantanés ont pour origine commune le centre du triangle équilatéral et on recueille leurs projections sur les côtés de ce triangle en mesurant les différences de potentiel entre ses sommets :

$$DI = VL - VR \quad DII = VF - VR \quad DIII = VF - VL \quad (1.1)$$

Théoriquement, on a la relation :

$$DI + DIII = DII \quad (1.2)$$

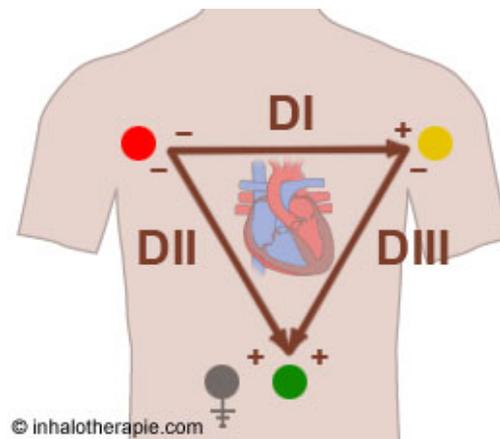


FIGURE 1.9 – triangle d'Einthoven. Le corps est supposé avoir une configuration

source : [26]

Les dérivations périphériques unipolaires

En 1934, Wilson introduit les dérivations unipolaires. Dans ce cas, il mesure la différence de potentiel entre un point de référence et chacun des points R, L et F. Dans le système dit de Wilson, ce point de référence appelé (borne centrale de Wilson) est virtuel, il est supposé demeurer à un potentiel invariable et de valeur pratiquement nulle. Goldberger propose en 1942, les (dérivations unipolaires augmentées des membres) [23] (aVR, aVL, aVF). Celles-ci permettent d'obtenir des signaux de plus grande amplitude que ceux du système de Wilson. Ces dérivations mesurent la différence de potentiel entre chacun des trois points et le potentiel moyen des deux autres.

$$aVR = V_R - \frac{V_L - V_F}{2} = 3 \frac{V_R}{2} \quad (1.3)$$

$$aVL = V_L - \frac{V_F - V_R}{2} = 3 \frac{V_L}{2} \quad (1.4)$$

$$aVLF = V_F - \frac{V_R - V_L}{2} = 3 \frac{V_F}{2} \quad (1.5)$$

Le système d'électrodes standard utilisant 12 dérivations est finalement constitué de l'ensemble des dérivations : aVR, aVL, aVF, DI, DII, DIII et V1 à V6.

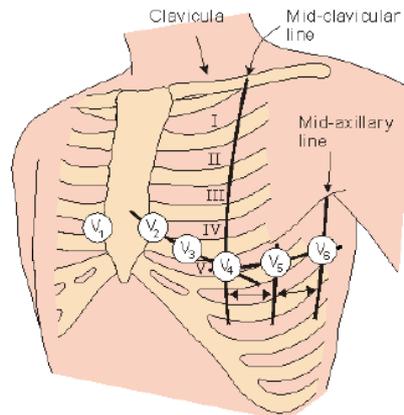


FIGURE 1.10 – positions d'électrodes recommandées pour l'enregistrement du système de dérivations unipolaires précordiales (V1-V6).

source : [28]

4 La commande rythmique du cœur

Le cœur est un muscle caractérisé par une activité automatique, spontanée et régulière. Le cœur isolé bat à une fréquence qu'on appelle (la fréquence cardiaque idiopathique)[18]. Elle est en moyenne de 70 battements par minute chez l'adulte. Cette fréquence diminue avec l'âge et l'entraînement physique. Sur le cœur normal, le nœud sinusal est soumis à une régulation extra cardiaque qui a pour effet de réduire la fréquence cardiaque au repos et de l'augmenter au cours de l'effort physique. Cette régulation est assurée en grande partie par le système nerveux autonome.

4.1 Le système nerveux autonome

Le système nerveux autonome joue un rôle clé dans la régulation de l'activité cardiaque, de la fréquence, de la force des battements cardiaques, de la pression artérielle (dilatation ou vasoconstriction des vaisseaux sanguins). Il comprend deux systèmes d'effet inverses :

Le système parasympathique

C'est l'élément dominant de la régulation de la fréquence cardiaque chez l'homme. Il permet le ralentissement de la fréquence cardiaque.

Le système sympathique

Il est surtout relié aux processus qui impliquent une dépense d'énergie. Lorsque l'organisme est en homéostasie, la fonction principale du système sympathique est de combattre les effets du système parasympathique. A l'inverse de la situation au repos, lors d'une tension extrême par exemple, le système sympathique domine le système parasympathique, surtout dans des situations de stress.

5 Rythme cardiaque physiologique.

La fréquence cardiaque est l'un des signes vitaux ou indicateurs importants de la santé du corps humain. Il mesure le nombre de fois où le cœur bat par minute. La vitesse du rythme cardiaque varie en fonction de l'activité physique et des réactions émotionnelles, mais pour examiner la santé d'une personne, on se réfère généralement à la fréquence cardiaque au repos, c'est-à-dire lorsque la personne est détendue. Le rythme cardiaque normal (figure 1.11) de quelqu'un dépend de facteurs comme l'âge de la personne et de son

mode de vie actif [18].

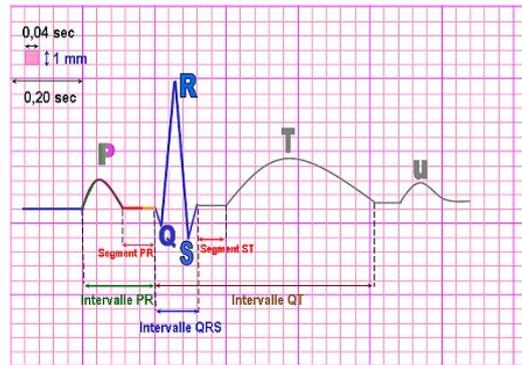


FIGURE 1.11 – ECG D'un rythme cardiaque normal

source : [29]

6 Les troubles de rythme ou les arythmies cardiaques (rythme cardiaque à l'état pathologique)

6.1 Qu'est-ce qu'une arythmie cardiaque ?

Au repos, le cœur bat normalement entre 60 et 70 fois par minute, jusqu'à 180 fois lors d'efforts physiques ou de stress psychique, en fonction de l'âge. Chaque jour, le cœur bat en moyenne 100'000 fois et à chaque battement, il pompe environ 60 ml de sang dans le corps, soit 6000 litres par jour, lui fournissant ainsi l'oxygène et les nutriments dont il a besoin[35].

Pour assurer cette fonction de pompe, il faut que les ventricules se contractent et se détendent de manière synchronisée. Le rythme provient du nœud sinusal, situé en haut de l'oreillette droite, qui représente le métronome naturel de ce phénomène. Il produit des impulsions électriques qui sont transmises au nœud auriculo-ventriculaire ou nœud AV. À partir de là, l'impulsion électrique passe par le faisceau de His et d'autres fibres conductrices pour aller stimuler le myocarde de manière coordonnée : les ventricules se contractent et éjectent le sang dans l'artère pulmonaire et dans l'aorte[10].

On parle de trouble du rythme cardiaque (arythmie) lorsque le cœur bat trop lentement , trop vite ou irrégulièrement. Si l'impulsion électrique est émise

régulièrement dans le nœud sinusal mais transmise trop lentement ou trop rapidement aux ventricules

les facteurs de risque des arythmies cardiaques

De nombreux facteurs peuvent affecter le rythme cardiaque, tels qu'une crise cardiaque, le tabagisme, les drogues[30], des malformations cardiaques congénitales et le stress. Certaines substances ou médicaments peuvent également provoquer des arythmies[31]

Les symptômes d'arythmie cardiaque

- Rythme cardiaque rapide ou lent [31]
- Sauter des temps
- Étourdissements ou vertiges
- Douleur à la poitrine
- Essoufflement
- Transpiration

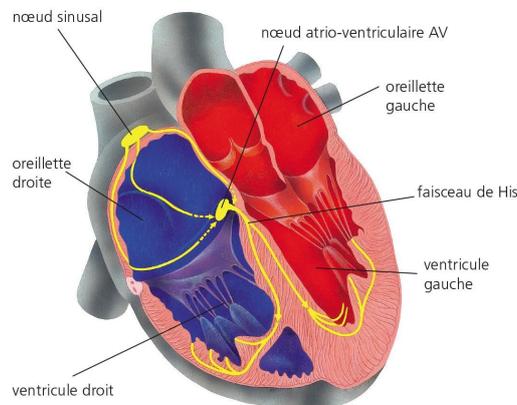


FIGURE 1.12 – système de conduction cardiaque

source : [32]

Il existe des troubles du rythme cardiaque très divers [10]. Les cardiologues distinguent entre autres deux grands groupes : d'une part les arythmies (les plus fréquentes) qui naissent dans les oreillettes. Celles-ci sont situées en amont des ventricules, d'où l'appellation d'arythmies supraventriculaires. Elles ne sont en général pas directement dangereuses. D'autre part les arythmies qui naissent dans les ventricules (arythmies ventriculaires). En l'absence

de traitement, celles-ci sont souvent des urgences aiguës qui peuvent entraîner un arrêt cardio-circulatoire et la mort subite cardiaque.

6.2 Les troubles du rythme supraventriculaire

Fibrillation auriculaire

La fibrillation auriculaire représentée à la figure 1.13 correspond au plus fréquent des troubles du rythme cardiaque. Elle se caractérise par une action non coordonnée des cellules auriculaires aboutissant à des battements des oreillettes cardiaques, irréguliers et rapides[33].

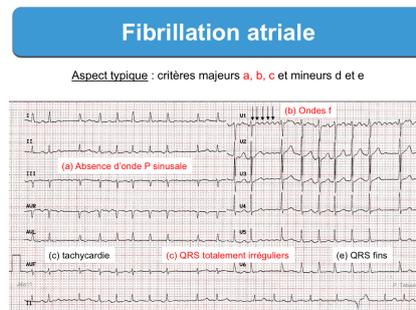


FIGURE 1.13 – ECG d'une Fibrillation auriculaire

source : [34]

Flutter auriculaire

Le flutter auriculaire est un rythme auriculaire rapide régulier secondaire à une macro circuit de réentrée intra-auriculaire. Les symptômes comprennent des palpitations et parfois de l'asthénie, une intolérance à l'effort, une dyspnée et un pré syncope. Des thrombi auriculaires peuvent se former et embolies. Le diagnostic repose sur l'ECG[35]. Le traitement implique un contrôle de la fréquence cardiaque par l'administration de médicaments, la prévention des accidents thromboemboliques avec des anticoagulants et, souvent, une conversion du rythme sinusal par les médicaments, la cardio-version, ou l'ablation du substrat de la fibrillation auriculaire.illustré sur la figure 1.14.

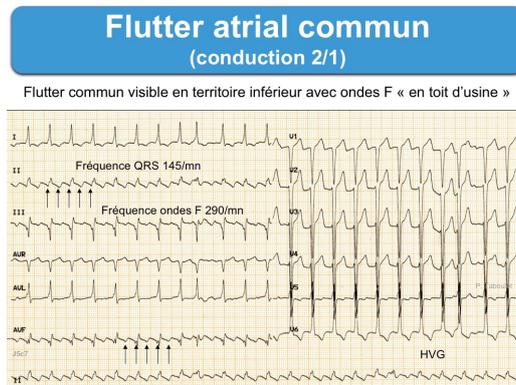


FIGURE 1.14 – Tracé ECG cas d'un Flutter auriculaire

source : [36]

Les troubles du rythme jonctionnels

Les tachycardies supraventriculaires par réentrée utilisent un circuit de réentrée dont une des voies au moins est située au-dessus de la bifurcation d'His.[35] Le patient a des épisodes paroxystiques...

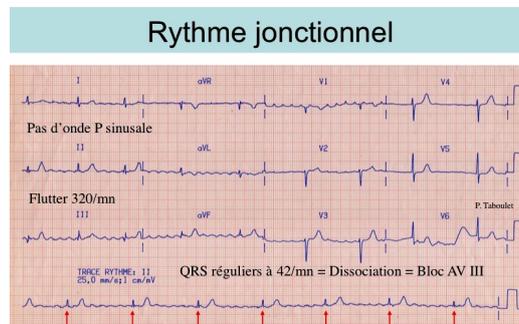


FIGURE 1.15 – ECG d'un trouble du rythme jonctionnels

source : [37]

6.3 Les troubles du rythme ventriculaire

Tachycardie ventriculaire

Le tachycardie ventriculaire correspond par définition au minimum supérieure ou égale à 3 extrasystoles consécutives à une fréquence supérieure ou

égale à 120 battements/minute. Les symptômes dépendent de la durée de la tachycardie...

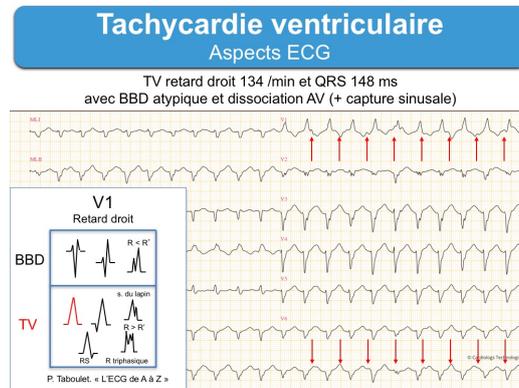


FIGURE 1.16 – ECG d'une Tachycardie ventriculaire

source : [38]

Fibrillation ventriculaire

La fibrillation ventriculaire est une arythmie ventriculaire totalement anarchique sans contraction cardiaque efficace[35]. Elle entraîne rapidement une perte de connaissance et le décès du patient...

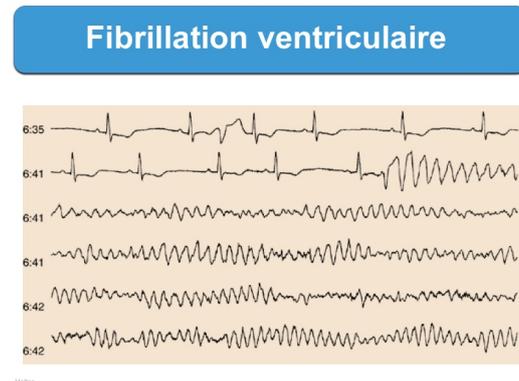


FIGURE 1.17 – ECG Fibrillation ventriculaire

source : [39]

7 Conclusion

La majorité des arythmies ne nécessitent pas une Prise en charge médicamenteuse, à moins qu'elles ne provoquent une décompensation hémodynamique voir un état de choc cardiogénique ou que l'arythmie soit symptomatique. Des études à long terme ont montré une augmentation de la mortalité chez les patients coronariens qui développent des troubles de rythme soutenues et non encore traités.

Le traitement peut conduire au retour d'un rythme de base normal (cardioversion pharmacologique), par différentes molécules comme L'amiodarone, la flécaïne ou les bêtabloquants pour réduire les troubles du rythme supraventriculaires et empêcher la récurrence d'épisodes d'arythmie, mais beaucoup plus pour empêcher l'apparition des complications cardiaques précoces et tardives.

Chapitre 2

État de l'art

1 Introduction

Les signaux physiologiques sont des signaux électriques délivrés par le corps humain. Parmi ces signaux, les plus couramment effectués pour la détection des maladies cardiovasculaires on distingue l'électrocardiogramme (ECG), qui est un signal électrophysiologie permettant de traduire l'activité électrique du cœur. La détection de ce type de signal peut être capté par des électrodes placées à la surface du corps.

Ce chapitre est consacré à la description des propriétés du signal de l'électrocardiogramme (ECG) à travers la base de données, afin d'extraire certains paramètres à travers le traitement numérique des différents signaux ECG de la base de données.

Dans ce chapitre, nous discutons : -traitement de signal,
-la description de la base de données,
-les travaux reliés,
-le travail proposé.

2 Traitement de signal

L'informatique joue un grand rôle dans la reconnaissance des différents cas pathologiques (arythmies cardiaques) du cœur à partir des signaux ECG à travers plusieurs étapes :

-l'extraction des données (paramètres) nécessaires, ou les paramètres qui définissent chaque partie de signal,
-la ressemblance des données dans une base de données,
-la classification : regrouper chaque cas dans une classe à partir des paramètres,

-l'évaluation des résultats.

La figure 2.1 illustre un battement ECG standard, plus précisément le rythme cardiaque normal à quatre paramètres : une onde P, qui représente la dépolarisation auriculaire, un complexe QRS associé à la dépolarisation ventriculaire, une onde T, qui représente la repolarisation ventriculaire, et enfin une onde connectée à la repolarisation du muscle papillaire.

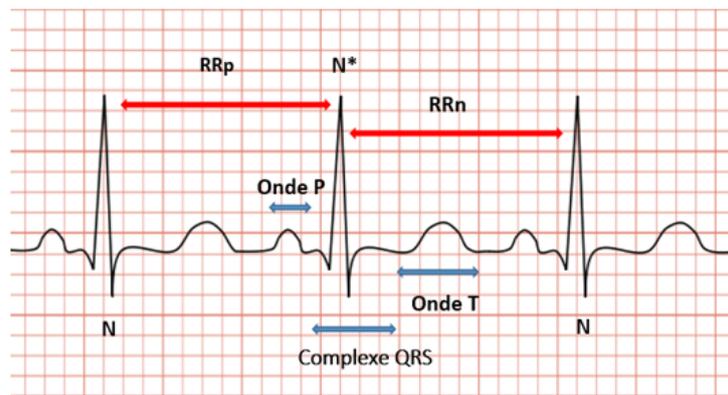


FIGURE 2.1 – battement ECG standard

source : [40]

3 la base de données

Un ensemble de données des clics de 22 patients différents ont été pris en compte dans les expériences. Chaque patient a au moins un cas pathologique qui peut aller jusqu'à plusieurs cas. Le nombre final est de 1032 cas montré par le tableau 2.1.

Dans cette base de données, le nombre de patients prenant part aux expériences pour fournir au moins différents cas d'arythmie présenté par le tableau 2.2 :

- Battement normal (N),
- Contraction ventriculaire prématurée (PVC),
- Bloc de branche droite (RBBB),
- Bloc de branche gauche (LBBB).

Cette base de données contient deux parties (matrices) :

yasrnsc : c'est la matrice sc sans classe avec patient et son arythmie par exemple yasrnsc100N : c'est la matrice sc sans classe avec patient 100 et l'arythmie N Cette matrice contient 11 descripteurs (paramètres)

yasrnc :c'est la matrice classe qui sont 4 classes montré dans le tableau 2.3.

Enregistrements	Normal	PVC	RBBB	LBBB
100	62	0	0	0
101	5	0	0	0
103	58	0	0	0
105	10	0	0	0
106	27	34	0	0
109	0	0	0	129
111	0	0	0	41
113	6	0	0	0
115	10	0	0	0
116	45	0	0	0
118	0	0	12	0
119	50	34	0	0
122	5	0	0	0
123	5	0	0	0
124	0	0	33	0
200	0	25	0	0
203	0	15	0	0
207	0	0	0	40
208	0	152	0	0
212	5	0	0	26
214	0	50	0	50
215	103	0	0	0

TABLE 2.1 – Données d'évaluation extraites de la base de données

classe	Normal	PVC	RBBB	LBBB	Total
Nombre d'échantillons	391	310	45	286	1032

TABLE 2.2 – Détails du jeu de données de la base de données

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

TABLE 2.3 – les paramètres ou descripteurs de la base de données

4 Travaux reliés

Les approches existantes en matière de détection automatique de l'arythmie cardiaque utilisent différentes machines et méthodes d'apprentissage, y compris les machines à vecteurs de support [43], Arbres de décisions [42], Réseaux de neurones artificiels[41]-[1], classificateurs flous [48], ensemble de modèles d'apprentissage automatique [44], règles de partition semi automatiquement floues [45],régression logistique polynomiale [46]-[47] , et autres recherches et études dans le domaine de médecine, pour en nommer quelques-unes.

Acharya et ses collaborateurs (2003) [41] ont étudié l'utilisation des réseaux de neurones artificiels (RNA) formés via propagation et un classificateur fuzzy pour classer automatiquement les signaux de fréquence cardiaque en 8 classes différentes (c.-à-d. contraction normale pré-ventriculaire, bloc cardiaque complet, syndrome des sinus, syndrome du faisceau gauche) cardiomyopathie ischémique / dilatée, fibrillation auriculaire et fibrillation ventriculaire).Pour l'évaluation, ils ont utilisé la base de données sur les arythmies MIT-BIH et ont échantillonné 1 000 cas. Les résultats obtenus ont montré que la machine proposée par l'apprentissage donne des modèles qui obtiennent une précision de classification d'environ 80% à 85%, avec le classificateur flou légèrement surperformant, le ANN.

Benchaib et ses collaborateurs (2014),[1] ont introduit une méthode intelli-

gente pour classer avec précision le rythme cardiaque des signaux ECG par le biais de la plastification artificielle du multi épaisseur et du perceptron multiple (AMMLP). La base de données MIT-BIH est utilisée pour classer les arythmies en trois types différents : Contraction ventriculaire prématurée (PVC), bloc de branche droit(RBBB) et bloc de branche gauche (LBBB) ; les signaux ECG normaux sont également utilisés dans l'étude. La précision de classification AMMLP obtenue de 98,25% est un excellent résultat par rapport au MLP classique et aux techniques de classification récentes appliquées à la même base de données.

Exarchos et ses collaborateurs (2007) [42] ont mis au point un modèle hybride d'apprentissage automatique intégrant la prise de décisions arborescents avec un classificateur flou : le modèle hybride utilise d'abord des arbres de décision pour extraire automatiquement des règles d'association entre les entités en entrées, qui sont ensuite utilisées comme entrées au classificateur flou. Le modèle de classification hybride a été évalué sur deux tâches étroitement liées, à savoir la détection de battement ischémique et arythmique. Les résultats expérimentaux ont indiqué que le modèle hybride de classification pour la et détection de battement arythmique atteint une précision de classification très élevée de 92% et 96%.

Tsipouras et al(2004) [43] ont présenté un ensemble de réseaux de neurones pour l'arythmie automatique. Dans leur approche, ils ont utilisé plusieurs réseaux de neurones formés de différentes combinaisons de fonctions de fréquence cardiaque. Un ensemble de règles construites manuellement a ensuite été utilisé pour combiner la sortie des réseaux de neurones. Les expériences menées ont montré que le réseau de neurones d'ensemble obtenait de solides performances de 87% de sensibilité et de 89% de spécificité. En outre, les auteurs ont démontré que le modèle proposé est capable d'identifier une large gamme de types d'arythmies. Une limitation du modèle de réseau neuronal d'ensemble est que la performance de la classification dépend du réglage fin d'un grand nombre d'hyper paramètres (par exemple, nombre de couches cachées, dimensionnalité des couches cachées, nombre d'époques d'entraînement, etc.). De plus, les réseaux de neurones ne sont pas facilement interprétables experts médicaux en raison du fait que les couches masquées ne sont pas mappées linéairement sur les entités en entrée.

Tsipouras et ses collaborateurs (2004) [44] ont présenté un ensemble de réseaux de neurones pour l'arythmie automatique. Dans leur approche, ils ont utilisé plusieurs réseaux de neurones formés de différentes combinaisons de fonctions de fréquence cardiaque. Un ensemble de règles construites manuellement a ensuite été utilisé pour combiner la sortie des réseaux de neurones. Les expériences menées ont montré que le réseau de neurones d'ensemble obtenait de solides performances de 87% de sensibilité et de 89% de spécificité.

En outre, les auteurs ont démontré que le modèle proposé est capable d'identifier une large gamme de types d'arythmies. Une limitation du modèle de réseau neuronal d'ensemble est que la performance de la classification dépend du réglage fin d'un grand nombre d'hyper paramètres (par exemple, nombre de couches cachées, dimensionnalité des couches cachées, nombre d'époques d'entraînement, etc.). De plus, les réseaux de neurones ne sont pas facilement interprétables experts médicaux en raison du fait que les couches masquées ne sont pas mappées linéairement sur les entités en entrée.

Behadada et ses collaborateurs (2015) [45] ont introduit une nouvelle méthode pour définir des règles de partition semi automatiquement floues de fins d'aperçu puissant et précis de l'arythmie cardiaque. En particulier, L'évaluation effectuée montre un taux de précision de 93% et une interprétabilité de 0,646, ce qui montre clairement que cette méthode offre un excellent équilibre entre précision et précision. Et la transparence du système. De plus, cela contribue considérablement à la découverte des connaissances et offre un outil puissant pour faciliter le processus de prise de décision, Utilisé dans sa classification cinq classes (RR précédent : RR0, RR suivant : RRn, complexe QRS, comp : Le rapport RR0 / RRs, PP, Energie du complexe QRS).

Behadada et ses collaborateurs (2016 a et b) [46]-[47] ont introduit une méthode basée sur la régression logistique multi-classes (c'est-à-dire les classes suivantes RR précédent : RR0, RR suivant : RRn, complexe QRS, comp : Le rapport RR0 / RRs, PP, Energie du complexe QRS) en tant que classificateur pour fournir un aperçu puissant et précis de l'arythmie cardiaque, l'un des facteurs prédictifs des maladies cardiovasculaires. Cette évaluation, fournit un système robuste, évolutif et précis, qui peut gérer avec succès les défis posés par l'utilisation du Big Data dans le secteur médical. Les expériences menées ont montré que la régression logistique multinomiale obtenait des solides performances de 92% de sensibilité et de 94% de spécificité et d'un meilleur taux de classification.

Il y a aussi d'autres études dans le domaine de la santé : Précision de la synchronisation d'intervalle d'impulsion dans la pression artérielle ambulatoire mesure [49], Lésion hépatique d'origine médicamenteuse : vers une prédiction précoce et le risque stratification [50].

5 Travail proposé

A partir de la programmation en Matlab 2015b, on classifera la base de données concerné par le classificateur de régression logistique multinomiale (multi classes) divisé en deux phases : 2/3 d'une phase de L'apprentissage et 1/3 d'une phase de test pour la détection des facteurs de risque des arythmies cardiaques en plusieurs étapes :

Phase d'apprentissage, phase de test et l'évaluation de résultats,

On prend juste 3 classes Normal, RBBB, LBBB pour la classification.

6 Conclusion

Dans ce chapitre nous avons présenté des notions générales sur les propriétés des signaux ECG de la base de données. Le traitement numérique des signaux est devenu une étape essentielle dans le diagnostic et le suivi des troubles ventilatoires dans l'orientation thérapeutique ainsi que la consultation pré-anesthésique. Le chapitre suivant est consacré à la présentation de classificateur de régression logistique .

Chapitre 3

régression logistique

1 Introduction

M. Georges Buffon, Histoire naturelle [51] , 1749 dit : « Le seul moyen de faire une méthode instructive et naturelle, est de mettre ensemble les choses qui se ressemblent et de séparer celles qui diffèrent les unes des autres . »

Évidemment, le processus général selon leur classification dans le domaine informatique est d'essayer de l'appliquer à des données numériques (points, tableaux, images, sons, etc.) et à ne pas exclure la règle imposée par ce célèbre naturaliste George Buffon.

Le travail général des méthodes de classification depuis 1749 est la tradition et l'automatisation, le principe d'utilisation et d'invention de moyens suffisants (calculatrices, théories classificatoires... Etc.).

Partons de ce principe, nous présentons dans ce chapitre des généralités sur la classification et la régression logistique multinomiale en particulier.

2 Généralité sur la classification

2.1 La classification

La classification est une discipline reliée de plusieurs domaines (la santé, l'industrie... etc.), elle est connue aussi sous plusieurs noms différents (classification, clustering, segmentation,...) en fonction des choses qu'il traite et des objectifs qu'il vise à atteindre.

Pour attribuer une définition au terme « classification », il faudrait d'abord définir ses racines, ça vient du verbe “classer” qui désigne plus une action qu'un domaine, ou plutôt une série de méthodes qu'une théorie unifiée [52].

En mathématique, On appelle classification, la catégorisation algorithmique d'objets. Elle consiste à attribuer une classe ou catégorie à chaque objet (ou individu) à classer, en se basant sur des données statistiques. Elle fait couramment appel aux méthodes d'apprentissage et est largement utilisée en reconnaissance de formes.

Il est important de noter qu'il ne faut pas confondre entre ces deux termes : « classification » et « classement ».

Dans un classement on affecte les objets à des groupes préétablis, c'est le but de l'analyse discriminante, qui est de fixer des règles pour déterminer la classe des objets. La classification est donc, en quelque sorte, le travail préliminaire au classement, savoir la recherche des classes "naturelles" dans le domaine étudié, en anglais « Cluster Analysis ».

La classification automatique est la répartition ou le regroupement d'un ensemble donné de n observations en différents groupes, catégories, classes, taxons, clusters de façon à regrouper les données similaires et à séparer les données dissimilaires.

2.2 Approche Paramétrique versus non-paramétrique

Non paramétrique

Les approches dites non paramétriques (classification hiérarchique, méthode des centres mobiles) basée sur l'hypothèse : plus deux individus sont proches, plus ils ont de chances de faire partie de la même classe, en plus ce que distingue cette approche est qu'on ne fait aucune hypothèse sur le modèle que suit les données, C'est le cas des plus proches voisins (k-PPV), donc Il suffit donc de rechercher des propriétés de proximité lorsque le nombre de données est grand.

Paramétrique : « Probabilistes »

C'est la seconde grande famille des méthodes de classification, ce sont les approches probabilistes, utilisant une hypothèse sur la distribution des individus à classer, c'est-à-dire, on suppose que l'on connaît la forme du modèle qui a généré les données. Par exemple, on peut considérer que les membres de chaque catégorie sont conformes au droit commun. Le problème se pose, le savoir-faire ou l'appréciation des lois (moyenne, variance) et la catégorie d'individus plus susceptibles d'appartenir au plein apprentissage.

On peut classer les méthodes classificatoires en deux grands types, cette

fois-ci, on prend en considération l'intervention ou non d'un « attribut classe » au fur et à mesure du processus de la classification, ces deux types sont : « supervisée (Classement) » et « non supervisée (Classification, Clustering) ».

supervisé (classement) si les classes sont prédéterminées et les exemples connus[53], le système apprend à classer selon un modèle de classement ; on parle alors d'apprentissage supervisé. Un expert doit préalablement labéliser des exemples. Le processus se passe en deux phases. Lors de la première phase, il s'agit de déterminer un modèle des données étiquetées. La seconde phase consiste à prédire l'étiquette d'une nouvelle donnée, connaissant le modèle préalablement appris[60]. Parfois il est préférable d'associer une donnée non pas à une classe unique, mais une probabilité d'appartenance à chacune des classes prédéterminées (on parle alors d'apprentissage supervisé probabiliste).

non supervisé (classification) les classes sont inconnues, Apprendre une méthode pour prédire la classe d'un élément à partir d'éléments déjà classés[53].

Cependant, Il existe d'autres types de classification qui s'appuient sur d'autres types de méthodes d'apprentissages comme « l'apprentissage semi-supervisé » et « l'apprentissage par renforcement ». En effet, l'apprentissage semi-supervisé est un bon compromis entre les deux types d'apprentissage « supervisé » et « non-supervisé », car il permet de traiter une des données, dont une bonne partie sont des données non étiquetées et peu de données étiquetées, et il profite des avantages des deux types mentionnés. Alors que L'apprentissage par renforcement est fort utilisé dans le cas d'apprentissage interactif.

3 Régression logistique

La régression logistique est une technique de classification supervisée permettant d'ajuster une surface de régression à des données lorsque la variable dépendante est dichotomique [54]. Il s'agit en fait de connaître les facteurs associés à un phénomène en élaborant un modèle de prédiction. La popularité de cette méthode est bien connue dans plusieurs domaines et sciences, surtout en médecine « science d'épidémiologie » [55]-[56] et en sciences humaines, où la variable à prédire est la présence ou l'absence d'une maladie.

Par exemple, il peut s'agir d'une étude sur la détection de l'arythmie cardiaque il faut connaître les facteurs, la prédisant le mieux, en étudiant des variables telles que le complexe QRS, l'onde P et tout ce qui est important dans l'enregistrement ECG.

3.1 Définition mathématique du modèle de régression logistique

La pratique de la régression logistique est très proche de celle de la régression linéaire[57]. La régression linéaire permet de caractériser les liens entre une variable à expliquer (Y) quantitative et des variables explicatives ($X_1, X_2, X_3, \dots, X_n$) au moyen du modèle présenté par la formule en équation 3.1. À l'évidence, ce modèle ne s'applique pas aux variables qualitatives et notamment binaires où Y s'exprime en termes de oui/non. Il est donc nécessaire d'utiliser un modèle adapté permettant de relier les variables explicatives à la variable qualitative (Y) à prédire. L'astuce de la régression logistique consiste non pas à modéliser la variable qualitative Y mais la probabilité que celle-ci se réalise. Le modèle logistique permet une expression non linéaire, variant de façon monotone entre 0 et 1, de cette probabilité en fonction des variables explicatives (X_i).

$$(p) + \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_n * X_n + \varepsilon \quad (3.1)$$

Modèle de régression linéaire multiple : la variable à expliquer (Y) est exprimée en fonction d'un intercept (ou ordonnée à l'origine) 0, des variables explicatives (X_i) rattachées à leurs coefficients β_i et à un terme de bruit [57] .

$$\ln \left(\frac{p}{1-p} \right) = \text{logit}(p) + \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_n * X_n + \varepsilon \quad (3.2)$$

le logit de la probabilité (p) de la réalisation de la variable à expliquer (Y) est exprimé en fonction d'un intercept (ou ordonnée à l'origine) β_0 , des variables explicatives (X_i) rattachées à leurs coefficients β_i et à un terme de bruit ε [57].

3.2 Intérêts de la régression logistique

Comme pour la régression linéaire et le modèle de Cox, le but de la régression logistique est de caractériser les relations entre une variable dépendante (ou variable à expliquer) et une seule (régression logistique simple) ou plusieurs variables prises en compte simultanément (régression logistique multiple). Il s'agit donc d'un modèle permettant de relier la variable dépendante

(Y) à des variables explicatives ($X_1, X_2, X_3, \dots, X_n$) [58].

À la différence de la régression linéaire (où la variable à expliquer est une variable quantitative) et du modèle de Cox (où la variable à expliquer est une variable censurée), la régression logistique s'applique lorsque la variable à expliquer (Y) est qualitative. Dans la recherche biomédicale, il est très fréquent de rencontrer de telles variables, le plus souvent de type binaire comme par exemple la présence (ou l'absence) d'une maladie, d'une récurrence ou d'une complication. Les variables explicatives (X_i) peuvent être, quant à elles, qualitatives ou quantitatives. Ces variables indépendantes sont susceptibles d'influencer la survenue ou non de la maladie, de la récurrence ou de la complication.

Dans le cas d'une variable explicative qualitative[?], une propriété très intéressante de la régression logistique est qu'elle permet d'estimer un odds ratio (OR) qui fournit une information sur la force et le sens de l'association entre la variable explicative (X_i) et la variable à expliquer (Y). L'OR (ou rapport des cotes), est une mesure de dépendance entre deux variables, il est toujours positif et compris entre 0 et $+\infty$. Lorsqu'il vaut 1, les deux variables sont indépendantes. Au contraire, plus l'OR est proche de 0 ou de $+\infty$, plus les variables sont liées entre elles. Si l'OR n'est pas un résultat aussi intuitif que le risque relatif, il présente l'avantage d'être utilisable quel que soit le design expérimental (cohorte ou cas témoin). Qui plus est, si la maladie étudiée est rare (prévalence inférieure à 5–10 %), l'OR devient une bonne approximation du risque relatif. En pratique, si l'OR est supérieur à 1, on parle de facteur de risque, et si l'OR est inférieur à 1, on parle de facteur protecteur.

Ainsi, la régression logistique tient compte de l'effet des autres variables X_i intégrées dans le modèle et permet de réaliser un ajustement de l'OR sur des covariables (on parle d'OR ajusté). La régression logistique constitue alors une méthode de choix pour rechercher et déterminer les facteurs de risque ou les facteurs protecteurs d'une maladie, tout en tenant compte des facteurs de confusion. Il faut cependant garder à l'esprit que l'identification d'une liaison entre une variable expliquée (Y) et une ou des variables explicatives (X_i) ne témoigne pas forcément d'un lien de causalité et il convient de garder une certaine prudence dans l'interprétation des résultats.

3.3 Les types de régression logistique

La régression logistique binaire

La régression logistique binaire est un type spécial de régression dans lequel la variable de réponse binaire est liée à un ensemble de variables ex-

plicatives, qui peuvent être discrètes et / ou continues. Le point important à noter ici est que, dans la régression linéaire, les valeurs attendues de la variable de réponse sont modélisées en fonction de la combinaison de valeurs prises par les prédicteurs. Dans la régression logistique, la probabilité ou les chances que la réponse prenne une valeur particulière sont modélisées en fonction de la combinaison de valeurs prises par les prédicteurs. Comme la régression, nous établissons une distinction explicite entre une variable de réponse et une ou plusieurs variables explicatives [59].

La régression logistique ordinale

La régression logistique ordinale s'applique lorsque la variable à expliquer possède trois ou plus de modalités qui sont ordonnées modéré, moyen ou fort [59].

La Régression logistique multinomiale

La régression logistique est très répandue pour les problèmes de prédiction ou d'explication d'une variable dépendante binaire (malade oui/non) à partir d'une série de variables explicatives continues, binaires ou binarisées. On parle dans cas de régression logistique binaire.

Lorsque la variable dépendante possède plusieurs catégories non ordonnées ($K > 2$), on parle de régression logistique multinomiale [61] (on parle aussi de Régression logistique polytomique à variable dépendante nominale). Elle est peu (ou moins) connue, pourtant cette configuration est finalement assez courante. De plus, elle est directement traitée par les autres méthodes d'apprentissage telles que l'analyse discriminante prédictive, les arbres de décision, etc.

La régression logistique multinomiale consiste à désigner une catégorie de référence, la dernière ($K^{\text{ème}}$) par exemple pour fixer les idées, et à exprimer chaque logit (ou log-odds) des $(K-1)$ modalités par rapport à cette référence à l'aide d'une combinaison linéaire des variables prédictives.

4 Modèle de régression logistique multinomiale proposé

L'algorithme d'apprentissage supervisé basé sur la régression logistique multinomiale vise à concevoir un classificateur capable de distinguer K classes, en utilisant les échantillons d'apprentissage L étiquetés, lorsque les vecteurs de caractéristiques sont donnés comme entrée pour la classification , ce algorithme comporte une phase d'apprentissage et un phase de test [47] , Les échantillons d'apprentissage L ayant des étiquettes de classe connues sont indiqués par :

$$D_L = \{(X_1, Y_1), \dots, (X_L, Y_L)\} \quad (3.3)$$

Qui s'appelle l'ensemble de formation et la distribution de classe postérieure Le modèle MLR est calculé à l'aide de l'estimation MAP des régresseurs [63]. En tant que :

$$P(y_1, = k|x_1, w) = \frac{\exp(w^{(k)}x_i)}{\sum_{k=1}^K \exp(w^{(k)}x_i)} \quad (3.4)$$

W^k est l'ensemble des régresseurs logistiques pour la classe k où $w = (w^{(1)T}, \dots, w^{(K-1)T})$ et la valeur de $w^{(K)}$ est généralement mis à zéro puisque le K^{ème} probabilité conditionnelle est trouvée en soustrayant la somme des régresseurs estimés de $(K - 1)$ classes à partir de l'unité. Représente la fonctionnalité vectrice sélectionnée pour la formation du modèle. la fonction de base radiale gaussienne (RBF) est représenté par l'équation suivante 3.5 :

$$K(x_i, y_j) \equiv \exp\left(\frac{-x_i - x_i^2}{(2\sigma^2)}\right) \quad (3.5)$$

Il représente les vecteurs d'apprentissage offrant une meilleure séparabilité des données dans l'espace transformé . La densité de probabilité postérieure de \mathbf{w} avec X_L et Y_L (ensemble de vecteurs caractéristiques dans les échantillons d'apprentissage étiquetés donnés) est représenté par l'équation 3.6 :

$$P(w|Y_L, X_L) \propto p(Y_L|X_L, w) (w|X_L) \quad (3.6)$$

l'utilisation de la maximisation de l'attente (EM) est montré par l'équation 3.7, les expressions pour l'estimation MAP de \mathbf{w} , qui maximise la probabilité des données de journal conditionnelles, est définie comme suit :

$$\hat{w} = \arg \max \{l(w) + \log p(w|X_L)\} \quad (3.7)$$

Où la fonction de log-vraisemblance de \mathbf{w} est exprimée en utilisant :

$$l(w) \equiv \log p(Y_L|X_L, w) \equiv \log \prod_{i=1}^L P(y_i, x_i, w) \equiv \sum_{i=1}^L (x_i^T w^{(y_i)}) - \log \sum_{j=1}^K \exp(x_i^T w^{(y_j)}) \quad (3.8)$$

Pendant la phase de test, les estimations des coefficients de régression $\mathbf{w}()$ sont entrées dans la modèle de régression logistique multinomiale permettant de calculer les densités de probabilité de classes postérieures de chaque vecteur de caractéristiques dans les classes K . L'étiquette de classe d'un vecteur de caractéristiques est déterminée à partir de l'indice correspondant à la probabilité de classe postérieure maximale du vecteur de pixel de test donné [64]-[65].

5 Conclusion

La régression logistique est une méthode vaste de classification et d'analyse multivariée puissante permettant d'obtenir une quantification de l'association entre une maladie étudiée et chacun des facteurs l'influençant, tout en tenant compte de l'effet simultané des autres facteurs. Et à partir de là, nous pouvons connaître le moment grâce au facteur de risque, ce qui nous permet d'éviter ces risques.

Chapitre 4

Résultats et discussion

1 Introduction

Nous présentons dans ce chapitre, notre étude expérimentale répartie en différentes parties appliquant, la technique de la régression logistique multinomiale à MATLAB pour la classification des cas pathologiques des maladies cardiovasculaires ; en trois classes différentes à partir des caractéristiques de chaque patient. La démarche adaptée pour notre travail est résumée dans les points suivants :

- préparation des données partielles à partir de notre base de données,
- classification « phase d'apprentissage et de test »,
- évaluation des performances.

2 Préparation de données

Dans cette section, nous prenons trois types « classes » d'arythmie cardiaque pour notre classification dans une nouvelle base de données, et l'élimination du quatrième type, ainsi que l'élimination de la maladie de type d'arythmie « Contraction ventriculaire prématurée (PVC) » qui est marqué par sa classe 2. Le nombre de patients est de 19, sur 22 patients qui ont 722 cas pathologiques montré dans le tableau 4.1

Les trois types choisis sont signifiés par les classes représentées par le tableau 4.2 :

On divise cette base de données en deux bases de données pour une classification supervisée : Une base d'apprentissage de la 1^{ère} phase d'apprentissage et une base de test de la 2^{ème} phase de test.

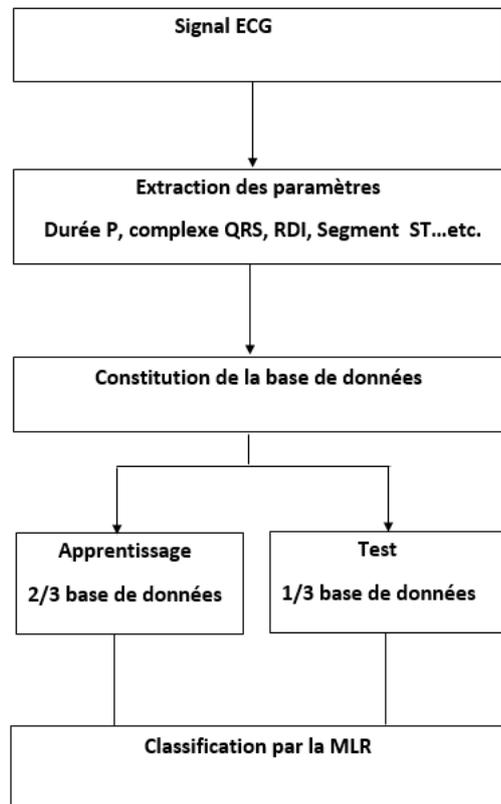


FIGURE 4.1 – plan de travail

classe	Normal	RBBB	LBBB	Total
Nombre d'échantillons	391	45	286	722

TABLE 4.1 – Détails du jeu de données, de la base de données après l'élimination de la deuxième classe

classe	Normal	RBBB	LBBB
signification	1	3	4

TABLE 4.2 – les types d'arythmie

3 Classification

elle divise en deux phase : apprentissage et test.

3.1 phase d'apprentissage

Dans cette partie, notre classificateur de la régression logistique multinomiale doit déterminer un modèle des données étiquetées. On prend 2/3 de données de la base préparée, c-à-dire pour chaque classe on prend les deux tiers 2/3 des échantillons de **yasrnsc**, la matrice sc sans classe avec patient et son arythmie (chapitre 02) et même travail pour le vecteur classe **yasrnc**.

3.2 Phase de test

Dans cette étape, notre classificateur de la régression logistique multinomiale doit prédire une classe de chaque échantillon à partir des échantillons de la phase précédente. On prend 1/3 de données de la base préparée, c-à-dire pour chaque classe prend les deux tiers 1/3 des échantillons de **yasrnsc**, la matrice sc sans classe avec patient et son arythmie (chapitre 02) et même travail pour le vecteur classe **yasrnc**.

3.3 La distribution des descripteurs en boîtes à moustaches

On représente en boîte à moustaches la distribution des différentes fonctionnalités (descripteurs) pour montrer que chaque descripteur est important ou non dans trois classes différentes qui sont : Normal, RBBB, LBBB

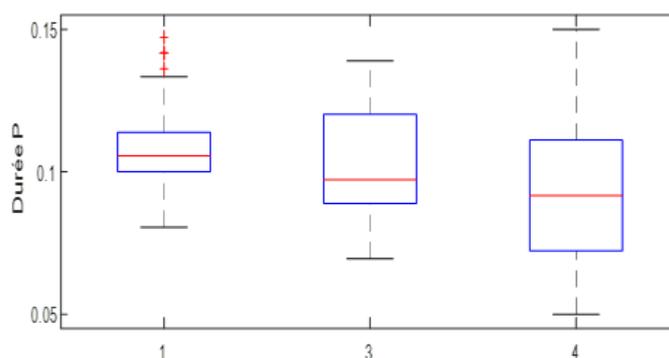


FIGURE 4.2 – la distribution de la fonctionnalité durée P dans différentes classes

D'après la figure 4.2, nous remarquons que le descripteur P n'a pas d'influence dans la classification des LBBB et RBBB, donc le descripteur P n'a

pas trop d'importance.

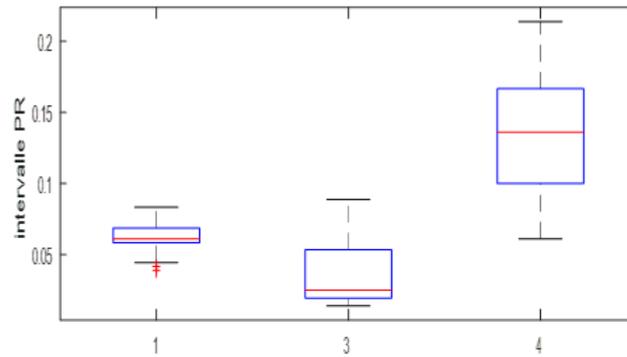


FIGURE 4.3 – la distribution de la fonctionnalité intervalle PR dans différentes classes

D'après les résultats de la boîte à moustaches illustré par la figure 4.3 nous constatons que le descripteur intervalle PR a d'influence dans la classification des LBBB et RBBB, donc le descripteur intervalle PR a trop d'importance.

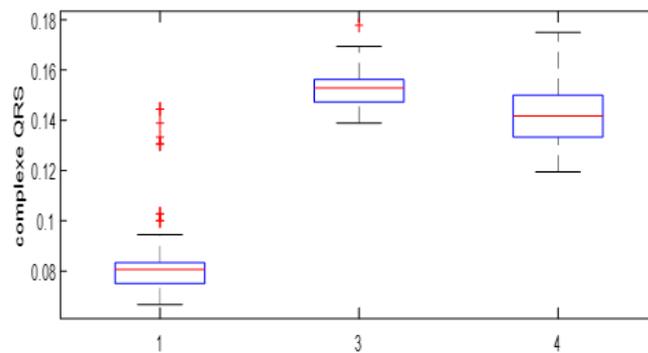


FIGURE 4.4 – la distribution de la fonctionnalité complexe QRS dans différentes classes

D'après la figure 4.4, nous remarquons que le descripteur complexe QRS a d'influence dans la classification des LBBB et RBBB, donc le descripteur complexe QRS a trop d'importance.

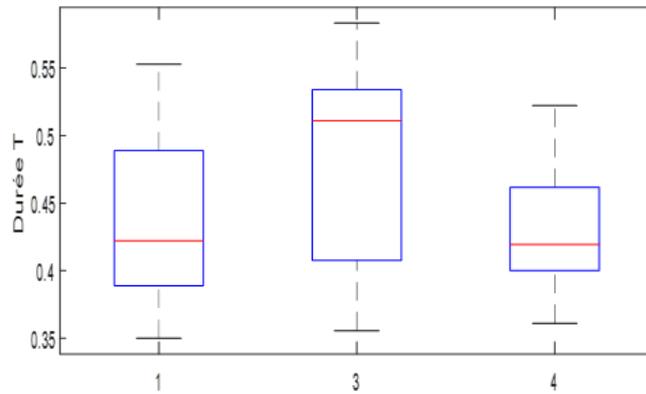


FIGURE 4.5 – la distribution de la fonctionnalité durée T dans différentes classes

D'après les résultats de la boîte à moustaches illustré par la figure 4.5 nous constatons que le descripteur T n'a pas d'influence dans la classification des LBBB et RBBB, donc le descripteur T n'a pas trop d'importance.

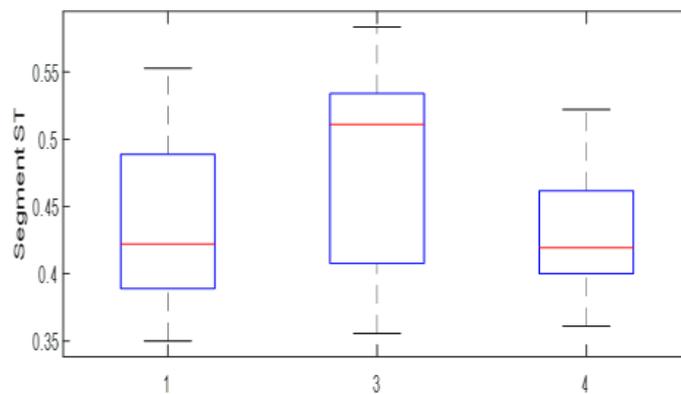


FIGURE 4.6 – la distribution de la fonctionnalité segment ST dans différentes classes

D'après la figure 4.6, nous remarquons que le descripteur ST n'a pas d'influence dans la classification des LBBB et RBBB, donc le descripteur ST n'a pas trop d'importance.

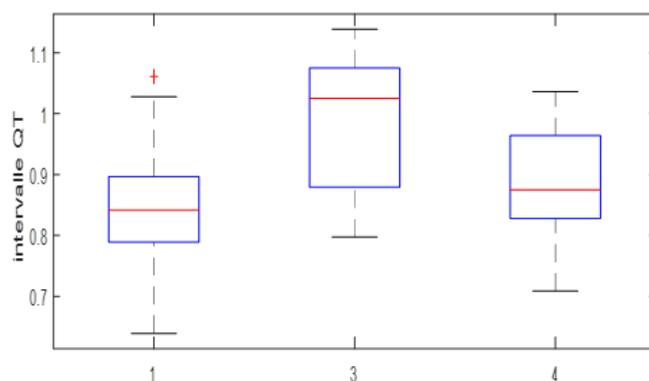


FIGURE 4.7 – la distribution de la fonctionnalité l'intervalle QT dans différentes classes

D'après les résultats de la boîte à moustaches illustré par la figure 4.7 nous constatons que le descripteur QT n'a pas d'influence dans la classification des LBBB et RBBB, donc le descripteur QT n'a pas trop d'importance.

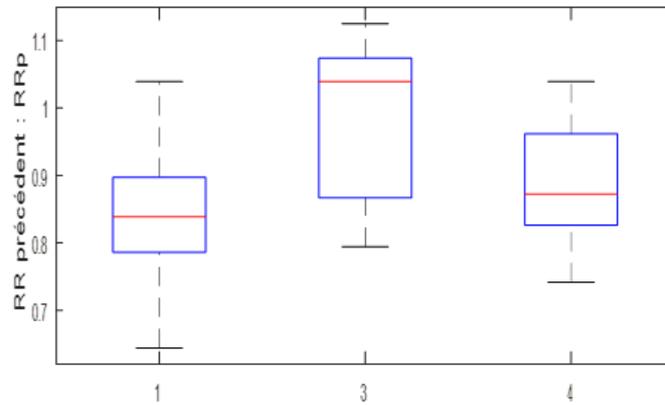


FIGURE 4.8 – la distribution de la fonctionnalité RRp dans différentes classes

D'après la figure 4.8, nous remarquons que le descripteur RRp n'a pas d'influence dans la classification des LBBB et RBBB, donc le descripteur RRp n'a pas trop d'importance.

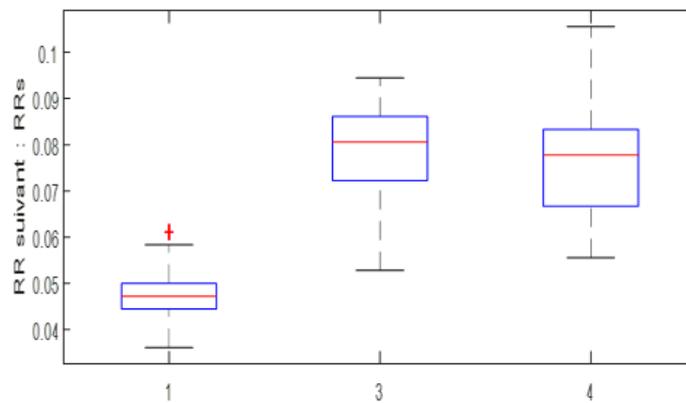


FIGURE 4.9 – la distribution de la fonctionnalité RRs dans différentes classes

D'après la figure 4.9, nous remarquons que le descripteur RRs a d'influence dans la classification des LBBB et RBBB, donc le descripteur RRp a trop d'importance.

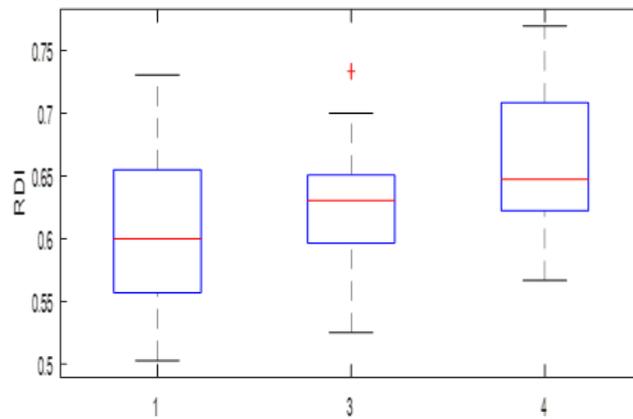


FIGURE 4.10 – la distribution de la fonctionnalité RDI battement dans différentes classes

D'après les résultats de la boîte à moustaches illustré par la figure 4.10 nous constatons que le descripteur RDI n'a pas d'influence dans la classification des LBBB et RBBB, donc le descripteur RDI n'a pas trop d'importance.

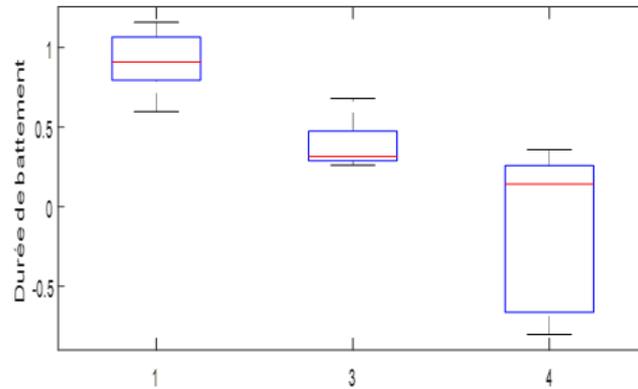


FIGURE 4.11 – la distribution de la fonctionnalité durée battement dans différentes classes

D'après la figure 4.11, nous remarquons que le descripteur durée battement a d'influence dans la classification des LBBB et RBBB, donc le descripteur durée battement a trop d'importance.

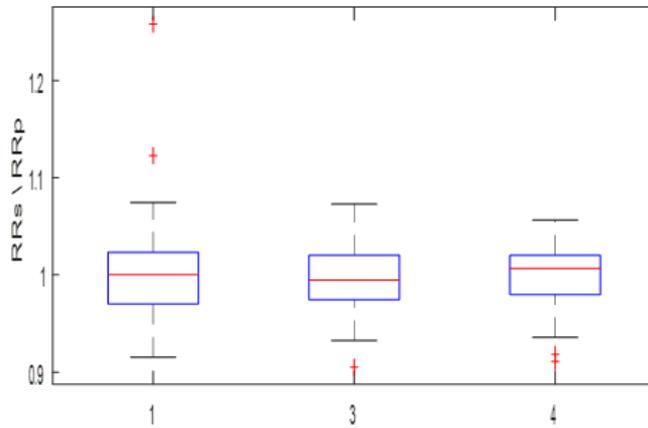


FIGURE 4.12 – la distribution de la fonctionnalité RRs/RRp dans différentes classes

D'après les résultats de la boîte à moustaches illustré par la figure 4.12 nous constatons que le descripteur RRs/RRp n'a pas d'influence dans la classification des LBBB et RBBB, donc le descripteur RRs/RRp n'a pas trop d'importance.

3.4 Étude de corrélation

Le but de cette section est d'évaluer toute relation existant entre les différentes caractéristiques. Le tableau 4.3 présente la distribution des échantillons. Pour évaluer les relations entre les entités en entrée d'évaluation et la catégorie de classification sous-jacente, nous avons examiné les propriétés de corrélation correspondantes. En particulier, le calcul des paramètres R est analysé individuellement, ainsi que dans les classes correspondantes. Cela a permis de déduire que les caractéristiques moyennes maximisant les relations au sein des classes correspondantes qui sont la Durée T et la durée du battement. Ceci confirme la pertinence de ces deux paramètres en ce qui concerne la détection des arythmies, suggérant également la présence de corrélation entre les caractéristiques, car la plupart d'entre eux sont liés au rythme cardiaque et à l'évolution temporelle du signal ECG.

Marquez les descripteurs de la base de données : Durée P, intervalle PR ,
complexe QRS, Durée T, segment ST , intervalle QT , RR précédent : RRp,
RR suivant : RRs, RDI , durée battement , RRs RRp avec les abréviations
respectivement :A1 ,A2 ,A3,A4 ,A5,A6,A7,A8 ,A9 ,A10 ,A11

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
A1	1	-0.6713	0.5578	0.6110	0.6703	0.6253	0.6419	0.5358	-0.0729	0.6678	-0.0134
A2	-0.6713	1	-0.6686	-0.8773	-0.7039	-0.6250	-0.6743	-0.7863	0.6539	-0.9830	0.0876
A3	0.5578	-0.6686	1	0.5653	0.6704	0.5123	0.5825	0.8442	-0.2247	0.6492	-0.1204
A4	0.6110	-0.8773	0.5653	1	0.8058	0.7151	0.7704	0.6977	-0.4019	0.9124	-0.0957
A5	0.6703	-0.7039	0.6704	0.8058	1	0.8027	0.8903	0.6854	0.0488	0.7524	-0.1577
A6	0.6253	-0.6250	0.5123	0.7151	0.8027	1	0.8826	0.5319	-0.0202	0.6935	0.2596
A7	0.6419	-0.6743	0.5825	0.7704	0.8903	0.8826	1	0.5928	-0.0122	0.7379	-0.2237
A8	0.5358	-0.7863	0.8442	0.6977	0.6854	0.5319	0.5928	1	-0.3964	0.7752	-0.1132
A9	-0.0729	0.6539	-0.2247	-0.4019	0.0488	-0.0202	-0.0122	-0.3964	1	-0.5873	-0.0103
A10	0.6678	-0.9830	0.6492	0.9124	0.7524	0.6935	0.7379	0.7752	-0.5873	1	-0.0770
A11	-0.0134	0.0876	-0.1204	-0.0957	-0.1577	0.2596	-0.2237	-0.1132	-0.0103	-0.0770	1

TABLE 4.3 – Coefficients de corrélation

4 Résultats et évaluation

Dans ce travail, nous avons choisi un modèle multinomiale hiérarchique, implémenté dans Matlab, qui était testé sur la base de données décrite ci-dessus, et les résultats sont représentés par la matrice de confusion dans Tableau 4.4. C'est un outil de mesure de la qualité d'un système de classification. Chaque colonne de la matrice représente le nombre d'occurrences d'une classe estimée. Tandis que chaque ligne représente le nombre d'occurrences d'une classe réelle (ou référence) la performance de la MLR .

Dans notre approche, nous adaptons l'algorithme MLR de sorte que la sortie ne soit pas seulement une répartition des probabilités, mais une classe. Diverses expériences ont été réalisées et nous avons obtenu un taux de classification de 81.50%, ce qui indique clairement que notre méthode proposée convient à ce type de données. En outre, il suggère de relever avec succès le défi de la classification des arythmies cardiaques, avec une sensibilité de 93.18% et une spécificité de 55.39%. Nous notons également que notre approche améliore le critère de transparence et d'interprétabilité du processus, en ayant un impact positif sur les résultats et le traitement des probabilités. En outre, la lisibilité des résultats a également été améliorée, aspect important du processus d'interprétation mené par des cardiologues experts, il est donc clair que notre approche fournit un avantage par rapport aux autres méthodes de classification.

Les résultats de la classification des trois types de battements cardiaques (N, RBBB et LBBB) par le MLR que nous avons implémenté sont illustrés sous forme d'une matrice de confusion illustrée dans le tableau 4.4 ci-dessous.

classe	Normal	RBBB	LBBB	Autres
Normal	84	0	1	0
RBBB	14	21	11	0
LBBB	1	0	0	0
Autres	0	0	23	0

TABLE 4.4 – matrice de confusion

Pour l'évaluation de la performance et la qualité de la classification du MLR, nous avons calculé le taux de classification, la spécificité et la sensibilité pour N et chaque type de pathologie (RBBB et LBBB). Ces résultats sont résumés dans tableau 4.5

Paramètres d'évaluation	Type de battements	Valeur (%)
Spécificité(%)	N	98.21
	RBBB	81.34
	LBBB	100
Sensibilité(%)	N	84.85
	RBBB	81.34
	LBBB	0
Taux de classification(%)	N	83.87
	RBBB	83.23
	LBBB	77.42
Taux de classification	Total	81.50

TABLE 4.5 – Résultats de Classification de la MLR pour les différents types de battements cardiaques

A partir des résultats résumés dans le tableau 4.4 , nous remarquons que la plupart des battements normaux N non identifiés sont ceux qui sont classés par la MLR comme des battements du type RBBB (14 battements au total). Aussi pour les battements du type RBBB sont bien classifiés. Pour les battement du type LBBB sont non identifiés.

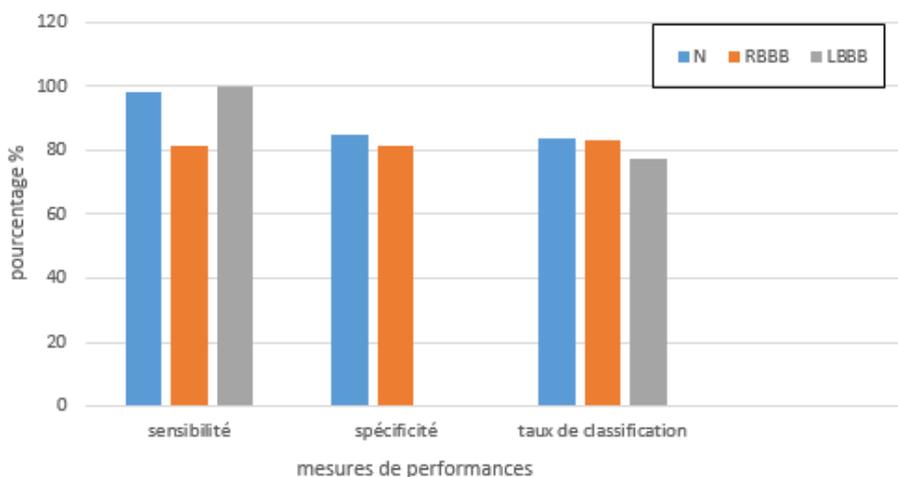


FIGURE 4.13 – histogramme de mesures de performances de la MLR.

5 Études comparatives avec d'autres techniques de classification

Cette section est destinée à fournir une étude comparative entre nos travaux et différentes méthodes largement utilisées dans la détection des arythmies cardiaques, basé sur la base de données étudiée. Dans notre étude, nous nous concentrons sur la robustesse, l'évolutivité, la précision et la fiabilité, en tant que paramètres objectifs principaux.

Le tableau 1.8 montre que notre méthode proposée présente un taux de classification de 81.50%. Les réseaux de neurones L'ARN et la carte de Kohonen [66] atteignent une performance de classification comparable de 95% et 90%, respectivement. Cependant, l'un des principaux inconvénients des modèles de réseau de neurones est qu'ils agissent comme des «boîtes noires» et ne permet pas une interprétation directe des résultats. La conséquence est que de telles méthodes ne facilitent pas une utilisation efficace par les cardiologues. En comparant notre méthode proposée avec les HMM (chaines de Markov cachés) et les systèmes d'inférence floue SIF [67].

Technique	Taux de classification (%)
HMM [67]	84.00
RNN [66]	70.00
MLR [47]	95.00
notre MLR	81.50

TABLE 4.6 – comparaison avec les techniques de classification de l'état de l'art.

5.1 Les mesures de performances

Les performances de classification des données sont évaluées par le calcul de critères suivants : vrais positifs (VP), vrais négatifs (VN), faux positifs (FP) et faux négatifs (FN) [1].

Le pourcentage de sensibilité (SE), la spécificité (SP) et le taux de classification (TC), leurs définitions respectives sont les suivantes :

- VP : Vrai Positif : nombre de positifs classés positifs.
- VN : Vrai Négatif : nombre de positifs classés négatifs.
- FP : Faux Positif : nombre de négatifs classés positifs.
- FN : Faux Négatif : nombre de négatifs classés négatifs.

La Sensibilité : c'est la capacité de donner un résultat positif quand la maladie est présente, elle est calculée par :

$$Se = \frac{VP}{VP + VN} * 100 \quad (4.1)$$

La Spécificité : c'est la capacité de donner un résultat négatif quand la maladie est absente, elle est calculée par :

$$Sp = \frac{VN}{VN + FP} * 100 \quad (4.2)$$

Le Taux de classification : c'est le pourcentage des exemples correctement classés, il est calculé par :

$$Tc = \frac{VP + VN}{VP + VN + FP + FN} * 100 \quad (4.3)$$

6 Conclusion

L'objectif de ce chapitre était d'appliquer les principes théoriques de la régression logistique multinomiale pour la classification des différentes arythmies cardiaques présentes dans notre base de données , et les comparer avec d'autres classifieurs appliqués dans le même domaine. les résultats obtenus par notre classifieur sont très prometteurs et peuvent être utilisés dans un système de détection automatique des arythmies cardiaques pour aider les médecins.

Conclusion générale

Dans ce mémoire, nous avons présenté une approche d'apprentissage automatique basée sur la régression logistique multinomiale (MLR), afin de détecter automatiquement les arythmies cardiaques. Les modèles de classification MLR offrent un avantage majeur pour la détection, le diagnostic et la prédiction de l'arythmie cardiaque, étant donné que dans le domaine médical, les experts ont besoin d'une aide au diagnostic automatique pour faciliter et justifier leurs décisions, ce qui tend à faire défaut dans plusieurs techniques citées dans la littérature, notamment les réseaux de neurones, support vector machine (SVM), k plus proches voisins, l'arbre de décision....etc. La contribution au processus de classification est démontrée par une précision de 81.50% et notre méthode offre une bonne flexibilité et une grande transparence dans le système de détection de l'arythmie cardiaque.

Annexe

Logiciel MATLAB

MATLAB, c'est un logiciel permettant d'effectuer des opérations complexes sur des matrices et vecteurs directement avec la syntaxe matricielle (presque la même écriture que dans les livres et publications). Matlab, c'est un interpréteur d'opérations matricielles. C'est aussi un langage de programmation de haut niveau. Matlab est écrit en C, mais il ne demande aucune connaissance du C de l'utilisateur.

Avec MATLAB, on peut effectuer tout ce que l'on peut faire avec un langage de programmation, mais beaucoup plus rapidement et efficacement. C'est un outil de recherche et d'application fantastique pour qui maîtrise bien l'écriture matricielle (et ses notions d'algèbre linéaire).

En Matlab, il y a 3 types de fichiers principaux :

- 1- fichiers « .m » : fichiers de commandes de Matlab
- 2- fichiers « .mat » : fichiers contenant des données (format spécifique à Matlab).
- 3- fichiers « .fig » : fichiers non-lisible contenant ce qu'il faut pour recréer une figure lors d'une séance ultérieure.

Les fichiers .mex sont des fichiers écrits dans une autre langage de programmation (fortran et C) et compilés et liés à Matlab. Outre cette distinction, ils jouent le même rôle que les fichiers « .m »

parmais les trucs et les conseils :

1. Matlab distingue les majuscules et minuscules. la matrice « M » et la matrice « m » ne sont donc pas la même entité pour Matlab. Règle générale j'évite l'utilisation des majuscules et je ne donne jamais le même nom (1 en majuscules et l'autre en minuscule) à deux entités différentes.
2. Il ne faut pas donner des noms aux variables qui correspondent à des noms de fonction (soit celles de Matlab, soit celles de l'utilisateur), sinon Matlab ne

peut plus accéder à ces fonctions.

3. J'utilise au maximum les fichiers « .m » (scripts et fonctions). Ceci sauve énormément de temps en bout de ligne. Rapidement, on doit apprendre à utiliser la page de base de Matlab comme une feuille de papier brouillon et réserver le travail sérieux aux fichiers « .m ».

4. J'insère les commentaires au fur et à mesure de l'écriture des fichiers « .m » et non seulement lorsque laprogrammation est terminée.

5. Il faut « vectoriser » le plus possible les opérations et éviter les boucles. Souvent, naturellement, on est porté à tout écrire sous forme de boucles. Ce n'est pas un problème pour de petites opérations mais un gain de temps considérable peut être obtenu en « vectorisant » ces opérations.

Il est préférable, surtout pour les grandes matrices remplies à l'intérieur de boucles, de les initialiser au début d'une procédure.

« Toolbox » (bibliothèque de fonctions) : Matlab vient avec un nombre impressionnant de fonctions de base. Le fabricant distribue aussi des bibliothèques de fonctions supplémentaires destinées à des usages spécifiques. Ces « toolbox » couvrent des domaines comme :

- analyse d'images
- statistique
- optimisation
- analyse de signal
- réseaux neuronaux
- analyse symbolique
- etc.

De plus, un grand nombre de fonctions et de toolbox sont disponibles gratuitement sur le WEB, y inclut sur le site de « mathworks » (www.mathworks.com). Ces fonctions sont parfois supérieures à celles du produit original. Il existe, entre autres, d'excellentes bibliothèques pour produire des cartes contours de meilleure qualité.

Bibliographie

- [1] BENCHAIIB, Yasmine et CHIKH, Mohamed Amine. Cardiac arrhythmias classification using artificial metaplasticity algorithm. *International Journal of Biomedical Engineering and Technology*, 2014, vol. 14, no 3, p. 209-224.
- [2] Dey, N., Prasad Dash, T. and Dash, S. (2011) ‘ECG signal denoising by Functional Link Artificial Neural Network (FLANN)’, *International Journal of Biomedical Engineering and Technology*, Vol. 7, No. 4, pp.377–389.
- [3] Richard, N. and Fogoros, M.D. (2013) Heart Health Center Guide. Available online at : <http://heartdisease.about.com/>.
- [4] World Health Statistics 2018 : Monitoring health for the SDGs. https://www.who.int/gho/publications/world_health_statistics/2018/en/
- [5] Alonso JM, Luis ML. An experimental study on the interpretability of fuzzy systems. *Proceedings of the Joint 2009 International Fuzzy Systems Association World Congress and 2009 European Society of Fuzzy Logic and Technology Conference, Lisbon, Portugal, 2009* ; 125–130.
- [6] Gacto MJ, Alcalá R, Herrera F. Interpretability of linguistic fuzzy rule-based systems : an overview of interpretability measures. *Information Sciences* 2011 ; 181(20) :4340–4360. doi :10.1016/j.ins.2011.02.021. Alonso JM. Interpretable fuzzy systems modeling with cooperation between expert and induced knowledge. PhD Thesis, 2007.
- [7] Rumelhart D, Hinton G, Williams R. Learning internal representations by error propagation. In *Parallel Distribution*
- [8] *Proceeding : Exploration in the Microstructure of Cognition, Foundations*, Vol. 1, Rumelhart D, McClelland J (eds). MIT Press : Cambridge, M.A, 1986 ; 318–362.
- [9] Fédération Française de Cardiologie. <https://www.fedecardio.org/Je-m-informe/Le-coeur/le-systeme-cardio-vasculaire>
- [10] Fondation Suisse de Cardiologie. Les arythmies cardiaques : Brochure d’information à l’intention du patient. <https://www.swissheart>.

- ch/fileadmin/user_upload/Swissheart/Shop/PDF_Broschueren/Herzrhythmusstoerungen_FR_web.pdf accéder le 12-02-2019.
- [11] HuguesJacobs. 28/12/2016.Futura Santé. <https://www.futura-sciences.com/sante/dossiers/medecine-coeur-organe-vie-1474/page/2/> accéder le 13-02-2019.
- [12] HuguesJacobs. 28/12/2016.Futura Santé. <https://www.futura-sciences.com/sante/dossiers/medecine-coeur-organe-vie-1474/> accéder le 13-02-2019.
- [13] TORTORA, Gerard J. et DERRICKSON, Bryan. Principles of anatomy & physiology. John Wiley & Sons, Incorporated, 2017.
- [14] irmcardiaque.<https://irmcardiaque.wordpress.com/physiopathologie-du-coeur/>. accéder le 16-02-2019.
- [15] HuguesJacobs. 28/12/2016.Futura Santé. <https://www.futura-sciences.com/sante/dossiers/medecine-coeur-organe-vie-1474/page/3/> accéder le 15-02-2019.
- [16] BHC "Brussels Heart Center". <http://www.brusselsheartcenter.be/fr/chirurgie-cardiaque/interventions/chirurgie-valvulaire-et-chirurgie-de-l-aorte-thoracique.html>. accéder le 18-02-2019.
- [17] Campus Cardiologie et maladies vasculaires.http://campus.cerimes.fr/cardiologie-et-maladies-vasculaires/enseignement/cardio_1/site/html/1.html. accéder le 25-02-2019.
- [18] Dr David ATTIAS . Pr Nicolas LELLOUCHE En collaboration scientifique avec le COLLEGE NATIONAL DES CARDIOLOGUES FRANÇAIS .cardiologie vasculaire 8^{ème} édition 2018.ISBN : 978-2-8183-1697-9
- [19] Campus Cardiologie et maladies vasculaires. http://campus.cerimes.fr/cardiologie-et-maladies-vasculaires/enseignement/cardio_1/site/html/2.html.. accéder le 26-02-2019.
- [20] Campus Cardiologie et maladies vasculaires.http://campus.cerimes.fr/cardiologie-et-maladies-vasculaires/enseignement/cardio_1/site/html/2_2.html .accéder le 28-02-2019.
- [21] CABASSON, Aline. Estimation et analyse des intervalles cardiaques. 2008. Thèse de doctorat. Université Nice Sophia Antipolis.
- [22] BOUCHIKHI, Sarra. Classification des Données Médicales par Les Modèles de Markov Cachés et La Logique Floue. Thèse de doctorat.
- [23] BENALI, Radhwane. Analyse du signal ECG par réseau adaptif d'ondelettes en vue de la reconnaissance de pathologies cardiaques. 2013. Thèse de doctorat.

- [24] Analyse du signal ECG par réseau adaptatif d'ondelettes en vue de la reconnaissance de pathologies cardiaques - Scientific Figure on ResearchGate. https://www.researchgate.net/figure/Etapes-successives-de-depolarisation-repolarisation-du-coeur-qui-se-traduisent-sur-le_fig3_281237494 . accéder le 09-03-2019.
- [25] Futura Santé. <https://www.futura-sciences.com/sante/definitions/medecine-electrocardiogramme-3353/> accéder le 09-03-2019.
- [26] Assistant Inhalo .<https://www.inhalotherapie.com/apprendre/electrocardiographie/derivations-peripheriques/> accéder le 11-03-2019.
- [27] TALBI, Mohamed Lamine. Analyse et traitement du signal électro cardiographique (ECG). 2011.
- [28] Modeling bipolar ECG signal strength with thorax models and validation of the modeling method - Scientific Figure on ResearchGate.https://www.researchgate.net/figure/Location-of-the-precordial-electrodes-V1-V6-in-the-standard-12-lead-system-ECG-recording_fig1_254558911 . accéder le 15-03-2019.
- [29] Le Repaire des Sciences.http://www.lerepairedessciences.fr/sciences/questions_sciences/ecg.htm accéder le 15-03-2019.
- [30] LAZZERINI, Pietro Enea, CAPECCHI, Pier Leopoldo, GALEAZZI, Mauro, et al. Biologic drugs and arrhythmic risk in chronic inflammatory arthritis : the good and the bad. Immunologic research, 2017, vol. 65, no 1, p. 262-275.
- [31] Arrhythmia (National Library of Medicine [NLM], MedlinePlus) <https://medlineplus.gov/arrhythmia.html> bibitem48 Heart Health Tests : MedlinePlus Health Topic From the National Institutes of Health (National Library of Medicine)
- [32] Fondation Suisse de Cardiologie.Les arythmies cardiaques :Brochure d'information à l'intention du patient. www.swissheart.ch examiné le 15-02-2019.
- [33] NIH "National Heart, Lung, and Blood Institute".<https://www.nhlbi.nih.gov/health-topics/arrhythmia>.accéder le 20-02-2019.
- [34] HUANG, Shoei K. Stephen et WOOD, Mark A. Catheter Ablation of Cardiac Arrhythmias E-book. Elsevier Health Sciences, 2019.
- [35] ZIPES, Douglas P., LIBBY, Peter, BONOW, Robert O., et al. Braunwald's Heart Disease E-Book : A Textbook of Cardiovascular Medicine. Elsevier Health Sciences, 2018.

- [36] PIRRE,Taboulet.ECG de A à Z.http://www.e-cardiogram.com/ecg-lexique_alpha.php?terme_lex=f&id_lex=192.accéder le 20-02-2019.
- [37] PIRRE,Taboulet.ECG de A à Z.http://www.e-cardiogram.com/?&id_quiz=51&rep=5&PHPSESSID=6a69f2088d0d876204abf527a949f93c.accéder le 20-02-2019.
- [38] PIRRE,Taboulet.ECG de A à Z.http://www.e-cardiogram.com/ecg-lexique_theme.php?id_th=5&id_lex=451.accéder le 23-02-2019.
- [39] PIRRE,Taboulet.ECG de A à Z..accéder le 25-02-2019.
- [40] Online Biology Notes.Electrocardiogram (ECG).<https://www.onlinebiologynotes.com/electrocardiogram-ecg-working-principle-normal-ecg-wave\application-of-ecg/> .accéder le 27-02-2019.
- [41] Acharya, U. R., Bhat, P. S., Iyengar, S. S., Rao, A., & Dua, S. 2003. Classification of heart rate data using artificial neural network and fuzzy equivalence relation.
- [42] Exarchos, T. P., Tsipouras, M. G., Exarchos, C. P., Papaloukas, C., Fotiadis, D. I., & Michalis, L. K. 2007. A methodology for the automated creation of fuzzy expert systems for ischaemic and arrhythmic beat classification based on a set of rules obtained by a decision tree. *Artificial Intelligence in Medicine*, 403, 187-200. PMID : 17543512. 10.1016/j.artmed.2007.04.001
- [43] Asl, B. M., Setarehdan, S. K., & Mohebbi, M. 2008. Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal. *Artificial Intelligence in Medicine*, 441, 51-64. PMID : 18585905. 10.1016/j.artmed.2008.04.007
- [44] Tsipouras, M. G., & Fotiadis, D. I. 2004. Automatic arrhythmia detection based on time and time-frequency analysis of heart rate variability. *Computer Methods and Programs in Biomedicine*, 742, 95-108. PMID : 15013592.
- [45] Behadada, O., Trovati, M., Chikh, M. A., & Bessis, N. 2016. Big data-based extraction of fuzzy partition rules for heart arrhythmia detection : A semi-automated approach. *Concurrency and Computation*, 282, 360-373. 10.1002/cpe.3428
- [46] Behadada, O., Trovati, M., Chikh, M., Bessis, N., & Korkontzelos, Y. 2016a. Logistic regression multinomial for arrhythmia detection. Paper presented at the IEEE International Workshops on Foundations and Applications of Self Systems pp. 133-137.

- [47] Behadada, O., Trovati, M., Chikh, M., Bessis, N., & Korkontzelos, Y. 2016b. Logistic regression multinomial for arrhythmia detection. Paper presented at the IEEE International Workshops on Foundations and Applications of Self Systems pp. 133-137.
- [48] Tsipouras, M., Goletsis, Y., & Fotiadis, D. 2004. A method for arrhythmic episode classification in ECGs using fuzzy logic and Markov models. Paper presented at the Computers in Cardiology '04 pp. 361-364.
- [49] Kane, S. A., Blake, J. R., McArdle, F. J., Langley, P., & Sims, A. J. 2016. Accuracy of pulse interval timing in ambulatory blood pressure measurement. *Scientific Reports*, 61, 37472. PMID : 27876841.
- [50] Raschi, E., & De Ponti, F. 2017. Drug-induced liver injury : Towards early prediction and risk stratification. *World J. Hepatol.*, 91, 30-37. PMID : 28105256.
- [51] BUFFON, Georges Louis Leclerc, GUNEAU DE MONTBLLIARD, Philibert, LACPDE, Bernard Germaine Etienne de La, et al. *Histoire naturelle, générale et particulière*. 1749.
- [52] AGGARWAL, Charu C. (ed.). *Data classification : algorithms and applications*. CRC press, 2014.
- [53] HUANG, Gao, SONG, Shiji, GUPTA, Jatinder ND, et al. Semi-supervised and unsupervised extreme learning machines. *IEEE transactions on cybernetics*, 2014, vol. 44, no 12, p. 2405-2417.
- [54] PAMPEL, Fred C. *Logistic regression : A primer*. Sage, 2000.
- [55] MANSOURNIA, Mohammad Ali, GEROLDINGER, Angelika, GREENLAND, Sander, et al. Separation in logistic regression : causes, consequences, and control. *American journal of epidemiology*, 2017, vol. 187, no 4, p. 864-870.
- [56] VATCHEVA, Kristina P., LEE, MinJae, MCCORMICK, Joseph B., et al. Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology (Sunnyvale, Calif.)*, 2016, vol. 6, no 2.
- [57] EL SANHARAWI, Mohamed et NAUDET, F. Comprendre la régression logistique. *Journal français d'ophtalmologie*, 2013, vol. 36, no 8, p. 710-715.
- [58] RAKOTOMALALA, R. *Pratique de la régression logistique, régression logistique binaire et polytomique, version 2.0*. Cours de l'Université Lumière Lyon, 2009, vol. 2.
- [59] RAKOTOMALALA, Ricco. *Pratique de la régression logistique. Régression Logistique Binaire et Polytomique*, Université Lumière Lyon, 2011, vol. 2, p. 258p.

- [60] LISON, Pierre. An introduction to machine learning. 2015.
- [61] A. Slavkovic, « Multinomial Logistic Regression Models – Baseline-Category Logit Model », in « STAT 504 – Analysis of Discrete Data », Pennsylvania State University, 2007.
- [62] Karthikeyani, V., & Begum, I. P. 2013. Comparison a performance of data mining algorithms CPDMA in prediction of diabetes disease. International Journal on Computer Science and Engineering, 53, 205.
- [63] Exarchos, T. P., Tsipouras, M. G., Exarchos, C. P., Papaloukas, C., Fotiadis, D. I., & Michalis, L. K. 2007. A methodology for the automated creation of fuzzy expert systems for ischaemic and arrhythmic beat classification based on a set of rules obtained by a decision tree. Artificial Intelligence in Medicine, 403, 187-200. PMID : 17543512. 10.1016/j.artmed.2007.04.001
- [64] Duda, R. O., Hart, P. E., & Stork, D. G. 2012. Pattern classification. John Wiley & Sons.
- [65] Camps-Valls, G., & Bruzzone, L. 2005. Kernel-based methods for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 436, 1351-1362.
- [66] Kohonen, T. 1993. Generalizations of the self-organizing map. Paper presented at the 1993 International Joint Conference on Neural Networks IJCNN'93, Nagoya. Vol. 1, pp. 457-462.
- [67] Rabiner, L., & Juang, B. 1986. An introduction to hidden Markov models. IEEE ASSP Magazine, 31, 4-16.

Résumé :

Les arythmies électrocardiographiques (ECG) telles que les arythmies ventriculaires et auriculaires ... sont l'une des causes les plus courantes du décès. Ces anomalies de l'activité électrique cardiaque peuvent entraîner la mort immédiate ou peuvent être responsables des séquelles graves nécessitant une prise en charge précoce et efficace afin de récupérer l'activité électrique de base. Si les symptômes anormaux peuvent être détectés et diagnostiqués tôt, le temps sera gagné pour prévenir la survenue d'une crise cardiaque. Par conséquent, il est nécessaire de disposer d'une méthode efficace de la détection et du traitement précoce. Nous proposons, dans ce mémoire, une méthode intelligente pour classer le rythme cardiaque des signaux ECG avec précision par la régression logistique multinomiale (MLR). La base de données est utilisée pour classer les arythmies en trois types différents: battements normales (N), Bloc de branche droit (RBBB) et Bloc de branche gauche (LBBB). Le taux de classification (MLR) obtenu est de 81,50%, ce qui est considéré un résultat très satisfait pour la détection des arythmies cardiaque.

Mots clés : arythmies cardiaque, électrocardiogramme, régression logistique multinomiale, classification.

Abstract:

Electrocardiographic arrhythmias (ECGs) such as ventricular and atrial arrhythmias... are one of the most common causes of death. These abnormalities in cardiac electrical activity can lead to immediate death or can be responsible for serious sequelae requiring early and effective management to restore basic electrical activity. If abnormal symptoms can be detected and diagnosed early, time will be saved to prevent a heart attack. Therefore, it is necessary to have an effective method of detection and early treatment. In this article, we propose an intelligent method to accurately classify the heart rate of ECG signals by multinomial logistic regression (MLR). It is therefore necessary to have a method for early detection and treatment. In this article, we propose an intelligent method to accurately classify the heart rate of ECG signals by multinomial logistic regression (MLR). The classification rate (MLR) obtained being 81.50%, it is considered a very satisfied result for the detection of cardiac arrhythmias.

Keywords: cardiac arrhythmias, electrocardiogram, multinomial logistic regression, classification.

ملخص الأطروحة

يعد عدم انتظام ضربات القلب في المخطط الكهربائي (ECG) مثل عدم انتظام ضربات البطين والأذين ... أحد أكثر أسباب الوفاة شيوعاً. قد تؤدي هذه التشوهات في النشاط الكهربائي إلى الوفاة الفورية أو قد تكون مسؤولة عن عواقب خطيرة تتطلب إدارة مبكرة وفعالة لاستعادة النشاط الكهربائي الأساسي. إذا كان يمكن اكتشاف الأعراض غير الطبيعية و تشخيصها في وقت مبكر يتم توفير الوقت لمنع ظهور نوبة قلبية. لذلك ، من الضروري أن يكون لديك طريقة فعالة للكشف والعلاج المبكرين. في هذه المذكرة ، نقترح طريقة ذكية لتصنيف معدل ضربات القلب لإشارات تخطيط القلب بدقة من خلال الانحدار اللوجستي (MLR) متعدد الحدود. تُستخدم قاعدة البيانات لتصنيف عدم انتظام ضربات القلب إلى ثلاثة أنواع مختلفة: الضربات الطبيعية للقلب (N) ، ضربات كتلة الفرع الأيمن للقلب (RBBB) ، ضربات كتلة الفرع الأيسر للقلب (LBBB). دقة تصنيف MLR التي تم الحصول عليها من 81.50 ٪ هي نتيجة جيدة للكشف عن عدم انتظام ضربات القلب.

الكلمات المفتاحية : عدم انتظام ضربات القلب, مخطط كهربية القلب, الانحدار اللوجستي متعدد الحدود, تصنيف .