

THÈSE DE DOCTORAT

de

l'Université d'Aix-Marseille

préparée au

Laboratoire d'Informatique et Systèmes

(LIS UMR 7020)

École Doctorale en Mathématiques et Informatique de Marseille

(ED 184)

Spécialité de doctorat : Apprentissage automatique, analyse et diagnostic

Par

Youssef TRARDI

Analyse Multivariée pour le Diagnostic de l'Arythmie Cardiaque

Soutenue publiquement, le 14 Décembre 2020

Devant le jury composé de :

Mme Mitra FOULADIRAD	Professeur des Universités Université de Technologie de Troyes, CNRS, UMR 6281	Rapporteure
M. Didier THEILLIOL	Professeur des Universités Université de Lorraine, Faculte des Sciences et Techniques, CNRS, CRAN UMR 7039	Rapporteur
M. Ahmed EL HAJJAJI	Professeur des Universités Université de Picardie Jules Verne, MIS EA 4290	Examineur
M. Vincent COCQUEMPOT	Professeur des Universités Université de Lille, CNRS, UMR 9189	Examineur
M. Stéphane DELLIAUX	Maître de conférences des universités – praticien hospitalier Aix-Marseille Université, APHM, INSERM, INRAE, C2VN	Examineur
M. Mustapha OULADSINE	Professeur des Universités Aix-Marseille Université, LIS UMR 7020	Directeur de thèse
Mme Bouchra ANANOU	Maître de conférences Aix-Marseille Université, LIS UMR 7020	Co-directeur de thèse

Youssef TRARDI : *Analyse Multivariée pour le Diagnostic de l'Arythmie Cardiaque*
Thèse de Doctorat



Cette œuvre est mise à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

ABSTRACT

Cardiovascular disease (CVD) is one of the most pressing health concerns and the leading cause of death worldwide. According to the World Health Organization (WHO), CVD is responsible for 17.9 million deaths worldwide each year, or 31% of all deaths. In France, CVD is the second leading cause of death after cancer, with approximately 150,000 deaths per year. Myocardial infarction also called a heart attack, is the most deadly form of CVD in the world. It causes 18,000 deaths per year in France, i.e., 10% of total mortality. In this thesis, we are interested in CVD, and more precisely in one of its leading causes, namely cardiac arrhythmias. Traditional laboratory methods for arrhythmia diagnosis on Holter ECG recordings are time-consuming and cost-intensive. Besides, permanent ectopic beats, motion artifacts, and noise complicate the interpretation and identification of the symptoms associated with cardiac arrhythmias. Considering these constraints and the enormous amount of time invested in the manual analysis of Holter ECG signals, doctors have emphasized the use of automated diagnostic support systems for arrhythmia. Academic research and industry rely on technological advances to develop IT tools for arrhythmia detection. In this work, we address this issue by proposing a new diagnostic strategy to distinguish healthy subjects in the presence of ectopic beats from subjects with atrial fibrillation. This strategy is based on the analysis of additional derivatives extracted from the R-R interval time series. This approach is composed of a process of derivative calculation and feature extraction. We applied different classification algorithms to build the diagnostic model, including support vector machine and multi-kernel learning. Also, we have developed a high-performance variable selection algorithm based on multi-kernel programming. The developed approach has been validated on different cardiac arrhythmia databases. The results obtained demonstrate the efficiency and robustness of the developed method.

RESUMÉ

Les maladies cardiovasculaires (MCV) sont parties des préoccupations sanitaires les plus pressantes. Les MCV présentent la première cause de décès dans le monde. Selon l'Organisation mondiale de la santé (OMS), les MCV sont à l'origine de 17,9 millions de décès dans le monde chaque année, soit 31% de l'ensemble des décès. En France, les MCV sont la deuxième cause de décès après le cancer, avec environ 150 000 décès par an. L'infarctus du myocarde, également appelé crise cardiaque, est la forme de MCV la plus meurtrière au monde. Il provoque 18 000 décès par an en France, soit 10% de la mortalité totale. Dans cette thèse, nous nous intéressons aux MCV, et plus précisément à l'une de ses principales causes, à savoir les arythmies cardiaques. Les méthodes d'analyse conventionnelles pour le diagnostic de l'arythmie par inspection visuelle des enregistrements ECG Holter sont coûteuses en termes de temps et d'argent. Ainsi, la présence de battements ectopiques permanents, d'artefacts de mouvement et de bruit superposés aux formes d'onde de l'ECG complique considérablement l'interprétation et l'identification des signes associés aux arythmies cardiaques. Compte tenu de ces contraintes et du temps énorme investi dans les analyses manuelles de signaux ECG Holter, les médecins ont privilégié le recours à des systèmes automatisés d'aide au diagnostic de l'arythmie. Les recherches académiques et les industriels s'appuient sur les avancées technologiques pour développer des outils informatiques pour la détection de l'arythmie. Dans ce travail, nous discutons cette problématique en proposant une nouvelle stratégie de diagnostic qui permet de distinguer les sujets sains en présence de battements ectopiques des sujets atteints de la fibrillation auriculaire. Cette stratégie est basée sur l'analyse de dérivés complémentaires extraits de la série chronologique des intervalles R-R. Cette approche est composée d'un processus de calcul de dérivées et d'extraction de caractéristiques. Pour construire le modèle de diagnostic, nous avons appliqué différents algorithmes de classification, notamment les séparateurs à vaste marge et l'apprentissage multinoyaux. En outre, nous avons développé un algorithme de sélection de variables très performant, basé sur la programmation multinoyaux. L'approche développée a été validée sur différentes bases de données d'arythmies cardiaques. Les résultats obtenus démontrent l'efficacité et la robustesse de la méthode développée.

TRAVAUX PUBLIÉS

-
-
- | | | |
|------|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2020 | [1] | Youssef Trardi , Bouchra Ananou, and Mustapha Ouladsine. An Advanced Arrhythmia Recognition Methodology Based on R-waves Time-Series Derivatives and Benchmarking Machine-Learning Algorithms. In <i>ECC 2020 – Proceedings of European Control Conference</i> |
| 2019 | [2] | Youssef Trardi , Bouchra Ananou, Zouhair HADDI, and Mustapha Ouladsine. An Effective Data-Driven Diagnostic Strategy for Cardiac Pathology Screening. In <i>ACD 2019 – Proceedings of 15th European Workshop on Advanced Control and Diagnosis</i> |
| 2018 | [3] | Youssef Trardi , Bouchra Ananou, Zouhair HADDI, and Mustapha Ouladsine. A Novel Method to Identify Relevant Features for Automatic Detection of Atrial Fibrillation. In <i>MED 2018 – Proceedings of 26th Mediterranean Conference on Control and Automation</i> |
| 2018 | [4] | Youssef Trardi , Bouchra Ananou, Zouhair HADDI, and Mustapha Ouladsine. Multi-Dynamics Analysis of QRS Complex for Atrial Fibrillation Diagnosis. In <i>CoDIT 2018 – Proceedings of 5th International Conference on Control, Decision and Information Technologies</i> |
-
-

TRAVAUX PUBLIÉS EN COLLABORATION

-
-
- | | | |
|------|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2019 | [5] | Zouhair HADDI, Bouchra Ananou, Youssef Trardi , Jean-François Pons, Stéphane Delliaux, Jean-Claude Deharo, and Mustapha Ouladsine. Advanced Machine Learning Coupled with Heart-Inter-beat derivatives for Cardiac Arrhythmia Detection . In <i>ACC 2020 – Proceedings of American Control Conference</i> |
| 2019 | [6] | Zouhair HADDI, Bouchra Ananou, Youssef Trardi , Mustapha Ouladsine, Jean-François Pons, Stéphane Delliaux, and Jean-Claude Deharo. Relevance Vector Machine as Data-Driven Method for Medical Decision Making . In <i>ECC 2019 – Proceedings of European Control Conference</i> |
| 2018 | [7] | Zouhair HADDI, Bouchra Ananou, Youssef Trardi , Jean-François Pons, Stéphane Delliaux, Mustapha Ouladsine, and Jean-Claude Deharo. An Efficient Pattern Recognition Kernel-Based Method for Atrial Fibrillation Diagnosis . In <i>CinC 2018 – Proceedings of Computing in Cardiology Conference</i> |
-
-
- | | | |
|------|-----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2020 | [8] | Zouhair HADDI, Bouchra Ananou, Youssef Trardi , and Mustapha Ouladsine. Fusion of Multiple Univariate Data Analysis based Detectors to Build a Specific Fingerprint of Atrial Fibrillation . In <i>CinC 2020 – Proceedings of Computing in Cardiology Conference</i> |
|------|-----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
-
-

TABLE DES MATIÈRES

Abstract	iii
Resumé	v
Travaux publiés	vii
Table des figures	xvii
Liste des tableaux	xix
Nomenclature	xxi
Introduction générale	1
I SIGNAL ECG	7
I.1 Introduction	9
I.2 Introduction au système cardiovasculaire	9
I.2.1 Système cardiovasculaire	9
I.2.1.1 Les vaisseaux sanguins	10
I.2.1.2 Le cœur	10
I.2.2 Structure et fonction du cœur	11
I.2.2.1 Structure du cœur	11
I.2.2.2 Fonction du cœur	11
I.2.3 Activité mécanique	12
I.2.4 Activité électrique	13
I.2.5 Cohérence cardiaque	15
I.2.5.1 Système Nerveux Autonome (SNA)	15
I.2.5.2 Fréquence Cardiaque (FC)	15
I.3 Électrocardiogramme (ECG)	17
I.3.1 Objectif et principe de l'ECG	17
I.3.2 Forme d'onde et composantes d'un ECG normal	18
I.3.2.1 Ondes ECG	18
I.3.2.2 Segments et intervalles ECG	19

I.4	Vue d'ensemble du rythme cardiaque	20
I.4.1	Rythme sinusal	20
I.4.2	Blocs cardiaques	22
I.4.2.1	Blocs auriculoventriculaires cardiaques	22
I.4.2.2	Bloc de branche	22
I.4.3	Arythmies cardiaques	23
I.4.4	Les formes d'arythmie	24
I.4.4.1	Arythmies auriculaires	24
I.4.4.2	Arythmies ventriculaires	25
I.4.5	Diagnostic de l'arythmie	25
I.4.6	Base de données sur l'arythmie	26
I.5	Contexte et objectifs	27
I.5.1	Fibrillation auriculaire : l'arythmie la plus courante	27
I.5.1.1	types et caractéristiques	27
I.5.1.2	FA vs. RSN	28
I.5.2	Défis de la détection de la FA	28
I.5.3	Diagnostic automatique de la FA	30
I.5.4	Méthodes de diagnostic existantes	30
I.5.4.1	Métriques d'évaluation	31
I.5.4.2	Méthodes de diagnostic univariées	32
I.5.4.3	Méthodes de diagnostic multivariées	35
I.6	Conclusion	38
II	TRAITEMENT DES SIGNAUX ECG	39
II.1	Introduction	40
II.2	ECG : étape de prétraitement	41
II.2.1	Aperçu général	41
II.2.2	La dérive de la ligne de base	43
II.2.2.1	Cas d'étude	43
II.2.2.2	Élimination de la dérive de la ligne de base : méthode de la régression linéaire multiple	44
II.2.3	Filtrage des interférences et artefacts	46
II.2.3.1	Filtre numérique	47
II.2.3.2	Filtre numérique à Réponse Impulsionnelle Infinie (RII)	47
II.2.4	Algorithme de Pan-Tompkins	49
II.3	ECG : détection du complexe QRS	50
II.3.1	Synthèse générale	50
II.3.2	Motivations	52

II.3.3	Méthodologie de détection : analyse préliminaire	53
II.3.3.1	Analyse préliminaire du signal en fonction d'une amplitude minimale	53
II.3.3.2	Analyse préliminaire du signal en fonction d'une distance minimale	54
II.3.4	Limites de la méthode	55
II.4	ECG : méthodologie de correction	56
II.4.1	Correction modérée basée sur la variabilité des données	56
II.4.2	Correction modérée en présence des intervalles R-R étroits	57
II.4.3	Correction modérée destinée aux intervalles R-R contenant un pic d'amplitude très faible	58
II.4.4	Correction exhaustive	59
II.5	Conclusion	61
III	Méthodes mathématiques d'analyse des données VFC	63
III.1	Introduction	64
III.2	Méthodes mathématiques appliquées aux VFC	65
III.2.1	Méthode conventionnelle	65
III.2.1.1	Stratégie de segmentation des signaux VFC	65
III.2.1.2	Principaux indicateurs de la VFC	65
III.2.1.3	Approches de classification appliquées aux indicateurs de la VFC	66
III.2.2	Méthode proposée	66
III.2.2.1	Objectifs	66
III.2.2.2	Dynamiques RR	67
III.3	Analyse de la variabilité des dynamiques RR	69
III.3.1	Objectifs	69
III.3.2	Statistique descriptive unidimensionnelle	69
III.3.2.1	Mesures de tendance centrale	69
III.3.2.2	Mesures de variabilité	70
III.3.3	Analyse statistique unidimensionnelle appliquée aux dynamiques RR^m	73
III.3.3.1	Mesures de la corrélation linéaire	73
III.3.3.2	Mesures d'association entre RSN sans battement ectopique et FA	73
III.3.3.3	Mesures d'association entre RSN avec battement ectopique et FA	76
III.3.4	Discussion	77
III.4	Extraction des Caractéristiques	78
III.4.1	Méthode linéaire	78
III.4.1.1	Caractéristiques du domaine temporel	78
III.4.1.2	Caractéristiques du domaine fréquentiel	79
III.4.2	Méthode géométrique	80
III.4.3	Méthode non linéaire	81

III.5 Conclusion	83
IV Diagnostic de la fibrillation auriculaire	85
IV.1 Introduction	87
IV.2 Diagnostic de la FA à l'aide de trois dynamiques RR	88
IV.2.1 Méthodologie	88
IV.2.2 Prétraitement des données	89
IV.2.2.1 Bases de données	89
IV.2.2.2 Détection et segmentation des ondes R	89
IV.2.3 Phase de réduction de la dimensionnalité	89
IV.2.3.1 Ensemble de données d'évaluation	89
IV.2.3.2 Processus d'entraînement : Association d'analyse factorielle et de classificateurs	89
IV.2.3.3 Résultats préliminaires	90
IV.2.4 Modèle de prise de décision	91
IV.3 Diagnostic de l'arythmie à l'aide de cinq dynamiques RR	92
IV.3.1 Méthodologie	92
IV.3.2 Prétraitement des données	94
IV.3.3 Analyse préliminaire	94
IV.3.3.1 Ensemble de données d'évaluation	94
IV.3.3.2 ANOVA	94
IV.3.4 Pipeline machine learning	94
IV.3.4.1 Apprentissage des pipelines	94
IV.3.4.2 Résultats : La validation croisée à 10 blocs	95
IV.3.5 Modèle idéal	97
IV.3.6 Motivations	98
IV.4 Nouvelle approche de sélection de variables	99
IV.4.1 Introduction aux enjeux de la réduction de la dimensionnalité	99
IV.4.1.1 Vue d'ensemble	99
IV.4.1.2 Aperçu technique	100
IV.4.2 Méthodologie	102
IV.4.3 Prétraitement des données	103
IV.4.4 Présentation de l'algorithme MKL	103
IV.4.5 MKL appliqué à la sélection des variables	105
IV.4.5.1 Formulation mathématique du problème	105
IV.4.5.2 Méthode de résolution	106
IV.4.6 Modèle idéal	108
IV.5 Conclusion	110

Conclusion	111
IV.6 Introduction	116
IV.7 Contexte et objectifs	117
IV.8 Modélisation de la propagation de l'épidémie de COVID-19	118
IV.8.1 Intérêt des modèles mathématiques en épidémiologie	118
IV.8.2 Modélisation mathématique des maladies infectieuses	119
IV.8.2.1 Modèle SEIR généralisé	119
IV.8.2.2 Modèle proposé	119
IV.8.3 Données COVID-19	122
IV.8.4 Implantation du modèle	122
IV.8.4.1 Phase de mise en route	122
IV.8.4.2 Phase d'apprentissage	123
IV.9 Résultats de prédiction	124
IV.9.1 Les effets de la relaxation du confinement	127
IV.10 Conclusion	129
Annexe A	131
A.1 Modèle SEIR généralisé	131
A.2 Modèle proposé	132
Bibliographie	133

TABLE DES FIGURES

I.1	Le système cardio-vasculaire : il s'agit d'un système circulatoire d'alimentation sanguine, composé du cœur et des vaisseaux sanguins (veines et artères). La zone rouge symbolise la circulation sanguine riche en oxygène (réseau artériel). La zone bleue symbolise la circulation sanguine pauvre en oxygène (réseau veineux).	10
I.2	Diagramme de l'anatomie du cœur	11
I.3	Schéma de la circulation sanguine (cycle cardiaque)	12
I.4	Les événements électriques du cycle cardiaque : (1) La dépolarisation auriculaire, déclenchée par le nœud SA et provoquant l'onde P. (2) La dépolarisation auriculaire terminée, l'impulsion est retardée au niveau du nœud AV. (3) La dépolarisation ventriculaire commence à l'apex, provoquant le complexe QRS. Une repolarisation auriculaire apparaît. (4) La dépolarisation ventriculaire est terminée. (5) La repolarisation ventriculaire commence à l'apex, provoquant l'onde T. (6) La repolarisation ventriculaire est complète.	14
I.5	Le tracé normal d'un ECG indique l'onde P, le complexe QRS et l'onde T, les intervalles PR, RR, et QT, ainsi que les segments PR et ST.	18
I.8	Schéma synoptique de l'ensemble des opérations du système de diagnostic.	31
II.1	Synoptique d'un système de détection des pics R.	41
II.7	Les opérations de prétraitement proposées dans l'algorithme de Pan-Tompkins [9] . .	49
II.9	Algorithme de détection des complexes QRS	52
II.10	Les résultats de détection obtenus après le seuillage en amplitude et en durée.	55
II.11	Exemple 1 - Résultats obtenus après la quatrième filtration des faux pics.	60
II.12	Exemple 2 - Résultats obtenus après la quatrième filtration des faux pics.	60
III.1	Schéma synoptique des opérations effectuées dans le système de diagnostic envisagé. .	67
III.2	Calcul de corrélation appliqué aux dynamiques normales et absolues	68
III.3	Diagramme synoptique présentant l'organigramme d'extraction des statistiques descriptives.	72
IV.1	Diagramme synoptique de la réalisation du diagnostic considéré en utilisant trois formes dynamiques.	88

IV.2	Diagramme synoptique de la réalisation du diagnostic considéré en utilisant trois formes dynamiques.	93
IV.3	ANOVA appliquée aux 65 caractéristiques. Les valeurs p (p-values) sont mesurées sous la forme $-\log_{10}(p_{valeurs})$. Chaque couleur reflète les propriétés des Feature Statistics (FS) obtenues respectivement sur chacune des cinq dynamiques.	95
IV.4	Diagramme synoptique de la réalisation du diagnostic considéré en utilisant six formes dynamiques et une approche de sélection de variables basée sur l'apprentissage de multinoyaux.	103
IV.5	Le diagramme synoptique du modèle proposé.	120
IV.6	1er cas : Écart absolu entre les estimations et les valeurs observées du modèle national.	125
IV.7	1er cas : Écart absolu entre les estimations et les valeurs observées du modèle régional.	125
IV.8	2e cas : Écart absolu entre les estimations et les valeurs observées du modèle national.	126
IV.9	2e cas : Écart absolu entre les estimations et les valeurs observées du modèle régional.	126
IV.10	1er cas : Impact de la relaxation du confinement sous différents scénarios possibles pour le modèle national et régional.	127
IV.11	2e cas : Impact de la relaxation du confinement sous différents scénarios possibles pour le modèle national et régional.	128

LISTE DES TABLEAUX

I.1	Fréquence cardiaque au repos chez les hommes, en fonction de l'âge (ans) et de la condition physique.	16
I.2	Fréquence cardiaque au repos chez les femmes, en fonction de l'âge (ans) et de la condition physique.	16
I.3	Caractéristiques des rythmes sinusaux.	21
I.4	Interprétation électrocardiographique anormale.	24
I.5	Présentation des bases de données des rythmes variés. Les chiffres indiqués représentent le nombre total d'épisodes de 30 secondes extraites de chaque rythme.	27
III.1	Résultats des mesures d'association entre RSN sans battement ectopique et FA	74
III.2	Résultats des mesures d'association entre RSN avec battement ectopique et FA. . . .	76
IV.1	Tableau des abréviations des approches	87
IV.2	Performance atteinte lors de la première expérience. La précision des classificateurs est calculée en pourcentages, en utilisant la technique de validation croisée à 10 blocs.	90
IV.3	Performance atteinte lors de la deuxième expérience. La précision des classificateurs est calculée en pourcentages, en utilisant la technique de validation croisée à 10 blocs.	90
IV.4	Évaluation SVM en pourcentage, en utilisant la technique de validation croisée à 10 blocs.	91
IV.5	Matrice de confusion	92
IV.6	Les résultats obtenus en combinant "l'analyse factorielle" et les classificateurs. La précision des classificateurs est calculée en pourcentages, en utilisant la technique de validation croisée à 10 blocs.	96
IV.7	Les résultats obtenus en combinant l'ACP et les classificateurs. La précision des classificateurs est calculée en pourcentages, en utilisant la technique de validation croisée à 10 blocs.	96
IV.8	Les résultats obtenus en combinant le sélecteur RFE et les trois classificateurs ETC, WKNN et SVM. La précision des classificateurs est calculée en pourcentages, en utilisant la technique de validation croisée à 10 blocs.	97
IV.9	Détails du modèle optimal sur la base de tests.	97
IV.10	Matrice de confusion	98

IV.11 Détails du modèle optimal sur la base de validation et test.	108
IV.12 Matrice de confusion	109
IV.13 1er expérience : Pourcentage d'erreurs d'estimation atteint en phase d'apprentissage et test en utilisant les données COVID-19 pour la France.	124
IV.14 1er expérience : Pourcentage d'erreurs d'estimation atteint en phase d'apprentissage et test en utilisant les données COVID-19 pour la région PACA.	124
IV.15 2e expérience : Pourcentage d'erreurs d'estimation atteint en phase d'apprentissage et test en utilisant les données COVID-19 pour la France.	124
IV.16 2e expérience : Pourcentage d'erreurs d'estimation atteint en phase d'apprentissage et test en utilisant les données COVID-19 pour la région PACA.	124

NOMENCLATURE

Abréviation

ACP	analyse en composantes principales
ADL	analyse discriminante linéaire
ANOVA	Analyse de la variance
ApEn	approximate entropy
AV	auriculo-ventriculaire
AVC	accident vasculaire cérébral
BPM	battements par minute
CoSEn	coefficient of sample entropy
CV	coefficient de variation
DWT	discrete wavelet transform
ECG	électrocardiogramme
EnSh	Shannon Entropy
ETC	classificateur d'arbres extras (extra trees classifier)
FA	fibrillation auriculaire
FLA	flutter auriculaire
FV	fibrillation ventriculaire
GDA	generalized discriminant analysis
LR	régression logistique
MKL	apprentissage multi-noyaux
MLP	perceptron multicouche

RCP	réanimation cardio-pulmonaire
RF	forêt d'arbres décisionnels (random-forest)
RFE	élimination des caractéristiques récursives
RMSSD	root mean square of the successive differences
RNA	réseau de neurones artificiels
ROC	Receiver Operating Characteristics
RSN	rythme sinusal normal
SA	sino-auriculaire
SampEn	sample entropy
SVM	support vector machines
SVR	support vector regression
TPR	turning point ratio
TPSV	tachycardie paroxystique supraventriculaire
TSV	tachycardie supraventriculaire
TV	tachycardie ventriculaire
VAI	vector angle index
VFC	variabilité de la fréquence cardiaque
VLT	vector length index
WKNN	K plus proches voisins pondérés

INTRODUCTION GÉNÉRALE

Les systèmes d'aide au diagnostic médical basés sur des algorithmes d'apprentissage automatique, appelé machine learning en anglais, représentent une forme d'Intelligence artificielle (IA) qui permet d'améliorer les performances, la qualité et la rapidité des soins de santé. Les améliorations apportées aux applications d'apprentissage automatique embarquées ont conduit à la conception de nombreux appareils de surveillance médicale. Ces derniers sont dotés de capteurs de signaux biologiques pour mesurer l'activité de l'organe cible d'un sujet. L'objectif essentiel de ces appareils est d'enregistrer, de sauvegarder et d'analyser les signaux acquis pour établir un diagnostic approprié et/ou identifier les signes de symptômes pathologiques. Les travaux de ma thèse s'inscrivent dans ce contexte et ont pour objectif principal de mettre en place de nouvelles stratégies pour l'analyse et le diagnostic des signaux d'Électrocardiogramme (ECG), axées sur la détection des épisodes d'arythmie cardiaque.

Contexte et défis du diagnostic de l'arythmie

Les maladies cardiovasculaires (MCV) sont parties des préoccupations sanitaires les plus pressantes. Les MCV présentent la première cause de décès dans le monde. Selon l'Organisation mondiale de la santé (OMS), les MCV sont à l'origine de 17,9 millions de décès dans le monde chaque année, soit 31% de l'ensemble des décès [10]. En France, les MCV sont la deuxième cause de décès après le cancer, avec environ 150 000 décès par an et 400 décès par jour. L'infarctus du myocarde, également appelé crise cardiaque, est la forme de MCV la plus meurtrière au monde. Il provoque 18 000 décès par an en France, soit 10% de la mortalité totale [11].

Dans cette thèse, nous nous intéressons aux MCV, et plus précisément à l'une de ses principales causes, à savoir les arythmies cardiaques. L'arythmie est une anomalie de la conduction cardiaque qui peut s'accompagner à diverses complications souvent mortelles. Une arythmie peut se produire lorsque les impulsions électriques qui coordonnent les battements du cœur ne fonctionnent plus correctement, provoquant une accélération, un ralentissement ou une irrégularité des battements du cœur [12–14]. Chez les personnes âgées, la fibrillation auriculaire (FA) est la forme d'arythmie la plus fréquente [12, 13]. Le diagnostic de la FA est une tâche délicate en raison de ses formes cliniques silencieuses et, en particulier, la forme de la FA paroxystique [12]. Les recommandations en vigueur consistent à utiliser un électrocardiogramme Holter (ECG) de 24-48 heures pour détecter la présence d'une FA paroxystique lorsqu'elle est suspectée, mais le taux de reconnaissance reste faible [15].

En cardiologie, l'ECG est une pratique courante pour la détection de nombreuses formes d'arythmies en raison de sa nature non invasive, et peu coûteuse. Le diagnostic conventionnel de l'arythmie consiste à examiner le tracé d'ECG, notamment ses intervalles et ses ondes, générées par les différents tissus cardiaques, et à vérifier sa conduction morphologique. Les principales caractéristiques morphologiques utilisées pour identifier les arythmies sont l'absence d'ondes P, l'irrégularité des intervalles R-R, ou la forme des ondes du tracé ECG.

Les méthodes d'analyse conventionnelles pour le diagnostic de l'arythmie par inspection visuelle des enregistrements ECG Holter sont coûteuses en termes de temps et d'argent. En outre, la présence de battements ectopiques permanents, d'artefacts de mouvement et de bruit superposés aux formes d'onde de l'ECG complique considérablement l'interprétation et l'identification des signes associés aux arythmies cardiaques. Compte tenu de ces contraintes et du temps énorme investi dans les analyses manuelles de signaux ECG Holter, les médecins ont privilégié le recours à des systèmes automatisés d'aide au diagnostic de l'arythmie. Ces systèmes sont conçus pour faciliter le diagnostic de l'arythmie, surmonter les difficultés liées à la prise de décision et réduire le temps d'évaluation.

Les recherches académiques et les industriels s'appuient sur les avancées technologiques pour développer des outils informatiques en cardiologie afin de répondre aux contraintes de la pratique médicale. De nombreux travaux de recherche se sont intéressés à la proposition de méthodes permettant de distinguer les sujets sains (normaux) des sujets atteints de la FA (anormaux). Toutes ces méthodes sont basées sur l'électrocardiogramme (c'est-à-dire le signal de l'ECG), puis la distinction repose sur le choix des caractéristiques morphologiques du tracé. La majorité se concentre sur l'analyse de la forme d'onde de l'ECG, notamment les ondes P et R. L'apparition de l'onde P constitue une meilleure caractéristique que celui de l'onde R. Cependant, la détermination exacte de l'onde P est très compliquée en raison de sa faible amplitude, voire de son absence totale. En effet, la mise en place d'un système de diagnostic basé sur l'onde P dans les dispositifs électroniques de faible puissance n'est pas réaliste. Pour ces raisons, l'analyse basée sur l'onde R — soit l'étude des intervalles R-R — suscite un intérêt plus marqué que l'absence d'onde P ou la combinaison des deux caractéristiques. En effet, l'intervalle R-R est un facteur clé pour le diagnostic des irrégularités cardiaques et contient des informations précieuses pour établir un diagnostic rapide et efficient de la FA. Néanmoins, la présence de battements ectopiques chez les sujets sains augmente considérablement la variabilité du rythme et imite le comportement d'un rythme anormal, mais ce dernier reste cohérent, ce qui n'est pas le cas de l'arythmie. Les études utilisant la variabilité du rythme cardiaque pour distinguer les sujets normaux des sujets anormaux atteignent leurs limites en présence de battements ectopiques. Par conséquent, l'analyse statistique (étude de variabilité) ne peut pas distinguer ces deux formes (normale et anormale) avec une précision suffisante. Cela constitue notre principal défi : développer un outil de diagnostic capable de distinguer un rythme sinusal normal (RSN) en présence des battements ectopiques d'un rythme de FA.

Dans ce travail, nous discutons cette problématique en proposant une nouvelle stratégie de diagnostic qui permet de distinguer un RSN en présence de battements ectopiques d'un rythme de FA.

Cette stratégie est basée sur l'analyse de dérivés complémentaires extraits de la série chronologique des intervalles R-R. Cette approche est composée d'un processus de calcul de dérivées et d'extraction de caractéristiques statistiques. Étant donné que les séries temporelles des ondes R dénotent la position des pulsations cardiaques, l'intervalle R-R, qui est la dérivée des ondes R, décrit la vitesse du rythme cardiaque, communément appelé variabilité de la fréquence cardiaque (VFC). De même, les deux premières dérivées successives de la série chronologique de l'intervalle R-R représentent l'accélération et les à-coups du rythme cardiaque. Nous démontrerons que les caractéristiques statistiques extraites des dérivées des intervalles R-R possèdent les facteurs clés pour ce résoudre problème.

Diagnostic de la FA : objectif et motivations

Cette thèse porte essentiellement sur l'étude des anomalies cardiaques. Une attention particulière a été consacrée à la détection de la fibrillation auriculaire (FA) par rapport aux rythmes sinusaux normaux (RSN), y compris les battements ectopiques. La FA est la forme d'arythmie la plus commune, surnommée le "tueur silencieux", car elle ne produit pas toujours les symptômes permettant d'alerter le patient ou le médecin dès son apparition. La FA est associée à des facteurs de risque critiques tels que les accidents vasculaires cérébraux, les caillots sanguins, un taux de morbidité et de mortalité plus élevé. La FA touche environ 1% de la population mondiale, avec environ 4,5 millions de cas confirmés dans l'UE et 2,3 millions aux États-Unis [16]. Plus de 5% des personnes âgées de 65 ans ou plus sont touchées. La prévalence de la FA se poursuit également pendant la vieillesse, où plus de 17% des personnes de plus de 85 ans en sont atteintes [17]. Les estimations de la FA prévoient que d'ici 2050, elle pourrait toucher environ 17,9 millions de personnes dans l'UE et 6,12 millions aux États-Unis [18, 19]. Le développement d'outils efficaces pour le diagnostic automatisé de la FA est indispensable sur le plan médical et représente un défi technologique.

La nécessité d'une stratégie appropriée permettant d'assurer un diagnostic précis de la FA est de plus en plus recherchée. Compte tenu des études publiées sur le diagnostic automatique de la FA, la plupart se sont principalement intéressées à l'analyse des caractéristiques issues des intervalles R-R. Cependant, ces intervalles ne contiennent pas forcément toutes les connaissances détaillées et les marqueurs essentiels pour décrire un rythme cardiaque, puisqu'ils ne décrivent que la fréquence cardiaque. Comme dans la sémantique physique, un état dynamique est mieux défini en déterminant sa position, sa vitesse et son accélération, ainsi que ses dérivés secondaires associés, plutôt que de se limiter à sa position dans le temps. L'hypothèse proposée est d'utiliser des variables induites — c'est-à-dire des variables extraites des dérivées temporelles des intervalles R-R — peut étendre le domaine de représentation du comportement cardiaque en intégrant de nouveaux indicateurs critiques et fournir un système d'aide au diagnostic de la FA plus efficace par rapport à ce qui est proposé dans la littérature.

La conception d'un système de diagnostic AF comprend généralement quatre parties principales : le prétraitement des données, le calcul de la dynamique et l'extraction des variables, la réduction

de la dimensionnalité et la construction du classificateur. En particulier, les étapes de prétraitement des données et l'extraction de variables sont primordiales et contribuent de manière significative à l'efficacité de diagnostic.

Le premier travail réalisé dans le cadre de cette thèse porte sur un algorithme de détection des ondes R. L'algorithme est structuré en deux parties principales : dans la première partie, un algorithme simple basé sur des seuils d'amplitude et de temps fixes est appliqué pour extraire les positions des pics maxima. Cependant, cette technique présente certaines contraintes liées à l'altération morphologique des signaux ECG, ce qui nous amène à introduire un processus de correction dans la deuxième partie, visant à éliminer la surdétection des pics induites par la première partie. L'avantage de la technique de correction réside dans sa capacité à éliminer toutes les fausses détections, malgré la forte fluctuation observée dans les signaux.

Les étapes de calcul des dynamiques et de l'extraction des variables consistent à convertir les séries chronologiques de l'onde R en variables. L'extraction des variables est une phase cruciale, qui décrit les irrégularités et les variations intégrées dans les formes dynamiques développées. La pertinence du modèle demandé dépend essentiellement de la qualité implicite des variables extraites et de leur cohérence. Dans une structure d'analyse multivariée, ces variables peuvent contenir des éléments inutiles ou insignifiants sans contribution significative au processus de diagnostic. À cet effet, une phase de réduction de la dimensionnalité (univariée et multivariée) est appliquée pour sélectionner les variables pertinentes afin de construire un modèle de diagnostic robuste. Toutes les variables sélectionnées sont ensuite formées par des méthodes d'apprentissage automatique afin de construire un système de diagnostic AF efficace.

Le nombre de variables sélectionnées est la préoccupation la plus importante. La complexité du modèle augmente avec le nombre de variables, ainsi que le risque de surapprentissage. L'objectif est de concevoir un système de diagnostic rapide et fiable. Néanmoins, nous avons toujours été confrontés au problème de dimensionnalité. En effet, le sous-ensemble de variables sélectionnées pour alimenter l'algorithme d'apprentissage n'est pas assez pertinent pour atteindre les performances souhaitées. Autrement dit, certaines variables pertinentes ne font pas partie du sous-ensemble final ce qui a pour conséquence la dégradation des performances de prédiction. De ce fait, nous avons pris la décision de développer une approche de sélection de variables pour atteindre les objectifs souhaités. Nous avons développé un algorithme de sélection de variables robuste, qui consiste à utiliser le principe de l'apprentissage multinoyaux pour concevoir une stratégie de sélection de caractéristiques pertinente. Cette approche est porteuse de perspectives inédites et permet une adaptation optimale à des problèmes de grande dimension, grâce au recours itératif de l'algorithme SVM.

Présentation de la thèse

La première partie de cette thèse I présente une introduction au système cardiovasculaire, nous décrivons la structure, la fonction et les activités mécaniques et électriques du cœur. Nous présenterons

sommairement l'électrocardiogramme - ses objectifs et son utilité dans le diagnostic des arythmies. Nous présentons également une description détaillée des composantes du tracé ECG, notamment les ondes, les intervalles et les segments. Ces notions sont essentielles pour une bonne compréhension de la problématique discutée. Elles permettent de comprendre les caractéristiques indicatrices des pathologies cardiaques typiques. Un état de l'art sur les différentes approches proposées dans la littérature est également exposé.

Le deuxième chapitre II est consacré au prétraitement des signaux ECG et aux étapes d'extraction des caractéristiques, qui constituent l'aspect le plus essentiel du traitement des données. Nous abordons tout d'abord les méthodes de filtrage utilisées pour corriger les signaux ECG. Certaines techniques de filtrage sont également abordées, notamment le filtre numérique à réponse impulsionnelle infinie, de type Butterworth, et le concept de filtrage utilisé dans l'algorithme de Pan-Tompkins. Ensuite, nous présentons une étude sur la détection des complexes QRS.

Le troisième chapitre III présente les méthodes mathématiques appliquées aux analyses de la variabilité de fréquence cardiaque (VFC). Comme la VFC reste le seul paradigme permettant de caractériser la variabilité des intervalles R-R, nous avons naturellement proposé d'autres paradigmes approfondis pour renforcer et élargir la portée du diagnostic de la FA. Celle-ci est présentée par différentes formes de dynamiques, la dynamique de base des intervalles R-R, et ses dérivées successives, normales et absolues. Nous avons également mené une étude d'analyse statistique descriptive unidimensionnelle pour évaluer la pertinence des dynamiques conçues. Enfin, la dernière partie de ce chapitre est consacrée à l'introduction de l'étape d'extraction des caractéristiques, y compris les caractéristiques du domaine temporel linéaire, du domaine fréquentiel, du domaine géométrique et du domaine non linéaire.

Le quatrième chapitre IV présente les différents systèmes de classification développés au cours de cette thèse. Nous avons introduit trois systèmes de diagnostic en adoptant des approches distinctes. En outre, nous avons présenté les différentes étapes constituant un système automatique de diagnostic de l'arythmie. La dernière partie de ce chapitre est consacrée à la discussion d'un nouvel algorithme de sélection de variables, basé sur le principe de l'apprentissage multinoyaux, permettant de définir un choix pertinent de caractéristiques.

Nous avons clôturé ce manuscrit par une conclusion générale, qui résume les travaux réalisés, les défis relevés et les résultats obtenus. Les principales contributions exposées au cours de cette thèse sont également mises en évidence. Pour conclure, des perspectives de recherche sur le diagnostic de l'arythmie en général et de la FA en particulier sont présentées.

SIGNAL ECG

I.1	Introduction	9
I.2	Introduction au système cardiovasculaire	9
I.2.1	Système cardiovasculaire	9
I.2.1.1	Les vaisseaux sanguins	10
I.2.1.2	Le cœur	10
I.2.2	Structure et fonction du cœur	11
I.2.2.1	Structure du cœur	11
I.2.2.2	Fonction du cœur	11
I.2.3	Activité mécanique	12
I.2.4	Activité électrique	13
I.2.5	Cohérence cardiaque	15
I.2.5.1	Système Nerveux Autonome (SNA)	15
I.2.5.2	Fréquence Cardiaque (FC)	15
I.3	Électrocardiogramme (ECG)	17
I.3.1	Objectif et principe de l'ECG	17
I.3.2	Forme d'onde et composantes d'un ECG normal	18
I.3.2.1	Ondes ECG	18
I.3.2.2	Segments et intervalles ECG	19
I.4	Vue d'ensemble du rythme cardiaque	20
I.4.1	Rythme sinusal	20
I.4.2	Blocs cardiaques	22
I.4.2.1	Blocs auriculoventriculaires cardiaques	22
I.4.2.2	Bloc de branche	22
I.4.3	Arythmies cardiaques	23
I.4.4	Les formes d'arythmie	24
I.4.4.1	Arythmies auriculaires	24
I.4.4.2	Arythmies ventriculaires	25
I.4.5	Diagnostic de l'arythmie	25
I.4.6	Base de données sur l'arythmie	26

I.5	Contexte et objectifs	27
I.5.1	Fibrillation auriculaire : l'arythmie la plus courante	27
I.5.1.1	types et caractéristiques	27
I.5.1.2	FA vs. RSN	28
I.5.2	Défis de la détection de la FA	28
I.5.3	Diagnostic automatique de la FA	30
I.5.4	Méthodes de diagnostic existantes	30
I.5.4.1	Métriques d'évaluation	31
I.5.4.2	Méthodes de diagnostic univariées	32
I.5.4.3	Méthodes de diagnostic multivariées	35
I.6	Conclusion	38

I.1 Introduction

Le système cardiovasculaire est constitué du cœur et des vaisseaux sanguins. La fonction principale du système cardiovasculaire est de maintenir un flux sanguin continu et suffisant aux organes et aux tissus et sous une pression suffisante pour satisfaire les besoins énergétiques et le renouvellement des cellules. Les pathologies affectant le cœur et la circulation sanguine sont appelées maladies cardiovasculaires. Ce sont des affections qui sont dues à des anomalies cardiaques entraînant un mauvais fonctionnement de l'irrigation sanguine du cœur. Parmi ces affections cardiovasculaires, il existe : les arythmies cardiaques ; les cardiopathies congénitales, les coronaropathies, les maladies du muscle cardiaque (cardiomyopathies), les maladies des valves cardiaques, les AVC, les maladies vasculaires (c.-à-d. des maladies des vaisseaux sanguins). Les arythmies sont l'une des principales causes de décès dans le monde et la plus commune des maladies cardiovasculaires. En cardiologie, l'électrocardiogramme est un outil essentiel pour détecter de nombreuses formes d'arythmies cardiaques, et cela grâce à sa nature non invasive, accessible et peu coûteuse.

Dans la première partie du chapitre, nous décrivons la structure, la fonction et les activités mécaniques et électriques du cœur. Ensuite, nous exposons les principaux aspects de la *cohérence cardiaque* tels que le système nerveux autonome ou encore la fréquence cardiaque. Dans la deuxième partie, nous présenterons l'électrocardiogramme - ses objectifs et son utilité dans le diagnostic des arythmies. Ensuite, nous décrivons de façon détaillée les composantes du tracé électrocardiographique, notamment les ondes, les intervalles et les segments. Dans la troisième partie, nous présentons les arythmies cardiaques les plus communes, leurs indicateurs, caractéristiques et marqueurs. Enfin, la quatrième partie de ce chapitre est consacrée à la formulation du contexte et de l'objectif de la thèse, en mettant l'accent sur la fibrillation auriculaire. Un état de l'art sur les différentes approches proposées dans la littérature est également exposé. Celles-ci sont introduites en deux parties : la première concerne les méthodes univariées et la seconde les méthodes multivariées. Dans le présent chapitre, l'introduction à la cardiologie est délibérément limitée aux concepts nécessaires à la compréhension des chapitres ultérieurs. Le lecteur intéressé peut cependant trouver des informations plus détaillées dans [20, 21].

I.2 Introduction au système cardiovasculaire

I.2.1 Système cardiovasculaire

Le pompage du sang s'effectue à travers le réseau d'artères et de veines appelé système cardiovasculaire. Le système cardiovasculaire est un système circulatoire fermé qui assure le transport du cœur vers les extrémités et les divers organes et, en retour, de ceux-ci vers le cœur. Il existe deux types de circulation sanguine : la grande circulation sanguine, chargée d'alimenter les tissus en oxygène et en nutriments, et la petite circulation sanguine (circulation pulmonaire) qui assure la réoxygénation du sang et l'élimination du dioxyde de carbone (Figure.I.1).

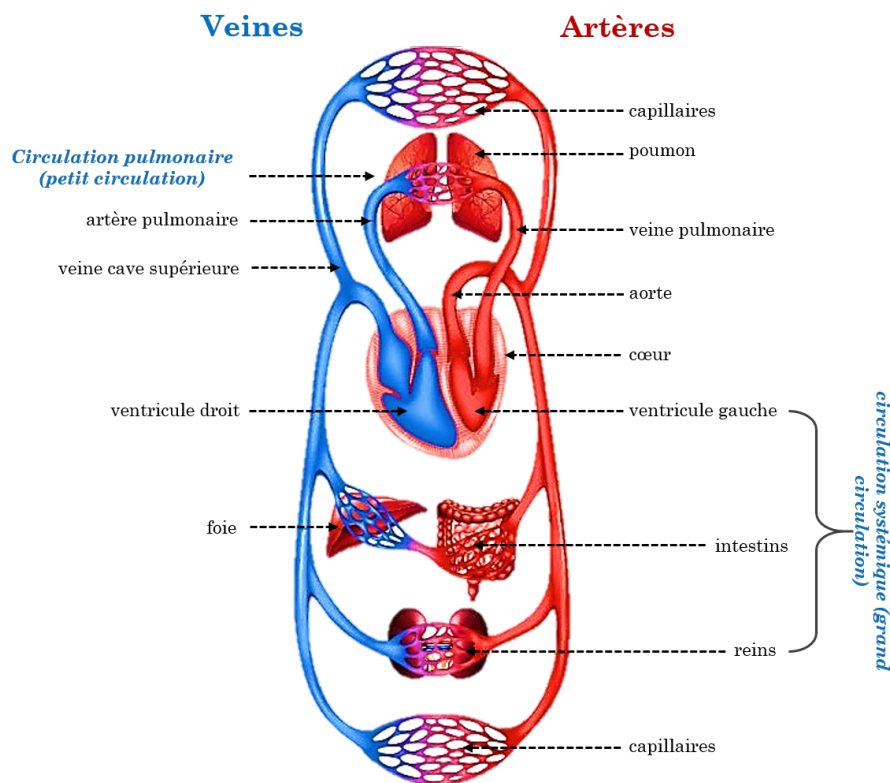


FIGURE I.1: Le système cardio-vasculaire : il s'agit d'un système circulatoire d'alimentation sanguine, composé du cœur et des vaisseaux sanguins (veines et artères). La zone rouge symbolise la circulation sanguine riche en oxygène (réseau artériel). La zone bleue symbolise la circulation sanguine pauvre en oxygène (réseau veineux).

I.2.1.1 Les vaisseaux sanguins

Les vaisseaux sanguins sont des conduits qui transportent le sang dans l'organisme. Les trois principaux types de vaisseaux sanguins sont : les artères, les veines et les capillaires.

- (a) Les veines sont des vaisseaux sanguins qui transportent le sang du corps vers le cœur.
- (b) Les artères sont des vaisseaux sanguins qui transportent le sang du cœur vers le corps.
- (c) Les capillaires sont des vaisseaux sanguins microscopiques à paroi mince qui relient les artères et les veines. Les capillaires assurent la transmission des nutriments et des déchets entre le sang et les tissus corporels.

I.2.1.2 Le cœur

Le cœur est le principal organe du système cardiovasculaire, soit la structure principalement responsable de la circulation sanguine et de la transportation des nutriments dans toutes les parties du corps à travers des contractions rythmiques. Le cœur transporte également les déchets métaboliques et le dioxyde de carbone via la circulation sanguine vers les poumons et les reins, entre autres organes, afin d'être évacués de l'organisme [22]. Elle permet ainsi l'échange entre la petite et la grande

circulation sanguine suivant un rythme régulier synchronisé par un stimulus électrique provenant du nœud sinusal. Le cœur est présenté sous trois aspects : structure et fonction ; activité mécanique ; activité électrique ; et cohérence cardiaque.

I.2.2 Structure et fonction du cœur

I.2.2.1 Structure du cœur

Le cœur est composé de quatre chambres, appelées les cavités. Les deux cavités supérieures (du haut) constituent les oreillettes, et les cavités inférieures (du bas) les ventricules. Les oreillettes sont alimentées par du sang désoxygéné. Les valves qui séparent ces oreillettes sont appelées valves atrio-ventriculaires, composées de la valve tricuspide à droite et de la valve mitrale à gauche. Les ventricules pompent le sang oxygéné dans tous les organes du corps, jusqu'aux plus petites cellules. Tout comme les oreillettes, les ventricules sont également séparés par des valves. Appelées collectivement valves semi-lunaires, elles comprennent la valve pulmonaire et la valve aortique (Figure.I.2).

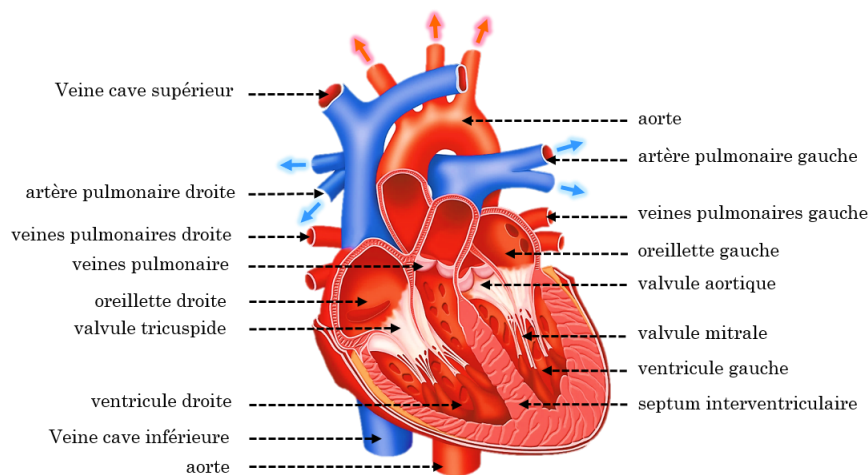


FIGURE I.2: Diagramme de l'anatomie du cœur

I.2.2.2 Fonction du cœur

La circulation sanguine dans l'organisme est assurée par la contraction et la dilatation du cœur. L'action de pompage au niveau du cœur se traduit par une alternance de mouvements de serrage et de relâchement du poing. À chaque battement, le cœur expulse le sang dans les artères, ce qui permet de créer le pouls. En général, cette circulation sanguine est assurée via quatre cavités :

1. L'oreillette droite reçoit le sang des veines et la pompe vers le ventricule droit.
2. Le ventricule droit reçoit le sang de l'oreillette droite et la pompe vers les poumons, où il est chargé d'oxygène.
3. L'oreillette gauche reçoit le sang oxygéné des poumons et la pompe vers le ventricule gauche.

4. Le ventricule gauche pompe le sang riche en oxygène vers le reste du corps. Les contractions vigoureuses du ventricule gauche créent notre pression sanguine.

Les artères cérébrales longent les parois du cœur et fournissent du sang riche en oxygène au muscle cardiaque. Un réseau nerveux circule également dans le cœur, conduisant les signaux complexes qui régissent sa contraction et relaxation.

I.2.3 Activité mécanique

Les performances cardiaques peuvent être facilement surveillées en fonction de l'analyse des activités mécaniques lorsqu'un problème ou un trouble cardiovasculaire est suspecté. En effet, un battement de cœur est une manifestation du processus de recharge en oxygène du cœur qui se compose de quatre phases majeures : systole auriculaire, systole ventriculaire, diastole auriculaire, et diastole ventriculaire. La systole est la phase durant laquelle le cœur se contracte et éjecte le sang dans les artères ; et la diastole est la phase durant laquelle le cœur se relâche et se remplit de sang. Le diagramme I.3 illustre les différentes phases de la systole-diastole du cycle cardiaque.

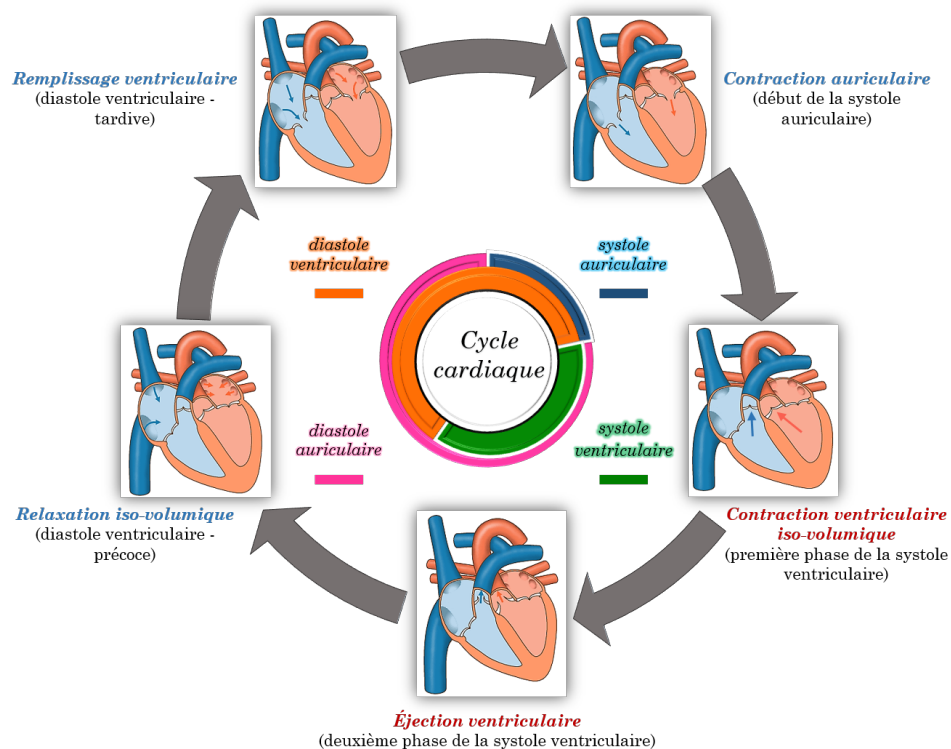


FIGURE I.3: Schéma de la circulation sanguine (cycle cardiaque)

1. Pendant la systole auriculaire, les oreillettes se contractent et pompent le sang dans les ventricules (remplissage actif). Une fois que le sang est éjecté des oreillettes, les valves auriculo-ventriculaires (valves entre les oreillettes et les ventricules) se ferment et le sang poursuit son écoulement vers les ventricules pour éviter le reflux du sang dans les oreillettes. La fermeture de ces valves provoque la résonance des pulsations cardiaques.

2. La systole ventriculaire provoque la contraction des ventricules, expulsant le sang vers le système circulatoire. En très peu de temps, les valvules sigmoïdes sont fermées. Ensuite, celle-ci s'ouvre dès que la pression à l'intérieur des ventricules dépasse la pression sanguine. Puis le sang est expulsé et les deux valves sigmoïdes (la valve pulmonaire à droite et la valve aortique à gauche) se ferment. Cela empêche le sang de refluer dans les ventricules. Lorsque les valvules sigmoïdes se ferment, une deuxième résonance cardiaque plus aiguë se produit. Cela provoque une augmentation de la pression sanguine.
3. Enfin, la diastole traduit le processus de relaxation de toutes les parties du cœur, permettant le remplissage (passif) des ventricules, par les oreillettes droites et gauches et par la veine cave et les veines pulmonaires. Les oreillettes sont remplies progressivement et le sang s'écoule dans les ventricules.

Un cœur en état de repos passe un tiers du temps en systole et deux tiers du temps en diastole. La relation systole-diastole est la référence pour mesurer la pression artérielle. Par ce mécanisme, le sang désoxygéné issu du ventricule droit est pompé vers les poumons, et le sang oxygéné est pompé pour être distribué dans l'organisme. Pendant que les ventricules se vident, les oreillettes sont déjà remplies de sang neuf et le processus recommence. La fréquence de pompage du sang dans le corps est appelée rythme cardiaque. L'examen du rythme cardiaque (battements par minute) est une méthode qui permet également de déterminer le fonctionnement régulier du cœur.

I.2.4 Activité électrique

L'excitation des cellules musculaires est le résultat d'un phénomène électrique appelé dépolarisation (la dépolarisation des cellules responsables de la systole, la phase de contraction). Cette excitation n'est pas permanente, donc après une période donnée, les cellules musculaires se remettent au repos, un phénomène électrique nommé repolarisation (la repolarisation des cellules responsables de la diastole, la phase de relaxation où le sang remplit les cavités auriculaires et ventriculaires). L'activité électrique du cœur est la somme de toutes les activités électriques de toutes les cellules du cœur. La contraction du muscle cardiaque est régulée par des impulsions électriques, qui sont générées par le nœud sinoauriculaire ou sinusal (NS), connue sous le nom de "stimulateur cardiaque naturel". Ces impulsions provoquent l'excitation des oreillettes et des ventricules. Les oreillettes et les ventricules interagissent, se contractant et se détendant alternativement pour pomper le sang dans tout l'organisme. Le système de conduction électrique comprend le nœud NS, le nœud auriculo-ventriculaire (NAV) et le réseau His-Purkinje.

1. Nœud NS - l'impulsion électrique est émise par le nœud SA, situé dans l'oreillette droite. L'activité électrique se propage à travers les parois des oreillettes en entraînant leurs contractions. Cela force le sang à pénétrer dans les ventricules. Le nœud SA détermine le rythme et la fréquence des battements de cœur. Le rythme cardiaque normal est souvent appelé rythme sinusal normal (RSN) parce que le nœud NS se déclenche régulièrement.

2. Nœud AV - le nœud AV se situe au centre du cœur, entre la paroi qui sépare les oreillettes et les ventricules, et agit comme une barrière qui ralentit le signal électrique de pénétrer dans les ventricules. Ce délai fourni à l'oreillette le temps de se contracter et que les ventricules se remplissent de sang avant de se contracter.
3. Réseau His-Purkinje - ce cheminement des fibres est responsable de déclenchement de l'impulsion électrique aux parois musculaires des ventricules et les fait se contracter. Cela force la circulation du sang du cœur vers les poumons et le corps.
4. Le nœud NS envoie une autre impulsion et le cycle recommence.

Le graphique I.4 présente les événements de l'activité électrique d'une contraction normale du cœur mis en évidence par un tracé ECG. Chaque tracé ECG correspond à un événement du cycle cardiaque. En bas : Position anatomique de l'activité électrique. En haut : son électrocardiogramme correspondant.

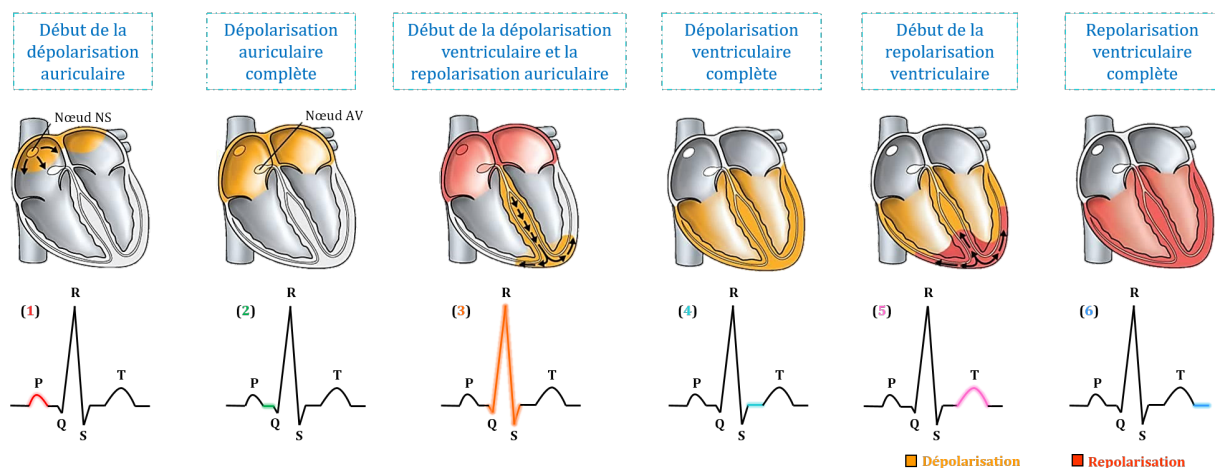


FIGURE I.4: Les événements électriques du cycle cardiaque : (1) La dépolarisation auriculaire, déclenchée par le nœud SA et provoquant l'onde P (2) La dépolarisation auriculaire terminée, l'impulsion est retardée au niveau du nœud AV. (3) La dépolarisation ventriculaire commence à l'apex, provoquant le complexe QRS. Une repolarisation auriculaire apparaît. (4) La dépolarisation ventriculaire est terminée. (5) La repolarisation ventriculaire commence à l'apex, provoquant l'onde T. (6) La repolarisation ventriculaire est complète.

En général, l'ampleur des variations électriques, la durée des épisodes et l'analyse vectorielle fine permettent d'obtenir une vision complète de la fonction cardiaque. Par exemple, une onde P amplifiée peut indiquer un élargissement des oreillettes, une onde Q élargie peut indiquer la présence d'un infarctus du myocarde, et une onde Q élargie atténuée ou inversée indique souvent un élargissement des ventricules. Les ondes T paraissent souvent plus plates lorsque l'apport d'oxygène au myocarde est insuffisant. L'infarctus du myocarde (IM), également connu sous le nom de crise cardiaque, se produit lorsque le flux sanguin diminue ou s'arrête dans une partie du cœur, ce qui provoque des troubles du muscle cardiaque.

I.2.5 Cohérence cardiaque

Le terme de cohérence cardiaque, récemment utilisé dans la littérature, correspond à l'état d'équilibre du cœur, connu médicalement sous le nom de rythme sinusal respiratoire. La cohérence cardiaque permet de renforcer l'adaptabilité du système nerveux autonome aux exigences et au stress, assurant une meilleure flexibilité envers les pathologies et aidant l'organisme à maintenir un meilleur équilibre cardio-respiratoire et cérébral. Le système nerveux autonome est responsable du contrôle des activités automatiques de l'organisme, telles que le rythme cardiaque ou la contraction musculaire, ainsi que la régulation de la fréquence cardiaque.

I.2.5.1 Système Nerveux Autonome (SNA)

Le rythme cardiaque constitue un biomarqueur intéressant pour l'activité du SNA. Le SNA régule l'activité des organes internes, y compris le cœur et les vaisseaux sanguins. Le SNA se compose de deux organes : le sympathique et le parasympathique. Chacun de ces deux systèmes a un effet opposé sur l'organisme humain. Dans les situations qui nécessitent de la vigilance et de l'énergie, comme faire face à un danger ou pratiquer des activités physiques. Le SNA active sa division sympathique afin de mobiliser le corps pour l'action. En revanche, dans les situations ordinaires, la division parasympathique est activée pour conserver et restaurer.

- (a) **Système Nerveux ParaSympathique (SNPS)** : il permet de ralentir les fonctions de l'organisme, par exemple, la fréquence cardiaque au repos et abaisse la pression sanguine.
- (b) **Système Nerveux Sympathique (SNS)** : il agit à l'inverse, son fonctionnement s'enclenche lorsque la personne est extrêmement active ou en état de stress émotionnel. Ce système a pour effet d'augmenter le rythme cardiaque (accélère le rythme respiratoire) et la pression artérielle, et de réduire la variabilité de la fréquence cardiaque.

Toute distorsion du fonctionnement du SNA peut entraîner des pathologies, voire des décès.

I.2.5.2 Fréquence Cardiaque (FC)

La fréquence cardiaque est l'un des "signes vitaux", ou indicateurs importants de la santé du corps humain, exprimée en nombre de battements par minute (bpm) - ce nombre est un indicateur de la charge de travail de l'appareil cardiovasculaire (système cardiovasculaire). La fréquence cardiaque varie en fonction des besoins en oxygène selon la nature des activités pratiquées. Il peut être classifié en fonction qu'il soit faible, normal ou élevé :

- (a) **Bradycardie** : fréquence cardiaque lente
- (b) **Normo fréquent** : fréquence cardiaque normale
- (c) **Tachycardie** : fréquence cardiaque élevée

Une fréquence cardiaque normale ne suffit pas à assurer l'absence de problèmes de santé, mais elle permet d'identifier un certain nombre de pathologies. Selon l'American Heart Association [23],

la fréquence cardiaque normale d'un adulte au repos est de 60 à 100 bpm. La tachycardie est un rythme cardiaque rapide, défini comme supérieur à 100 bpm au repos. La bradycardie est un rythme cardiaque lent, défini comme étant inférieur à 60 bpm au repos.

Surveiller sa fréquence cardiaque peut aider à identifier les problèmes de santé. Les anomalies du rythme cardiaque sont parfois révélatrices d'une maladie. Les tableaux I.1 et I.2 indiquent les valeurs standard de la fréquence cardiaque au repos en fonction du sexe, de l'âge et de la condition physique [23].

TABLE I.1: Fréquence cardiaque au repos chez les hommes, en fonction de l'âge (ans) et de la condition physique.

H/Condition physique	18-25	26-35	36-45	46-55	56-65	65+
Athlète	49-55	49-54	50-56	50-57	51-56	50-55
Excellente	56-61	55-61	57-62	58-63	57-61	56-61
Bonne	62-65	62-65	63-66	64-67	62-67	62-65
Au-dessus de la moyenne	66-69	66-70	67-70	68-71	68-71	66-69
Moyenne	70-73	71-74	71-75	72-76	72-75	70-73
En dessous de la moyenne	74-81	75-81	76-82	77-83	76-81	74-79
Mauvaise	82+	82+	83+	84+	82+	80+

TABLE I.2: Fréquence cardiaque au repos chez les femmes, en fonction de l'âge (ans) et de la condition physique.

F/Condition physique	18-25	26-35	36-45	46-55	56-65	65+
Athlète	56-60	54-59	54-59	54-60	54-59	54-59
Excellente	61-65	60-64	60-64	61-65	60-64	60-64
Bonne	66-69	65-68	65-69	66-69	65-68	65-68
Au-dessus de la moyenne	70-73	69-72	70-73	70-73	69-73	69-72
Moyenne	74-78	73-76	74-78	74-77	74-77	73-76
En dessous de la moyenne	79-84	77-82	79-84	78-83	78-83	77-84
Mauvaise	85+	83+	85+	84+	84+	84+

Une fréquence cardiaque élevée dépassant les recommandations de la fréquence du pouls au repos peut être un signe de stress ou de mauvaise condition physique, ce qui nécessite un entraînement du muscle cardiaque. La pratique d'une activité physique et l'adoption d'une alimentation saine peuvent entraîner la circulation et donc le muscle cardiaque. Si le rythme cardiaque demeure élevé, cela pourrait être dû à une arythmie cardiaque dans certains cas, tel que la fibrillation auriculaire.

I.3 Électrocardiogramme (ECG)

Le cœur possède un rythme intégré de contraction et de relaxation, les cellules du muscle cardiaque, appelées pacemaker, contribuent à ce rythme. Le stimulateur cardiaque génère une impulsion électrique qui se propage dans les oreillettes, provoquant une contraction de celles-ci. Cette impulsion se propage ensuite aux ventricules, causant leurs contractions. Les changements électriques propagés au niveau du cœur peuvent être détectés à la surface du corps à l'aide d'un instrument appelé électrocardiographe. Le médecin utilise cet instrument pour évaluer la santé cardiaque du patient. À cette fin, un examen spécifique est mené en utilisant des électrodes placées dans un certain nombre de points sur la poitrine du patient afin d'enregistrer les changements électriques du cœur sur un papier graphique mobile, ce tracé est connu sous le nom d'électrocardiogramme (ECG) [24].

I.3.1 Objectif et principe de l'ECG

L'ECG est la méthode classique d'enregistrement des activités électriques à court terme du cœur pendant les cycles cardiaques (dépolarisation et repolarisation des oreillettes et des cavités cardiaques). Ceci est effectué en utilisant un ensemble d'électrodes connectées à différents points du corps humain. Ce procédé constitue une pratique de diagnostic clinique indispensable qui permet aux cardiologues d'observer les détails des fonctions cardiaques normales et anormales. L'électrocardiographe standard (l'instrument qui génère un ECG) utilise 3, 5 ou 12 dérivations. Plus l'électrocardiographe utilise un grand nombre de dérivations, plus l'ECG fournit davantage d'informations. Le terme "dérivation" peut être employé comme référence à la liaison entre l'électrode et l'enregistreur électrique, mais il décrit généralement la différence de tension entre deux électrodes. L'électrocardiographe à 12 dérivations utilise 10 électrodes placées à des endroits standard sur la peau du patient. Dans les électrocardiographes ambulatoires continus, le patient porte un petit appareil portable fonctionnant sur batterie, appelé moniteur Holter, ou simplement Holter, qui surveille en permanence l'activité électrique du cœur, pendant une période de 24 heures au cours de la routine normale du patient [24].

Le signal ECG est utilisé pour vérifier que l'impulsion provient bien du nœud sinusal - le rythme cardiaque est appelé rythme sinusal normal - et sa distribution dans le cœur est conforme à une séquence rigoureusement ordonnée. Le flux électrique doit être conduit sans ralentissement ni retard. Les amplitudes et les durées des ondes électriques peuvent être mesurées pour diagnostiquer une hypertrophie de la paroi cardiaque, une dilatation des cavités ou un manque de dépolarisation dans certaines zones (infarctus du myocarde, etc.). L'ECG peut également être utilisé pour diagnostiquer de nombreux problèmes cardiaques tels que des accélérations anormales du rythme cardiaque (tachycardie) ou des ralentissements anormaux (bradycardie), blocages auriculo-ventriculaires à travers l'examen de l'onde P, ainsi que les fibrillations auriculaires et ventriculaires qui sont symbolisées par une contraction anarchique des ventricules [25].

Le signal ECG se compose généralement de trois volets principaux : une onde P, un complexe QRS et une onde T [26, 24]. L'onde P est une onde lisse de faible amplitude générée par la dépolarisation de

l'oreillette, qui fait référence à la caractéristique électrique responsable de la contraction auriculaire. Les oreillettes commencent à se contracter environ 25 ms après le début de l'onde P. Le complexe QRS de forte amplitude fait référence à la dépolarisation ventriculaire, formée lors de la réduction des cavités cardiaques en état d'excitation. Le complexe QRS est responsable des contractions des ventricules gauche et droit. Les ventricules commencent à se contracter lorsque le QRS atteint le pic de l'onde R. Enfin, la repolarisation ventriculaire est représentée par l'onde T d'amplitude moyenne. La repolarisation des oreillettes se produit pendant le complexe QRS, ce qui la masque sur un ECG.

La figure I.5 présente un tracé ECG normal. Chaque onde, segment et intervalle est étiqueté et décrit les événements électriques importants dans un cycle cardiaque.

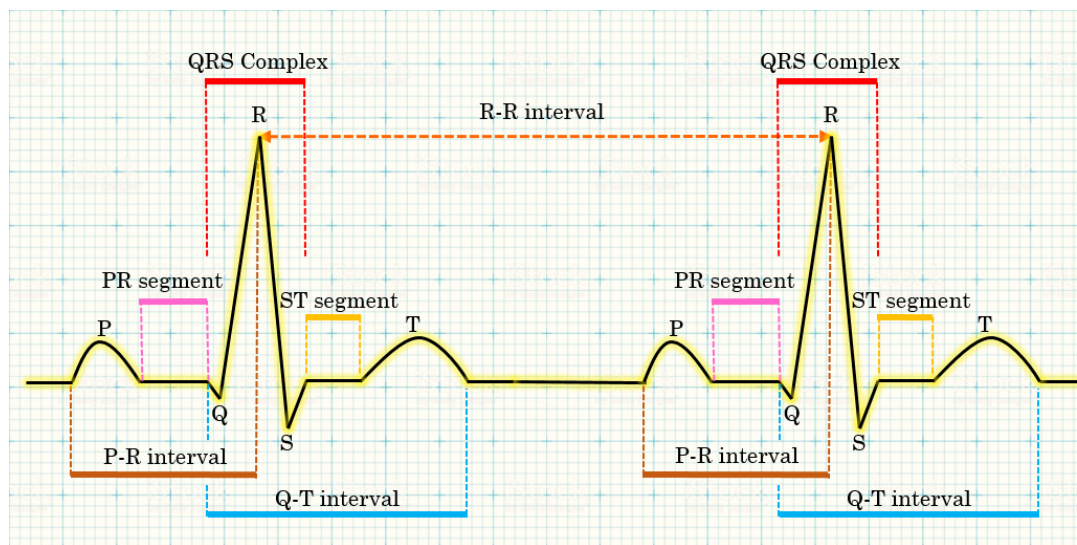


FIGURE I.5: Le tracé normal d'un ECG indique l'onde P, le complexe QRS et l'onde T, les intervalles PR, RR, et QT, ainsi que les segments PR et ST.

I.3.2 Forme d'onde et composantes d'un ECG normal

I.3.2.1 Ondes ECG

I.3.2.1.1 L'onde P : la première onde détectable, l'onde P, apparaît à l'arrivée du flux dans le tissu myocardique de l'oreillette, dépolarisant les oreillettes (phase de la systole auriculaire). L'onde P a une durée normale < 120 ms et peut être positive ou négative. En général, son observation reste délicate, notamment en présence de bruit. Il existe de nombreux rythmes cardiaques dans lesquels les ondes P ne peuvent pas être identifiées, notamment la fibrillation auriculaire et parfois les rythmes jonctionnels. Parfois, les ondes P peuvent être enterrées à la fin du complexe QRS, ce qui provoque un scénario de "RP court", tel qu'on le voit dans la tachycardie supraventriculaire [27, 28]. En cas d'ondes P similaires et de forme normale, le noeud SA est le principal stimulateur cardiaque. Dans ce cas, le rythme est de nature sinusale. En revanche, si les ondes P ont une forme variable ou sont absentes, d'autres tissus assurent le fonctionnement du stimulateur, entraînant des anomalies cardiaques [28].

I.3.2.1.2 Le complexe QRS : comme son nom l'indique, le complexe QRS est composé de l'onde Q, de l'onde R et de l'onde S. Les trois ondes se succèdent rapidement. Le complexe QRS représente l'impulsion électrique qui se propage dans les ventricules et indique la dépolarisation ventriculaire. Les complexes QRS ne comprennent pas nécessairement les ondes Q, R et S. La convention est que l'onde Q est toujours négative et que l'onde R est la première onde positive du complexe. Dans les conditions normales, la durée du complexe QRS est comprise entre 0,06 et 0,10 seconde chez un adulte. Généralement, la durée maximale du complexe QRS ne doit pas dépasser 0,10 seconde. En effet, un complexe élargi indique un élargissement ventriculaire (hypertrophie) ou une dépolarisation ventriculaire provoquée par le tissu du stimulateur cardiaque sous le nœud AV. La forme du complexe QRS dans la figure (I.5) est idéalisée. En fait, la forme change en fonction des électrodes d'enregistrement utilisées. Elle change également lorsqu'il y a une conduction anormale des impulsions électriques dans les ventricules [25].

I.3.2.1.3 L'onde T : suit le complexe QRS pour indiquer la repolarisation ventriculaire. Par rapport à une onde P, l'onde T est légèrement asymétrique et son pic est plus proche de sa fin que de son début. Sa durée varie de 0,20 à 0,25 seconde et suit généralement la même direction que le complexe QRS qui le précède. Lorsqu'une onde T se produit dans la direction opposée du complexe QRS, elle reflète généralement une forme de pathologie cardiaque. Parfois, une onde de faible amplitude peut se produire entre l'onde T et l'onde P, qui peut être une onde U. Le fondement biologique d'une onde U est inconnu. L'onde T a une durée plus longue que le complexe QRS. La raison est que la conduction de l'onde de repolarisation est plus lente que celle de l'onde de dépolarisation.

I.3.2.2 Segments et intervalles ECG

I.3.2.2.1 L'intervalle P-R (ou P-Q) : on appelle intervalle P-R la distance entre le début de l'onde P et le début du complexe QRS (Q ou R), qui est normalement comprise entre 0,12 et 0,20 seconde. Compte tenu des phases du cycle cardiaque, l'intervalle P-R mesure le temps entre le début de la dépolarisation auriculaire et le début de la dépolarisation ventriculaire (c.-à-d., le temps de propagation de l'impulsion électrique entre le nœud sinusal et les ventricules). Si l'intervalle P-R est $> 0,20$ second, cela signifie que l'impulsion est retardée pour pénétrer dans les ventricules, ce qui entraîne une conduction AV est bloqué (bloc cardiaque du premier degré) [27, 29].

I.3.2.2.2 L'intervalle Q-T : désigne la durée de la systole ventriculaire, soit de l'excitation des ventricules (début du complexe QRS) jusqu'à la fin de leur relaxation (fin de l'onde T). Sa durée normale varie de 0,20 à 0,44 seconde en fonction de la fréquence cardiaque. À des fréquences cardiaques élevées, les potentiels d'action ventriculaires diminuent en durée, ce qui réduit l'intervalle Q-T. Étant donné que des intervalles Q-T prolongés peuvent être considérés comme des signes de susceptibilité à certains types de tachyarythmie. Il est donc important de déterminer si un intervalle Q-T est excessivement long [25].

I.3.2.2.3 L'intervalle R-R : désigne la période du cycle cardiaque mesurée entre deux ondes R successifs du complexe QRS. Les propriétés intrinsèques du nœud sinusal ainsi que les facteurs du système nerveux autonome (SNA) affectent la durée de l'intervalle R-R. En rythme sinusal, ces intervalles doivent être réguliers et constants. L'analyse des intervalles R-R peut servir à détecter les arythmies ainsi qu'à étudier la variabilité de la fréquence cardiaque [30, 31]. La morphologie et les caractéristiques de l'intervalle R-R sont décrites plus en détail dans le chapitre suivant.

I.3.2.2.4 Le segment PR (ou PQ) : ce segment est compris dans l'intervalle P-R, qui commence à la fin de l'onde P et se termine au début du complexe QRS. Compte tenu des phases du cycle cardiaque, ce segment décrit le temps entre la fin de la systole auriculaire gauche et le début de la systole électrique ventriculaire. Toute irrégularité dans sa durée peut entraîner un dysfonctionnement ventriculaire dû au remplissage anormal induit. Le segment PR est généralement isoélectrique puisqu'elle coïncide avec la partie initiale de la repolarisation auriculaire, mais elle peut être déviée vers le bas ou vers le haut (selon la polarisation de l'onde P) [32].

I.3.2.2.5 Le segment ST : la période isoélectrique (segment ST) qui suit le QRS et qui se termine au début de l'onde T indique la période de dépolarisation complète des deux ventricules. En principe, un segment ST normal est isoélectrique à la ligne de base, bien qu'il puisse être légèrement élevé ou abaissé par rapport à la ligne de base [33].

I.4 Vue d'ensemble du rythme cardiaque

I.4.1 Rythme sinusal

Les pulsations cardiaques peuvent être générées par 5 stimulateurs cardiaques : le nœud SA, le nœud AV, un foyer ectopique auriculaire, les cellules ventriculaires, ou un pacemaker artificiel. Le rythme sinusal prend naissance au niveau du nœud sinusal, au cours duquel la NS active la contraction du rythme cardiaque en émettant des impulsions à une fréquence de 60 à 100 bpm pour un rythme sinusal normal (RSN). En pratique, la fréquence cardiaque s'apparente à la fréquence ventriculaire. Celle-ci est mesurée en comptant le nombre de complexes QRS par minute. En fonction de la fréquence cardiaque, le rythme sinusal peut être qualifié de RSN, de bradycardie sinusale, de tachycardie sinusale ou d'arythmie sinusale. En effet, le cœur se ralentit (bradycardie) pendant le sommeil et il s'accélère (tachycardie) en fonction de l'émotion, lors de l'activité physique, dans des circonstances physiologiques qui nécessitent une surcharge métabolique ou sous l'effet d'émotions ou existantes [34]. En outre, l'arythmie sinusale désigne un rythme irrégulier variable avec la respiration, c'est-à-dire que le rythme cardiaque augmente fréquemment avec l'inspiration et diminue avec l'expiration. Le rythme sinusal possède certaines caractéristiques qui peuvent être observées dans le tracé de l'ECG. En effet, pour assurer une conduction cardiaque normale, plusieurs facteurs doivent être satisfaits :

1. Les intervalles entre les complexes QRS (intervalles R-R) doit être constants, ce qui indique que le rythme ventriculaire est régulier.
2. La morphologie de l'onde P auriculaire et du complexe QRS ventriculaire doit être normale.
3. Chaque onde P doit être suivie d'un complexe QRS, qui est suivi d'une onde T. Cela assure une séquence normale pour chaque cycle cardiaque. La détection des anomalies de conduction entre les oreillettes et les ventricules passe par l'analyse de cette séquence.
4. Les intervalles entre les ondes P (intervalles P-P) doit être cohérents, ce qui indique que le rythme auriculaire est régulier.

L'analyse du signal ECG consiste à observer la forme d'onde pour détecter toute anomalie telle que des altérations des complexes QRS, la durée du segment ST ou toute divergence comme des ondes P ou T manquantes. L'analyse des signaux ECG doit être méthodique et systématique, et les caractéristiques temporelles des formes d'onde doivent être soigneusement vérifiées. Concernant l'analyse du signal, il convient d'étudier :

1. La fréquence cardiaque
2. Durée et amplitude de l'onde P
3. Durée de l'intervalle P-R
4. Durée et morphologie du complexe QRS
5. Durée et morphologie de l'intervalle S-T, plus l'amplitude de l'onde T

Les caractéristiques des rythmes sinusaux sont indiquées dans le tableau I.3.

TABLE I.3: Caractéristiques des rythmes sinusaux.

Onde & Intervalle	Caractéristiques			
	RSN	Tachycardie	Bradycardie	Arythmie sinusale
FC (bpm)	60 – 120	> 120	< 60	fréquence variable en fonction de l'inspiration
Rythme	régulier	régulier	régulier	irrégulier, variant en fonction de la respiration
onde P	< 120 ms 0.15 – 0.4 mV	normal	normal	normal
intervalle P-R (ms)	120 – 200	120 – 200	120 – 200	120 – 200
durée du QRS (ms)	60 – 100	60 – 100	60 – 100	60 – 100

I.4.2 Blocs cardiaques

Les rythmes de bloc cardiaque se produisent lorsque l'impulsion électrique cardiaque est retardée ou bloquée dans le nœud AV, le faisceau de His ou le système de Purkinje. Les rythmes du bloc cardiaque comprennent :

I.4.2.1 Blocs auriculoventriculaires cardiaques

Le système de conduction électrique du cœur transmet les signaux électriques de l'oreillette via le nœud AV (nœud auriculoventriculaire). Le nœud AV est responsable de la circulation des signaux entre les cavités supérieures du cœur (oreillettes) et les cavités inférieures (ventricules). Les signaux provenant du nœud AV passent par plusieurs faisceaux de fibres (branches) avant de se propager à travers les ventricules. Les blocs auriculoventriculaires (BAV) peuvent se produire dans n'importe quelle partie de ce système de conduction [27, 29]. Un bloc AV est caractérisé par le ralentissement ou l'interruption de transmission électrique entre les oreillettes et les ventricules. Les blocs AV sont répartis en trois niveaux, en fonction de leur gravité :

1. Le bloc AV du premier degré, le moins grave, se caractérise par un intervalle P-R prolongé. Il s'agit de la forme de bloc qui ne nécessite aucun traitement particulier et ne provoque aucun symptôme perceptible. Lors d'un bloc AV du 1er degré, le passage du stimulus à travers le nœud AV ou le système His-Purkinje subit un retard qui repousse l'apparition du complexe QRS. Sur l'ECG, ce type de bloc se traduit par la présence d'ondes P suivies de complexes QRS, sauf que l'intervalle P-R est prolongé sur une durée de 0,20 seconds ou plus.
2. Lors d'un bloc AV du deuxième degré, la circulation du stimulus des oreillettes vers les ventricules est interrompue de façon discontinue. Cela se traduit par des ondes P non conduites, c'est-à-dire non suivies d'un complexe QRS. Ce type de bloc peut provoquer des symptômes gênants tels que les vertiges nécessitant la mise en place d'un stimulateur cardiaque.
3. On parle de bloc AV complet ou du troisième degré, le type le plus grave, lorsque les signaux électriques ne se propagent pas entre les cavités supérieures et inférieures du cœur (interruption complète de la conduction AV). Ce type de blocage est plus fréquent chez les patients atteints de maladies cardiaques. Sans stimulateur cardiaque, le patient est exposé à un risque sérieux de crise cardiaque.

I.4.2.2 Bloc de branche

Un bloc de branche est un trouble de la conduction lié à un blocage d'une des branches du faisceau de His localisé dans les ventricules cardiaques. Le faisceau de His est un groupe de fibres qui conduisent les impulsions électriques du nœud AV, et est composé de deux branches : la branche gauche et la branche droite. La branche gauche est chargée de conduire les impulsions vers le ventricule gauche, tandis que la branche droite assure la conduction des impulsions vers le ventricule droit. La conduction

des impulsions peut être bloquée dans la branche gauche ou dans la branche droite. Le blocage de la branche droite ne provoque généralement aucun symptôme et peut survenir chez des personnes apparemment en bonne santé. En revanche, le blocage de la branche gauche se produit le plus souvent chez les personnes âgées, et sa gravité est plus grande. Ce blocage est souvent le signe d'une maladie coronarienne due à l'athérosclérose.

I.4.3 Arythmies cardiaques

Les arythmies constituent l'une des principales causes de mortalité à l'échelle mondiale et contribuent essentiellement au problème des maladies cardio-vasculaires. L'arythmie est une anomalie qui affecte le comportement normal du rythme du cœur, c'est-à-dire que le cœur a tendance à manquer des battements ou à battre irrégulièrement, ou encore à battre très vite ou très lentement, ce qui entraîne une diminution de l'efficacité du pompage cardiaque. L'arythmie peut être caractérisée soit par l'irrégularité d'un seul battement de cœur (battement arythmique), soit par l'irrégularité d'un ensemble de battements de cœur (épisode arythmique). La présente étude est consacrée au diagnostic des épisodes arythmiques. En cas d'arythmie, le rythme cardiaque est perturbé, entraînant des battements irréguliers, tels que des battements lents ou rapides. Certaines formes d'arythmies peuvent se manifester chez des personnes en bonne santé et ne provoquer aucune complication grave, néanmoins, elles sont parfois accompagnées de divers signes, tels que l'anxiété, la fatigue, les étourdissements, la transpiration, les évanouissements ou les douleurs thoraciques. Par exemple, l'arythmie sinusale respiratoire, un phénomène naturel de variation périodique du rythme cardiaque, correspondant à l'activité respiratoire. Sa fréquence augmente pendant l'inspiration et diminue pendant l'expiration. Sur l'ECG, l'intervalle RR devient plus court pendant la phase d'inspiration et plus longue pendant la phase d'expiration. D'autres formes d'arythmie présentent un danger majeur, qui affecte la circulation sanguine et nécessite une intervention médicale immédiate. Sur l'ECG, les arythmies les plus courantes sont caractérisées par une variabilité irrégulière des intervalles RR, ainsi qu'une séquence cardiaque désordonnée et déséquilibrée. Une arythmie non traitée peut entraîner un accident vasculaire cérébral, une crise cardiaque, une insuffisance cardiaque ou une mort subite. Les signes d'arythmie, qui peuvent être observés sur un tracé ECG, sont présentés dans le tableau I.4. L'arythmie apparaît lorsque :

- (a) Le stimulus électrique se produit à l'extérieur du nœud sinusal, c'est-à-dire, dans le nœud AV.
- (b) Le signal électrique se propage sur d'autres circuits que ceux de la propagation normale
- (c) Le stimulateur cardiaque naturel du cœur développe des pulsations anormales.
- (d) Le stimulateur cardiaque naturel du cœur est actionné par une autre partie du cœur.
- (e) La conduction sanguine normale est bloquée.

Les arythmies sont classées en fonction de leur origine dans le cœur et de leurs effets sur le rythme cardiaque. Si elles se produisent dans les oreillettes ou le nœud AV, on parle d'arythmie supraventriculaire [12]; si elles se produisent dans les ventricules, on parle d'arythmie ventriculaire [13, 14].

TABLE I.4: Interprétation électrocardiographique anormale.

Composante	Description	Causes
onde P	anormal	hypertrophie auriculaire gauche ou droite, extrasystoles
onde P	Absent	fibrillation auriculaire, hyperkaliémie
intervalle P-R	long	bloc AV du premier degré
complexe QRS	large	flutter ou fibrillation ventriculaire, hyperkaliémie
intervalle Q-T	long	infarctus du myocarde, hyperkaliémie, hypocalcémie, accident vasculaire cérébralsyndrome congénital du QT long
intervalle Q-T	court	hypercalcémie, hypermagnésémie
onde T	aplatis ou inversé	troubles de la conduction, troubles électrolytiques, hypertrophie ventriculaire gauche

I.4.4 Les formes d'arythmie

I.4.4.1 Arythmies auriculaires

Les arythmies auriculaires se produisent dans la zone située au-dessus des ventricules, généralement dans les cavités supérieures du cœur, appelées oreillettes. Parmi les formes d'arythmie auriculaires les plus courantes, on peut citer : la fibrillation auriculaire (FA), le flutter auriculaire (FLA), la tachycardie supraventriculaire (TSV). Il existe également d'autres formes d'arythmie auriculaire, notamment : la tachycardie auriculaire ectopique et l'extrasystole auriculaire.

I.4.4.1.1 Tachycardie supraventriculaire : La tachycardie supraventriculaire (TSV) est un rythme cardiaque rapide et régulier caractérisé par une fréquence cardiaque de 150 à 250 bpm dans les oreillettes. La TSV désigne également la tachycardie paroxystique supraventriculaire (TPSV), où le mot "paroxystique" signifie occasionnellement ou de temps en temps. La TSV ou TPSV se produit lorsque les signaux électriques des cavités supérieures du cœur se déclenchent de manière anormale, ce qui interfère avec les signaux électriques provenant du nœud SA (le stimulateur cardiaque naturel du cœur) [12, 13].

I.4.4.1.2 Flutter auriculaire : Cette forme d'arythmie se caractérise par la présence de nombreuses ondes entre les complexes QRS sans pouvoir déterminer s'il s'agit d'ondes P ou T. Cette forme se produit lorsqu'un stimulateur cardiaque ectopique émerge dans le muscle auriculaire et se déclenche plus rapidement que le nœud SA. Cela génère de multiples dépolarisations dans le muscle auriculaire, qui se traduisent par des ondes appelées flutter. En principe, le nœud AV ne permet le passage que d'une seule de ces ondes à travers le ventricule à chaque cycle, ce qui entraîne une réponse ventriculaire régulière [13].

I.4.4.1.3 Fibrillation auriculaire : La fibrillation auriculaire (FA) se produit lorsque plusieurs stimulateurs ectopiques apparaissent dans le muscle auriculaire et se déclenchent à une vitesse supérieure à celle du nœud sinusal. Cela génère de multiples dépolarisations dans le muscle auriculaire, beaucoup plus fréquentes que celles associées au flutter auriculaire. Le nœud AV est tellement surchargé d'impulsions qu'il ne peut en transmettre aucune de façon régulière [12, 13].

I.4.4.2 Arythmies ventriculaires

Les arythmies ventriculaires se produisent dans les cavités inférieures du cœur, appelées ventricules. Deux principaux types d'arythmies ventriculaires graves sont diagnostiqués, à savoir la tachycardie ventriculaire et la fibrillation ventriculaire. Il existe également d'autres formes d'arythmie ventriculaire moins grave comme l'extrasystole ventriculaire [27, 29, 14].

I.4.4.2.1 Tachycardie ventriculaire : La tachycardie ventriculaire (TV) est un rythme cardiaque rapide et régulier provenant de signaux électriques anormaux émis par les ventricules. La rapidité du rythme cardiaque empêche les ventricules de se remplir et de se contracter efficacement pour pomper suffisamment de sang vers l'organisme. Cela du au fait que le nœud SA ne contrôle plus les battements des ventricules. À la place, d'autres parties du circuit électrique inférieur jouent le rôle de stimulateur cardiaque. La tachycardie ventriculaire ne provoque pas nécessairement des troubles graves chez les personnes ayant un cœur sain, mais elle peut constituer une urgence médicale qui nécessite un traitement rapide en cas de cardiopathie ou d'affaiblissement du cœur [14].

I.4.4.2.2 Fibrillation ventriculaire : La fibrillation ventriculaire (FV) est une arythmie cardiaque mortelle, caractérisée par le remplacement de la contraction coordonnée du myocarde ventriculaire par une excitation désorganisée à haute fréquence, ce qui se traduit par une insuffisance de pompage sanguin du cœur. La FV est l'arythmie la plus couramment identifiée chez les patients victimes d'un arrêt cardiaque. En effet, la FV provoque une perte de conscience et sans traitement rapide, la mort s'ensuit. La réanimation cardio-pulmonaire (RCP) doit être effectuée en quelques minutes, suivie d'une défibrillation (un choc électrique administré à la poitrine) pour rétablir un rythme cardiaque normal [14].

I.4.5 Diagnostic de l'arythmie

Pour diagnostiquer une arythmie, le médecin procède à un examen visuel standard des enregistrements des données ECG. L'analyse des enregistrements ECG et Holter-ECG restent la technique la plus couramment utilisée pour détecter les arythmies. Les principales caractéristiques morphologiques permettant de détecter les arythmies cardiaques sont l'absence d'ondes P, la présence d'irrégularités dans les intervalles R-R ou encore une combinaison des deux, ainsi que la morphologie des composantes ECG (ondes et intervalles). Le traitement conventionnel des enregistrements ECG demande

un temps considérable et est peu rentable. En effet, une grande masse de données électrocardiographiques réduira inévitablement l'efficacité du diagnostic de l'arythmie et augmentera la charge de travail des médecins. En outre, la présence de bruit et de battements cardiaques ectopiques dans l'ECG est susceptible de compliquer l'interprétation et l'identification des symptômes associés à des formes anormales du rythme cardiaque. Dans ce cas, les inspections visuelles peuvent donc aussi conduire à un mauvais diagnostic de l'arythmie, même dans des épisodes courts que le cardiologue peut gérer confortablement. Nous sommes donc confrontés à un problème de santé très sensible et d'une importance capitale, qui peut être efficacement traité via l'adoption de modèles automatisés intelligents permettant de détecter une arythmie et d'apporter un soutien quotidien aux médecins.

I.4.6 Base de données sur l'arythmie

Les études proposées pour le diagnostic de l'arythmie sont principalement validées en fonction des performances obtenues sur les signaux fournis par Physionet, une base de données ECG accessible en ligne [35]. Six ensembles de données sont couramment exploités :

1. Long term AF database (LTAFDb) [36] : cette base de données comprend 84 enregistrements ECG à long terme provenant de sujets atteints de FA paroxystique et persistant, ainsi que d'autres arythmies. Chaque enregistrement est échantillonné à 128 Hz et leurs durées varient entre 24 à 25 heures.
2. MIT-BIH Atrial Fibrillation Database (MITAFDb) [37] : cette base de données comprend 25 sujets, chacun à une durée d'enregistrement de 10 heures, échantillonné à 250 Hz. Cette base de données comprend les sujets RSN, FA, AFL et J.
3. MIT-BIH Arrhythmia Database (MITArrDb) [38] : cette base de données se compose de 48 segments de 30 minutes d'enregistrements ECG, échantillonnés à 360 Hz chacun. Cette base de données contient des données de RSN, de FA et de nombreuses autres pathologies.

La composition de ces trois premières bases de données est présentée dans le tableau I.5 avec les rythmes cardiaques inclus. En outre, il existe également trois bases de données distinctes, composées d'un seul rythme cardiaque, couramment utilisé, appelé :

4. AF Termination Challenge Database (AFTCDb) [39] : cette base de données contient 80 enregistrements, chacun étant un segment d'une minute de FA, échantillonné à 128 Hz.
5. MIT-BIH Normal Sinus Rhythm Database (MITNSRDb) [40] : cette base de données comprend 18 enregistrements ECG, d'une durée comprise entre 20 et 24 heures chacune, échantillonnés à 128 Hz. Celles-ci sont prélevées sur des personnes qui ne présentent aucune anomalie cardiaque.
6. Normal Sinus Rhythm RR Interval Database (NSRRRIDb) [41] : cette base de données comprend des enregistrements cardiaques de 54 sujets ayant un rythme sinusal normal. Chaque enregistrement a été échantillonné à 128 Hz, et les annotations des battements cardiaques ont été obtenues par l'analyse automatisée d'un examen et d'une correction manuels.

TABLE I.5: Présentation des bases de données des rythmes variés. Les chiffres indiqués représentent le nombre total d'épisodes de 30 secondes extraites de chaque rythme.

Abréviation	Description	LTAfDb	MITArrDB	MITAFDb
RSN	rythme sinusal normal	72 214	1872	17 587
FA	fibrillation auriculaire	69 632	223	11 006
BA	bigéminine auriculaire	123	2	
AFL	flutter atrial		16	163
B	bigéminine ventriculaire	32	33	
P	rythme du stimulateur cardiaque		193	
PREX	préexcitation ventriculaire		4	
SBR	bradycardie sinusale	2633	58	
SVTA	tachyarythmie supraventriculaire	77	3	
T	trigonométrie ventriculaire	11	11	
VFL	Flutter ventriculaire		3	
VT	tachycardie ventriculaire		3	
IVR	rythme idioventriculaire		3	
J	fibrillation jonctionnelle			611

I.5 Contexte et objectifs

I.5.1 Fibrillation auriculaire : l'arythmie la plus courante

I.5.1.1 types et caractéristiques

La forme la plus commune d'arythmie est la fibrillation auriculaire (FA), surnommée "tueur silencieux" du fait qu'elle ne produit pas toujours de symptômes pour prévenir le patient ou le médecin de son apparition. La FA est associée à des facteurs de risque critiques, tels que les accidents vasculaires cérébraux, les caillots sanguins, une morbidité et une mortalité accrue. En fonction de la durée de son apparition, un épisode de FA peut être classé en trois formes principales.

- La FA paroxystique : il s'agit d'une forme d'anomalie qui apparaît soudainement puis s'arrête spontanément. Selon sa sensibilité et l'état de santé du patient, sa durée peut varier de quelques secondes, minutes, heures ou jusqu'à 7 jours avant que le rythme normal ne soit rétabli.
- FA persistante : une forme d'anomalie qui persiste plus de sept jours et ne s'arrête pas spontanément. Pour rétablir un rythme normal, un traitement approprié est donc nécessaire, à savoir la prise de médicaments, une ablation par cathéter ou un choc électrique (appelé cardioversion électrique). Sans traitement, le rythme cardiaque reste irrégulier.

- FA permanente : il s'agit d'une arythmie cardiaque permanente qui est irrémédiable. La prise de médicaments et les chocs électriques contrôlés ne permettent pas non plus de rétablir un rythme normal. La FA permanente fait référence à la décision mutuelle du patient et du médecin de renoncer à toute autre tentative de restauration du rythme sinusal.

La plupart de ces formes de FA se présentent chez les personnes âgées et peuvent avoir de nombreuses causes différentes.

I.5.1.2 FA vs. RSN

En pratique, un rythme cardiaque normal peut être identifié dans le signal ECG par un Rythme Sinusal Normal (RSN), où les intervalles RR sont réguliers, ainsi que par l'existence d'ondes P. En revanche, le diagnostic de la FA, par exemple, est confirmé sur l'ECG par l'absence de l'onde P, un intervalle RR irrégulier et, pour certains patients, une onde fibrillaire apparaissant sur la ligne de base de l'ECG (voir Fig.I.6). La plupart des systèmes ambulatoires s'appuient sur l'analyse de l'intervalle RR pour la détection de la FA. Cela est dû au fait que l'onde P, et en particulier l'onde fibrillaire, a un faible rapport signal/bruit en phase ambulatoire. Toutes ces connaissances constituent les principales références morphologiques nécessaires pour distinguer un épisode de FA d'un épisode de RSN.

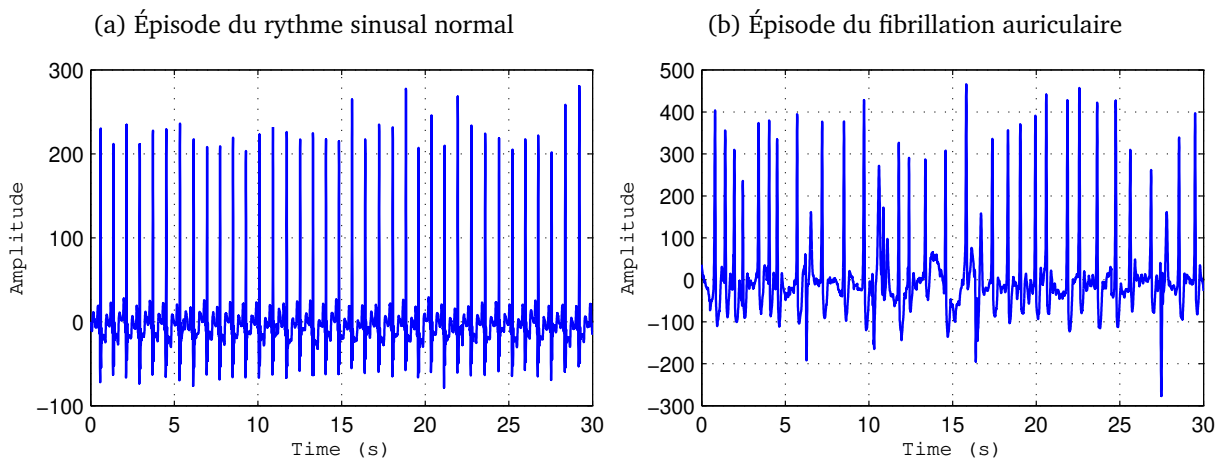


FIGURE I.6: Comparaison entre les intervalles RR sur un épisode NSR et FA de 30 secondes.

I.5.2 Défis de la détection de la FA

Les battements de cœur normaux naissent du stimulateur cardiaque naturel nommé nœud sinusal, situé dans la chambre supérieure droite (oreillettes droites). Parfois, des battements supplémentaires peuvent être induits en dehors du stimulateur normal, on parle alors de "battements ectopiques". Un battement ectopique est un battement supplémentaire provenant de la chambre supérieure (les oreillettes) ou de la chambre inférieure du cœur (le ventricule). Il se manifeste avant que le battement

normal du cœur ne puisse se former. La détection automatique précise de la FA à partir de l'ECG reste un véritable défi à cause de la présence de battements ectopiques permanents, de la présence d'artefacts de mouvement et de la surdétection de l'onde T. Dans de nombreux cas, un épisode de Rythme Sinusal Normal (RSN) peut présenter des battements ectopiques, ce qui perturbe sa régularité, comme s'il s'agissait d'un épisode de FA (voir Fig.I.7).

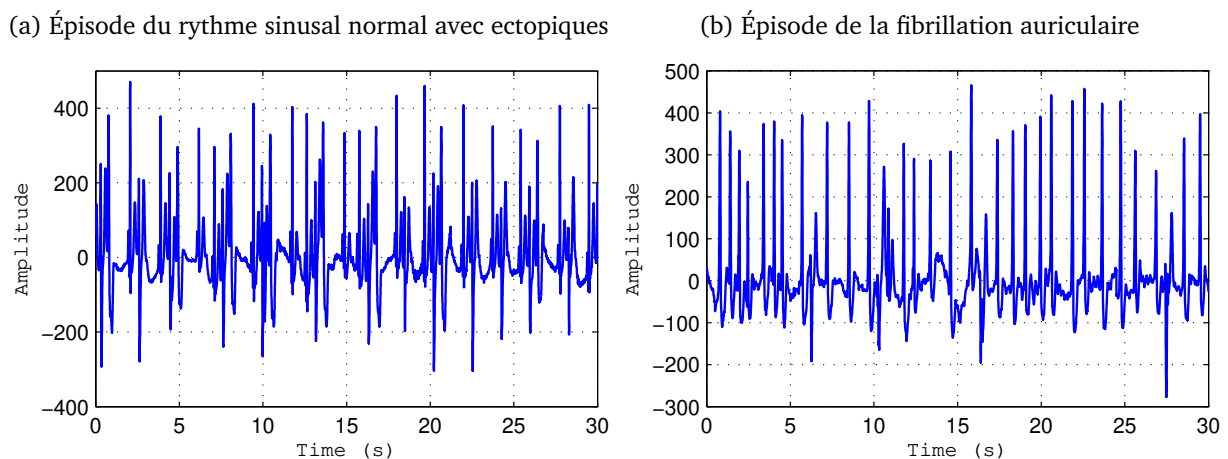


FIGURE I.7: Comparaison entre les intervalles RR sur un épisode NSR avec battements ectopiques et FA de 30 secondes.

Les battements ectopiques sont de caractère irrégulier et sont généralement inoffensifs sans cause particulière. Leurs incidences compliquent souvent l'analyse du VFC. Lors de l'activité électrique du cœur, ces ectopiques battements perturbent le cycle d'impulsions déclenché par le nœud SA et impliquent que les intervalles RR adjacents à un battement ectopique ne peuvent pas être retenus pour l'analyse de la VRC. Dans ce cas, la régulation autonome du nœud SA est temporairement perdue tandis qu'un focus ectopique déclenche prématurément le battement suivant. Le focus ectopique entraîne différents types de perturbation au niveau de l'intervalle RR. En effet, un battement ventriculaire empêche le battement sinusal consécutif, ce qui introduit une interruption compensatoire après le battement ectopique, tandis qu'un battement supraventriculaire décharge le nœud SA prématurément et provoque le battement sinusal suivant. D'autres perturbations apparaissent, dues à un battement ectopique intercalé, qui se manifeste sous la forme de deux courts intervalles RR adjacents au battement ectopique. Les perturbations du rythme peuvent également être dues à des battements absents ou mal détectés. Ces erreurs sont dues à de mauvaises décisions effectuées par le détecteur QRS, par exemple une surdétection de l'onde T. Les battements ectopiques se produisant à la fois chez les sujets normaux et chez les sujets atteints de maladies cardiaques, leurs présences représentent une source d'erreur de premier ordre à prendre en compte avant l'analyse du VRC.

I.5.3 Diagnostic automatique de la FA

La détection automatique des arythmies est principalement basée sur l'analyse du complexe QRS et de la variabilité de la fréquence cardiaque (VFC) de l'ECG. La variabilité morphologique due à la déformation du signal de l'ECG se produit dans le cas de problèmes cardiaques. L'extraction des caractéristiques morphologiques et dynamique du complexe QRS est une approche couramment adoptée pour la classification des arythmies cardiaques, ainsi que pour détecter les différents battements arythmiques telle que le battement sinusal normal (N), le battement ectopique supraventriculaire (S), les battements ectopiques ventriculaires (V), le battement de fusion (F), etc. [42].

Les algorithmes de classification automatisés ont toujours été et continuent d'être un axe de recherche très actif, notamment avec la contribution des techniques avancées de traitement et d'analyse des données. La détection automatique de la FA repose principalement sur deux approches fondamentales : l'analyse univariée, qui consiste à évaluer une seule variable ou une seule mesure pour chaque sujet, tandis que l'analyse multivariée consiste à évaluer plusieurs variables à la fois. Dans certaines études, la durée et la morphologie des ondes P ont été utilisées pour détecter et/ou prédire les épisodes de FA [voir 28, 43, et les références qui s'y trouvent]. Néanmoins, les algorithmes disponibles pour l'analyse et la détection des ondes P sont encore limités. Tout d'abord, ces algorithmes nécessitent des calculs massifs et coûteux en temps. En outre, la présence de bruit dans les enregistrements ECG réduit l'efficacité de l'extraction de ses caractéristiques, y compris les ondes P [44]. Pour éviter ces limitations et garantir un traitement fiable, l'analyse des irrégularités à base de RR est plus abordée par les chercheurs que l'absence d'ondes P ou la combinaison des deux caractéristiques.

I.5.4 Méthodes de diagnostic existantes

Nous présentons ici uniquement les études développées dans le cadre de l'analyse des intervalles RR, également connue sous le nom d'analyse de la variabilité de la fréquence cardiaque (VFC). Deux grands axes de recherche sont abordés, à savoir les études univariées et les études multivariées. En général, ceux-ci sont constitués de quatre parties essentielles :

1. Le prétraitement des données, regroupe l'ensemble des opérations nécessaires à la constitution des intervalles RR.
2. L'extraction des variables, consiste à calculer les indicateurs de la VFC.
3. La réduction de la dimensionnalité, appliquée notamment à l'analyse multivariable.
4. La construction du classificateur, qui consiste à établir la fonction caractéristique permettant de détecter la présence de la FA.

La figure I.8 présente les opérations et les séquences comprises dans les études dont nous parlerons ultérieurement. Avant de décrire les méthodes qui sont citées, nous allons d'abord présenter les mesures essentielles utilisées pour évaluer les performances de celles-ci.

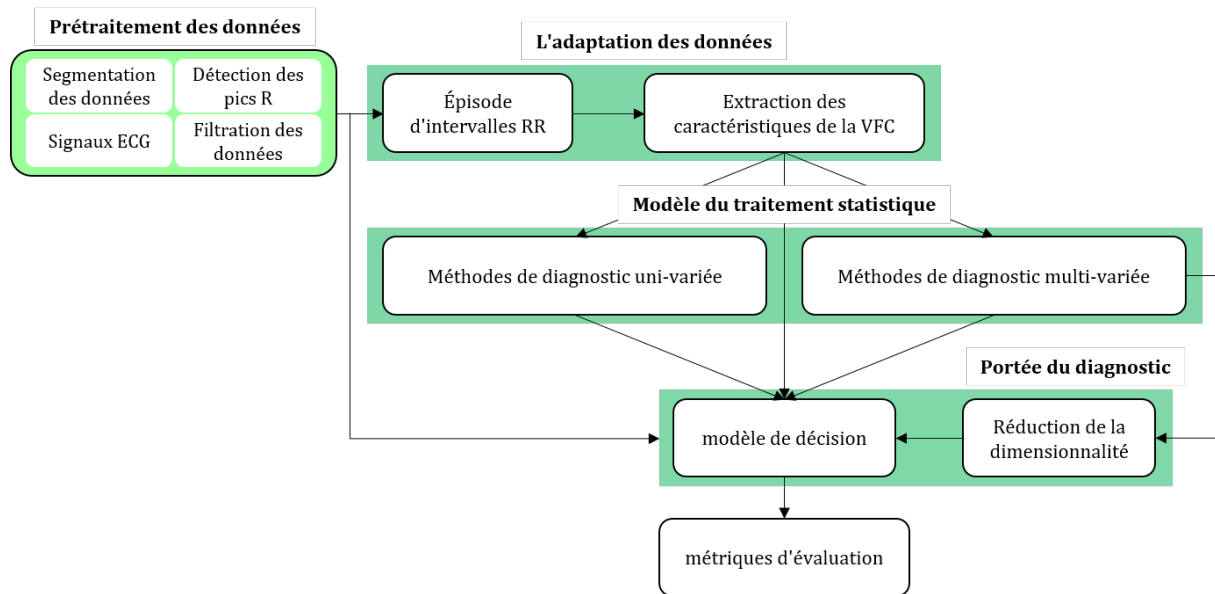


FIGURE I.8: Schéma synoptique de l'ensemble des opérations du système de diagnostic.

I.5.4.1 Métriques d'évaluation

En apprentissage automatique supervisé, la fiabilité d'un classificateur peut être évaluée à l'aide de diverses mesures présentées sous forme graphique, telles que les courbes *ROC* (Receiver Operating Characteristics) et la courbe de Précision - rappel (Precision-Recall Curve), ou par une mesure de la surface de l'aire sous la courbe *ROC* (Area Under the Curve - *AUC*). Ainsi, certaines métriques peuvent être déduites de la matrice de confusion, telles que la sensibilité, la spécificité, la justesse et la précision. Considérant une classification binaire supervisée composée de deux classes, à savoir P pour la classe positive et N pour la classe négative, la matrice de confusion est composée de 2×2 terminologies établies en fonction des étiquettes réelles et prédites. Parmi celles-ci, on trouve :

- Vrai Positif (VP , correctement prédit FA), qui représente le nombre d'échantillons positifs qui ont été correctement prédits comme positif.
- Vrai négatif (VN , correctement prédit RSN), qui représente le nombre d'échantillons négatifs qui ont été correctement prédits comme négatif.
- Faux négatif (FN , RSN incorrectement prédit), qui représente le nombre d'échantillons positifs qui ont été prédits comme négatif.
- Faux positif (FP , FA incorrectement prédit), qui est le nombre d'échantillons négatifs qui ont été prédits comme étant positifs.

La sensibilité (Sen , taux vrais positifs) désigne la capacité du classificateur à détecter tous les échantillons positifs (épisodes de FA).

$$Sen = \frac{VP}{VP + FN} \quad (I.1)$$

La spécificité (Spe , taux vrais négatifs) désigne la capacité du classificateur à détecter tous les échantillons négatifs (épisodes RSN).

$$Spe = \frac{VN}{VN + FP} \quad (I.2)$$

La précision (Pre) désigne la capacité du classificateur à ne pas confondre un échantillon vrai positif avec un vrai négatif.

$$Pre = \frac{VP}{VP + FP} \quad (I.3)$$

La justesse (Acc , Accuracy) désigne le nombre de prédictions correctes faites par le modèle sur toutes sortes de données prédites.

$$Acc = \frac{VP + VN}{VP + VN + FP + FN} \quad (I.4)$$

Le taux de faux positifs (TFP) correspond au nombre de décisions déclarées positives, alors qu'il est en réalité négatif.

$$TFP = \frac{FP}{FP + VN} \quad (I.5)$$

La valeur prédictive positive (VPP) est liée au TFP par : $VPP = 1 - TFP$.

I.5.4.2 Méthodes de diagnostic univariées

Une analyse univariée des données permet d'obtenir un aperçu général sur la faisabilité des séries temporelles RR pour évaluer la présence d'irrégularités. De nombreuses études ont préféré cette approche à l'analyse de données multivariées en raison de ses avantages précieux, tels que la simplicité, la rapidité et l'efficacité. Dans cette sous-section, nous présentons les différentes études qui ont opté pour l'analyse univariée des données pour le diagnostic de la FA.

En effet, S. Hargittai et al. [45] ont testé la faisabilité de différentes caractéristiques permettant la reconnaissance de la FA, à savoir l'entropie de Shannon (Shannon Entropy, $EnSh$), la moyenne quadratique des intervalles 80 RR successives (root mean square of the successive differences, $RMSSD$), le rapport des points de retournement (turning point ratio, TPR), le diagramme de dispersion (scatter plot), l'échantillon d'entropie (sample entropy, $SampEn$), ainsi que d'autres mesures statistiques, telles que la valeur moyenne (μ_{RR}), l'écart-type (σ_{RR}), etc. L'utilisation du diagramme de dispersion des intervalles RR successives (dRR Lorenz Plot) et le $SampEn$ ont atteint un taux d'erreur global d'environ 5% sur les données MITArrDb, MITAFDb et LTAfDb.

Langley et al. [46] ont présenté une autre application de la méthode de l'entropie pour évaluer l'impact de la durée d'enregistrement sur la fiabilité de la détection des épisodes FA. Trois caractéris-

tiques ont été évaluées, à savoir le coefficient de variation (coefficient of variance, *CV*), la moyenne des intervalles de temps successifs (Δ) et le coefficient de l'échantillon d'entropie (coefficient of sample entropy, *CoSEn*). Pour une durée d'enregistrement de 10 secondes. Les résultats obtenus par l'analyse ROC indiquent une sensibilité supérieure à 94% et une spécificité d'environ 93%. Selon les estimations, les durées d'enregistrement inférieures à 5 secondes sont possibles, mais moins pratiques, car à des fréquences cardiaques basses, le nombre de battements est insuffisant à la détermination des caractéristiques. D'autre part, la fiabilité du *CoSEn* a démontré moins de corrélation avec le rythme cardiaque et il s'avère être un détecteur robuste en présence de battements ectopiques.

D'autres applications de l'entropie ont été étudiées également par DE Lake et al. [47] pour évaluer les épisodes de FA de courte durée (12 battements). Cette étude a donné une sensibilité de 91% et une spécificité de 94% sur les données de MITAFDb.

D'autre part, trois caractéristiques linéaires et non linéaires ont été proposées par S. Cerutti et al. [48], pour caractériser la dynamique RR et détecter la présence d'épisodes FA. En effet, deux caractéristiques linéaires ont été extraites via un modèle autorégressif basé sur les séries temporelles RR. Ces caractéristiques sont utilisées pour estimer la présence ou l'absence de RSN et de FA à travers le tachogramme. Ensuite, une entropie conditionnelle corrigée est calculée pour vérifier la régularité des séquences RR. Les résultats ont montré que les trois indicateurs partagent la même sensibilité (93,3%) et présentent une valeur prédictive positive différente, avec une performance optimale de 94,4%.

En outre, Petrénas et al. [49], ont proposé une méthode de détection de la FA relativement simple. Cette méthode implique de nombreux blocs, y compris un filtrage médian à 3 points, une mesure des irrégularités RR, une fusion de signaux, une fonction de prise de décision et enfin une détection basée sur un seuil. Un taux de sensibilité et de spécificité de 97,1% et 98,3%, respectivement, a été obtenu pour la base de données MITAFDb et MITNSRDb.

Plutôt que de se baser sur les variables extraites des intervalles RR, L. Hong-wei et al. [50], ont reconstruit un espace de phase des intervalles RR selon le théorème de Takens. Une analyse de la distribution de la fonction de densité de probabilité (probability density function, PDF) de la distance entre deux points de l'attracteur est effectuée. Les auteurs estiment que la PDF peut contenir des indications importantes sur la distribution spatiale des points de phase dans l'attracteur reconstruit. Des tests sur l'ensemble de données MITAFDb ont démontré la possibilité d'identifier correctement la FA avec une sensibilité de 97.80% et une spécificité de 99% pour les intervalles 40-RR. Cependant, en diminuant le nombre d'intervalles RR, la précision de la détection se dégrade.

K. Tateno & L. Glass [51], ont étudié la similarité des histogrammes de densité standard de 100-RR et la différence des intervalles RR successives (ΔRR). Tout d'abord, un histogramme de densité de RR et ΔRR a été compilé. Les deux histogrammes ont ensuite été utilisés comme modèle pour la détection du FA. Les résultats obtenus ont démontré que les histogrammes de densité de ΔRR sont plus pertinents que les histogrammes de densité des intervalles RR, avec une sensibilité moyenne de 93,2% et spécificité moyenne de 96,7%.

Moody G & Mark R [37] ont été les premiers chercheurs à proposer un détecteur automatique FA basé sur le modèle de Markov. Selon les résultats obtenus, le modèle de Markov couplé à un filtrage et à une interpolation présente de meilleures performances et atteint 96,09% et 86,79% de sensibilité et valeur prédictive positive sur les données MITArrDb.

Ce modèle de Markov a été également étudié par Brian Young et al. [52], qui ont comparé plusieurs algorithmes de détection de la FA et défini l'algorithme le mieux adapté à un usage clinique réel. Les algorithmes étudiés étaient basés sur le modèle de Markov caché (MMC), les mesures de variance, le codage prédictif linéaire et la mesure de l'entropie approximative (approximate entropy, *ApEn*). Le taux d'erreur a été estimé à 7% pour l'ensemble de données d'entraînement (MITArrDb) et à 5,97% pour l'ensemble de données de test (MITAFDb).

Similaire, Duverney, et al. [53] ont présenté une cascade composée de deux analyses séquentielles des intervalles RR pour la détection automatisée de la FA. Une première analyse est basée sur une transformation en ondelettes discrètes (discrete wavelet transform, *DWT*) pour identifier les épisodes VFC à coefficients élevés. Ensuite, une analyse fractale basée sur la transformation de Fourier a été appliquée. Les résultats obtenus sur les données locales ont abouti à une spécificité de 99,9% pour la SR, une sensibilité de 99,2% pour la FA chronique et une sensibilité et spécificité de 96,1% et 92,6% respectivement pour la FA paroxystique.

G. Hindricks et al. [54] ont proposé un moniteur cardiaque implantable sous-cutané pour détecter la FA. Ce moniteur a été optimisé pour détecter les ondes R et construit les séries temporelles VFC. La VRC a été évaluée sur une période de 2 minutes à l'aide d'un diagramme de Lorenz des intervalles RR. La sensibilité moyenne de détection de la FA était de 90,6% et 98,1% pour les épisodes ≥ 2 minutes et ≥ 4 minutes, respectivement.

Idem, Ghodrati et al. [55] ont étudié deux mesures d'irrégularités RR pour détecter la présence de FA via les moniteurs ECG, à savoir la déviation absolue normalisée des intervalles RR et la différence absolue normalisée des intervalles RR successives. Selon les résultats de la mesure ROC, les sensibilités obtenues sont de 87,33% et 89,33% respectivement pour trois ensembles de données, comprenant MITArrDb, MITAFDb et une base de données privée sur la FA.

Le comportement irrégulier des séries temporelles de l'intervalle RR a été étudié par R. Xiuhua et al. [56], qui ont tenté de distinguer entre la FA et la RSN en utilisant le diagramme de dispersion des intervalles RR d'une minute. Quatre caractéristiques géométriques ont été appliquées, à savoir l'indice angulaire vectoriel (vector angle index, *VAI*), l'indice de longueur vectorielle (vector length index, *VLI*), la dispersion des points sur la ligne diagonale (*SD2*) et perpendiculaire à la ligne diagonale (*SD1*). Tous ces indices ont été évalués à l'aide du test de Mann-Whitney pour déterminer les écarts significatifs entre les deux classes. En effet, les Box plots ont démontré une différence marquée de chaque indice entre la FA et RSN. Les seuils optimaux ont été déterminés par l'analyse des courbes ROC. Les paramètres qui donnent les meilleures performances sont *SD1* et *SD2* dont la sensibilité et la spécificité atteignent respectivement 98,3% et 100%. Ces résultats ont été obtenus sur des ensembles de données contenant seulement 120 enregistrements d'une durée d'une minute (60 provenant de la

base de données AFTCDB et 60 enregistrements ECG normaux provenant de personnes en bonne santé), ce qui n'est pas significatif pour un outil à portée étendue.

I.5.4.3 Méthodes de diagnostic multivariées

Les techniques de mesure simples sont destinées à détecter et à quantifier la nature aléatoire des intervalles RR, tandis que les approches de modélisation sont conçues pour construire un modèle basé sur les irrégularités RR. Compte tenu de la nature irrégulière de la FA, les techniques de mesure simples sont moins performantes. Cependant, les approches de modélisation sont généralement très utiles pour distinguer l'irrégularité RR associée à la FA des autres arythmies cardiaques. Les travaux de recherche menés sur la détection automatisée de la FA à l'aide de l'analyse multivariée sont abordés dans la sous-section suivante.

Pour démontrer la faisabilité de ces méthodes, Roberta Colloca et al. [57] ont comparé l'analyse univariée et multivariée afin de construire un détecteur FA fiable. Au total, dix caractéristiques ont été extraites de séries temporelles RR et évaluées à l'aide d'une analyse univariée, ainsi que par un algorithme des machines à vecteurs de support (SVM), en tant qu'analyse de donnée multivariée robuste. À l'issue du processus d'entraînement, le SVM, construit à partir de neuf des dix caractéristiques, a abouti à des scores remarquables, avec une sensibilité et une justesse de 99,07% et 98,84%, respectivement, en utilisant MITAFDB, et une sensibilité et spécificité de 96,35% et 98,91%, respectivement, comme résultats des tests utilisant les bases de données MITNSRDB et MITArrDB.

M. Mohebbi et H. Ghassemian [58] ont comparé deux méthodes d'analyse multivariée pour détecter les épisodes de FA à 32 intervalles RR. L'idée était de réduire la dimensionnalité des caractéristiques d'entrée par la technique de l'analyse en composantes principales (ACP) et de sélectionner les caractéristiques discriminantes en utilisant la technique de l'analyse discriminante linéaire (ADL), puis de procéder à une classification SVM basée sur les caractéristiques résultantes. L'ACP et la LDA ont permis de réduire l'ensemble initial des caractéristiques de 9 à 4, soit respectivement les composantes principales et les caractéristiques discriminantes. La sensibilité et la spécificité du test (LDA+SVM) ont atteint respectivement 99,07% et 100% sur la base de données MITArrDB. Ces résultats ont dépassé ceux de l'approche PCA+SVM avec 99,28% de spécificité, ainsi que ceux obtenus sur les neuf caractéristiques initiales (97,22% de la sensibilité et 98,57% de la spécificité).

La même démarche a été adoptée par Mohammadzadeh et al. [59] afin de discriminer six classes d'arythmie différentes, dont le NSR et la FA. La réduction des caractéristiques par GDA (generalized discriminant analysis) a été appliquée dans cette étude et a permis d'obtenir cinq caractéristiques pertinentes au lieu de quinze, dérivées du signal VFC à 32 intervalles RR. Les auteurs ont comparé les performances de deux techniques de décomposition de SVM : la technique "un contre un" et la technique "un contre tous". Cette dernière a donné des résultats intéressants en termes de discrimination du RSN avec une justesse globale de 98,94%, et de 98,53% pour la FA.

Dans une autre enquête, M. Mohebbi & H. Ghassemian [60] ont extrait 14 caractéristique du signal VFC. Celles-ci se composent de 4 variables spectrales, 6 variables transspectrales et 4 variables

non linéaires, dont $SD1$, $SD2$, $SD1/SD2$ et $SampEn$. Un test d'ANOVA a été effectué pour évaluer les écarts statistiques entre les variables extraites des épisodes de FA paroxystique (PAF) et non PAF. Pour la base de données MITAFDb, ces variables ont montré des différences statistiquement significatives entre les deux groupes, en excluant les pics de fréquence dans les bandes LF et HF , qui ont été éliminés avant la construction du classificateur SVM. La sensibilité et la spécificité obtenues étaient respectivement de 96,30% et 93,10%.

SG. Artis et al. [61] sont parmi les premiers à avoir exploité les méthodes de reconnaissance de formes pour diagnostiquer la FA. Ils ont développé et évalué un réseau neuronal artificiel (artificial neural network, ANN) d'apprentissage par rétropropagation pour détecter la FA en présence d'autres arythmies cardiaques. Une matrice de transition d'intervalle généralisée, similaire à celle décrite précédemment par Moody et Mark en [37], a été appliquée en entrée du ANN. Un post-traitement, basé sur un moyen mobile de 30 points et deux seuils, appliqué aux sorties de l'ANN, a permis d'obtenir une sensibilité de 92,86% et une valeur prédictive positive de 92,34% pour l'ensemble des données de l'AFDB.

La classification basée sur l'approche Perceptron multicouche (MPL) a été adoptée par Acharya, et al. [62] pour distinguer le NSR et huit arythmies, dont la FA. Les paramètres de la variabilité de l'intervalle RR ont été extraits à l'aide d'un algorithme de transformation de Fourier rapide (fast Fourier transform, FFT) et de trois techniques de modélisation, à savoir le processus autorégressif (AR), le processus moyen mobile (moving average, MA) et le modèle autorégressif et moyenne mobile (ARMA). Les résultats ont démontré une meilleure précision avec le modèle ARMA par rapport aux autres approches de modélisation. Le modèle ARMA permet d'obtenir une très bonne reconnaissance du NSR (100% de classification correcte), tandis que la FA n'est pas correctement identifiée (66,67% de classification correcte).

F. Yaghouby et ses collègues [63] ont comparé deux approches de la programmation génétique, à savoir la programmation génétique linéaire (linear genetic programming, LGP) et la programmation multiexpression (multi-expression programming, MEP). Une analyse de régression des moindres carrés pondérés (weighted least squares, WLS) a également été effectuée comme benchmark de ces modèles. Quatre caractéristiques statistiques, dont la moyenne, l'écart-type, $RMSSD$ et $pNN50$ (nombre d'intervalles RR successifs supérieurs à 50 ms divisé par le nombre total d'intervalles), et deux caractéristiques géométriques, telles que l'indice triangulaire VFC et l'interpolation triangulaire des histogrammes d'intervalles NN, ont été extraits des 64-RR segments. Ensuite, une technique de sélection de variables IFFS (improved forward floating selection) a été appliquée, permettant de définir la moyenne, le STD et le $RMSSD$ comme trois variables pertinentes. Les résultats ont montré que le modèle LGP surpassait les modèles MED et WLS en termes de performances de formation et de tests. La sensibilité et la spécificité du modèle LGP en phase de formation étaient respectivement de 99,56% et 99,19%. De même que 99,11% et 98,91% respectivement pour les données des tests.

Carrara et al. [64] ont cherché à caractériser l'AF, le NSR et le SR à travers une analyse basée sur cinq caractéristiques, évaluées sur un intervalle RR de 10 minutes. Ces caractéristiques incluent

l'écart-type (SD), la valeur moyenne des intervalles RR (μ_{RR}), le $CoSEn$, le score de la dynamique locale (local dynamics score LDs) et l'analyse des fluctuations détendues (detrended fluctuation analysis, DFA). Toutes ces caractéristiques sont combinées à l'aide du classificateur Random-Forest (RF), où les résultats obtenus ont atteint des valeurs prédictives positives de 97%, 98% et 90% pour AF, NSR et SR, respectivement.

Ce même classificateur a été utilisé par Kennedy et al. [65] pour renforcer la détection automatique de la FA à l'aide de quatre caractéristiques d'irrégularité RR, tels que le $CoSEn$, le CV , la $RMSSD$ et la médiane de la déviation absolue (median absolute deviation, MAD). Ces caractéristiques ont été calculées sur une période d'intervalle de 30 RR. Les résultats obtenus ont atteint une spécificité de 98,3%, une sensibilité de 92,8% et une valeur prédictive positive de 92,1%.

Dans une autre optique, Ebrahimzadeh et al. [66] ont étudié la FA paroxystique afin de permettre une détection précoce et de prévenir la poursuite des troubles. Cette étude est basée sur une analyse multivariée de 28 caractéristiques extraites des séries temporelles des intervalles RR, dont neuf du domaine temporel, 11 du domaine fréquentiel et huit caractéristiques non linéaires. Puis, une approche de sélection des variables a été conduite pour réduire la dimensionnalité. Pour la classification, la technique d'apprentissage automatique le Mélange d'Experts a été utilisée composée de la méthode des k plus proches voisins (K-nearest-Neighbors, KNN), SVM et MLP. En tout, 50 segments de 5 minutes ont été employés pour entraîner les classificateurs et 56 segments pour le test. La justesse, la sensibilité, la spécificité et la précision obtenues ont atteint respectivement 98,21%, 95,55%, 100% et 96,42%.

Récemment, des modèles avancés ont été aussi développés pour distinguer les épisodes FA et NSR à l'aide de réseaux neuronaux approfondis (NN). En effet, Wang [67] a proposé un NN à 11 couches composées d'un réseau de neuronal convolutif (convolutional neural network, CNN) et du NN d'Elman modifié (modified Elman neural network, MENN). Celui-ci a été évalué à l'aide de 38 312 segments FA et 74 038 segments RSN. Les résultats montrent que 97,9% des segments FA et 97,1% des segments RSN ont été correctement classés. Faust et al. [68] ont réalisé un MENN profondi récurrent (recurrent neural network, RNN) avec mémoire à court terme (long short-term memory, LSTM) pour détecter automatiquement les segments FA en utilisant les données des intervalles 100 RR, ce qui a donné des résultats remarquables avec une justesse de 98,51%, une sensibilité de 98,32% et une spécificité de 98,67%. Andersen et al. [69] ont combiné CNN et RNN pour extraire les caractéristiques pertinentes des intervalles RR et fournir une fonction de décision capable de distinguer entre FA et RSN. Les résultats rapportés ont atteint une sensibilité et une spécificité de 98,98% et 96,95%, respectivement.

I.6 Conclusion

Les problèmes de rythme cardiaque (arythmies cardiaques) surviennent lorsque les impulsions électriques qui coordonnent les battements du cœur ne fonctionnent plus correctement, ce qui entraîne des battements trop rapides, trop lents ou irréguliers. L'ECG est l'un des outils de diagnostic les plus efficaces pour détecter les maladies cardiaques. C'est une méthode qui permet de mesurer et d'enregistrer les différentes activités électriques du cœur. Les signaux ECG peuvent être classés en deux catégories : les signaux normaux et les signaux anormaux. Pour ce faire, il est nécessaire d'analyser chaque battement de cœur, c'est-à-dire la combinaison des formes d'onde produites par les différents tissus cardiaques, ainsi que leurs morphologies et les intervalles entre ces ondes. Les dernières décennies ont vu l'émergence d'un nombre important de propositions et d'avancées scientifiques en lien avec le diagnostic des arythmies cardiaques. Parmi celles-ci, on retrouve des propositions de systèmes de diagnostic par ECG, comprenant le prétraitement des signaux, des techniques de segmentation des battements cardiaques, des méthodes d'extraction de caractéristiques et des algorithmes d'apprentissage automatique. La phase de prétraitement des signaux ECG, ainsi que les étapes d'extraction des caractéristiques constituent l'aspect le plus important de l'exploitation des données. Le processus de prétraitement des signaux permet d'éliminer le bruit, de lisser et de décomposer les signaux afin de les adapter à une analyse plus approfondie. Dans le prochain chapitre, nous présenterons les techniques de prétraitement utilisées pour la construction d'un système de diagnostic.

 TRAITEMENT DES SIGNAUX ECG

II.1 Introduction	40
II.2 ECG : étape de prétraitement	41
II.2.1 Aperçu général	41
II.2.2 La dérive de la ligne de base	43
II.2.2.1 Cas d'étude	43
II.2.2.2 Élimination de la dérive de la ligne de base : méthode de la régression linéaire multiple	44
II.2.3 Filtrage des interférences et artefacts	46
II.2.3.1 Filtre numérique	47
II.2.3.2 Filtre numérique à Réponse Impulsionnelle Infinie (RII)	47
II.2.4 Algorithme de Pan-Tompkins	49
II.3 ECG : détection du complexe QRS	50
II.3.1 Synthèse générale	50
II.3.2 Motivations	52
II.3.3 Méthodologie de détection : analyse préliminaire	53
II.3.3.1 Analyse préliminaire du signal en fonction d'une amplitude minimale	53
II.3.3.2 Analyse préliminaire du signal en fonction d'une distance minimale	54
II.3.4 Limites de la méthode	55
II.4 ECG : méthodologie de correction	56
II.4.1 Correction modérée basée sur la variabilité des données	56
II.4.2 Correction modérée en présence des intervalles R-R étroits	57
II.4.3 Correction modérée destinée aux intervalles R-R contenant un pic d'amplitude très faible	58
II.4.4 Correction exhaustive	59
II.5 Conclusion	61

II.1 Introduction

En raison de sa simplicité et de sa nature non invasive, l'ECG est largement utilisé pour la détection des maladies cardiaques. Certaines irrégularités peuvent être détectées en analysant la forme d'onde des impulsions émises par les différents tissus cardiaques. Toutefois, ces signaux ECG sont généralement influencés par la présence de bruit, ce qui perturbe et complique la détermination de ses caractéristiques, en particulier le complexe QRS, une composante cruciale du signal ECG, largement utilisée dans le dépistage des anomalies. Pour mieux exploiter les données d'ECG volumineuses, qui sont très coûteuses en temps de révision manuelle, des modèles de diagnostic automatisés ont été introduits. Ces modèles permettent d'améliorer la qualité des signaux (filtrage du bruit), de mettre en évidence les caractéristiques pertinentes, d'extraire les données non visibles par les examens manuels et surtout d'aider les médecins dans leur prise de décision concernant les signaux ECG. Cependant, le diagnostic automatisé des anomalies à partir du signal ECG est assez complexe, du fait que nous ne pouvons pas extraire toutes les mesures nécessaires au médecin pour établir un bon diagnostic. En outre, la mise en place d'un système doté de toutes les connaissances est une tâche extrêmement complexe et coûteuse à réaliser. Pour répondre aux besoins de la communauté médicale et pour garantir une évaluation ECG fiable, précise et rapide, des approches basées sur l'étude de la variabilité de la fréquence cardiaque sont recommandées. Ces approches permettent de décrire et d'analyser mathématiquement le comportement du cœur à l'aide des emplacements temporels du QRS. Pour ce qui est du rythme cardiaque, l'apparition de l'onde P constitue un meilleur indicateur que celui du complexe QRS. Cependant, la détermination exacte de l'onde P est extrêmement difficile à réaliser en raison de sa faible amplitude, voire de son absence totale. De ce fait, un indicateur lié au complexe QRS est souvent utilisé.

En général, le diagnostic automatisé des anomalies cardiaques est composé de quatre étapes : le prétraitement du signal (filtrage du bruit), la détection des complexes QRS, l'extraction des indicateurs pertinents et enfin la classification/discrimination des anomalies. Ce chapitre est consacré aux démarches de prétraitement des données et de détection des complexes QRS, le reste des étapes sera abordé dans les chapitres suivants. L'opération de filtrage consiste à éliminer les composantes indésirables affectant le signal ECG pendant la phase d'acquisition. Il s'agit notamment du bruit à haute fréquence, causé par l'activité musculaire et les interférences électriques, et du bruit à basse fréquence dû aux mouvements respiratoires, aux interférences du contact électrodes-peau et aux variations du débit sanguin. Puis, la phase de détection, qui a pour but de localiser les complexes QRS inclus dans le signal acquis.

Dans ce chapitre, nous abordons en premier lieu les méthodes de filtrage permettant de corriger les signaux ECG et de les adapter au processus de détection du complexe QRS. Pour cela, nous avons utilisé un filtre numérique à réponse impulsionnelle infinie, de type Butterworth. Nous avons également appliqué la partie filtrage proposée dans l'algorithme de Pan-Tompkins, l'un des algorithmes les plus populaires utilisant la méthode du double seuil adaptatif. Dans la deuxième partie de ce

chapitre, nous présentons une étude sur la détection des complexes QRS. Celle-ci est structurée en deux parties principales. Au cours de la première partie, nous exposons un algorithme simple basé sur des seuils fixes d'amplitude et de temps pour extraire ces complexes QRS. Cependant, cette technique présente certaines limitations en cas de présence d'une variabilité imprévisible du signal. Ce qui nous amène à introduire, dans la seconde partie, un processus de correction visant à rectifier la surdétection induite par la partie initiale. Cela implique l'utilisation d'un ensemble d'algorithmes capables d'éliminer les fausses détections.

II.2 ECG : étape de prétraitement

II.2.1 Aperçu général

L'acquisition de l'ECG est corrompue par divers types d'artefacts et d'interférences tels que la dérive de la ligne de base, interférences des lignes électriques, bruit des artefacts de mouvement des électrodes, bruit dus aux activités musculaires, ou bruit des appareils d'électromyographies (EMG). De nombreuses études ont été menées sur la conception des filtres pour atténuer ces différents types de bruits et améliorer la lisibilité du signal. Les filtres passe-bande sont largement utilisés pour traiter différents types de bruits [9, 70]. Cependant, l'élimination du bruit dû aux activités musculaires représente un problème de filtrage majeur, beaucoup plus difficile à traiter à cause du chevauchement spectral entre l'ECG et le bruit de ces activités musculaires [71, 72]. Plus, les artefacts de mouvement des électrodes peuvent être évités en minimisant les mouvements effectués par la personne. La dérive de la ligne de base est un bruit de basse fréquence, un filtre passe-haut avec une fréquence de coupure de 0,5 à 0,6 Hz permet de le neutraliser. Les interférences des lignes électriques peuvent être atténuées en utilisant un filtre passe-bande à une fréquence de coupure de 50 ou 60 Hz. Le bruit des EMG est de haut fréquence (supérieure à 100 Hz) et est susceptible d'être atténué par un filtre passe-bas à fréquence de coupure adéquate [72].

Dans le cadre actuel, l'opération de filtrage en question vise principalement à améliorer l'adéquation du signal afin de faciliter l'extraction des complexes QRS. Ici, nous présentons des techniques simples de traitement du signal pour corriger un signal ECG, un algorithme d'extraction des complexes QRS, ainsi que des techniques simples pour corriger la détection. La Figure II.1 présente la structure généralisée de la détection QRS couramment utilisée dans de nombreuses études. Les données d'entrée

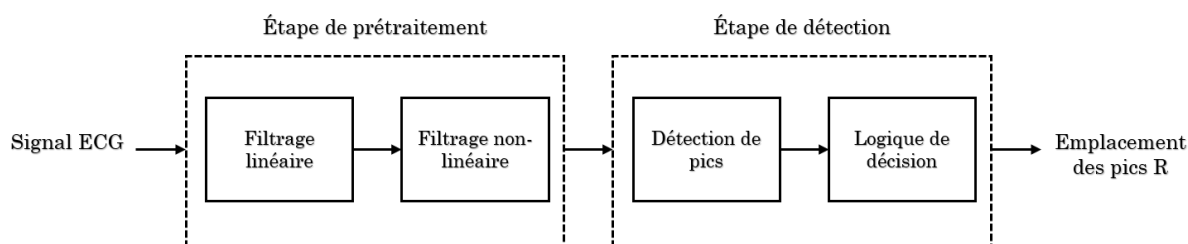


FIGURE II.1: Synoptique d'un système de détection des pics R.

sont les signaux ECG et les données de sortie indiquent les emplacements du pic R inclus dans ces signaux. En général, la première et la plus importante opération, c'est l'étape de prétraitement du signal ECG. Plusieurs techniques de filtrage sont utilisées en combinaison avec les algorithmes de détection, notamment le filtrage linéaire, le filtrage non-linéaire et la transformation en ondelettes. Parmi les techniques de filtrage linéaire, on trouve les méthodes de filtrage numérique. Les filtres numériques sont de plus en plus adoptés en raison de leurs précisions, flexibilités et adaptabilités au traitement des signaux en temps réel et en différé. Dans ce chapitre, nous présenterons brièvement les différentes méthodes de filtrage qui sont couramment utilisées dans la littérature, notamment, la technique de filtrage numérique à réponse impulsionnelle infinie type Butterworth et l'algorithme de Pan-Tompkins. Ensuite, une fois que nous avons éliminé toutes les composantes de basse et de haute fréquence du signal, on obtient un signal propre adapté aux opérations de détection des complexes QRS. L'extraction de la morphologie et de la dynamique des QRS est une méthodologie largement utilisée pour la reconnaissance de plusieurs formes de maladies cardiaques. En effet, le diagnostic automatisé de ces maladies cardiaques est principalement basé sur l'efficacité de la détection du complexe QRS. Or, outre le bruit, il existe également d'autres phénomènes associés aux signaux ECG, susceptibles de réduire la précision de la détection du QRS, tels que la variabilité et les altérations de la morphologie des complexes QRS [73], voir Fig.II.2.

(a) Morphologie du QRS et variations d'amplitude (b) Bruit musculaire et artefacts similaires au QRS

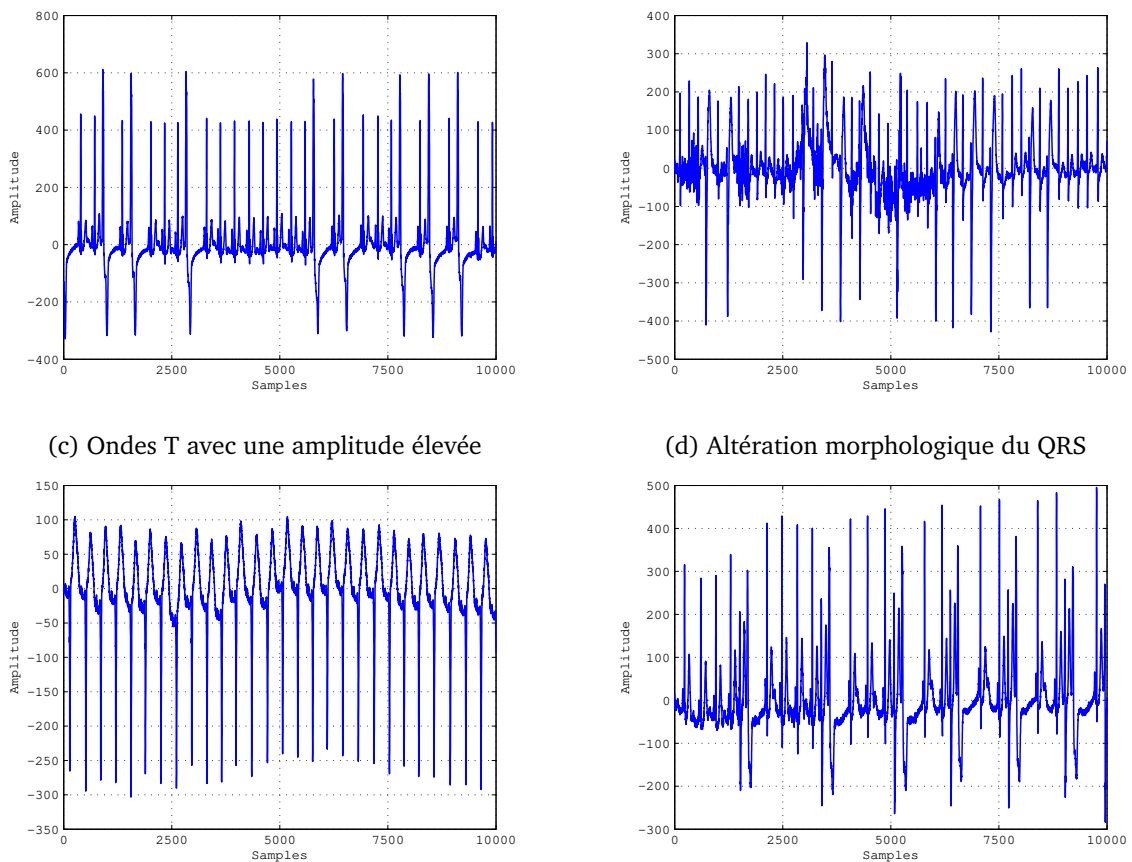


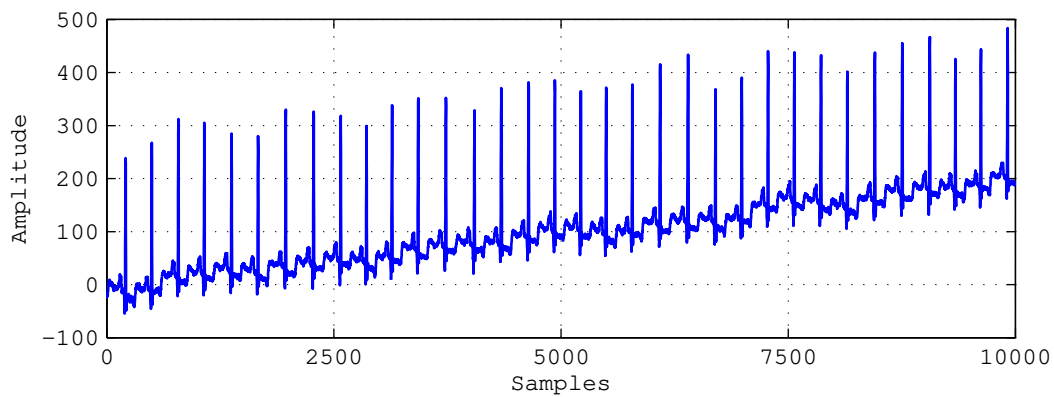
FIGURE II.2: Les différents types de problèmes rencontrés lors de la détection des complexes QRS.

II.2.2 La dérive de la ligne de base

II.2.2.1 Cas d'étude

Les signaux ECG doivent être corrigés au préalable avant de procéder à une analyse ou un traitement. Cette correction consiste à éliminer la dérive de la ligne de base causée par la ligne isoélectrique du cœur. Celle-ci est généralement due à une impédance peau-électrode inadéquate, à des artefacts liés au mouvement corporel du patient ou à sa respiration. Ce type de distorsion peut entraîner des difficultés dans l'analyse morphologique du tracé ECG [74]. Dans ce contexte, nous présentons deux cas d'étude de signaux ECG affectés par des interférences de dérive de la ligne de base : le premier cas concerne un signal affecté par une tendance linéaire, c'est-à-dire que la distorsion a une forme linéaire. Dans un second cas, nous étudions un signal ECG présentant une tendance non-linéaire (voir Fig. II.4).

(a) Signal ECG avec une tendance linéaire



(b) Signal ECG avec une tendance non linéaire

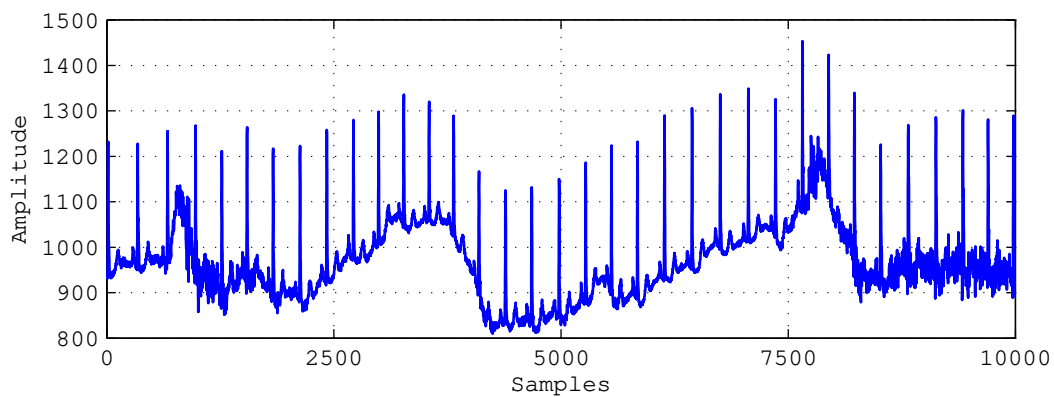


FIGURE II.3: Signal ECG affecté par la dérive de la ligne de base.

II.2.2.2 Élimination de la dérive de la ligne de base : méthode de la régression linéaire multiple

Pour éliminer la tendance linéaire, un modèle de régression linéaire multiple (RLM) du premier ordre a été adopté. D'autre part, pour éliminer la tendance non linéaire, un modèle de RLM d'ordre supérieur peut être utilisé.

La RLM est une méthode statistique qui utilise plusieurs variables explicatives (prédicteurs) pour prédire le résultat d'une variable réponse (cible). Le but de RLM est de modéliser la relation linéaire entre les variables prédicteurs et les variables cibles. Cette méthode est une extension de la méthode de régression des moindres carrés ordinaires, où le modèle est ajusté de manière à minimiser la somme des carrés des différences entre les valeurs observées et prédites. La formule de RLM est donnée par :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (\text{II.1})$$

avec, y variable dépendante (dite variable endogène), x_i variables explicatives, β_0 terme constant, β_i coefficients de pente pour chaque variable explicative.

La forme matricielle du modèle s'écrit comme suit :

$$y = \begin{bmatrix} 1 & x_1 & x_2 & \dots & x_p \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} = X\beta \quad (\text{II.2})$$

Le vecteur β peut être estimé en utilisant la technique du moindre carré (MC) :

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (\text{II.3})$$

Aussi, le vecteur ajusté (ou estimé) peut être exprimé sous la forme :

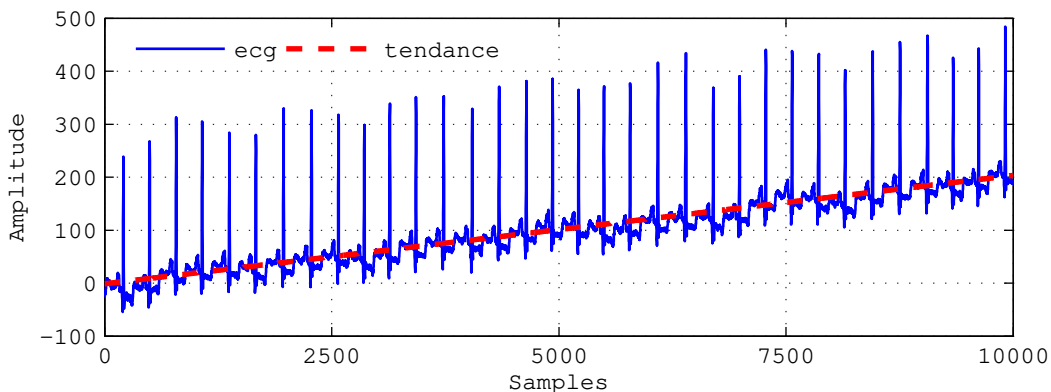
$$\hat{y} = X\hat{\beta} = X(X^T X)^{-1} X^T y \quad (\text{II.4})$$

Ensuite, l'expression du signal corrigé peut être obtenue en utilisant la formule suivante :

$$y^c = y - \hat{y} \quad (\text{II.5})$$

La Figure II.4 illustre les deux types de tendances extraites de chaque tracé ECG. Ici, y est indiqué par le label "ecg" et \hat{y} par le label "tendance".

(a) Détermination de la tendance linéaire



(b) Détermination de la tendance non linéaire

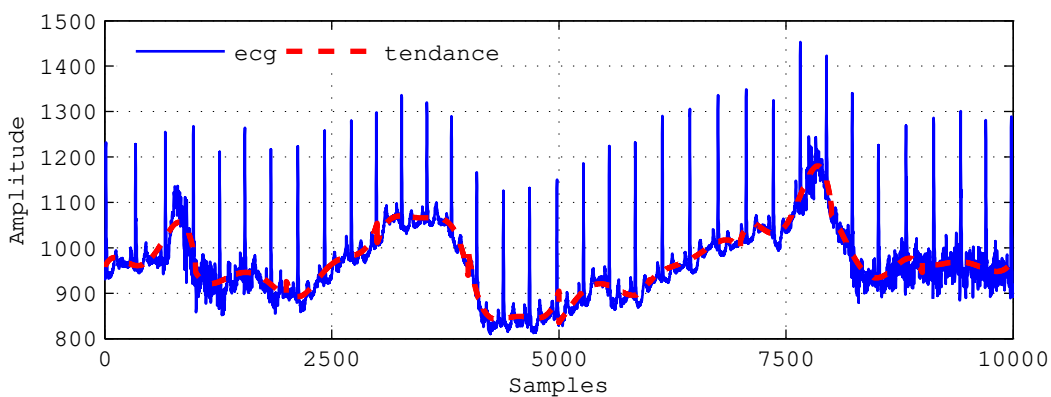


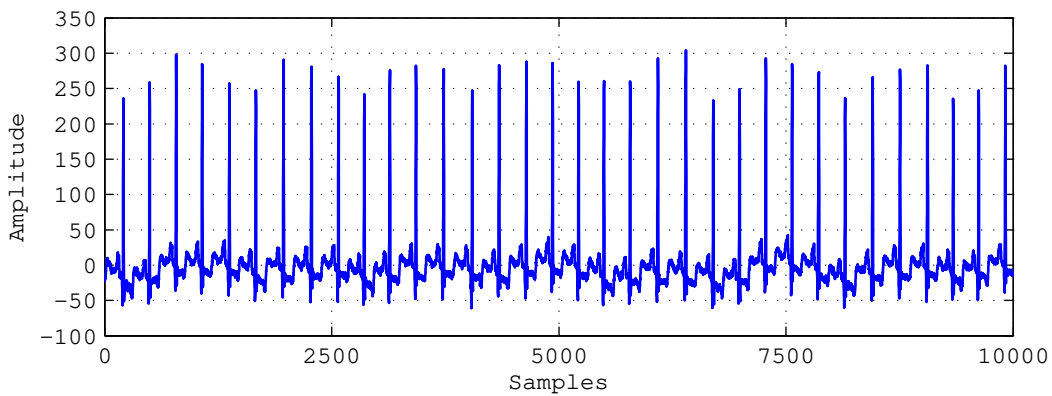
FIGURE II.4: Signal ECG affecté par la dérive de la ligne de base.

Outre les outils de modélisation mathématique, il est également possible d'utiliser des techniques de filtrage pour éliminer la tendance. Gustavo Lenis et al. [74] ont comparé cinq techniques de filtrage utilisées fréquemment dans la littérature : filtre passe-haut de Butterworth, le médian mobile, le filtrage par approximation spline, le filtrage par ondelettes et le filtrage passe-haut. Les paramètres de filtrage de chaque technique ont été abordés, ainsi que le paramétrage optimal pour éliminer un artefact dans le signal ECG.

Les techniques de filtrage ne permettent pas toujours de supprimer complètement la ligne de base d'un signal ECG. Malgré les nombreuses techniques mises au point, les méthodes numériques sont souvent appliquées pour neutraliser la ligne de base. Ces méthodes considèrent que le signal observé est la somme de la ligne de base et du signal utile. La correction consiste donc à estimer la ligne de base puis à la soustraire du signal observé. Une méthode simple est l'application de la dérivée première ou seconde pour corriger la composante linéaire ou continue de la ligne de base. Cependant, la plupart des signaux, en particulier les signaux électrocardiographiques, n'ont pas toujours

une ligne de base linéaire et, de plus, ils sont souvent trop bruyants et ne permettent donc pas d'utiliser une telle méthode. De ce fait, nous avons appliqué la technique RLM, qui est une approche très rapide et efficace, capable de modéliser correctement la plupart des lignes de base, soit linéaires ou non-linéaires (Voir Fig. II.4). La Figure II.5 illustre les deux types de tracés ECG post-corrigés calculés en utilisant l'Eq. II.5. Ces signaux ne contiennent aucun décalage dans la ligne de base ; la tendance linéaire et la tendance non-linéaire ont été parfaitement supprimées.

(a) L'ECG corrigé dans le cas linéaire



(b) L'ECG corrigé dans le cas non linéaire

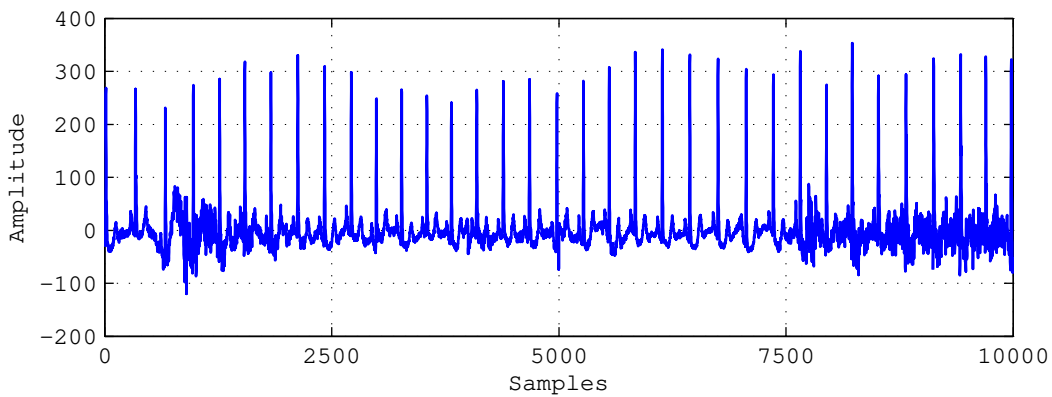


FIGURE II.5: Illustration des deux signaux ECG après suppression de la dérive de la ligne de base.

II.2.3 Filtrage des interférences et artefacts

Généralement, les champs électromagnétiques provoqués par une ligne électrique représentent une source de bruit commune dans l'ECG, ainsi que tout autre signal bioélectrique enregistré à la surface du corps. Ce bruit est caractérisé par une interférence sinusoïdale de 50 ou 60 Hz accompagnée d'un certain nombre d'harmoniques. En outre, les artefacts physiques sont principalement causés

par la déformation au niveau de la peau, qui modifie l'impédance cutanée autour de l'électrode. Les artefacts de mouvement sont semblables aux dérives de la ligne de base, mais plus compliqués à traiter en raison de leur contenu spectral qui se superpose à celui du complexe PQRST. Ils surviennent principalement dans la gamme de 1 à 10 Hz. Dans l'ECG, ces artefacts se manifestent en forme d'onde de grande amplitude, parfois confondue avec la morphologie des ondes P-Q-R-S-T. Pour éliminer le bruit des interférences et artefacts, nous utilisons un filtre numérique.

II.2.3.1 Filtre numérique

La technique de filtrage numérique s'appuie sur une séquence d'opérations mathématiques pour filtrer un signal, c'est-à-dire qu'elle transforme la composante spectrale du signal d'entrée en atténuant ou en supprimant certaines composantes spectrales indésirables. Ce filtrage numérique élimine les signaux à haute fréquence générés par l'activité des muscles non cardiaques et les interférences des équipements électriques. Par rapport aux filtres analogiques, qui sont construits en utilisant un montage de composants physiques (résistance, condensateur, etc.), les filtres numériques sont fabriqués soit par des circuits intégrés adaptés, des processeurs programmables (microprocesseur, microcontrôleur, etc.), soit avec un software sur ordinateur. Différents types de filtres numériques sont utilisés pour filtrer le signal ECG, tels que les filtres à réponse impulsionnelle finie (RIF) et à réponse impulsionnelle infinie (RII). Les filtres RIF possèdent une haute exactitude temporelle et un faible déphasage. Les filtres RII sont simples à implémenter, ils se caractérisent par une bande de fréquence étroite, et possèdent une réponse impulsionnelle, théoriquement infinie. Nikita R Pai et al. [75] ont présenté une étude comparative entre le filtre RIF et le filtre RII. Selon Nikita R Pai, le filtre RII du 2e ordre a une meilleure performance et une puissance de calcul plus faible que le filtre RIF du même ordre. Le filtre RIF est beaucoup plus complexe à réaliser que le filtre RII, et nécessite plus de mémoire et de puissance de calcul dans le temps. Dans le cadre de cette thèse, nous utilisons un filtre numérique RII.

II.2.3.2 Filtre numérique à Réponse Impulsionnelle Infinie (RII)

Dans cette étude, un filtre numérique RII a été conçu et proposé. Le filtre passe-bas RII, type butterworth, est appliqué pour filtrer les bruits à haute fréquence. Les filtres RII, des filtres récursifs, sont de type numérique. Dans ce cas, le signal de sortie du filtre est réinjecté à l'entrée de celui-ci, constituant un circuit récursif [76]. D'où, le filtre RII peut être exprimé sous forme d'une équation de récurrence où $x(n)$ représente les valeurs du signal d'entrée, $y(n)$ représente les valeurs du signal de sortie et b_k , a_k sont les coefficients du filtre selon la formule suivante :

$$y[n] = b_0x[n] + b_1x[n-1] + b_2x[n-2] + \dots + b_Nx[n-N] - a_1y[n-1] - a_2y[n-2] - \dots - a_My[n-M] \quad (\text{II.6})$$

D'où :

$$y[n] = \sum_{k=0}^N b_k x[n-k] - \sum_{k=1}^M a_k y[n-k] \quad (\text{II.7})$$

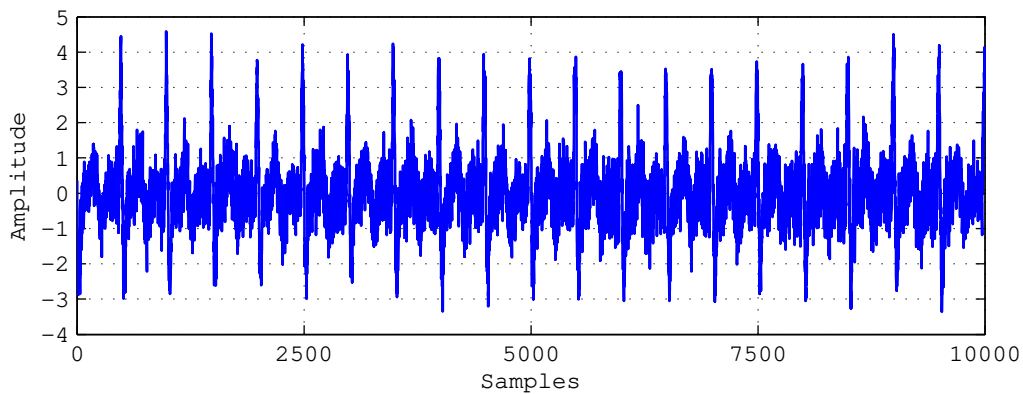
En général, N est inférieur à M et le filtre est dit d'ordre M . Les filtres RII sont caractérisés par une fonction de transfert exprimée sous la forme d'une fraction rationnelle en z^{-1} . Le filtre passe-bas de Butterworth d'ordre n est défini par le module carré de sa fonction de transfert comme suit :

$$H(z) = \frac{\sum_{k=0}^N b_k z^{-k}}{1 + \sum_{k=1}^M a_k z^{-k}} \quad \rightarrow \quad |H(j\Omega)| = \frac{H_0}{\sqrt{1 + \left(\frac{\Omega}{\Omega_c}\right)^{2n}}} \quad (\text{II.8})$$

où H_0 est le gain du filtre, Ω_c est la fréquence de coupure à 3 dB et n est l'ordre du filtre.

Dans cette étude, nous avons développé un filtre passe-bas avec une ondulation maximale à 3 dB dans la bande passante, définie de 0 et 50 Hz, et une atténuation minimale de 60 dB à la fréquence de coupure, définie de 150 Hz, pour un échantillonnage de données à 1000 Hz. En l'occurrence, un filtre d'ordre 3 a été construit, dont sa réponse est illustrée dans la Fig II.6.

(a) ECG bruité



(b) ECG filtré

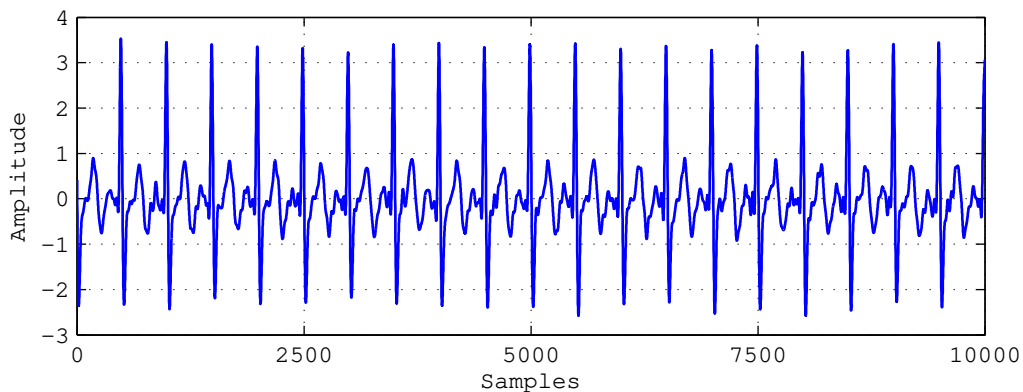


FIGURE II.6: L'ECG original et après filtration.

II.2.4 Algorithme de Pan-Tompkins

L'algorithme de Pan & Tompkins [9] consiste en une analyse numérique des pentes, des amplitudes et des intervalles des différentes ondes incluses dans le signal ECG : elle comprend

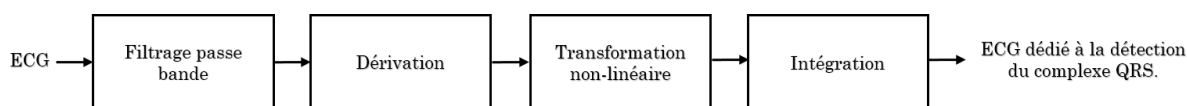


FIGURE II.7: Les opérations de prétraitement proposées dans l'algorithme de Pan-Tompkins [9]

Ces opérations illustrées dans Fig.II.7 sont mises en œuvre pour obtenir une meilleure représentation du signal, facile à traiter pour extraire les pics R. En fonction des données physiologiques, les ondes R possèdent une énergie maximale dans la bande 5-15 Hz. De ce fait, un filtre passe-bande RII a été utilisé pour atténuer les artefacts de contraction musculaire, les interférences des lignes électriques de 50 Hz et les interférences des ondes T susceptibles de perturber la détection de ces ondes R.

Pan & Tompkins ont également adopté une structure en cascade composée d'un filtre passe-bas et d'un filtre passe-haut pour analyser les signaux à une fréquence de 200 Hz. Le signal passe ensuite par un filtre dérivateur numérique destiné spécifiquement à mettre en évidence les pentes élevées distinguant les complexes QRS des composantes ECG à basse fréquence telles que les ondes P et T. Après différenciation, le signal est élevé au carré (transformation non-linéaire), soit point par point. Cette opération consiste à mettre en évidence les pentes élevées attribuées principalement aux complexes QRS. Le signal au carré passe ensuite à travers un intégrateur à fenêtre mobile. Cela permet d'obtenir un signal contenant des informations pertinentes sur la pente et la largeur du complexe QRS. La taille de la fenêtre d'intégration doit être adaptée à la largeur moyenne du complexe QRS. Si la taille de la fenêtre est trop grande, on obtient un maximum décalé dans le temps par rapport à la position du pic R ; si la taille de la fenêtre est trop petite, on obtient plusieurs pics pour la même onde R.

Pan & Tompkins ont empiriquement sélectionné une fenêtre de 150 ms, qui correspond à deux fois la largeur moyenne d'un complexe QRS. Après intégration du signal, celui-ci présente des maxima absolus lisses et uniques pour chaque cycle ECG, ainsi que d'autres maxima locaux, généralement de faible amplitude, correspondant soit au bruit, soit aux ondes T. La sortie de l'intégrateur à fenêtre mobile peut être utilisée pour détecter les complexes QRS et mesurer les intervalles RR. À ce stade, Pan & Tompkins réalise une opération de seuillage par amplitude pour éliminer les maxima faibles qui peuvent correspondre au bruit, et une opération de seuillage temporel pour éliminer les maxima proches qui peuvent être introduits par les ondes T. Ces deux types de seuillage sont décrits en détail dans [9, 77]. Les résultats de chaque étape du processus sont présentés dans Fig. II.8.

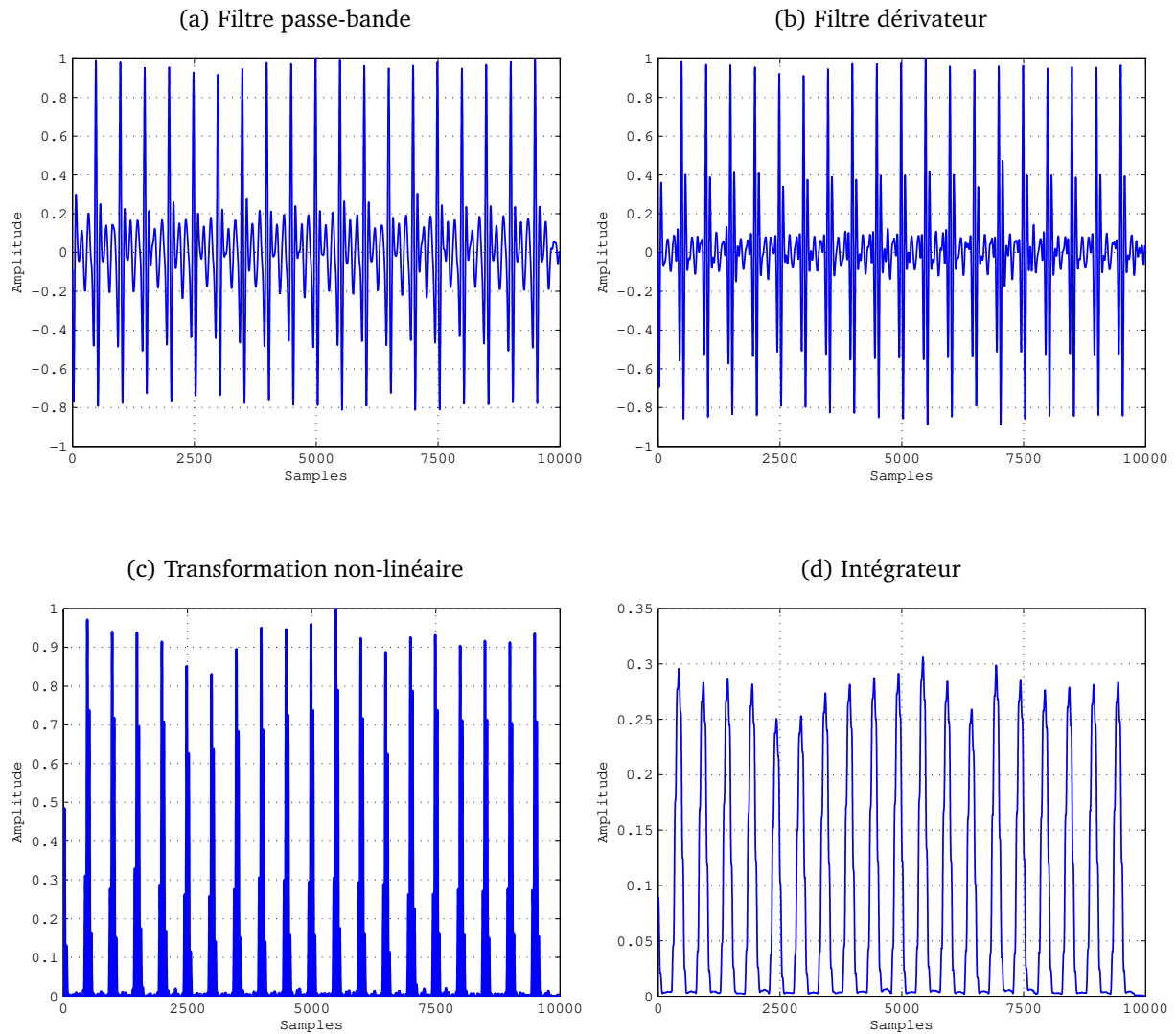


FIGURE II.8: Les résultats du filtrage en utilisant l'algorithme de Pan & Tompkins [9].

II.3 ECG : détection du complexe QRS

II.3.1 Synthèse générale

Le complexe QRS est la caractéristique la plus significative du tracé ECG et la moins sensible aux artefacts musculaires et/ou autres bruits [78, 79]. En général, la forte magnitude de l'onde R rend la détection automatisée du QRS plus facile que pour les autres ondes caractéristiques. Cependant, la détection automatique précise du QRS est parfois compliquée en raison d'aspects morphologiques et physiologiques, de l'apparition de maladies cardiaques et de l'incidence du bruit. La fiabilité de la détection automatique du QRS est également primordiale pour une extraction rigoureuse des intervalles beat-to-beat (séries chronologiques R-R) des longs enregistrements ECG. Les séries

chronologiques R-R sont d'une importance cruciale pour une analyse correcte de la VRC, ce qui permet notamment de mesurer quantitativement l'activité cardiaque en fonction de l'état de santé [72].

Au cours des dernières décennies, de nombreuses études ont été consacrées à la détection des QRS. La plupart de ces études se sont axées autour de trois aspects : les techniques de détection dans le domaine temporel [80], les techniques de détection par transformation [81, 82], les techniques de filtrage morphologique [83] et les techniques de filtrage par template-matching [84]. Ces techniques sont appliquées dans le cadre des analyses de VFC, la classification de l'arythmie, le calcul de la fréquence cardiaque, l'extraction de caractéristiques ECG, et l'analyse de l'intervalle R-R. Les techniques de détection QRS comportent deux étapes essentielles :

1. Le renforcement du QRS, une technique appliquée pour amplifier le complexe QRS par rapport aux autres composantes du tracé ECG, telles que P et T.
2. La détection du QRS, un ensemble adapté de règles et de normes bien établies à l'appui d'une identification robuste du QRS.

En principe, la première partie consiste en un filtrage passe-bande pour éliminer le bruit et les ondes P et T. Puis, une transformation non linéaire est effectuée, notamment une dérivation du signal pour identifier les pentes fortes autour de l'onde R, et un quadrillage pour quantifier le QRS. La deuxième partie concerne la sélection basée sur des critères de seuil. Pan & Tomkins ont développé l'un des algorithmes les plus populaires basés sur la méthode du double seuil adaptatif pour déterminer le complexe QRS [9]. Cette méthode est basée sur la pente extrême de l'onde R, permettent de détecter le complexe QRS en fonction du seuil de la dérivée première ou second du signal ECG. Ils utilisent également un filtre passe-bande numérique (BPF) adapté pour réduire les erreurs de détection. Yongli et Huilong [83], appliquent un filtrage mathématique de la morphologie pour améliorer la qualité du QRS suivi d'une opération de seuillage. Cuiwei Li et al. [85] ont proposé la méthode des transformations d'ondelettes pour détecter le complexe QRS. Les travaux de recherches [86–89] employaient une série de filtres numériques adaptés pour détecter et classer un signal ECG dans le domaine temporel ou fréquentiel. Kleydis V. Suarez et al. [90] ont proposé l'algorithme des correspondances géométriques pour déterminer les battements électrocardiographiques. Basé sur l'estimation de la dérivée du premier ordre du signal ECG, Xiaomin Xu et al. [91] ont proposé l'algorithme SVW (Slope Vector Waveform) pour déterminer les positions QRS. Certains travaux de recherche ont appliqué la dérivée première combinée à la dérivée seconde, suivie d'une opération de seuillage [92]. Dinh et al. [93], ont appliqué une transformée en ondelettes au signal de l'ECG suivi d'un seuil. En dehors des algorithmes mentionnés ci-dessus, un certain nombre d'algorithmes de détection des complexes QRS performants basés sur les techniques de démodulation complexe à fréquence variable [94], dérivation numérique [95], transformée de Fourier rapide (FFT) [96], modulateur delta [97], filtre de Kalman [98] sont également proposés dans la littérature récente.

II.3.2 Motivations

Afin de réduire les coûts de calcul, de tenir compte de la variation des signaux ECG, de prendre en compte les exigences de fabrication industrielle et de fournir une méthode de détection QRS adaptée, nous proposons une approche de détection simple basée sur deux opérations préliminaires : la première opération est effectuée en fonction de l'énergie du signal (amplitude), et la seconde opération est effectuée en fonction des distances minimales séparant deux pics R. Dans une deuxième section, une analyse exhaustive est également proposée pour corriger les éventuelles fausses détections issues de l'analyse préliminaire. Ce processus de correction peut être très pratique dans le cas où le signal n'est pas convenablement filtré, ce qui fait que certaines ondes de forte amplitude ayant l'apparence d'un QRS et satisfaisant aux critères R_{min} et D_{min} , sont systématiquement détectées par la première analyse comme des pics R. La figure II.9 illustre les résultats définitifs obtenus pour R_{pic} .

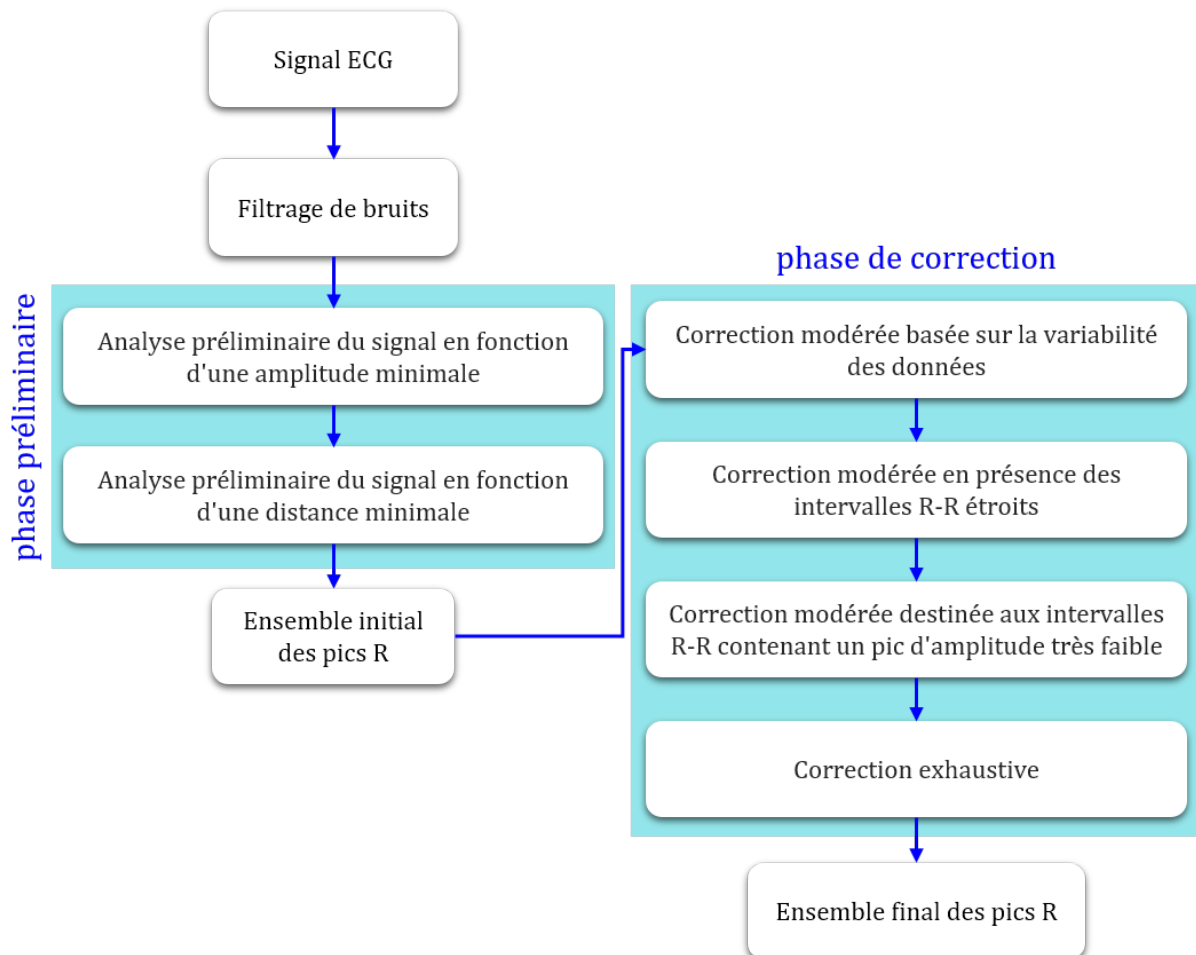


FIGURE II.9: Algorithme de détection des complexes QRS

II.3.3 Méthodologie de détection : analyse préliminaire

L'algorithme de détection des pics R repose sur deux étapes critiques essentielles :

1. Analyse préliminaire du signal en fonction d'une amplitude minimale : nous sélectionnons uniquement les points d'échantillonnage supérieurs à une amplitude minimale " R_{min} ". Il faut savoir qu'une bonne estimation de la valeur R_{min} permet de réduire considérablement le temps de calcul.
2. Analyse préliminaire du signal en fonction d'une distance minimale : nous sélectionnons uniquement les points qui sont séparés par une distance minimale " D_{min} ". Ce paramètre peut être adapté pour exclure les pics moins importants susceptibles de se situer à proximité d'un sommet plus élevé. Par exemple, si un pic important se produit à l'indice N , tous les pics secondaires dans la plage $[N - D_{min} \quad N + D_{min}]$ sont exclus.

II.3.3.1 Analyse préliminaire du signal en fonction d'une amplitude minimale

Suite au prétraitement, le signal filtré possède un maximum absolu pour chaque complexe QRS. En outre, le signal comporte d'autres maxima locaux, souvent de faible amplitude, qui correspondent soit au bruit, soit aux ondes T. De ce fait, une recherche de maxima est effectuée en procédant à un seuillage d'amplitude. Cela permet d'éviter la prise en compte des maxima faibles qui peuvent correspondre au bruit. Par ailleurs, l'amplitude des battements est susceptible de varier dans le temps et d'un enregistrement à l'autre, ce qui fait que le seuil utilisé ne peut être fixé a priori sur toute la durée de l'enregistrement cardiaque : il est donc recalculé, puis fixé sur chaque segment de taille prédéfinie tout au long de l'analyse. Pour commencer, on définit un seuil

$$R_{min} = \left\{ \frac{1}{m} \sum_{i=1}^m x_i \mid x = [\max(y^f) \rightarrow \min(y^f)]; m = \lfloor 10\% \times \text{Dim } y^f \rfloor \right\} \quad (\text{II.9})$$

où x désignent les données du signal filtré, rangé en ordre croissant et m correspond à 10% de la taille du signal filtré, qui peut être ajusté.

En fonction de la valeur R_{min} déterminée, le nombre d'échantillons sélectionnés varie, pour une faible valeur R_{min} , on s'assure de ne pas éliminer un vrai pic R de basse amplitude, induit par une raison morphologique ou numérique (système d'acquisition). En revanche, une faible valeur R_{min} augmente considérablement le nombre de pics retenus, qui doivent être traités par le second bloc d'analyse. Pour une valeur R_{min} moyen (pas trop élevée ou trop basse), peut être très pratique, permet de capturer toutes les zones de vrais pics, et minimise le nombre d'échantillons sélectionnés, ce qui réduit automatiquement le nombre d'opérations mathématiques effectuées dans le second bloc.

Après avoir déterminé le seuil R_{min} , on extrait tous les échantillons dont l'amplitude est supérieure à ce seuil. Le signal y^f peut être représenté par un vecteur de n échantillons comme suite :

$$y^f = [x_1 \quad x_2 \quad \cdots \quad x_{n-1} \quad x_n], \quad n = \text{Dim } y^f, \quad (\text{II.10})$$

D'où, les échantillons dont l'amplitude est supérieure à R_{min} sont identifier comme suite :

$$R_0 = \{x_i \in y^f \mid x_i \geq R_{min} ; \forall i = 1, \dots, n\} \quad (\text{II.11})$$

Les échantillons sélectionnés dans R_0 sont considérés comme la première collection possible d'échantillons de l'onde R. L'ensemble des maxima est ensuite obtenu en procédant comme suit :

1. On dérive le signal y^f

$$\dot{y}^f = \{x_{i+1} - x_i ; \forall i = 1, \dots, n-1\} \quad (\text{II.12})$$

2. On calcule le signe du dérivé \dot{y}^f

$$S = \text{signe}(\dot{y}^f) = \begin{cases} +1 & \text{si } \dot{y}^f(i) \geq 0 \\ -1 & \text{si } \dot{y}^f(i) < 0 \end{cases} \quad \forall i = 1, \dots, n-1 \quad (\text{II.13})$$

3. On dérive le signal S

$$\dot{S} = \{S(i+1) - S(i) ; \forall i = 1, \dots, n-2\} \quad (\text{II.14})$$

Cette technique permet de localiser les points d'inflexion dans le signal ECG (y^f). Ces points d'inflexion indiquent les changements dans la concavité du signal. Un signe positif de la dérivée \dot{S} signifie que la fonction est convexe et un signe négatif de la dérivée \dot{S} signifie que la fonction est concave.

Notre objectif consiste à extraire tous les points pour lesquels la dérivée $\dot{S} = -2$, indiquant que le signal y^f est concave, i.e., la présence d'un pic maximal.

4. On extrait les données des échantillons à une amplitude \dot{S} égal à -2 .

$$\dot{S}_{-2} = \{S(i) \in \dot{S} \mid S(i) = -2 ; i = 1, \dots, n-2\} \quad (\text{II.15})$$

5. On détermine les points d'intersection entre les échantillons R_0 et \dot{S}_{-2} .

$$R_1 = R_0 \cap \dot{S}_{-2} \quad (\text{II.16})$$

II.3.3.2 Analyse préliminaire du signal en fonction d'une distance minimale

À l'issue du traitement précédent, le signal obtenu comporte une amplitude maximale pour chaque complexe QRS et d'autres crêtes locales, généralement d'amplitude faible, correspondant soit au bruit, soit aux ondes T. Le seuillage d'amplitude est utilisé pour distinguer les maxima correspondant aux pics R de ceux correspondant aux ondes T ou bruits, qui sont généralement beaucoup plus faibles. Dans cette partie, nous avons besoin de l'amplitude des échantillons précédemment sélectionnés $R_{amp} = y^f(R_1)$ et de leurs positions en abscisse R_1 . Nous trions d'abord les amplitudes R_{amp} par ordre croissant, puis nous commençons à traiter les pics les plus importants afin d'éviter de sélectionner accidentellement un faible sommet et d'en éliminer un important à proximité. Le processus de sélection se déroule comme suit : si un pic n'est pas proche d'un autre pic plus élevé, alors nous éliminons tous

les pics dans la plage $\pm D_{min}$. Nous considérons une distance minimale de $D_{min} = 200ms$ entre chaque deux pics R, de sorte que d'un point de vue physiologique, aucun intervalle R-R ne se présente à moins de $200ms$ de distance. L'ensemble des maxima extraits du processus initial peut être représenté par un vecteur comme suit :

$$R_1 = [r_1 \ r_2 \ \cdots \ r_m] \quad (\text{II.17})$$

En effet, un pic est retenu s'il satisfait à l'équation suivante :

$$R_{pic} = \{r_i \in R_1 \mid r_i = \max(r_j \in r_i \pm D_{min}) ; i, j = 1, \dots, m\} \quad (\text{II.18})$$

II.3.4 Limites de la méthode

L'approche de détection présentée est basée sur deux paramètres principaux, la R_{min} et D_{min} . Le premier doit être ajusté en fonction du signal étudié, mais le second peut être réglé à $0.2 \times Fs$ pour rester dans une norme de fiabilité, avec Fs désigne la fréquence d'échantillonnage. Compte tenu de la simplicité des techniques de filtrage et de détection des pics R, la démarche adoptée est limitée, c'est-à-dire que, parfois, il est possible qu'un pic remplisse à la fois les critères d'amplitude et de durée, mais qu'il ne corresponde pas à un véritable pic R. Comme le montre la figure II.10, décrivant l'étape finale de détection, certains pics sélectionnés ne font pas partie du complexe QRS, il s'agit des "faux pics R".

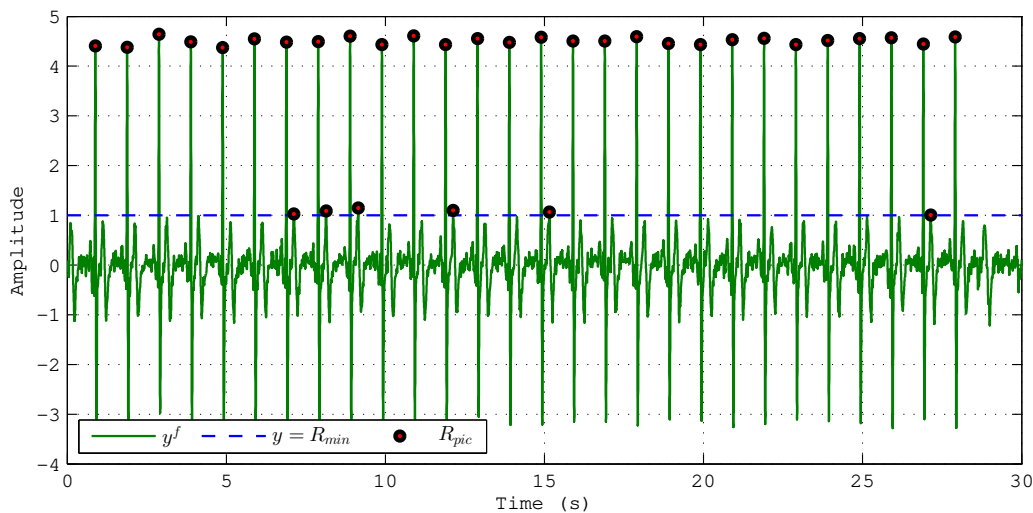


FIGURE II.10: Les résultats de détection obtenus après le seuillage en amplitude et en durée.

En réalité, les durées entre les complexes QRS et leurs amplitudes sont très variées, raison pour laquelle on choisit une valeur seuil basse, afin d'éviter la suppression de certains véritables pics. Toutefois, un seuil bas peut induire de faux pics R et compliquer le processus de détection. Néanmoins, une bonne adaptation des paramètres de seuillage permet d'obtenir des résultats fiables avec moins d'opérations mathématiques et sans avoir recours à des opérations de correction.

De nombreuses techniques de seuillage adaptatif sont disponibles dans la littérature [99–101], bien qu'elles soient très coûteuses en termes d'opération mathématique exécutée. De ce fait, nous suggérons une technique basée sur un seuillage fixe, qui est très pratique quand on connaît les marges de tension des complexes QRS, permettant de réduire le nombre d'opérations mathématiques et d'obtenir des résultats précis. À défaut, nous pouvons utiliser le bloc de correction, qui comporte des techniques de seuillage adaptatif servant à éliminer tous les autres pics qui ne font pas partie des complexes QRS.

II.4 ECG : méthodologie de correction

Le prétraitement des données est généralement suivi d'une étape de décision qui consiste à extraire et à localiser les complexes QRS définitifs. L'étape de décision s'appuie généralement sur des méthodes heuristiques pour détecter l'emplacement exact du complexe QRS. Dans le cas des tracés ECG qui comportent des variations physiologiques, la détermination des bons critères de sélection devient problématique. Pour tenir compte des variations morphologiques des formes de signaux ECG, la technique du seuil adaptatif peut être utilisée.

La technique du seuil adaptatif est un outil très sensible lors de la détection de complexes QRS. L'adaptation du seuil améliore la précision de la détection au prix de la complexité du calcul, ce qui complique la détection des QRS en temps réel. Pour réduire les coûts de calcul, tenir compte des variations morphologiques des signaux ECG et améliorer la précision de la détection des QRS, nous proposons une technique de détection rigoureuse des QRS basée sur 4 niveaux d'évaluation distincts. Dans notre contexte, le bloc de correction proposé consiste à effectuer quatre étapes de correction des pics R. Chaque étape est adaptée à une certaine marge d'évaluation pour confirmer ou supprimer un pic R. Dans ce contexte, nous avons défini un seuil R_{min} égal à

$$R_{min} = \left\{ \frac{1}{m} \sum_{i=1}^m x_i \mid x = [max(y^f) \rightarrow min(y^f)]; m = \lfloor 30\% \times Dim y^f \rfloor \right\}, \quad (II.19)$$

afin de renforcer la complexité du problème et de tenir compte des situations délicates où le choix des paramètres du seuillage est flou, ce qui exige de retenir une limite très basse. En réduisant la valeur R_{min} , nous obtenons un nombre très élevé de faux pics, raison pour tester la capacité de l'algorithme à éliminer tous ces faux pics.

II.4.1 Correction modérée basée sur la variabilité des données

L'objectif de cette étape est de supprimer les fausses détections causées par le bruit ou les ondes T à haute intensité à proximité du complexe QRS. Or, le complexe QRS est un pic unique de forte intensité jusqu'à au moins 2 ou 3 fois l'amplitude maximale des pics à proximité. Certains pics peuvent être détectés entre deux maxima R. Pour y remédier, une première étape de correction a été développée pour les éliminer en fonction du degré de variabilité dans leurs régions respectives. Ainsi, une forte

variabilité dans un tronçon donné implique que la densité d'oscillation est élevée, d'où la présence de pics détectés. Un calcul de variabilité locale est effectué pour chaque pic R détecté au cours du processus initial. La formule est la suivante :

$$V = Var \left(y^f \left(R(n) - \frac{1}{2} \min(RR) \rightarrow R(n) + \frac{1}{2} \min(RR) \right) \right), \quad (\text{II.20})$$

où $Var(\cdot)$ est l'opérateur de variance, y^f le signal filtré, le terme $\frac{1}{2} \min(RR)$ désigne la fenêtre temporelle, c'est-à-dire que pour chaque pic $R(n)$, nous calculons la variabilité du signal sur un segment de largeur

$$L = \left[R(n) - \frac{1}{2} \min(RR) , R(n) + \frac{1}{2} \min(RR) \right], \quad (\text{II.21})$$

et RR indique la dérive de la série temporelle des pics $R(n)$, calculé comme suit :

$$RR = \{ R(i+1) - R(i); \quad \forall i = 1, \dots, n-1 \} \quad (\text{II.22})$$

L'équation appliquée pour identifier les faux pics est donnée comme suit :

$$S_{pic1} = \{ R(i) \mid V(i) \leq \frac{1}{3} \bar{V} \text{ et } R_{amp}(i) < \theta, \quad \forall i = 1, \dots, n-1 \} \quad (\text{II.23})$$

où $R_{amp} = y^f(R_{pic})$ désigne l'ensemble des amplitudes des pics R sélectionnés précédemment.

Avec

$$\theta = \left[\frac{\overline{R_{amp}}}{\max(R_{amp})} \right] \times \overline{R_{amp}} \quad (\text{II.24})$$

D'où, nous supprimons de R_{pic} et R_{amp} tous les indices S_{pic1} .

$$\begin{cases} R_{pic} = R_{pic} \setminus R_{pic}(S_{pic1}) \\ R_{amp} = R_{amp} \setminus R_{amp}(S_{pic1}) \end{cases} \quad (\text{II.25})$$

Nous répétons ce processus jusqu'à l'élimination de tous les pics satisfaisant les critères de sélection. La suite du processus traite les pics retenus du premier cycle de sélection.

II.4.2 Correction modérée en présence des intervalles R-R étroits

Dans cette étape de correction, nous allons vérifier chaque pic en fonction de son intervalle R-R associé. En effet, un intervalle à faible portée peut avoir deux interprétations, soit la fréquence cardiaque (FC) est trop élevée, ce qui conduit à un ensemble étroit d'intervalles R-R, soit la présence de faux pics entre certains vrais, ce qui induit la présence de divergences dans leurs intervalles. Pour identifier si un intervalle R-R est composé de deux véritables pics R ou non, nous proposons la formule de calcul de seuil suivante :

$$S_{RR} = \min(RR) + \frac{1}{3} (\overline{RR} - \min(RR)), \quad (\text{II.26})$$

Après avoir défini les critères de correction, nous éliminons les pics qui satisfont aux conditions données dans l'équation suivante :

$$S_{pic2} = \{R(i) \mid RR_{min} < S_{RR} \text{ et } R_{amp}(i) < \theta \text{ et } RR'(i) < Me(RR), \forall i = 1, \dots, n-1\} \quad (II.27)$$

où θ peut être calculé en utilisant la même formule décrite dans l'équation II.24.

Avec

$$RR_{min} = \begin{cases} RR(i) & \text{si } i = 1 \\ \min\{RR(i), RR(i-1)\} & \text{si } 2 < i < n \end{cases} \quad (II.28)$$

Et

$$\begin{cases} RR' = RR(i+1) - RR(i-1) & \text{si } 2 < i < n-1 \\ RR' = 2 \times \overline{RR} & \text{si } i = 1 \vee i = n \end{cases} \quad (II.29)$$

D'où, nous supprimons de R_{pic} et R_{amp} tous les indices S_{pic2} .

$$\begin{cases} R_{pic} = R_{pic} \setminus R_{pic}(S_{pic2}) \\ R_{amp} = R_{amp} \setminus R_{amp}(S_{pic2}) \end{cases} \quad (II.30)$$

Nous répétons également ce processus jusqu'à l'élimination de tous les pics satisfaisant les critères de sélection.

II.4.3 Correction modérée destinée aux intervalles R-R contenant un pic d'amplitude très faible

Dans cette étape, nous cherchons à éliminer tout pic de faible amplitude situé à proximité d'un autre pic d'amplitude dominante. En effet, pour tout intervalle $RR(n)$ donné comportant un déséquilibre significatif de ses amplitudes ($R(n) \gg R(n+1)$ ou $R(n+1) \gg R(n)$), les critères de correction sont appliqués au pic le plus faible. Donc, si l'amplitude d'un pic est très faible par rapport à celle de son adjacent et par rapport à S'_{amp} , il sera pris en compte pour l'évaluation. Le critère du seuil appliqué dans cette partie est défini comme :

$$S'_{amp} = \min \left\{ \theta \quad \frac{1}{2} \max \{R_{amp}(n) \quad R_{amp}(n+1)\} \right\} \quad (II.31)$$

où θ peut être calculé en utilisant la même formule décrite dans l'équation II.24.

Après avoir défini les critères de correction, nous éliminons les pics qui satisfont aux conditions données dans l'équation suivante :

$$S_{pic3} = \{R(p) \mid R_{amp}(p) < S'_{amp} \text{ et } RR(i) < S_{RR}, \forall i = 1, \dots, n\} \quad (II.32)$$

où S_{RR} peut être calculé en utilisant la même formule décrite dans l'équation II.26.

Avec

$$p = \begin{cases} \underset{\min}{\operatorname{arg}} \{R_{amp}(i) \ R_{amp}(i+1)\} & \text{si } i < n \\ i & \text{si } i = n \end{cases} \quad (\text{II.33})$$

D'où, nous supprimons de R_{pic} et R_{amp} tous les indices S_{pic3} .

$$\begin{cases} R_{pic} = R_{pic} \setminus R_{pic}(S_{pic3}) \\ R_{amp} = R_{amp} \setminus R_{amp}(S_{pic3}) \end{cases} \quad (\text{II.34})$$

Nous répétons ce processus jusqu'à l'élimination de tous les pics satisfaisant les critères de sélection abordés.

II.4.4 Correction exhaustive

La présente étape propose une technique de correction approfondie pour assurer l'élimination totale des faux pics, qui sont peut-être en suspens à l'égard de tous les critères de correction. Les fausses détections qui se produisent au début ou à la fin du signal sont relativement compliquées à éliminer. L'étape actuelle a donc introduit des contraintes supplémentaires pour éliminer les faux pics dans les bornes. Dans ce contexte, la correction a été effectuée sur trois évaluations, pour chaque paire de pics $R(i)$ et $R(i+1)$. En effet, pour chaque itération, on définit le pic d'amplitude minimale et sa position, tel que :

$$j = \underset{\min}{\operatorname{arg}} \{R_{amp}(i) \ R_{amp}(i+1)\} \quad \text{et} \quad k = \underset{\max}{\operatorname{arg}} \{R_{amp}(i) \ R_{amp}(i+1)\} \quad (\text{II.35})$$

Puis, on procède aux tests suivants :

T1 : pour la première itération, nous vérifions :

$$S_{pic4} = \left\{ R(i) \mid R_{amp}(i) < \max \left\{ S \ \frac{1}{3} R_{amp}(i+1) \right\} \quad \text{et} \quad RR(i) < S_{RR} \right\} \quad (\text{II.36})$$

T2 : pour l'avant-dernière itération, nous vérifions :

$$S_{pic4} = \left\{ R(j) \mid R_{amp}(j) < \max \left\{ S \ \frac{1}{3} R_{amp}(k) \right\} \quad \text{et} \quad RR(i) < S_{RR} \right\} \quad (\text{II.37})$$

T3 : pour tous les itération $2 < i < n - 2$, nous vérifions :

$$S_{pic4} = \left\{ R(j) \mid R_{amp}(j) < \max \left\{ S \ \frac{1}{3} R_{amp}(j-1) \ \frac{1}{3} R_{amp}(j+1) \right\} \quad \text{et} \quad RR(i) < S_{RR} \right\} \quad (\text{II.38})$$

De même, nous répétons ce processus jusqu'à l'élimination de tous les pics satisfaisant les critères de sélection. L'avantage de cette technique de filtration réside dans sa capacité à éliminer toutes les fausses détections subsistant des étapes précédentes et qui sont produites à la suite du procédé de détection initiale. Comme le montrent les figures II.11 et II.12, l'approche adoptée est performante

et permet la détection de tous les faux pics restants, malgré la forte fluctuation observée dans les signaux. Le label indiqué par un " S_{pic1} ", " S_{pic2} ", " S_{pic3} " et " S_{pic4} " désigne les pics qui seront éliminés après chacune des quatre phases de correction.

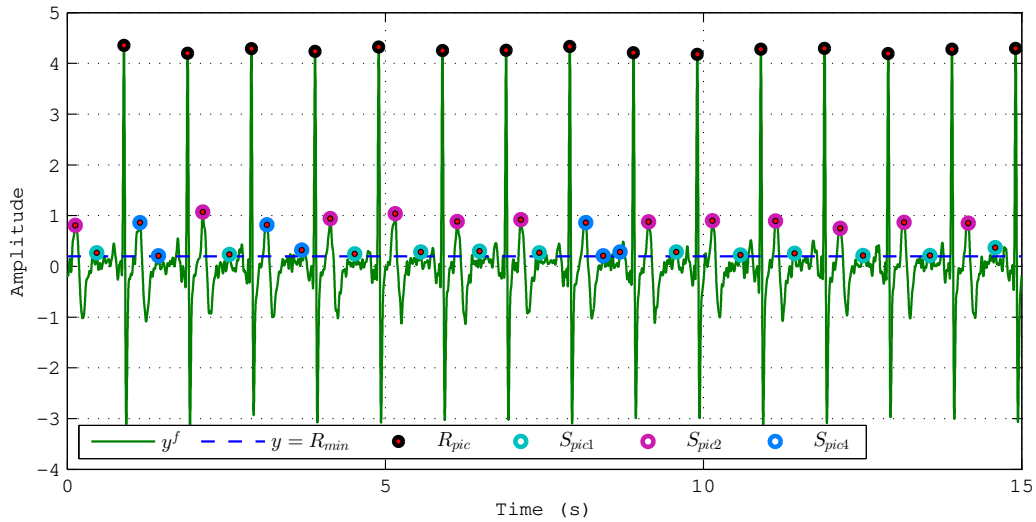


FIGURE II.11: Exemple 1 - Résultats obtenus après la quatrième filtration des faux pics.

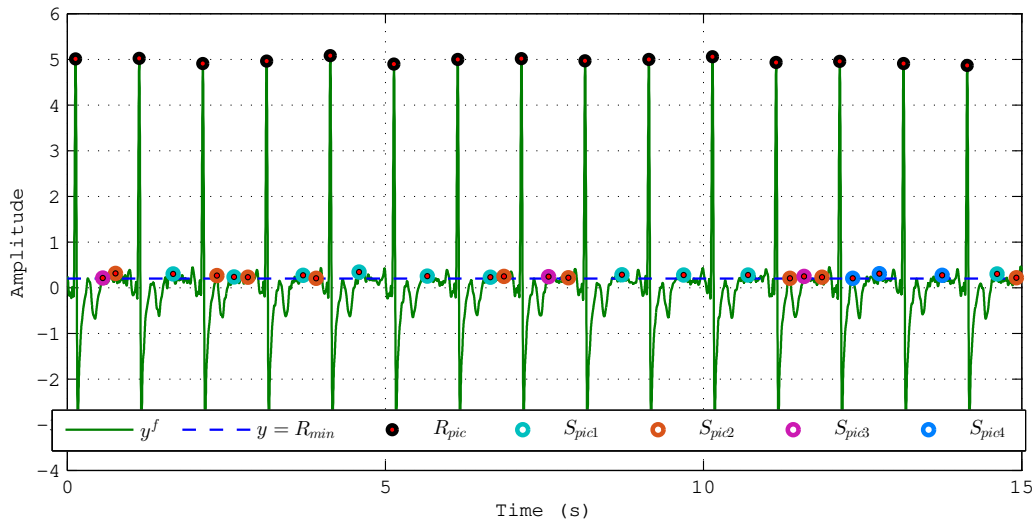


FIGURE II.12: Exemple 2 - Résultats obtenus après la quatrième filtration des faux pics.

Après avoir supprimé tous les faux pics, les intervalles R-R sont alors construits. Dans ce travail, les intervalles R-R sont utilisés pour distinguer entre un rythme sinusal normal et une arythmie. En effet, le rythme sinusal normal est caractérisé par des durées de R-R relativement similaires avec une variabilité régulière. En revanche, une arythmie est caractérisée par une variabilité irrégulière de ses intervalles R-R. De nombreux chercheurs ont développé des méthodes de détection de l'arythmie basées sur l'analyse des intervalles R-R. Le prochain chapitre donnera plus de détails à ce sujet.

II.5 Conclusion

Pendant l'enregistrement, les signaux ECG sont exposés à certaines altérations morphologiques et aux bruits superposés aux données des formes d'onde. Le filtrage des bruits est d'une importance capitale pour garantir une détection correcte des composantes de l'ECG, en particulier les complexes QRS, qui sont particulièrement intéressants pour le diagnostic des dysfonctionnements cardiaques. Dans ce chapitre, le traitement de l'ECG a été présenté en deux parties. La première partie comporte un procédé de filtrage, destiné à corriger la ligne de base en utilisant une méthode RLM, et à éliminer les bruits en utilisant un filtre IIR de type Butterworth. La seconde partie est consacrée à la détection des complexes QRS. La démarche de détection des QRS est introduite en deux sections. La première section consiste en une détection initiale, basée sur des opérations mathématiques de dérivés et de seuillage pour repérer les pics maximums, appelés pics R. La deuxième section inclut des stratégies de correction appliquées dans le cas où le choix des critères de seuillage est mal établi alors que de faux pics sont détectés. Pour une telle situation, un algorithme comprenant quatre étapes de filtration des faux pics a été développé. Ces quatre étapes consistent à éliminer tous les pics indésirables induits par le processus de filtrage initial et à produire les véritables pics R. Dans le prochain chapitre, nous présenterons l'application des pics R pour construire un modèle de diagnostic de l'arythmie.

MÉTHODES MATHÉMATIQUES D'ANALYSE DES DONNÉES VFC

III.1 Introduction	64
III.2 Méthodes mathématiques appliquées aux VFC	65
III.2.1 Méthode conventionnelle	65
III.2.1.1 Stratégie de segmentation des signaux VFC	65
III.2.1.2 Principaux indicateurs de la VFC	65
III.2.1.3 Approches de classification appliquées aux indicateurs de la VFC	66
III.2.2 Méthode proposée	66
III.2.2.1 Objectifs	66
III.2.2.2 Dynamiques RR	67
III.3 Analyse de la variabilité des dynamiques RR	69
III.3.1 Objectifs	69
III.3.2 Statistique descriptive unidimensionnelle	69
III.3.2.1 Mesures de tendance centrale	69
III.3.2.2 Mesures de variabilité	70
III.3.3 Analyse statistique unidimensionnelle appliquée aux dynamiques RR^m	73
III.3.3.1 Mesures de la corrélation linéaire	73
III.3.3.2 Mesures d'association entre RSN sans battement ectopique et FA	73
III.3.3.3 Mesures d'association entre RSN avec battement ectopique et FA	76
III.3.4 Discussion	77
III.4 Extraction des Caractéristiques	78
III.4.1 Méthode linéaire	78
III.4.1.1 Caractéristiques du domaine temporel	78
III.4.1.2 Caractéristiques du domaine fréquentiel	79
III.4.2 Méthode géométrique	80
III.4.3 Méthode non linéaire	81
III.5 Conclusion	83

III.1 Introduction

Au cours des dernières années, des demandes urgentes sont apparues concernant le développement de systèmes automatisés permettant d'évaluer des données ECG volumineuses et de détecter les épisodes arythmiques. En effet, les enregistrements ECG à long terme et la quantité de données connexes compliquent les interprétations manuelles faites par les médecins. Idéalement, des algorithmes de dépistage automatique peuvent être utilisés pour réduire le temps consacré aux examens manuels aux seules interventions nécessaires. La plupart des algorithmes qui s'inscrivent dans ce contexte ont mis l'accent sur l'étude de la variabilité des séries temporelles des intervalles R-R. L'intervalle R-R désigne l'oscillation entre deux battements cardiaques successifs. Elle peut être calculée comme l'intervalle de temps séparant deux ondes R adjacentes. La variation de la durée de ces intervalles est de nature dynamique et non linéaire, dont la représentation graphique en fonction du temps décrit la variabilité de la fréquence cardiaque (VFC). La VFC constitue l'un des outils diagnostiques clés de l'ECG, qui permet de modéliser et d'analyser mathématiquement le comportement cardiaque. De nombreuses études ont utilisé les indicateurs de VFC issus du biosignal ECG pour surveiller le rythme cardiaque. D'autres études ont également démontré le bénéfice des indicateurs de VFC pour détecter les arythmies, en particulier la fibrillation auriculaire (FA).

Dans ce chapitre, la première partie aborde les méthodes mathématiques appliquées aux VFC, en présentant les méthodologies conventionnelles appliquées ainsi que les principaux indicateurs des VFC présents dans la littérature. Nous avons également présenté les algorithmes d'apprentissage qui sont souvent appliqués pour combiner les indicateurs VFC et construire des modèles de diagnostic. Compte tenu des limitations des approches existantes, nous avons présenté la méthodologie adoptée pour décrire les comportements des battements de cœur. Celle-ci est présentée par différentes dynamiques, la dynamique de base des intervalles R-R, et ses dérivées successives, normales et absolues. Dans la deuxième partie, une étude de la variabilité a été menée sur chaque dynamique, en utilisant une analyse statistique descriptive unidimensionnelle. L'objectif consistait à tester la capacité de chaque dynamique en vue de distinguer un rythme sinusal normal (RSN) avec et sans battements ectopiques d'un rythme FA. Deux études distinctes ont été effectuées pour évaluer les dynamiques extraites. La première a été consacrée à la mesure du degré d'association entre un épisode de RSN sans battements ectopiques et un épisode de FA. La seconde consiste en un épisode RSN avec battements ectopiques et FA. Enfin, la troisième partie de ce chapitre est consacrée à introduire l'étape d'extraction de caractéristiques réalisée dans le cadre de cette thèse. Celle-ci comprend les caractéristiques du domaine temporel linéaire, du domaine fréquentiel, du domaine géométrique et du domaine non linéaire.

III.2 Méthodes mathématiques appliquées aux VFC

III.2.1 Méthode conventionnelle

Les approches de détection automatisée des épisodes FA sont généralement basées sur des indicateurs VFC caractérisant le comportement dynamique et/ou la variabilité des séries temporelles des intervalles R-R. En fonction des indicateurs mesurés, une analyse statistique ou une méthode de classification avancée est appliquée pour construire une fonction de discrimination permettant de distinguer un épisode de FA d'un épisode de RSN.

III.2.1.1 Stratégie de segmentation des signaux VFC

Les indicateurs estimés de la variabilité de la série temporelle R-R sont calculés sur une fenêtre glissante comprenant un nombre prédéterminé d'intervalles R-R consécutifs. Comme il n'existe pas de norme sur la longueur de la fenêtre, de nombreuses études ont essayé de définir une longueur adéquate, afin d'obtenir la meilleure performance de classification. Certains auteurs ont considéré que les épisodes de FA d'une durée inférieure à 30 secondes ne sont pas significatifs pour le diagnostic médical, cette hypothèse conduisant donc à augmenter le nombre de battements cardiaques à 64 [63, 102], 80 [45], 100 [51], 128 [102] et 150 [103]. D'autres auteurs affirment que les fenêtres étendues sont susceptibles de manquer les épisodes de FA courts, et ont donc appliqué des fenêtres très courtes : 40 [50], 32 [58, 59, 102], 12 [47] ou même 8 battements [49]. De toute évidence, la longueur des fenêtres utilisée pour la détection automatique de la FA dépend de l'approche employée.

III.2.1.2 Principaux indicateurs de la VFC

La mesure de la VFC s'appuie principalement sur des méthodes du domaine temporel et du domaine fréquentiel [104] :

- Les caractéristiques du domaine temporel consistent en trois mesures classiques communes, à savoir la valeur moyenne (μ_{RR}), écart-type (σ_{RR} or SD) et médiane (M_{RR}). En outre, des mesures très répandues, telles que TPN , CV , $SDSD$, $pNN50$ et $RMSSD$, sont souvent appliquées [45, 66, 104, 65]. Les intervalles R-R normalisés [105] et les différences R-R normalisées [55], ainsi que des mesures non linéaires, telles que $EnSh$ [45], $ApEn$ [52], $SampEn$ [60, 104], et $CoSEn$ [65] sont aussi adoptés. D'autres formes de caractérisation des irrégularités R-R comme l'histogramme de densité des ΔRR s [106], le diagramme des séries temporelles d'intervalles R-R en séquences symboliques binaires [103] et le score de Markov de l'intervalle R-R [44] ont montré des résultats intéressants.
- Les caractéristiques du domaine fréquentiel ont été également appliquées, dont la plus fréquente est la FFT [62]. D'autres techniques consistent à utiliser le modèle AR [62, 107] et la décomposition en ondelettes (DWT) [53]. La méthode de Burg a été couramment adoptée

pour établir la bande spectrale des fréquences VFC [104]. Deux bandes de fréquences sont généralement distinguées : la bande des hautes fréquences (*HF* : entre 0,15 Hz et 0,40 Hz) et la bande des basses fréquences (*LF* : entre 0,04 et 0,15 Hz). La puissance basse fréquence normalisée (*LF_n*) et la puissance haute fréquence normalisée (*HF_n*), ainsi que leur rapport (*LF/HF*) ont été largement exploités comme indices dans le domaine fréquentiel [108].

III.2.1.3 Approches de classification appliquées aux indicateurs de la VFC

La méthode classique de classification FA consiste à utiliser la courbe ROC afin de déterminer les seuils optimaux des variables d'entrée assurant la meilleure performance de classification [105, 102]. La méthode de test statistique (Kolmogorov-Smirnov) a été également appliquée dans [106] pour vérifier les histogrammes de densité des données de test par rapport aux histogrammes de densité standard établis comme référence pour les épisodes FA. Dans le cas des études multivariées, les indicateurs de la VFC ont été combinés et évalués afin de distinguer les épisodes de FA et de non-FA, en utilisant différentes méthodes de classification : modèle de forêt d'arbres décisionnels [64, 65], classificateur des *k* plus proches voisins [66], carte auto adaptative, machine à vecteurs de support [66, 57–60], ainsi que les algorithmes de réseau de neurones artificiels, avec perceptron multicouche [66, 62], l'apprentissage de la quantification vectorielle [109], filtre de Kalman étendu [110], réseau neuronal probabiliste [111], réseau neuronal de régression généralisée [52, 61], et de nombreuses autres approches de classification.

Un apprentissage profond inédit a été adopté également pour automatiser la détection de la FA dans les enregistrements ECG à long terme. En effet, le système de classification étudie les intervalles R-R sans avoir besoin d'en extraire les caractéristiques. En [68], un modèle basé sur le réseau neuronal récurrent avec mémoire à court terme a été appliqué, et en [69], une combinaison a été proposée avec les réseaux de neurones convolutifs et récurrents pour extraire les caractéristiques essentielles. Malgré les performances remarquables de la classification, la complexité de calcul du modèle d'apprentissage profond est bien plus élevée que celle du classificateur classique basé sur les caractéristiques.

III.2.2 Méthode proposée

III.2.2.1 Objectifs

Le déploiement de méthodes de diagnostic fiables de la FA est désormais indispensable à l'amélioration de la qualité de la prise de décision sur les enregistrements ECG à long terme. En raison du temps considérable investi dans les analyses de données ECG volumineuses, les médecins ont privilégié le recours à des systèmes d'aide au diagnostic automatisés, ce qui leur permet de réduire le temps d'évaluation, sauf si nécessaire. Considérant les travaux de recherche sur le diagnostic automatique de la FA, la plupart d'entre eux se sont principalement focalisés sur l'analyse des caractéristiques extraites des intervalles R-R (étude de la VFC) et sur les algorithmes d'apprentissage automatique modernes.

L'extraction des caractéristiques des intervalles R-R menée dans le cadre des algorithmes précités est principalement basée sur des mesures linéaires, telles que des indicateurs du domaine temporel et du domaine fréquentiel, des mesures géométriques et des mesures non linéaires. Cependant, ces séries d'intervalles R-R ne contiennent pas forcément toutes les connaissances pointues et les informations essentielles sur la variation du rythme cardiaque, puisqu'elles décrivent seulement la fréquence cardiaque. Tout comme dans la sémantique physique, où un état dynamique se décrit plus précisément en déterminant sa position, sa vitesse et son accélération, plutôt que de se limiter à sa position, nous émettons l'hypothèse que le recours à des variables induites - soit des variables extraits des dérivées temporelles d'intervalles R-R - peut étendre le champ de représentation du comportement cardiaque en intégrant de nouveaux indicateurs clés et fournir un système d'aide au diagnostic de l'arythmie plus performant par rapport à la littérature. La figure III.1 illustre la structure du système proposé pour le diagnostic de l'arythmie.

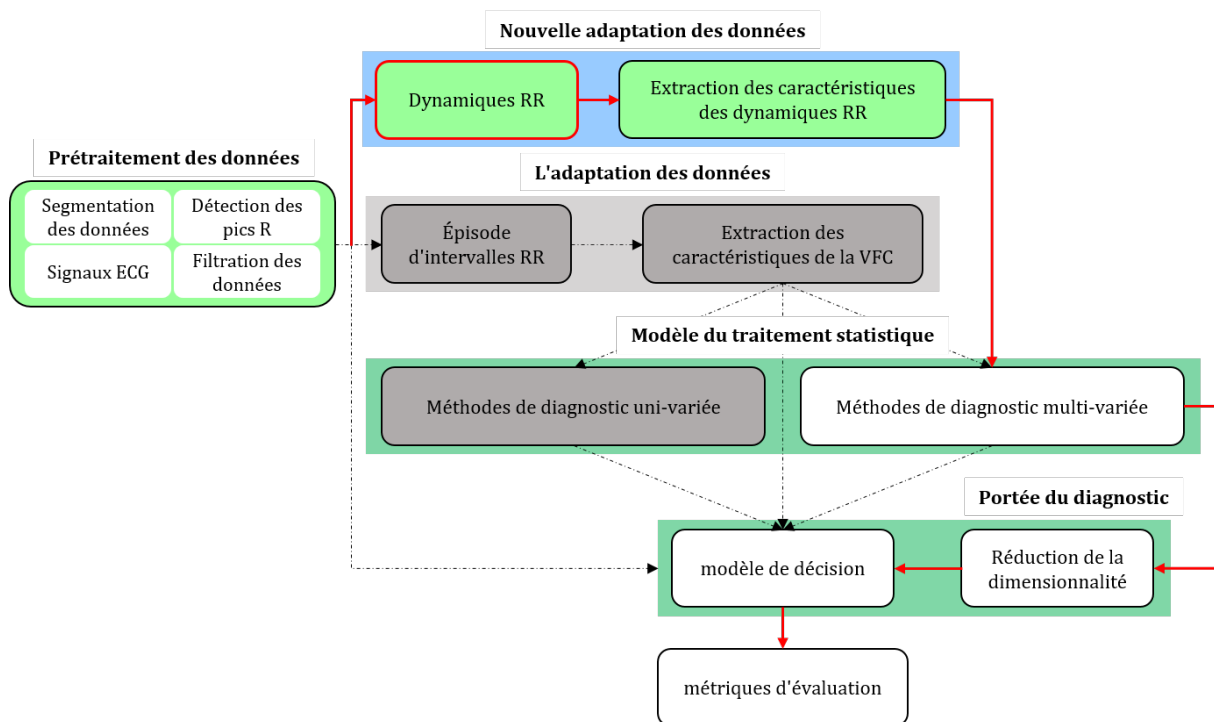


FIGURE III.1: Schéma synoptique des opérations effectuées dans le système de diagnostic envisagé.

III.2.2.2 Dynamiques RR

Dans ce travail, de nombreuses formes de dynamiques R-R sont introduites. Chaque forme est évaluée à l'aide des mesures de variabilité et des indicateurs statistiques associés, afin de décrire le comportement dynamique en présence de RSN, RSN avec battements ectopiques et FA. Les dynamiques R-R décrites ici sont dérivées des séries temporelles successives des ondes R. Deux formes de dérivées sont considérées : les dérivées normales et les dérivées absolues.

- Dynamique de base (série temporelle des intervalles R-R)

$$RR = RR^1 = \{R(i+1) - R(i) \quad \forall i = 1, \dots, n-1\} \quad (\text{III.1})$$

- Dérivées normales

$$RR^{m+1} = \left\{ \frac{R^m(i+1) - R^m(i)}{R^1(i+1)} \quad \forall i = 1, \dots, n-m-1 \right\} \quad \forall m = 1, \dots, p \quad (\text{III.2})$$

- Dérivées absolues

$$RR^{m+1} = \left\{ \frac{|R^m(i+1) - R^m(i)|}{R^1(i+1)} \quad \forall i = 1, \dots, n-m-1 \right\} \quad \forall m = 1, \dots, p \quad (\text{III.3})$$

En analyse, le calcul de la différence finie permet de mesurer le taux de croissance (ou taux de variation). Pour une série temporelle d'intervalles R-R réguliers, où la distance entre deux pics R successifs se ressemble, le taux de variation est souvent faible. En revanche, pour une série temporelle d'intervalles R-R irréguliers, la distance entre chaque deux pics R varie de manière aléatoire et le taux de variation augmente. L'étude de la variabilité menée dans le cadre de cette thèse est basée sur la variabilité des épisodes des intervalles R-R, et sur ses "p" dérivées temporelles. Cette démarche consiste à introduire de nouvelles sources de données, permettant de mieux distinguer les rythmes réguliers et irréguliers.

La figure III.2 montre les coefficients de corrélation mesurés entre les segments RR de RSN et AF. Selon les résultats présentés dans cette figure, nous avons retenu les dérivées des intervalles R-R jusqu'à l'ordre 4.

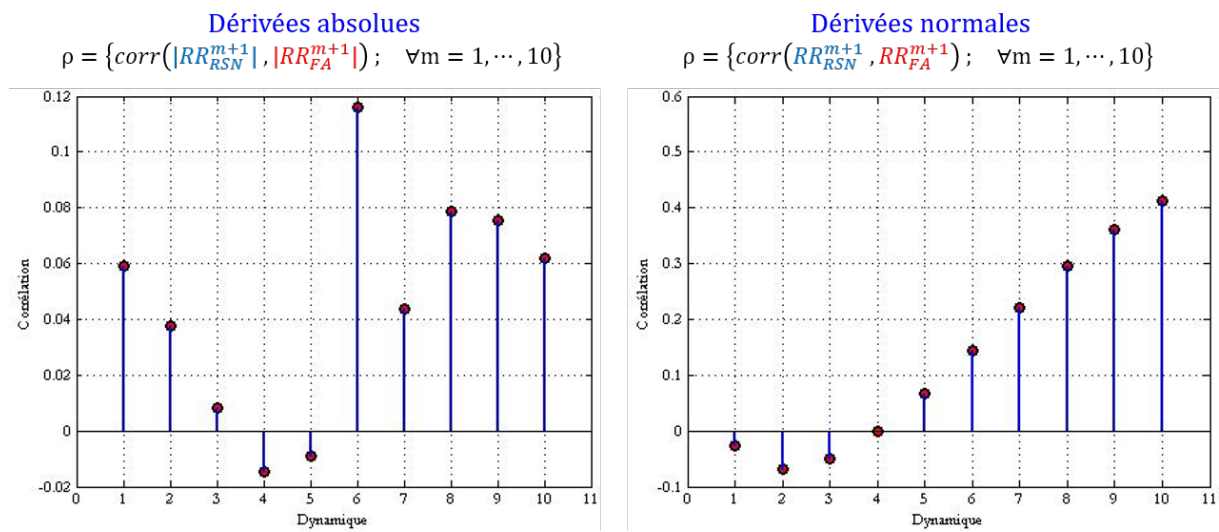


FIGURE III.2: Calcul de corrélation appliqué aux dynamiques normales et absolues

III.3 Analyse de la variabilité des dynamiques RR

III.3.1 Objectifs

Dans cette section, l'objectif est d'étudier la consistance de chaque dynamique RR implémentée. Pour ce faire, une analyse de variabilité a été adoptée, en utilisant des méthodes statistiques descriptives. Celle-ci consiste à examiner chaque dynamique RR et à démontrer sa pertinence à partir de la comparaison des différents indicateurs pris sur les épisodes de RSN avec ceux mesurés sur les épisodes de FA. Les dynamiques RR présentant des indicateurs de variabilité significatifs par rapport aux RSN et aux FA seront prises en compte dans le modèle de la prise de décision.

L'analyse statistique descriptive facilite l'interprétation des données et constitue une composante fondamentale de l'apprentissage automatique. En effet, l'apprentissage automatique consiste à établir des prédictions. D'autre part, l'analyse statistique consiste à établir des indicateurs à partir de données. Les statistiques descriptives décrivent les caractéristiques essentielles des données d'une étude. Elles permettent ainsi de synthétiser les mesures et la répartition des échantillons. Celles-ci sont classées en deux approches principales :

1. L'approche quantitative qui consiste à décrire et à présenter les données sous forme numérique.
2. L'approche graphique qui permet d'illustrer les données à l'aide de diagrammes, de graphiques, d'histogrammes et autres.

Les statistiques descriptives peuvent être appliquées à une ou plusieurs séries de mesures ou variables. La description d'une seule variable est une analyse dite univariée (unidimensionnelle). En revanche, la description de deux variables fait l'objet d'une analyse bivariée (bidimensionnelle). Ainsi, une analyse multivariée (multidimensionnelle) porte sur plusieurs variables à la fois.

III.3.2 Statistique descriptive unidimensionnelle

III.3.2.1 Mesures de tendance centrale

Les mesures de la tendance centrale constituent la description la plus basique et, souvent, la plus instructive des particularités d'une série temporelle. Il permet de décrire la répartition des données autour de la valeur centrale. Les trois principales mesures les plus courantes sont : la moyenne, la médiane et le mode. Dans notre contexte, le mode représentant la valeur la plus fréquente apparaissant dans les données est insignifiant puisque les données des dynamiques RR^m ne sont pas toutes identiques. Nous avons donc limité notre analyse à la moyenne et à la médiane.

III.3.2.1.1 La moyenne La moyenne, désigne la valeur moyenne calculée sur tous les éléments d'un ensemble de données. Elle s'exprime mathématiquement sous la forme suivante :

$$\overline{RR^m} = \frac{1}{Nm} \sum_{i=1}^{Nm} RR_i^m \quad (\text{III.4})$$

D'autres formes de moyennes peuvent également être incluses, telles que la moyenne harmonique, la moyenne géométrique et la moyenne quadratique.

- La moyenne harmonique

La moyenne harmonique est une forme de moyenne calculée en divisant le nombre d'observations par la réciproque de chaque mesure de la série temporelle. Elle s'exprime mathématiquement sous la forme suivante :

$$\overline{RR^m}^H = Nm \times \left(\sum_{i=1}^{Nm} \frac{1}{RR_i^m} \right) \quad (\text{III.5})$$

- La moyenne géométrique

La moyenne géométrique est la n -ième racine du produit des tous les éléments dans une série de données. Elle s'exprime mathématiquement sous la forme suivante :

$$\overline{RR^m}^G = \sqrt[n]{\prod_{i=1}^{Nm} RR_i^m} \quad (\text{III.6})$$

- La moyenne quadratique

La moyenne quadratique d'un ensemble de mesures est la racine carrée de la moyenne arithmétique des carrés de ces mesures. Elle s'exprime mathématiquement sous la forme suivante :

$$\overline{RR^m}^Q = \sqrt{\frac{1}{Nm} \sum_{i=1}^{Nm} (RR_i^m)^2} \quad (\text{III.7})$$

III.3.2.1.2 La médiane La médiane est la valeur mesurée à partir de la série ordonnée en ordre croissant ou décroissant. Si le nombre des données n est impair, la médiane est la valeur centrale de l'ensemble de données : $0.5(n + 1)$. Si n est pair, alors la médiane est la moyenne arithmétique entre les deux valeurs du milieu, c'est-à-dire les éléments aux positions $0.5n$ et $0.5n + 1$. Elle s'exprime mathématiquement sous la forme suivante :

$$\begin{cases} \text{si } n \text{ est impair} & M_e^m = RR_{(n+1)/2}^m \\ \text{si } n \text{ est pair} & M_e^m = \frac{1}{2} (RR_{(n/2)}^m + RR_{(n/2)+1}^m) \end{cases} \quad (\text{III.8})$$

III.3.2.2 Mesures de variabilité

Les mesures de la variabilité servent à évaluer la "répartition" des données. Il se peut que deux ensembles de données distincts aient exactement la même moyenne, mais possèdent des degrés de variance différents. Par conséquent, en se basant uniquement sur la moyenne, l'analyse aboutit à des

interprétations erronées. Dans ce contexte, nous considérons différentes mesures de variabilité, telles que l'Écart interquartile, Variance, Écart-type, Coefficient de variation et Coefficient d'asymétrie.

III.3.2.2.1 Écart interquartile L'écart interquartile (EI) est une mesure de la dispersion statistique entre les quartiles supérieurs (75e) et inférieurs (25e) de RR^m .

$$EI^m = Q_3^m - Q_1^m \quad (\text{III.9})$$

où le premier quartile Q_1 est défini comme la valeur intermédiaire entre la valeur minimale et la médiane du RR^m , et le troisième quartile Q_3 est la valeur intermédiaire entre la médiane et la valeur maximale du RR^m .

III.3.2.2.2 Variance La variance est la mesure de dispersion la plus courante. Elle indique la distance numérique entre les points de données et la valeur moyenne. La variance est exprimée mathématiquement sous la forme suivante :

$$VAR^m = \frac{1}{Nm-1} \sum_{i=1}^{Nm} (RR_i^m - \overline{RR^m})^2 \quad (\text{III.10})$$

Pour plus de détails sur le principe de la division par $(Nm - 1)$ au lieu de Nm , nous recommandons de consulter la correction de Bessel [112].

III.3.2.2.3 Écart-type L'écart-type (σ ou ET) est une autre mesure de la dispersion des données. Il est lié à la variance exprimée sous forme de racine carrée de celle-ci. L'écart-type est souvent plus approprié que la variance, car il a la même unité que les données.

$$ET^m = \sqrt{\frac{1}{Nm-1} \sum_{i=1}^{Nm} (RR_i^m - \overline{RR^m})^2} \quad (\text{III.11})$$

III.3.2.2.4 Coefficient de variation Le coefficient de variation (CV) permet d'évaluer la vraisemblance de la moyenne par rapport à l'ensemble des observations. Il indique le degré d'homogénéité d'une série de données.

$$CV^m = \frac{\sigma}{\overline{RR^m}} \times 100 \quad (\text{III.12})$$

III.3.2.2.5 Coefficient d'asymétrie Le coefficient d'asymétrie (Skewness en anglais) mesure l'asymétrie des données. Plusieurs définitions mathématiques des coefficients d'asymétrie sont proposées. L'une des expressions les plus courantes est :

$$Skew^m = \frac{\sum_{i=1}^{Nm} (RR_i^m - \overline{RR^m})^3}{(Nm-1)\sigma^3} \quad (\text{III.13})$$

Sept mesures ont été adoptées entre ceux de la tendance centrale et de variabilité. Ces mesures sont appliquées pour chaque ensemble de dynamique RR^m pour démontrer l'existence des indépendances intrinsèques. La figure III.3 ci-après résume les principales étapes de l'extraction des statistiques descriptives. L'ensemble de données ECG LTAfDb a été utilisé pour constituer l'ensemble de données d'analyse. Après segmentation en épisodes R-R de 30 secondes, nous avons sélectionné 5000 épisodes de chaque rythme RSN avec et sans battement ectopique et FA.

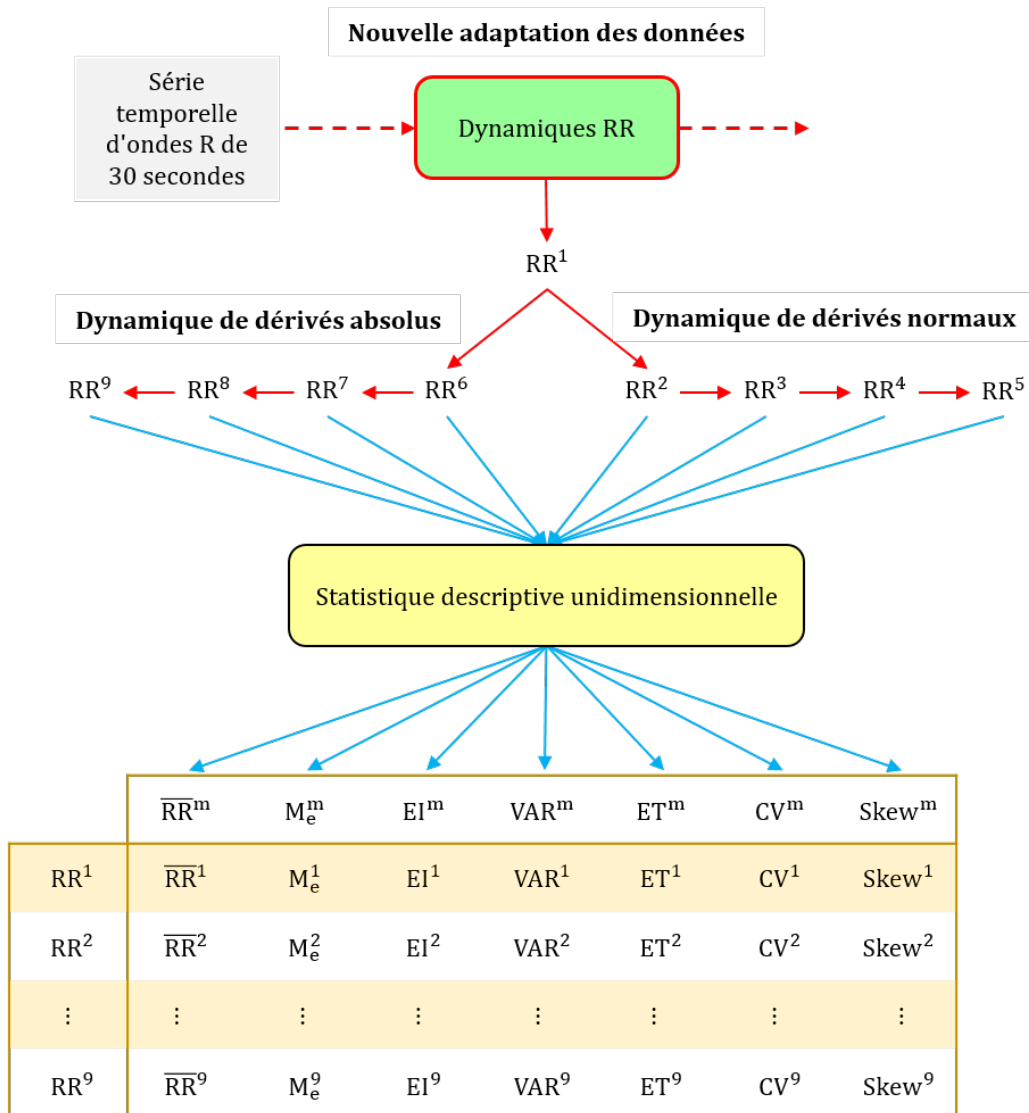


FIGURE III.3: Diagramme synoptique présentant l'organigramme d'extraction des statistiques descriptives.

III.3.3 Analyse statistique unidimensionnelle appliquée aux dynamiques RR^m

L'idée de l'analyse statistique unidimensionnelle consiste à estimer la fiabilité de chaque forme RR^m , basé sur le principe de la corrélation linéaire, c'est-à-dire que, si le degré de corrélation entre les variables extraites des épisodes RSN et FA est très élevé, la forme RR^m impliqué ne permet pas de distinguer les comportements de ses épisodes. D'autre part, si la mesure de corrélation est très faible, cela signifie que la forme RR^m implicite comporte elle-même une certaine caractéristique d'indépendance entre le comportement des épisodes RSN et celui de la FA.

III.3.3.1 Mesures de la corrélation linéaire

La détermination des corrélations linéaires entre les ensembles de données est très importante, et les coefficients de corrélation sont d'une grande utilité. Le coefficient de corrélation de Pearson, également appelé r de Pearson est l'un des outils les plus couramment utilisés pour interpréter la corrélation linéaire entre deux variantes. La corrélation est exprimée par un coefficient compris entre -1 et 1 , où

- $r > 0$ indique une corrélation positive
- $r < 0$ indique une corrélation négative
- $r = 1$ est la valeur maximale possible de r . Elle correspond à une corrélation linéaire positive parfaite entre les variables.
- $r = -1$ est la valeur minimale possible de r . Elle correspond à une corrélation linéaire négative parfaite entre les variables.
- $r \approx 0$, ou quand r s'approche de zéro, signifie que la corrélation entre les variables est faible.

La formule r de Pearson est donné comme suit : Soient n mesures quantitatives d'une même variable

$$X = [X_1 \quad X_2 \quad \cdots \quad X_k \quad X_{k+1} \quad X_{k+2} \quad \cdots \quad X_n]$$

de deux classes (X_i, X_j) avec $i = 1, 2, \dots, k$ et $j = k + 1, k + 2, \dots, n$ désigne les indices de la classe RSN et FA respectivement. Le coefficient de corrélation r entre les mesures X_i et X_j est

$$r = \frac{\text{cov}(X_i, X_j)}{\sigma(X_i) \sigma(X_j)} \quad (\text{III.14})$$

Le r de Pearson est constitué essentiellement de la covariance $\text{cov}(\cdot)$ divisé par le produit des écarts-types. Nous avons également introduit la valeur de probabilité pour améliorer les interprétations.

III.3.3.2 Mesures d'association entre RSN sans battement ectopique et FA

Le tableau III.1 présente les scores de corrélation et les valeurs p obtenues pour chaque variable entre deux classes RSN sans battements ectopiques et AF. Selon l'interprétation statistique de r , les

scores les plus faibles indiquent un faible degré de corrélation entre les deux classes, cette hypothèse caractérisant essentiellement les variables discriminantes. Selon le tableau des résultats, les valeurs r les plus basses sont associées à des valeurs p proches de 1, ce qui signifie qu'il n'y a pas de présomption contre l'hypothèse nulle. Sur la base des mesures d'association rapportées, nous pouvons souligner l'existence de 4 dynamiques ayant des signes d'indépendance significative (case en vert dans III.1).

Les mesures d'association ne se limitent pas au calcul des valeurs de corrélation, mais nous pouvons également démontrer l'existence des variables discriminantes par la visualisation de la distribution. Par la suite, les courbes de distribution de certaines variables sont fournies, notamment celles à caractère indicatif mises en évidence par la distribution des échantillons (voir Fig.III.4).

TABLE III.1: Résultats des mesures d'association entre RSN sans battement ectopique et FA

		\overline{RR}	Me	EI	VAR	ET	CV	$Skew$
RR^1	r	0.0015	0.0035	-0.0095	-0.0145	-0.0024	0.0112	-0.0119
	p	0.9142	0.8038	0.5026	0.3054	0.8634	0.4280	0.4004
RR^2	r	0.0023	0.0015	0.0031	-0.0094	-0.0049	-0.0019	0.0049
	p	0.8699	0.9136	0.8275	0.5070	0.7294	0.8912	0.7293
RR^3	r	-0.0151	-0.0059	0.0037	-0.0019	-0.0050	0.0052	-0.0178
	p	0.2854	0.6753	0.7935	0.8941	0.7230	0.7137	0.2091
RR^4	r	8.61e-4	0.0058	-0.0014	0.0091	-5e-4	5.76e-4	-0.0077
	p	0.9515	0.6828	0.9220	0.5215	0.9719	0.9675	0.5870
RR^5	r	-0.0110	0.0104	0.0255	-0.0033	0.0025	-2.14e-4	-0.0097
	p	0.4352	0.4616	0.0712	0.8131	0.8605	0.9879	0.4941
RR^6	r	0.0027	0.0056	0.0158	-0.0108	-0.0108	-0.0069	-0.0103
	p	0.8461	0.6901	0.2628	0.4461	0.4470	0.6262	0.4680
RR^7	r	0.0130	0.0129	0.0101	-0.0068	-0.0067	-0.0152	0.0084
	p	0.3596	0.3605	0.4751	0.6323	0.6367	0.2819	0.5538
RR^8	r	0.0135	0.0237	-0.0048	2.32e-4	-0.0011	0.0016	0.0087
	p	0.3391	0.0933	0.7340	0.9869	0.9364	0.9100	0.5378
RR^9	r	6.82e-4	-0.0038	-0.0045	-0.0026	-0.0035	0.0087	0.0159
	p	0.9615	0.7907	0.7478	0.8514	0.8073	0.5396	0.2616

D'après la figure III.4, la distribution de la mesure des variables CV^1 , VAR^2 , ET^3 , EI^4 , ET^5 et ET^9 démontre un modèle de comportement distinct entre le RSN et le FA. Du point de vue de leurs coefficients r , ceux-ci ne sont pas significatifs, alors qu'ils présentent deux tendances bien marquées. En effet, toute variable présentant un comportement de distribution distinct implique que la dynamique sous-jacente est en cause. Dans ce cas, cette dynamique est significative et peut aboutir à des résultats intéressants dans le cadre de l'analyse multidimensionnelle. Jusqu'à présent, nous avons discuté de l'association entre la RSN sans battement ectopique et la FA. Dans la suite, nous aborderons le cas de la RSN avec battement ectopique et de la FA.

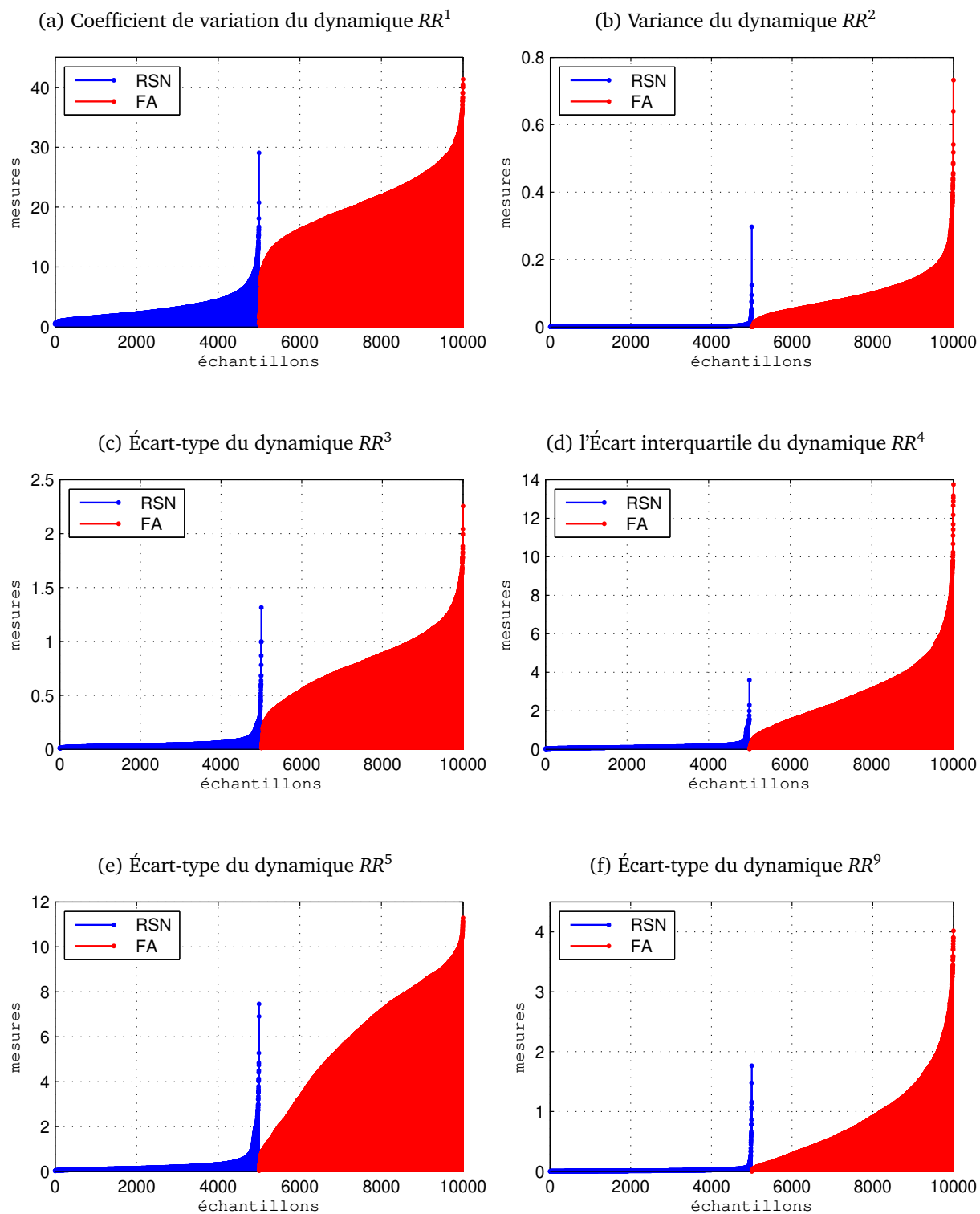


FIGURE III.4: Courbes de mesures d'association indiquant la distribution statistique du RSN sans battement ectopique et de la FA.

III.3.3.3 Mesures d'association entre RSN avec battement ectopique et FA

Le tableau III.2 présente les scores de corrélation et les valeurs p obtenues pour chaque variable entre deux classes RSN avec battements ectopiques et AF. Sur la base des mesures d'association rapportées, nous pouvons souligner également dans ce cas l'existence de 5 dynamiques ayant des signes d'indépendance significative (case en vert dans III.2). Cependant, les courbes de distribution n'ont pas apporté d'autres formes que celles déjà sélectionnées.

Selon les mesures de corrélation présentées dans le tableau III.2, on peut constater l'existence de scores significatifs traduisant la pertinence de certaines dynamiques susceptibles d'être utiles pour distinguer la variabilité des deux ensembles de données. En combinant les connaissances acquises dans les deux cas d'étude, nous concluons que la plupart des dynamiques ont des qualités essentielles qui peuvent être utiles pour caractériser le rythme cardiaque. En ce qui concerne la dynamique 6, aucun score de corrélation significatif n'a été détecté dans les deux cas, ce qui nous permet de l'éliminer. Une fois les dynamiques souhaitées retenues, la méthodologie consiste à remodeler chaque dynamique en une série de variables descriptives à travers un bloc d'extraction de variables (voir Fig. III.1). Ce processus est considéré comme un deuxième niveau d'évaluation de la pertinence de la dynamique via ses prédicteurs déduits. Dans ce cas l'évaluation est conduite par des méthodes de sélection de variables ou de réduction de la dimensionnalité.

TABLE III.2: Résultats des mesures d'association entre RSN avec battement ectopique et FA.

		\overline{RR}	Me	EI	VAR	ET	CV	$Skew$
RR^1	r	0.0112	0.0163	0.0057	-9.43e-4	-0.0098	-0.0094	0.0238
	p	0.4274	0.2493	0.6895	0.9469	0.4866	0.5073	0.0928
RR^2	r	-0.0051	-0.0055	-0.0118	-0.0086	-0.0192	-0.0040	-0.0218
	p	0.7197	0.6955	0.4026	0.5426	0.1738	0.7777	0.1237
RR^3	r	-0.0120	-3.43e-4	-0.0082	-0.0099	-0.0129	-2.33e-4	0.0040
	p	0.3957	0.9807	0.5614	0.4834	0.3615	0.9869	0.7793
RR^4	r	-0.0021	-0.0131	-0.0014	-0.0034	-0.0021	-9.36e-4	0.0107
	p	0.8799	0.3547	0.9185	0.8089	0.8800	0.9472	0.4477
RR^5	r	-0.0159	-0.0048	0.0125	0.0016	0.0031	0.0015	-0.0086
	p	0.2601	0.7353	0.3755	0.9113	0.8289	0.9169	0.5452
RR^6	r	-0.0159	-0.0132	-0.0048	-0.0091	-0.0206	-0.0065	-0.0095
	p	0.2609	0.3520	0.7326	0.5203	0.1449	0.6456	0.5035
RR^7	r	-0.0066	-2.79e-4	-0.0020	-0.0025	-0.0122	0.0129	-6.14e-4
	p	0.6426	0.9843	0.8858	0.8570	0.3874	0.3621	0.9654
RR^8	r	0.0021	-0.0075	-0.0024	-0.0035	-0.0036	-0.0064	-0.0170
	p	0.8815	0.5967	0.8664	0.8070	0.7970	0.6490	0.2281
RR^9	r	0.0046	0.0051	0.0045	3.93e-4	-0.0024	-0.0123	-0.0259
	p	0.7464	0.7184	0.7488	0.9779	0.8680	0.3840	0.0672

III.3.4 Discussion

La discrimination entre les épisodes de RSN avec battements ectopiques et de FA constitue le défi majeur du diagnostic de l'arythmie. En général, l'existence de battements ectopiques dans les signaux RSN est problématique, car leur présence perturbe la structure de conduction normale. Les chercheurs s'appuient sur l'analyse de la VFC ont parfois préféré de supprimer ces battements ectopiques avant de procéder à des mesures de variabilité [52, 113], tandis que d'autres ont tenté d'atténuer les effets des battements ectopiques en utilisant des méthodes de filtrage [49].

Dans la pratique, Il est préférable de tenir compte des épisodes à battements ectopiques pour préserver la fiabilité du modèle en temps réel. Considérant l'impact de ces battements ectopiques sur la dynamique des intervalles R-R qui limitent son utilité comme indicateur d'irrégularité. Ceci à été abordé en appliquant des formes dérivées des intervalles R-R pour réduire de tels impacts, de sorte qu'un épisode de RSN avec des battements ectopiques perturbateurs devienne une forme régulière de RSN normal.

L'objectif principal de l'étude sera maintenant de transformer (remodeler) les dynamiques calculées en indices de variabilité (variables descriptives ou indicateurs) que les algorithmes d'apprentissage automatique peuvent exploiter pour construire un modèle de décision. De nombreuses mesures peuvent être déduites sur un épisode de 30 secondes des dynamiques RR. Celles-ci peuvent être divisées en trois types :

1. Le premier type concerne les mesures extrait par les méthodes traditionnelles d'analyse linéaire pour refléter la variabilité des données, y compris la moyenne, la médiane, la variance, l'écart-type, le coefficient de variation et d'autres mesures destinées à caractériser la variabilité d'un épisode dans le domaine temporel, tandis que la puissance dans le domaine des basses fréquences (BF), la puissance dans le domaine des hautes fréquences (HF) et le rapport entre BF et HF sont appliqués pour refléter la variabilité dans le domaine fréquentiel.
2. Le second type consiste en des caractéristiques géométriques extraites par des méthodes non linéaires, reflétant les tendances de distribution des données, en utilisant les diagrammes de Poincaré et les mesures de la distribution de la densité d'échantillon.
3. Le troisième type, se sont les mesures extrait par des méthodes analytiques non linéaires, exprime la complexité des données, y compris les mesures d'entropie telles que l'entropie de l'échantillon, l'entropie de Shannon, l'entropie approximative, ainsi que les mesures de et les mesures non paramétriques, comme TPR (turning point ratio).

III.4 Extraction des Caractéristiques

Dans cette section, nous aborderons les principales méthodes utilisées pour extraire les mesures essentielles du rythme cardiaque, sur lesquelles nous n'avons aucune connaissance particulière. Il s'agit donc d'évaluer chaque série de battements cardiaque pour en identifier et en extraire ses caractéristiques.

L'extraction des caractéristiques est une phase cruciale, qui décrit les irrégularités et les variations incluses dans les modèles dynamiques élaborés précédemment. L'adéquation du modèle objectif est principalement basée sur la pertinence de ces caractéristiques retenues et fournies au processus d'apprentissage. Chaque dynamique RR^m peut être exprimée en utilisant des méthodes linéaires, géométriques et non linéaires. Tous les paramètres mentionnés dans cette section sont appliqués automatiquement à chaque dynamique RR^m . Si nous supposons que le processus d'extraction des caractéristiques comporte M variables, nous obtenons en sortie $M \times 9$ variables.

III.4.1 Méthode linéaire

III.4.1.1 Caractéristiques du domaine temporel

Les mesures statistiques dans le domaine temporel sont divisées en deux classes :

1. Mesures effectuées sur les intervalles R-R
2. Mesures à partir des différences entre les intervalles R-R

III.4.1.1.1 Mesures effectuées sur les intervalles R-R Il s'agit des mesures déduites directement à partir des intervalles R-R, comme le moyen \overline{RR} , la médiane M_e , l'écart interquartile EI , la variance VAR , l'écart-type ET , le coefficient de variation CV , et le coefficient d'asymétrie $Skew$. Ces dernières sont déjà décrites précédemment dans la section consacrée à l'étude statistique descriptive unidimensionnelle. En outre, la déviation absolue médiane (MAD) et la médiane des médianes (MM) sont deux indicateurs robustes de la variabilité qui peuvent également être mesurés sur les intervalles R-R.

- **MAD** (l'écart absolu médian)

$$MAD = M_e \left(\sum_{i=1}^N |RR_i - M_e(RR)| \right) \quad (III.15)$$

- **MM** (la médiane des médianes)

$$MM = M_e \left(\begin{array}{c} M_e(RR(1 \rightarrow Q)) \\ M_e(RR(Q + 1 \rightarrow 2 * Q)) \\ M_e(RR(2 * Q + 1 \rightarrow N)) \end{array} \right) \quad (III.16)$$

où Q désigne la partie entière de la division $\frac{N}{3}$ noté $\lfloor \frac{N}{3} \rfloor$.

III.4.1.1.2 Mesures à partir des différences entre les intervalles R-R Il s'agit des mesures calculées à partir des différences entre les intervalles R-R. La mesure la plus courante est

- **SDSD** (La déviation standard des différences entre les intervalles R-R adjacents)

$$\text{SDSD} = \sqrt{\frac{\sum_{i=1}^{N-1} (D_i - \bar{D})^2}{N-1}} \quad (\text{III.17})$$

où le coefficient D_i est défini par $D_i = RR_{i+1} - RR_i$ selon l'expression :

$$D = \begin{bmatrix} D_1 & D_2 & \cdots & D_{N-2} & D_{N-1} \end{bmatrix} \quad (\text{III.18})$$

$$= \begin{bmatrix} RR_2 - RR_1 & RR_3 - RR_2 & \cdots & RR_{N-1} - RR_{N-2} & RR_N - RR_{N-1} \end{bmatrix} \quad (\text{III.19})$$

et

$$\bar{D} = \frac{1}{N-1} \sum_{i=1}^{N-1} D_i \quad (\text{III.20})$$

Parmi les mesures standard les plus répandues aussi appliquées aux intervalles R-R, on peut citer NN50, qui correspond au nombre de différences d'intervalles R-R successives supérieures à 50 ms, pNN50, qui correspond à la proportion obtenue en divisant NN50 par le nombre total d'intervalles R-R et RMSSD. Ces mesures peuvent être utilisées pour estimer les variations à haute fréquence du rythme cardiaque.

- **RMSSD** (la moyenne quadratique des intervalles R-R successifs)

$$\text{RMSSD} = \sqrt{\frac{\sum_{i=1}^{N-1} (D_i)^2}{N-1}} \quad (\text{III.21})$$

- **pNN50** (pourcentage des D_i successifs dont la différence est supérieure à 50 ms)

$$\text{pNN50} = \frac{\sum_{i=1}^{N-1} \Delta_i}{N-1} \times 100 \quad (\text{III.22})$$

où

$$\Delta_i = \begin{cases} 1 & \text{si } (D_{i+1} - D_i > 50) \\ 0 & \text{sinon} \end{cases}$$

L'ensemble des Δ_i désigne la variable NN50.

III.4.1.2 Caractéristiques du domaine fréquentiel

Les méthodes de fréquence sont plus complexes que les méthodes temporelles en termes d'exécution et d'opérations de calcul. L'analyse spectrale appliquée aux épisodes RR consiste à décrire la composition fréquentielle des intervalles R-R comme une somme d'ondes sinusoïdales correctes d'amplitude variable. La mesure de la densité spectrale de puissance (DSP) dans des plages de fréquences individuelles permet d'évaluer la VFC dans l'analyse des fréquences [108].

La méthode de Burg a été couramment adoptée pour établir la bande spectrale des fréquences VFC à travers une mesure de la densité spectrale de puissance (DSP) dans des plages de fréquences individuelles [104]. Deux bandes de fréquences sont généralement distinguées :

- **LF** (la bande des basses fréquences entre 0,04 et 0,15 Hz)
- **HF** (la bande des hautes fréquences entre 0,15 Hz et 0,40 Hz)
- **LF/HF** (le rapport entre les puissances des bandes de basse et de haute fréquence.)

Les gammes de fréquences indiquées ci-dessus sont définies à la suite de la publication de la norme commune européenne et américaines en 1996 [114].

III.4.2 Méthode géométrique

Le diagramme de Poincaré est une méthode pertinente consacrée à l'analyse des signaux à partir d'un quantitatif géométrique et non linéaire permettant d'évaluer la dynamique des VFC [56, 115]. Le diagramme de Poincaré est une projection graphique de la corrélation temporelle entre des intervalles R-R successifs. Le diagramme de Poincaré peut être décrit par plusieurs indices, tels que SD1, SD2, SD1/SD2, VAI (vector angular index) et VLI (vector length index) [56, 116, 117].

- **SD1** (reflète la variabilité à court terme des battements cardiaques, définis comme la déviation standard de la projection du tracé Poincaré sur la droite perpendiculaire à la droite d'identité ($y = x$))

$$SD1 = \sqrt{\frac{1}{2}SDSD^2} \quad (III.23)$$

- **SD2** (définie comme la distance de l'ellipse mesure la variabilité à long terme des battements cardiaques et est calculée comme la déviation standard de la projection du tracé Poincaré sur la ligne d'identité ($y = -x + 2\overline{RR}$))

$$SD2 = \sqrt{2ET - \frac{1}{2}SDSD^2} \quad (III.24)$$

- **RSDSD** (est la mesure de la randomisé dans les séries temporelles des VFC et est définie comme le rapport entre **SD1** et **SD2**)

$$RSDSD = \frac{SD1}{SD2} \quad (III.25)$$

- **VAI** (mesure la dispersion angulaire des points de distribution)

$$VAI = \frac{1}{N} \left(\sum_{i=1}^N |\theta_i - 45| \right) \quad (III.26)$$

où θ_i est l'angle entre la ligne établie à partir de chaque point de dispersion jusqu'au point d'origine (0,0) et l'axe des d'identité.

- **VLI** (mesure la dispersion à distance des points de distribution)

$$VLI = \sqrt{\frac{1}{N} \left(\sum_{i=1}^N (l_i - L)^2 \right)} \quad (\text{III.27})$$

où l_i est la distance entre chaque point de dispersion et le point d'origine, et L est la moyenne de tous les l_i .

III.4.3 Méthode non linéaire

Pour remédier à la complexité des caractéristiques irrégulières de la VFC, certains indicateurs non linéaires sont implémentés pour quantifier ces effets. La complexité liée aux séries temporelles d'intervalles R-R est largement abordée à l'aide de mesures basées sur l'entropie. L'"entropie" permet de quantifier le degré de désordre dans un système. Lorsqu'elle est appliquée à des séries temporelles, elle permet de mesurer le degré de répétabilité et de prévisibilité de ceux-ci. Une faible entropie indique plus de régularité. Plusieurs variantes de l'entropie sont proposées dans la littérature, notamment l'entropie approximative (ApEn, approximate entropy), présenté par Pincus [118] pour mesurer la variabilité des enregistrements courts et bruyants de signaux physiologiques.

- **ApEn** (l'entropie approximative), est une mesure de la régularité et de la complexité des séries temporelles. ApEn, appliqué aux données des VFC, indique une faible prédictibilité aux fluctuations des intervalles R-R successifs. Les faibles indices ApEn indiquent que le signal est régulier et prévisible [118]. Cependant, l'indice d'entropie décrit par ApEn pour un signal est influencé par la combinaison de ses paramètres d'entrée.

$$\text{ApEn}(RR, m, r \times ET(RR)) \quad (\text{III.28})$$

Richman et al. [119] a développé une solide méthode pour mesurer la régularité, l'entropie de l'échantillon (SampEn), réduisant le biais causé par l'autocorrespondance de l'ApEn. D'autres mesures impliquent, le coefficient de l'entropie de l'échantillon (CoSEn, coefficient of sample entropy) [47] suivi de l'entropie floue (FuzzyEn, fuzzy entropy) [120]. Ceux-ci ont été adoptés comme mesures performantes permettant de caractériser les irrégularités des épisodes RR.

- **SampEn** (l'entropie de l'échantillon) : la fonction SampEn permet d'identifier la régularité en fonction de la probabilité qu'un fragment de longueur m prélevé sur un signal, soit identique à une seconde partie de ce même signal, avec une tolérance r . Cette correspondance doit également être établie si le fragment est de longueur $m + 1$. Le paramètre m définit donc la taille des séquences qui peuvent être comparées et r agit comme un filtre, tout pic de bruit supérieur à r n'affectant pas le résultat, qui est généralement défini par rapport à l'écart-type de la série temporelle. [119, 121].

$$\text{SampEn}(RR, m, r \times ET(RR)) \quad (\text{III.29})$$

- **CoSEn** (le coefficient de l'entropie de l'échantillon) : la fonction CoSEn est une estimation de l'entropie optimisée pour la détection de l'arythmie de la fibrillation auriculaire. Celui-ci est censé être plus élevé lorsqu'il est appliqué au signal de la VRC avec un taux d'irrégularité important.

$$\text{CoSEn} = \text{SampEn}(RR, m, r \times ET(RR)) - \ln(2r) - \ln(\overline{RR}) \quad (\text{III.30})$$

- **FuzzyEn** (l'entropie floue) : Le FuzzyEn est une fonction pertinente et efficace conçue pour refléter la complexité dynamique du signal de la VRC. Elle a été proposée et appliquée pour la première fois à la caractérisation du signal EMG de surface [120].

$$\text{FuzzyEn}(RR, m, r \times ET(RR), n) \quad (\text{III.31})$$

Généralement, pour les valeurs r inférieurs, les indices d'entropie sont plus élevés et moins précis en raison de la réduction des correspondances m et, plus encore, pour les correspondances $m + 1$. Pour surmonter ce problème, Lake et al. [47] ont développé une nouvelle mesure appelée entropie quadratique de l'échantillon (QSE), qui est basé sur des estimations de densité plutôt que sur des estimations de probabilité. Afin de normaliser la valeur de r , SampEn a été modifiée en divisant sa probabilité par la taille de la fenêtre de tolérance globale $2r$.

- **QSE** (l'entropie quadratique de l'échantillon)

$$\text{QSE} = \text{SampEn}(RR, m, r \times ET(RR)) + \ln(2r) \quad (\text{III.32})$$

Une étude comparative a été menée par Weiting Chen et al. [122] qui ont évalué trois indicateurs de complexité, à savoir ApEn, SampEn et FuzzyEn, et les algorithmes de chaque fonction ont également été décrits en détail. En outre, pour chaque mesure d'entropie, des paramètres standard spécifiques doivent être initialisés, tels que la dimension d'encastrement (m), le seuil de tolérance (r) et le gradient du périmètre de la fonction exponentielle (n).

III.5 Conclusion

Dans ce chapitre, nous avons développé une approche de modélisation pour décrire les séries temporelles des ondes R. Tout d'abord, l'étude a été réalisée sur un intervalle de temps de 30 secondes. À partir d'un épisode de 30 secondes d'ondes R, nous avons construit une série de dynamiques, dénommées RR^m . Celle-ci est composée de la série temporelle des intervalles R-R, et ses quatre dérivées successives, normales et absolues. Ces neuf dynamiques sont soumises à une analyse unidimensionnelle afin d'évaluer les propriétés essentielles de chacune d'entre elles et de sélectionner uniquement le sous-ensemble de dynamiques le plus pertinent. Dans un premier temps, une étude statistique descriptive a été réalisée sur la base de deux mesures de la tendance centrale et de cinq mesures de la variabilité. Celles-ci sont appliquées à chaque dynamique, puis une mesure de corrélation est effectuée pour estimer la force d'association entre les épisodes de RSN avec et sans ectopique contre les épisodes de FA. Ensuite, nous avons extrait de chaque dynamique sélectionnée une série de mesures à l'aide de méthodes linéaires, géométriques et non linéaires. L'objectif est de constituer une base de données d'entraînement propre à la conception de modèles de diagnostic destinés à discriminer entre un rythme RSN et FA. Dans le prochain chapitre, nous présenterons l'application de ces dynamiques RR^m pour construire les modèles de diagnostic de l'arythmie.

DIAGNOSTIC DE LA FIBRILLATION AURICULAIRE

IV.1 Introduction	87
IV.2 Diagnostic de la FA à l'aide de trois dynamiques RR	88
IV.2.1 Méthodologie	88
IV.2.2 Prétraitement des données	89
IV.2.2.1 Bases de données	89
IV.2.2.2 Détection et segmentation des ondes R	89
IV.2.3 Phase de réduction de la dimensionnalité	89
IV.2.3.1 Ensemble de données d'évaluation	89
IV.2.3.2 Processus d'entraînement : Association d'analyse factorielle et de classificateurs	89
IV.2.3.3 Résultats préliminaires	90
IV.2.4 Modèle de prise de décision	91
IV.3 Diagnostic de l'arythmie à l'aide de cinq dynamiques RR	92
IV.3.1 Méthodologie	92
IV.3.2 Prétraitement des données	94
IV.3.3 Analyse préliminaire	94
IV.3.3.1 Ensemble de données d'évaluation	94
IV.3.3.2 ANOVA	94
IV.3.4 Pipeline machine learning	94
IV.3.4.1 Apprentissage des pipelines	94
IV.3.4.2 Résultats : La validation croisée à 10 blocs	95
IV.3.5 Modèle idéal	97
IV.3.6 Motivations	98
IV.4 Nouvelle approche de sélection de variables	99
IV.4.1 Introduction aux enjeux de la réduction de la dimensionnalité	99
IV.4.1.1 Vue d'ensemble	99
IV.4.1.2 Aperçu technique	100
IV.4.2 Méthodologie	102
IV.4.3 Prétraitement des données	103

IV.4.4	Présentation de l'algorithme MKL	103
IV.4.5	MKL appliqué à la sélection des variables	105
IV.4.5.1	Formulation mathématique du problème	105
IV.4.5.2	Méthode de résolution	106
IV.4.6	Modèle idéal	108
IV.5	Conclusion	110

IV.1 Introduction

La détection automatisée de l'arythmie, assistée par des systèmes informatiques, est conçue pour améliorer et renforcer les prestations de services de santé et assurer un suivi de diagnostic en temps réel. Les progrès en matière de diagnostic automatique appliqué à l'arythmie ont donné naissance à des dispositifs de dépistage informatisés tels que des bracelets, des montres intelligentes et de nombreux autres instruments. Ces dispositifs comprennent des capteurs de signaux biologiques mesurant l'activité du coeur. Ces signaux sont ensuite analysés dans le but d'établir une stratégie de diagnostic et/ou d'identifier des signes de symptômes pathologiques. Ce chapitre s'inscrit dans ce contexte et a pour objectif de développer des approches automatisées pour le diagnostic des arythmies cardiaques, notamment la fibrillation auriculaire.

Dans la première partie de ce chapitre, nous présentons l'une des premières approches de diagnostic développées pour la fibrillation auriculaire. Il s'agit d'un système de diagnostic combinant trois modèles dynamiques, décrits par 11 variables descriptives, dont 4 variables linéaires, 4 variables géométriques et 3 variables non linéaires. La structure du modèle de décision est établie en évaluant les facteurs de chaque sous-ensemble fournis par la technique d'analyse factorielle à travers des approches d'apprentissage automatique, telles que SVM, WKNN et LR. Dans la deuxième partie, nous présentons un deuxième système de diagnostic conçu pour discriminer le RSN de quatre formes d'arythmie. Ce système est composé de cinq modèles dynamiques. Chaque modèle est décrit à l'aide de 13 variables, dont 4 variables linéaires, 6 variables géométriques et 3 variables non linéaires. La démarche proposée consiste initialement à évaluer l'ensemble des 65 variables via une méthode d'analyse de la variance. Ensuite, les données de l'analyse univariée sont appliquées à une évaluation multivariée approfondie. Enfin, la troisième partie est consacrée à la présentation du modèle de diagnostic ultime, construit à l'aide d'un nouvel algorithme de sélection de variables. Nous avons adopté le principe de l'apprentissage multinoyaux pour concevoir une stratégie de sélection de caractéristiques pertinente. Dans cette structure, 72 variables ont été étudiées à l'aide d'un algorithme d'apprentissage multinoyaux.

TABLE IV.1: Tableau des abréviations des approches

SVM	support vector machine	séparateurs à vaste marge
RF	random-Forest	forêt d'arbres décisionnels
MLP	multi-layer perceptron	perceptron multicouche
WKNN	weighted K-nearest neighbors	K plus proches voisins pondérés
ETC	extra trees classifier	classificateur d'arbres extras
LR	logistic regression	régression logistique
PCA	principal component analysis	analyse en composantes principales
RFE	recursive feature elimination	élimination des caractéristiques récursives

IV.2 Diagnostic de la FA à l'aide de trois dynamiques RR

IV.2.1 Méthodologie

Dans ce contexte, nous présentons un premier modèle de diagnostic, construit à partir de trois formes dynamiques, à savoir la série temporelle des intervalles R-R et ses première et deuxième dérivées absolues. Pour décrire les caractéristiques essentielles de chaque forme dynamique, 11 méthodes descriptives ont été appliquées, dont quatre méthodes du domaine temporel, quatre méthodes géométriques et trois méthodes non linéaires. Donc, de chaque dynamique, un ensemble de 11 caractéristiques est déduit, donnant lieu à 33 caractéristiques au total. En général, l'ensemble des 33 caractéristiques peut inclure des éléments inutiles, voire non indicatifs, sans contribution significative à la réalisation du diagnostic. De ce fait, une approche d'analyse factorielle a été appliquée pour réduire l'ensemble des caractéristiques et établir une combinaison homogène des principaux indicateurs pertinents. Enfin, le sous-ensemble de caractéristiques conduisant à un diagnostic optimal sera pris en compte. La figure IV.1 illustre la disposition du modèle adopté dans ce travail.

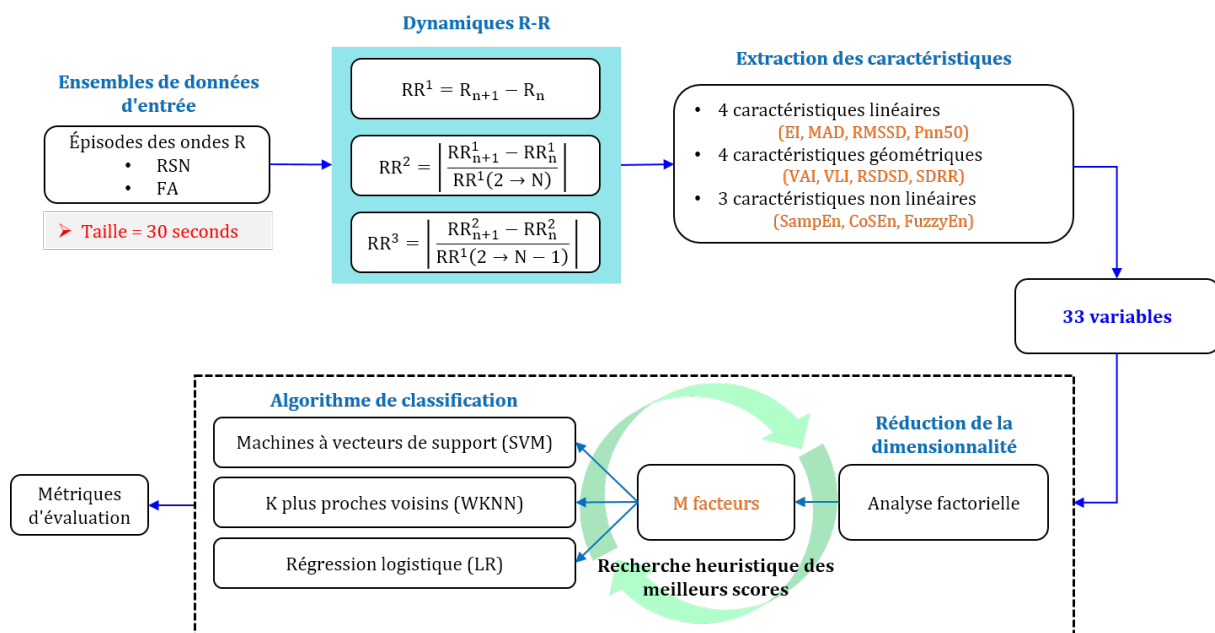


FIGURE IV.1: Diagramme synoptique de la réalisation du diagnostic considéré en utilisant trois formes dynamiques.

IV.2.2 Prétraitement des données

IV.2.2.1 Bases de données

Les modèles de diagnostic décrits dans cette étude ont été établis à partir de séries temporelles de la fibrillation auriculaire (FA) et le rythme sinusal normal (RSN). Ces dernières sont extraites des données ECG accessibles gratuitement sur PhysioBank [35]. Les bases de données utilisées sont : Long-Term AF Database [36], MIT-BIH Arrhythmia Database [38], MIT-BIH Normal Sinus Rhythm Database [40] et Normal Sinus Rhythm RR Interval Database [41].

IV.2.2.2 Détection et segmentation des ondes R

La préparation des données d'entrée comprend deux étapes principales : Identification et segmentation des ondes R. La première étape consiste à extraire les emplacements temporels de l'onde R. La seconde étape consiste à découper la séquence d'ondes R en un ensemble de "P" épisodes d'ondes R d'une durée moyenne de 30 secondes. La durée de la séquence est définie par la distance entre le premier et le dernier pic de la série temporelle de l'onde R. Cette opération nécessite donc une séquence d'onde R supérieure ou égale à 30 secondes. À partir des quatre bases de données précitées, nous avons extrait 243 876 épisodes de RSN et 76 146 épisodes de FA.

IV.2.3 Phase de réduction de la dimensionnalité

IV.2.3.1 Ensemble de données d'évaluation

La phase de réduction de la dimensionnalité a été effectuée à l'aide de 10 000 sujets, ce qui correspond à 5000 par classe, sélectionnés respectivement parmi des personnes atteintes de FA et d'autres de RSN. Chaque sujet est décrit en fonction des 33 caractéristiques qui ont été extraites de chaque épisode. Parmi les 10 000 sujets, deux tiers (soit 6 666 sujets) ont été étudiés par la technique de l'analyse factorielle afin d'établir les paramètres de mappage. Le tiers restant (soit 3334 sujets) a ensuite servi pour évaluer la pertinence des paramètres associés au modèle de mappage.

IV.2.3.2 Processus d'entraînement : Association d'analyse factorielle et de classificateurs

L'analyse factorielle a été effectuée sur les données d'apprentissage et tous les facteurs sélectionnés sont ensuite évalués à l'aide de trois classificateurs différents, à savoir SVM, LR et WKNN. Pour chaque modèle de classification, une validation croisée à 10 blocs est appliquée pour calculer les métriques d'évaluation moyennes. Après l'étape d'apprentissage, la validation se poursuit, les paramètres associés au modèle de mappage sont appliqués aux données de validation, puis les classificateurs interviennent pour évaluer l'association des nouveaux facteurs. Les classificateurs adoptés dans cette partie ont été exécutés selon leurs configurations prédéfinies. En d'autres termes, nous avons intégré un support vector machine (SVM) linéaire avec un paramètre de pénalité $C = 1$, un weighted K-nearest neighbors (WKNN) avec un facteur $K = 5$ et une logistique régression (LR) avec une régularisation L2.

IV.2.3.3 Résultats préliminaires

IV.2.3.3.1 Première expérience. Dans cette expérience, seuls les épisodes d'intervalle R-R sont utilisés, constituant un ensemble de 11 variables à travers le bloc d'extraction des caractéristiques. L'analyse factorielle est ensuite appliquée pour réduire les 11 variables à un nombre de facteurs plus restreint. Comme le montre le tableau IV.2, la plage de réduction comprend de cinq à neuf facteurs, et les pourcentages de précision et de variance ont été indiqués pour chaque exécution. Selon les résultats obtenus, une solution mutuelle a été obtenue entre la variance retenue et la précision en utilisant 6 variables.

TABLE IV.2: Performance atteinte lors de la première expérience. La précision des classificateurs est calculée en pourcentages, en utilisant la technique de validation croisée à 10 blocs.

Nombre de facteurs	Variation retenue	SVM linéaire	WKNN	LR
5	96.30	94.42	98.44	94.33
6	99.58	98.44	98.62	98.47
7	99.85	97.39	98.38	97.42
8	99.98	97.84	98.41	97.81
9	99.99	97.60	98.59	97.66

IV.2.3.3.2 Deuxième expérience. Pour démontrer l'intérêt de la méthodologie proposée, nous procédons à une seconde étude en utilisant les trois dynamiques évoquées précédemment. Ici, un total de 33 variables ont été extraites à travers le bloc d'extraction des caractéristiques puis analysées à l'aide de l'analyse factorielle. La démarche de réduction de la dimensionnalité pilotée par l'analyse factorielle comportait des sous-ensembles de dix à quatorze facteurs comme le montre le tableau IV.3. Les pourcentages de précision et de variance ont été indiqués pour chaque exécution. Conformément aux résultats observés, une solution mutuelle entre la variance retenue et la précision a été conclue pour 14 variables.

TABLE IV.3: Performance atteinte lors de la deuxième expérience. La précision des classificateurs est calculée en pourcentages, en utilisant la technique de validation croisée à 10 blocs.

Nombre de facteurs	Variation retenue	SVM linéaire	WKNN	LR
10	97.11	98.38	98.62	98.32
11	98.06	98.41	98.65	98.65
12	98.76	98.59	98.74	98.65
13	99.56	98.62	98.83	98.65
14	100.0	98.77	98.65	98.89

IV.2.4 Modèle de prise de décision

La faisabilité de la méthodologie appliquée dans cette étude est démontrée par les métriques illustrées dans le tableau IV.4. Ici, nous avons comparé les scores obtenus en utilisant une seule dynamique, à savoir les intervalles R-R, la démarche la plus répandue dans la littérature. Ainsi que les scores obtenus par la combinaison des trois dynamiques proposées, comme approche alternative possible d'amélioration.

TABLE IV.4: Évaluation SVM en pourcentage, en utilisant la technique de validation croisée à 10 blocs.

	Validation						échantillons
	F1-score		Précision		Recall		
	M.6	M.14	M.6	M.14	M.6	M.14	
RSN	98.62	98.77	98.68	98.62	98.56	98.92	1667
FA	98.62	98.77	98.56	98.92	98.68	98.62	1667
micro avg	98.62	98.77	98.62	98.77	98.62	98.77	3334
macro avg	98.62	98.77	98.62	98.77	98.62	98.77	3334
weighted avg	98.62	98.77	98.62	98.77	98.62	98.77	3334

	Test						échantillons
	F1-score		Precision		Recall		
	M.6	M.14	M.6	M.14	M.6	M.14	
RSN	98.94	99.21	99.14	99.27	98.74	99.14	238876
FA	96.47	97.35	95.81	97.14	97.13	97.57	71146
micro avg	98.37	98.78	98.37	98.78	98.37	98.78	310022
macro avg	97.70	98.28	97.48	98.21	97.93	98.36	310022
weighted avg	98.37	98.78	98.38	98.78	98.37	98.79	310022

Les résultats présentés dans le tableau 1 sont obtenus à l'aide de l'algorithme SVM, optimisé par une technique de recherche exhaustive par grille. Deux modèles de diagnostic sont comparés, le premier entraîné sur 6 facteurs et le second entraîné sur 14 facteurs, désignés respectivement par (M.6) et (M.14). Le modèle de diagnostic, M.6, reprend les 6 facteurs retenus de l'expérience 1, et M.14 reprend les 14 facteurs retenus de l'expérience 2. Pour chaque modèle SVM, certains hyperparamètres doivent être ajustés, tels que le choix du noyau, le coefficient du noyau et le paramètre de régularisation. Ces hyperparamètres ne sont pas directement appris des estimateurs. Donc, pour ajuster les hyperparamètres, une stratégie de recherche exhaustive par grille a été adoptée, un des algorithmes les plus importants et les plus répandus. Il s'agit d'un algorithme de recherche intensive qui consiste à tester toutes les combinaisons possibles d'hyperparamètres à partir d'une grille de valeurs. Chaque combinaison d'hyperparamètres est donc évaluée par une validation croisée à K-bloc. Enfin, la combinaison qui donne les meilleurs scores est sélectionnée. Dans ce contexte, une recherche exhaustive par grille avec une validation croisée décuplée a été effectuée afin d'établir les paramètres optimaux de chaque modèle. Les hyperparamètres optimisés sont le paramètre de régularisation et le coefficient de noyau. Plusieurs choix de paramètres sont testés et la meilleure

configuration est obtenue en définissant : le paramètre de régularisation $C = 0,8$ pour le modèle (M.6) et $C = 4$ pour (M.14), le coefficient du noyau ($\sigma = 0,1$) pour les deux et une fonction noyau RBF. Le tableau ci-dessous résume les performances des mesures statistiques données sous la forme d'une matrice de confusion. Celle-ci indique le nombre d'épisodes correctement classifiés et le nombre d'épisodes mal classifiés.

TABLE IV.5: Matrice de confusion

		Valeur prédite		
		RSN	FA	
Valeur réelle	RSN	VN = 236 833	FP = 2 043	238 876
	FA	FN = 1 732	VP = 69 414	71 146
		238 565	71 457	

IV.3 Diagnostic de l'arythmie à l'aide de cinq dynamiques RR

IV.3.1 Méthodologie

Dans ce travail, nous mettons l'accent sur la discrimination entre le rythme sinusal normal (RSN) et quatre rythmes cardiaques anormaux (RCA) tels que la fibrillation auriculaire (FA), le flutter auriculaire (AFL), la Bradycardie sinusale (SBR) et la tachycardie supraventriculaire (TSV). La description de chaque forme d'arythmie s'appuie sur une combinaison de cinq formes dynamiques, à savoir la série temporelle d'intervalles R-R et ses quatre premières dérivées absolues. La phase d'extraction des caractéristiques adoptée dans le cadre de l'étude actuelle comprend 13 méthodes de mesure, dont quatre méthodes temporelles, six méthodes géométriques et trois méthodes non linéaires. Par rapport au premier système de diagnostic, nous avons considéré deux méthodes géométriques et deux formes dynamiques supplémentaires. Au cours de la phase d'extraction, un ensemble de 13 caractéristiques sont extraites de chacune des 5 formes dynamiques, soit 65 caractéristiques au total. Le schéma fonctionnel de la méthode proposée est illustré dans la figure IV.4.

Le système de diagnostic adopté dans ce travail implique une phase préliminaire d'analyse univariée. L'analyse univariée sert à résoudre les problèmes des variables fortement corrélées et identifier les variables qui, seules ou multiplexées, sont susceptibles de discriminer entre des clusters

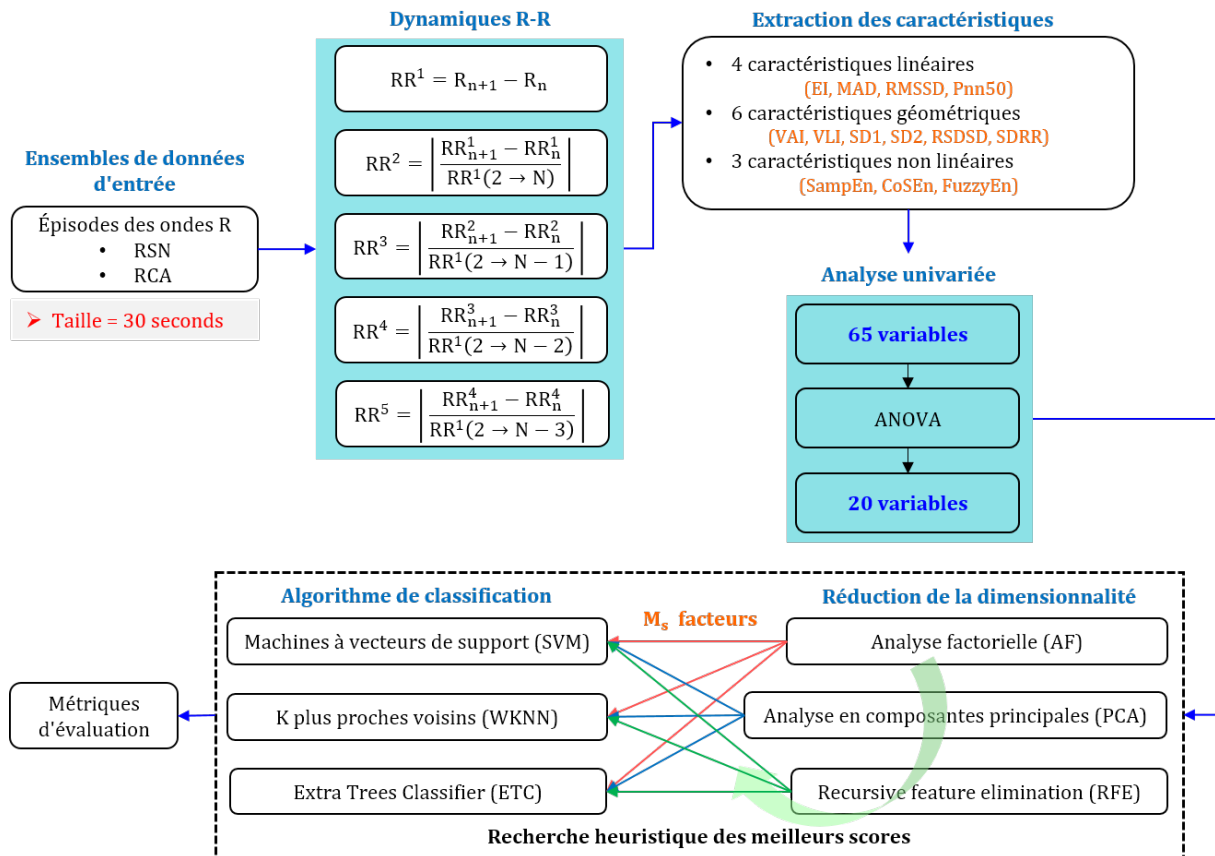


FIGURE IV.2: Diagramme synoptique de la réalisation du diagnostic considéré en utilisant trois formes dynamiques.

de données, par exemple, en distinguant un cluster de rythme cardiaque régulier d'un cluster de rythme cardiaque irrégulier. La méthode univariée appliquée pour cette analyse est l'ANOVA (Analyse de la variance). L'ANOVA est une approche de prétraitement qui permet d'éliminer les variables redondantes et non pertinentes contenues dans un ensemble de données de grande dimension. En appliquant l'ANOVA sur l'ensemble des 65 caractéristiques, on a pu identifier finalement que 20 variables significatives.

Après avoir déduit toutes les variables significatives, nous procédons à une analyse multivariée. Dans ce contexte, nous avons développé un pipeline de méthodes statistiques de réduction de la dimensionnalité et d'algorithmes d'apprentissage automatique. Les méthodes statistiques comprennent des méthodes non supervisées, telles que l'analyse en composantes principales et l'analyse factorielle, et une méthode supervisée, telle que le recursive feature elimination. En outre, pour les algorithmes d'apprentissage, les méthodes SVM, ETC et WKNN ont été utilisées pour évaluer la qualité des sous-ensembles fournis à une dimension inférieure par le processus de réduction de la dimensionnalité. Le pipeline construit a pour but de sélectionner un modèle optimal formé d'une stratégie de réduction + un algorithme de classification, permettant d'obtenir les meilleures métriques de discrimination entre les épisodes RSN et RCA.

IV.3.2 Prétraitement des données

Les données d'apprentissage utilisées dans cette étude sont extraites des signaux ECG de sujets atteints des anomalies cardiaques, tels que la FA, la AFL, la SBR et la TSV, ainsi que de sujets normaux RSN. Ces données proviennent de la banque PhysioBank [35], notamment Long-Term AF Database [36], MIT-BIH Arrhythmia Database [38], MIT-BIH AF Database [37]. À partir des trois bases de données précitées, nous avons extrait 135 899 épisodes de RSN et 137131 épisodes de RCA. En ce qui concerne la phase de détection et de segmentation des ondes R, elle est identique à celle de la démarche précédente.

IV.3.3 Analyse préliminaire

IV.3.3.1 Ensemble de données d'évaluation

L'analyse statistique des caractéristiques a été réalisée sur 10 000 sujets, soit 5 000 sujets par classe. Plus précisément, les 5 000 premiers sujets ont été prélevés à partir des signaux RSN et les 5 000 autres sujets ont été prélevés à partir des signaux FA, AFL, SBR et TSV. Chaque sujet est décrit en fonction de 65 caractéristiques issues des 5 dynamiques susmentionnées. L'ensemble des 10 000 sujets a été réparti comme suit : 3/4 (soit 7500 sujets) est consacré à l'apprentissage, et 1/4 (soit 2500 sujets) aux tests.

IV.3.3.2 ANOVA

L'approche ANOVA a été appliquée, en comparant chaque caractéristique avec la variable cible pour vérifier la présence d'une relation statistiquement significative entre les éléments. Pour chaque caractéristique, un score de test est attribué, puis tous les scores sont comparés et les caractéristiques ayant obtenu les scores les plus élevés sont sélectionnées. La figure IV.3 ci-dessous illustre les résultats obtenus après avoir effectué une ANOVA sur les 65 caractéristiques. Pour finir, les 20 premiers éléments ayant obtenu les meilleurs scores sont sélectionnés pour conduire le prochain processus.

IV.3.4 Pipeline machine learning

IV.3.4.1 Apprentissage des pipelines

Pour améliorer l'efficacité de l'apprentissage du classificateur et la fiabilité du modèle, les caractéristiques sélectionnées précédemment sont donc redéfinies et optimisées. Le bloc de réduction de la dimensionnalité est donc piloté pour établir une combinaison homogène des composantes les plus pertinentes et les plus informatives. Ainsi, les facteurs désormais acquis permettent la constitution d'un classificateur encore plus précis. En effet, l'objectif principal est de déterminer la meilleure stratégie de détection en évaluant toutes les combinaisons possibles d'analyse factorielle, d'analyse en composantes principales et de la recursive feature elimination avec les classificateurs SVM, ETC et WKNN.

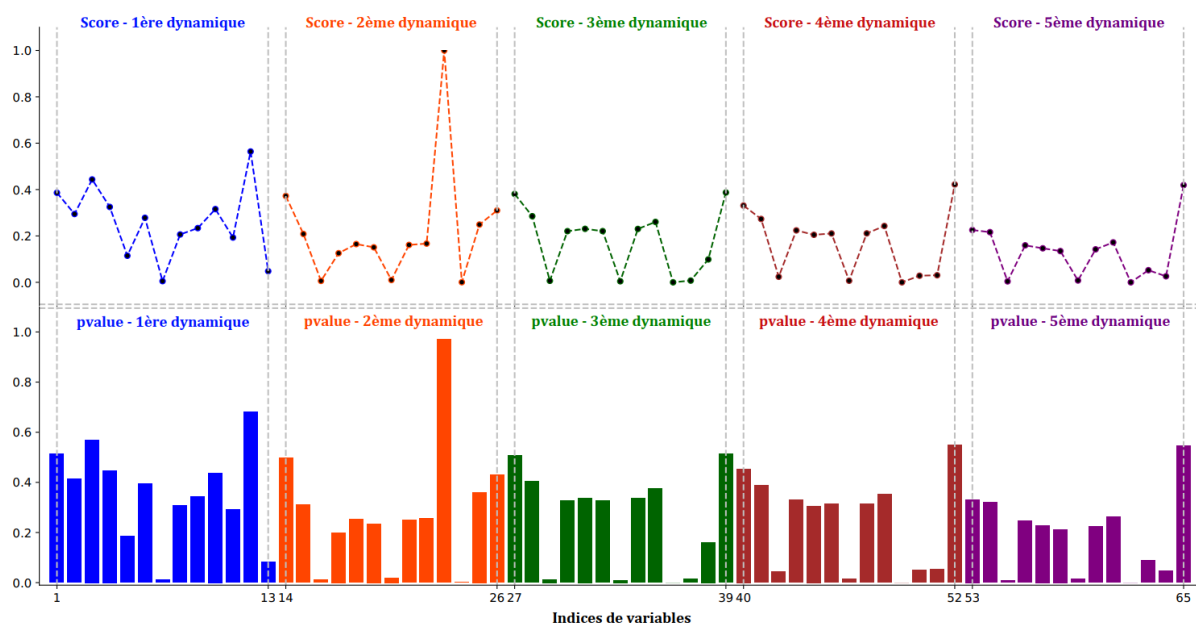


FIGURE IV.3: ANOVA appliquée aux 65 caractéristiques. Les valeurs p (p-values) sont mesurées sous la forme $-\log_{10}(p_{valeurs})$. Chaque couleur reflète les propriétés des Feature Statistics (FS) obtenues respectivement sur chacune des cinq dynamiques.

Pour chaque classificateur, certains hyperparamètres doivent être adaptés, tels que le choix du noyau, le coefficient du noyau et le paramètre de régularisation pour l'algorithme SVM. Pour l'algorithme ETC, trois paramètres doivent aussi être adaptés, notamment le nombre de divisions arborescentes, la fonction permettant de mesurer la qualité de la division et le nombre de caractéristiques à prendre en compte pour déterminer la meilleure division. Enfin, pour la technique WKNN, il faut ajuster le nombre K, et l'algorithme utilisé pour calculer les K plus proches. Donc, pour ajuster tous ces hyperparamètres, une stratégie de recherche exhaustive par grille a été adoptée. Chaque combinaison d'hyperparamètres est évaluée par une validation croisée à 10 blocs. Dans la prochaine section, nous présentons les scores obtenus pour chaque stratégie de réduction par rapport aux trois classificateurs prédéfinis du pipeline. Dans un premier temps, seulement les performances de précision obtenues pour la phase d'apprentissage sont indiquées. Le modèle qui sera retenu parmi les trois cas d'étude sera détaillé à la fin de la section.

IV.3.4.2 Résultats : La validation croisée à 10 blocs

IV.3.4.2.1 Analyse factorielle : L'analyse factorielle est une technique multivariée appliquée pour décrire les liaisons entre différentes variables examinées à travers de nouvelles variables appelées facteurs, où le nombre de facteurs est inférieur au nombre de variables initiales. Cette méthode permet d'obtenir une description efficace avec moins de facteurs si la corrélation entre les variables étudiées est importante. Le tableau IV.6 résume les résultats de tests exprimés par un score de précision.

TABLE IV.6: Les résultats obtenus en combinant "l'analyse factorielle" et les classificateurs. La précision des classificateurs est calculée en pourcentages, en utilisant la technique de validation croisée à 10 blocs.

Nombre de facteurs	ETC	WKNN	SVM
8	98.89	98.32	98.73
9	98.88	98.73	98.8
10	99.16	98.45	98.69
11	98.89	98.49	98.76
12	99.24	98.97	99.17
13	99.16	98.72	99.08
14	99.17	98.73	98.97
15	99.21	98.77	98.97

IV.3.4.2.2 Analyse en composantes principales : L'ACP est l'une des méthodes multivariées non supervisées les plus connues. C'est une technique utilisée pour transformer des variables corrélées en un nouvel ensemble de variables appelées composantes principales (CP) qui ne sont pas corrélées et où les premières composantes principales conservent l'essentiel de la variation. Le tableau IV.7 résume les résultats des tests exprimés par un score de précision.

TABLE IV.7: Les résultats obtenus en combinant l'ACP et les classificateurs. La précision des classificateurs est calculée en pourcentages, en utilisant la technique de validation croisée à 10 blocs.

Nombre de facteurs	ETC	WKNN	SVM
8	98.81	98.97	99.05
9	98.89	98.85	99.21
10	98.85	98.85	99.01
11	98.77	98.93	98.94
12	98.81	98.97	99.01
13	98.77	98.97	99.09
14	98.69	99.01	99.13
15	98.73	98.97	99.05

IV.3.4.2.3 Recursive feature elimination : Le tableau IV.8 résume les résultats des tests exprimés par un score de précision. La méthode RFE sélectionne un sous-ensemble optimal de caractéristiques à l'aide d'un estimateur prédéfini. Les caractéristiques sont éliminées récursivement en fonction du score de validation croisée atteint par l'estimateur appliqué au sous-ensemble actif. La solidité de cette technique dépend fortement de l'estimateur adopté pour évaluer les caractéristiques à chaque itération. Dans la présente analyse, nous avons utilisé les estimateurs SVM linéaire (LSVM) et la

régression logistique (LR), destinés à mesurer la pertinence des caractéristiques présentées avec différents paramètres de régularisation C .

TABLE IV.8: Les résultats obtenus en combinant le sélecteur RFE et les trois classificateurs ETC, WKNN et SVM. La précision des classificateurs est calculée en pourcentages, en utilisant la technique de validation croisée à 10 blocs.

Nombre de facteurs	Estimateur	ETC	WKNN	SVM
4	LR (C=0.001)	97.16	96.84	97.25
9	LR (C=0.01)	99.36	98.88	99
10	LSVM (C=0.001)	99.12	98.92	99.2
11	LR (C=0.1)	99.32	99.12	99.01
12	LSVM (C=0.1)	99.2	99.04	99.32
17	LSVM (C=0.01)	99.36	99.12	99.32

IV.3.5 Modèle idéal

Le modèle optimal a été sélectionné en fonction des scores obtenus en distinguant les épisodes de RSN de ceux de CAR. Le meilleur modèle a atteint une spécificité et une sensibilité moyennes de 99,36% sur une validation croisée de 10 blocs, réalisée à l'aide d'un classificateur ETC et de neuf caractéristiques sélectionnées par la technique de la RFE. La faisabilité de la méthodologie appliquée est démontrée par les métriques illustrées dans le tableau IV.9.

TABLE IV.9: Détails du modèle optimal sur la base de tests.

	Test			
	F1-score	Précision	Sensibilité	échantillons
RSN	98.24	98.08	98.40	130899
RCA	98.25	98.42	98.09	132131
	Sensibilité	Spécificité	Justesse	échantillons
	98.42	98.08	98.25	263030

Le tableau IV.10 résume les performances des mesures statistiques données sous la forme d'une matrice de confusion. Celle-ci indique le nombre d'épisodes correctement classifiés et le nombre d'épisodes mal classifiés.

TABLE IV.10: Matrice de confusion

		Valeur prédite		
		RSN	RCA	
Valeur réelle	RSN	VN = 128 813	FP = 2 086	238 876
	RCA	FN = 2 519	VP = 129 612	71 146
		131 332	131 698	

IV.3.6 Motivations

Les irrégularités de la variabilité de la fréquence cardiaque (VFC) contenue dans un épisode associé à une forme d'arythmie quelconque sont considérées comme des facteurs cruciaux pour les modèles de diagnostic. De nombreuses études ont porté sur la détection des arythmies à l'aide de caractéristiques extraites des séries de la VFC, c'est-à-dire des épisodes d'intervalles R-R. À titre d'amélioration potentielle, nous avons proposé une méthodologie de diagnostic inédite dédiée à décrire une arythmie, on étudiant non seulement la série temporelle des intervalles R-R mais aussi ses quatre premières dérivées absolues. Cette intégration de nouvelles sources de données pourrait en effet permettre de mieux comprendre les comportements de battements cardiaques et d'élargir l'espace de reconnaissance, favorisant ainsi une meilleure contribution à la prise de décision clinique. Dans l'approche précédente, nous n'avons considéré que 3 dynamiques, la série temporelle des intervalles R-R et ses première et deuxième dérivées absolues. Dans la présente étude, nous avons intégré deux formes adjacentes de dérivées successives complémentaires (la dynamique 4 et 5), dans le but de pouvoir examiner davantage d'indicateurs et d'aperçus sur les comportements cardiaques à travers ces dynamiques supplémentaires, qui peuvent être d'une importance capitale pour le diagnostic. En outre, la conception d'un modèle robuste de diagnostic de l'arythmie est déduite à partir d'une étude comparative, incluant différentes stratégies de réduction de dimensionnalité et d'algorithmes de classification, nommée pipeline. Un pipeline est constitué d'une séquence d'étapes comprenant une liste de transformateurs et d'estimateurs (classificateurs). Un transformateur prend un ensemble de données en entrée et fournit un ensemble de données réduit en sortie. Par exemple, la technique d'analyse factorielle est un transformateur qui prend un ensemble de données avec N variables en entrée et le traduit en un nouvel ensemble de données avec M facteurs en sortie, où $M \leq N$. Tandis qu'un estimateur exploite les M facteurs pour effectuer un apprentissage et construire un modèle de

prise de décision. Dans un pipeline, plusieurs transformateurs et estimateurs peuvent être combinés, de sorte que les scores les plus élevés peuvent être sélectionnés en fonction de l'association la plus appropriée établie entre un transformateur et un estimateur. Par rapport à l'approche précédente, ici nous avons évalué plusieurs associations potentielles. Ceci est très pratique pour déterminer le système de diagnostic le plus adéquat. Enfin, la principale particularité du modèle actuel par rapport à l'approche précédente est la prise en compte de plusieurs formes d'arythmie, ce qui élargit le spectre de reconnaissance pour distinguer les différentes formes d'arythmie du RSN.

IV.4 Nouvelle approche de sélection de variables

IV.4.1 Introduction aux enjeux de la réduction de la dimensionnalité

IV.4.1.1 Vue d'ensemble

Au cours de la dernière décennie, la santé numérique a connu une immense révolution avec les progrès technologiques et le développement d'outils de diagnostic intelligents et automatisés. À ce propos, l'équipe de recherche du laboratoire LIS a été investie dans la mise au point de systèmes de diagnostic automatisés, en particulier pour le diagnostic de l'arythmie. Depuis toujours, de nombreux critères ont été imposés par les experts, comme l'efficacité du diagnostic tout en minimisant les fausses alarmes (minimisation des faux positifs), et la rapidité de calcul tout en minimisant les opérations de traitement.

La conception d'un modèle de diagnostic d'arythmie fiable généralement nécessite la prise en compte de connaissances étendues mettant en évidence la présence d'une anomalie, exprimée sous forme de marqueurs ou d'indicateurs. Tandis que la combinaison de toutes les connaissances et caractéristiques essentielles (indicateur) décrivant une série de battements de cœur nécessite un calcul numérique massif et compliqué. Par exemple, l'extraction de toutes les ondes morphologiques et l'analyse de leurs intervalles et segments nécessitent un système de mesure de grande taille.

Pour répondre aux exigences des experts et pour éviter les modèles complexes, l'analyse des intervalles R-R est adoptée, vu que l'extraction des ondes R est beaucoup plus facile que l'extraction d'autres ondes et intervalles. Les méthodes de description des intervalles R-R sont très étendues, en fonction du nombre de variables extraites et du nombre de combinaisons de dynamiques utilisées.

La taille des composants construits et sélectionnés est la question la plus fréquemment évoquée. La complexité du modèle augmente au fur et à mesure que le nombre de caractéristiques augmente. Or, plus le nombre de caractéristiques augmente, plus le risque de surapprentissage persiste. Un modèle d'apprentissage automatique qui est entraîné sur une large base de données devient progressivement dépendant de celles-ci et suradapté, entraînant une mauvaise productivité sur les données réelles, contrairement à l'objectif souhaité. Éviter le surapprentissage est une des principales préoccupations pour la réduction de la dimensionnalité. Plus les données d'entraînement sont réduites, moins les estimations du modèle sont fiables et mieux adaptées.

Parmi les avantages éventuels de la réduction de la dimensionnalité, il y a tout d'abord l'amélioration de précision du modèle et l'extensibilité supérieure qui permet de surmonter la menace de la dimensionnalité en élimine les caractéristiques redondantes et le bruit. Les avantages secondaires sont notamment une meilleure visualisation et interprétation des données, une réduction des exigences en matière de mesure et de stockage, ainsi qu'une diminution du temps consacré à l'apprentissage et à l'inférence.

IV.4.1.2 Aperçu technique

Les améliorations apportées aux méthodes et techniques de mesure de la variabilité des séries temporelles, notamment les méthodes linéaires, fréquentielles, géométriques et non linéaires, élargissent la portée des descriptions des séries temporelles et contribuent à faciliter leurs interprétations. La pertinence et la qualité des mesures extraites sont essentielles pour construire un modèle de discrimination robuste. En théorie, le fait d'intégrer davantage de mesures devrait améliorer la capacité de discrimination. Cependant, de nombreuses expériences pratiques avec des algorithmes d'apprentissage automatique ont démontré que ce n'est pas toujours le cas. Typiquement, un jeu de caractéristiques peut contenir des éléments inutiles ou insignifiants sans contribution significative au processus de discrimination. Des études d'analyse univariées et multivariées (appliquées à des échantillons de mêmes substances associés à des classes différentes) sont donc nécessaires pour sélectionner les mesures adéquates à la réalisation de modèles de diagnostic fiable.

IV.4.1.2.1 Analyse variée : Les approches univariées, également appelées approches de sélection univariées, sont des techniques qui consistent à évaluer chaque variable individuellement à travers des tests statistiques unidimensionnels. Il s'agit notamment des tests de variabilité tels que le coefficient de corrélation de Pearson, le coefficient d'information maximal, et les métriques de la force de liaison (corrélation de distance ou la covariance de distance), à partir duquel les variables présentent une importance significative pour la détection de l'arythmie peut être identifié. Ces mesures de la variabilité mentionnées ci-dessus sont appliquées à des échantillons de même variables et appartenant à des classes différentes.

IV.4.1.2.2 Analyse multivariée : Les approches multivariées sont appliquées pour évaluer plusieurs variables à la fois, elles comprennent des méthodes de sélection des variables et de réduction de dimensionnalité.

1. Réduction de la dimensionnalité

Les méthodes de réduction de la dimensionnalité sont essentiellement axées sur les approches factorielles telles que l'analyse en composantes principales, l'analyse factorielle discriminante, l'analyse factorielle des correspondances simples et multiples, l'analyse canonique, multidimensionnelle scaling et la factorisation matricielle non négative ou NMF. Il s'agit d'approches destinées à transformer

une grande base de données en une nouvelle forme de données de dimensions réduites utilisant de nouvelles variables, appelées facteurs ou composantes principales.

2. Sélection de variables

Les approches de sélection de variables permettent d'obtenir un sous-ensemble de caractéristiques de l'ensemble initial qui constitue les caractéristiques les plus représentatives des données. Parmi les approches de sélection les plus répandues, on trouve les approches basées sur des modèles d'apprentissage machine, où les sous-ensembles de variables sélectionnés sont combinés et évalués à travers des algorithmes d'apprentissage automatique pour évaluer la pertinence du groupe. En outre, certains modèles d'apprentissage machine ont même une structure hiérarchique interne spécifique au classement des caractéristiques. Il s'agit notamment des modèles de régression, des SVMs, des arbres de décision, des forêts aléatoires, etc.

En littérature, les techniques de sélection des variables sont classées en trois types : (i) méthodes d'enveloppes (Wrapper methods), y compris les techniques de sélection forward, backward, et stepwise selection ; (ii) méthodes de filtrage (Filter methods), y compris l'ANOVA, la corrélation de Pearson, le seuillage de la variance ; (iii) les méthodes intégrées (Embedded methods), y compris les techniques Lasso, Ridge et les arbres de décision.

• Wrapper methods

Les méthodes d'enveloppes consistent à évaluer un sous-ensemble donné de variables à travers des opérations d'entraînement et de test effectués à l'aide d'un modèle de classification défini. L'idée consiste à itérer et tester plusieurs combinaisons des sous-ensembles de variables jusqu'à ce que le sous-ensemble optimal soit atteint. Or, comme l'espace du sous-ensemble de variables augmente exponentiellement en fonction du nombre de variables total, des méthodes de recherche heuristique sont utilisées pour guider la recherche d'un sous-ensemble optimal. Guyon et al. [123] ont proposé une approche de sélection variable appelée SVM-Recursive Feature Elimination (SVM-RFE) pour un noyau linéaire, ainsi qu'une approximation pour les noyaux non linéaires.

• Filter methods

Les méthodes de filtrage estiment la pertinence des variables en se basant uniquement sur les caractéristiques intrinsèques des données sans tenir compte de son adaptation à l'algorithme d'apprentissage. Ces méthodes utilisent des indicateurs qui permettent de hiérarchiser les variables en fonction de leurs mesures descriptives. Les méthodes de filtrage ont l'avantage d'avoir un temps de calcul très faible et surtout en évitant l'effet de surapprentissage les données.

• Embedded methods

Les méthodes intégrées effectuent la sélection des caractéristiques comme partie intégrante pour la création du modèle. Cette démarche permet généralement de faire un compromis entre les deux méthodes décrites précédemment, vu que la sélection est assurée en même temps que le processus

d'adaptation du modèle. Les méthodes de régression Lasso et Ridge sont les deux principales méthodes de sélection de caractéristiques utilisées pour ce type d'analyse.

Les approches modernes se concentrent donc sur la construction d'algorithmes de sélection de caractéristiques basés sur un apprentissage automatique, tel que l'algorithme SVM à base de noyau. Le SVM est un outil puissant pour l'analyse des données. À l'origine, l'application du SVM aux données biomédicales était limitée à des objectifs de classification et de régression puisqu'il n'était pas conçu pour évaluer l'importance des variables prédictives. Actuellement, des recherches ont été menées pour mesurer l'importance des variables associées à un modèle SVM, en mettant l'accent sur le modèle SVM à noyau linéaire.

IV.4.2 Méthodologie

Les séries temporelles R-R peuvent être décrites à l'aide de diverses caractéristiques. Puis, ces caractéristiques devaient être optimisées pour améliorer l'efficacité de l'apprentissage du classificateur et la fiabilité du modèle. Différents types d'analyse de variables sont utilisés pour résoudre les problèmes d'extraction ou de sélection des caractéristiques. Le but de ces méthodes est de supprimer les caractéristiques non informatives et non pertinentes de sorte que le sous-ensemble généré puisse être exploité pour construire un classificateur plus précis [124]. Dans de nombreux cas, la sélection d'un très grand nombre de variables complique le processus de réduction et de sélection des variables appropriées. Nous sommes principalement intéressés à réduire autant que possible le nombre de caractéristiques requises par le modèle de reconnaissance. En effet, pour des données de grande dimension, les méthodes de réduction et de sélection permettent d'identifier des sous-ensembles de caractéristiques significatives, cependant que ce nombre reste très élevé par rapport aux exigences du projet.

La démarche développée dans ce travail consiste à utiliser le principe de l'apprentissage multi-noyaux (MKL) pour supprimer les caractéristiques redondantes, en se basant sur un ensemble de noyaux combinés entre les caractéristiques. La méthode MKL a été proposée par [125], qui considère des combinaisons coniques de matrices de noyaux pour la classification, menant à un système quadratique convexe à contrainte quadratique. Le recours à la MKL s'explique par sa capacité à apprendre à partir d'un ensemble prédéfini plus large de noyaux et de paramètres une combinaison linéaire (ou non linéaire) optimale des noyaux. Plusieurs approches ont été présentées à cette fin, notamment l_1 -norm [126], l_p -norm [127], entropie [128] et normes mixtes [129]. Dans ce document, l'approche "SimpleMKL" a été adoptée dans le cadre d'un processus de sélection des caractéristiques [130].

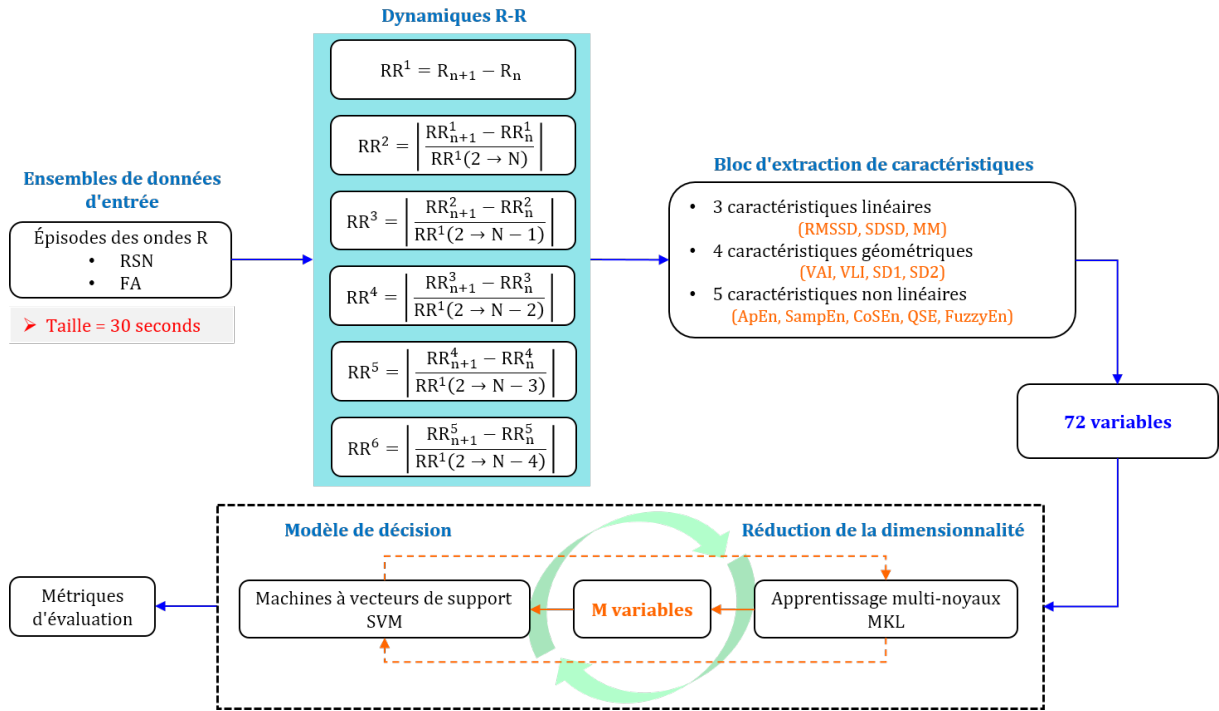


FIGURE IV.4: Diagramme synoptique de la réalisation du diagnostic considéré en utilisant six formes dynamiques et une approche de sélection de variables basée sur l'apprentissage de multinoyaux.

IV.4.3 Prétraitement des données

Les données d'apprentissage utilisées dans cette étude sont extraites des signaux ECG de sujets atteints de la fibrillation auriculaire AF, ainsi que de sujets normaux RSN. Ces données proviennent de la banque PhysioBank [35], notamment AF Termination Challenge Database [39] Long-Term AF Database [36], et MIT-BIH Arrhythmia Database [38]. À partir de ces bases de données, nous avons extrait 70 133 épisodes de RSN et 65 698 épisodes de RCA. En ce qui concerne la phase de détection et de segmentation des ondes R, elle est identique à celle des démarches précédentes.

IV.4.4 Présentation de l'algorithme MKL

Le recours à l'apprentissage multinoyaux s'explique par sa capacité à apprendre, à partir d'un ensemble plus large de noyaux et de paramètres prédéfinis, une combinaison optimale de noyaux, linéaire ou non linéaire, permettant une résolution optimale d'un problème d'apprentissage automatique. En outre, plutôt que de construire un nouveau noyau, les algorithmes à noyaux multiples peuvent être utilisés pour combiner des noyaux existants pour établir de nouvelles formes de représentation des données. Pour les modèles à base de noyau, la solution du problème d'apprentissage peut être exprimée comme suit :

$$f(x) = \sum_{s=1}^k \alpha^s y^s K(x, x_s) + b \quad (\text{IV.1})$$

où α^s et b sont des paramètres obtenus en résolvant le problème d'optimisation standard du SVM, et le noyau $K(\cdot)$ est défini par une combinaison convexe de M noyaux, calculée comme par :

$$\left\{ \begin{array}{l} K(x_i, x_j) = \sum_{m=1}^M \omega_m K_m(x_i, x_j) \\ \text{sous contraintes } \omega_m \geq 0, \quad \sum_{m=1}^M \omega_m = 1 \end{array} \right. \quad (\text{IV.2})$$

où M est le nombre total de noyaux et ω_m ses poids associés.

Ces noyaux peuvent simplement être une fonction RBF avec différents choix de paramètres σ . Le problème de la représentation des données à travers le noyau devient alors un problème de détermination des poids ω_m . Dans ce contexte, nous avons utilisé un algorithme, appelé "SimpleMKL", pour résoudre le problème de la représentation des données à travers une combinaison multiple de noyaux et déterminer les poids ω_m optimaux. L'algorithme SimpleMKL est une formulation de régularisation pondérée de la norme l_2 où la dispersion de la combinaison linéaire des noyaux est régulé par une contrainte de la norme l_1 sur les poids ω_m . La forme de la solution du problème d'apprentissage devient alors :

$$f(x) = \sum_{s=1}^k \alpha^s y^s \sum_{m=1}^M \omega_m K_m(x, x_s) + b \quad (\text{IV.3})$$

L'apprentissage des coefficients α^s et les poids ω_m dans un seul problème d'optimisation constitue ce qu'on appelle le problème de l'apprentissage à noyaux multiples (MKL). Dans le cadre du MKL, la fonction de décision donnée par

$$f(x) = \sum_{m=1}^M f_m(x), \quad (\text{IV.4})$$

est une combinaison de différentes fonctions $f_m(x)$ associés à chaque noyau $K_m(\cdot)$. La solution du problème MKL primaire est calculée en résolvant les équations convexes suivantes :

$$\begin{array}{l} \min_{f_m, b, \epsilon, \omega_m} \quad \frac{1}{2} \sum_m \frac{1}{\omega_m} \|f_m\|^2 + C \sum_i \epsilon_i \\ \text{sous contraintes } \left\{ \begin{array}{l} y_i (\sum_m f_m(x_i) + b) \geq 1 - \epsilon_i \quad \epsilon_i \geq 0 \quad \forall i \\ \sum_{m=1}^M \omega_m = 1 \quad \omega_m \geq 0 \quad \forall m \end{array} \right. \end{array} \quad (\text{IV.5})$$

où chaque ω_m contrôle la norme au carré de f_m de la fonction de coût.

Lorsque $\omega_m = 0$, f_m s'applique également à zéro afin de garantir une valeur de consigne finie. La contrainte de la norme l_1 appliqué aux poids ω consiste à forcer une partie des ω_m à être nulle, favorisant ainsi la dispersion des noyaux de base. Pour résoudre IV.5, nous considérons le problème

d'optimisation sous contrainte suivant :

$$\min_{\omega} J(\omega) \text{ tel que } \sum_{m=1}^M \omega_m = 1 \quad \omega_m \geq 0 \quad \forall m \quad (\text{IV.6})$$

$$\text{Où } J(\omega) = \begin{cases} \min_{f_m, b, \epsilon, \omega_m} & \frac{1}{2} \sum_m \frac{1}{\omega_m} \|f_m\|^2 + C \sum_i \epsilon_i \\ y_i (\sum_m f_m(x_i) + b) \geq 1 - \epsilon_i \quad \epsilon_i \geq 0 \quad \forall i \end{cases} \quad (\text{IV.7})$$

Pour résoudre le problème donné dans l'équation IV.6, nous utilisons la méthode du gradient simple. Ensuite, pour résoudre le problème formulé dans IV.7, nous adoptons une structure identique à celle appliquée pour résoudre un problème dual du SVM. En appliquant la formulation de Lagrange au problème IV.7, la valeur optimale de la fonction objectif $J(\omega)$ peut être calculée en utilisant le noyau combiné comme :

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j \sum_{m=1}^M \omega_m K_m(x_i, x_j) + \sum_{i=1}^n \alpha_i \\ \text{sous contraintes} \quad & \begin{cases} 0 \leq \alpha_i \leq C \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases} \quad \forall i \end{aligned} \quad (\text{IV.8})$$

La fonction $J(\omega)$ est définie comme la valeur objective optimale du problème IV.8. En raison de la forte dualité, $J(\omega)$ est aussi la valeur objective de la duale du problème :

$$J(\omega) = -\frac{1}{2} \sum_{i,j=1}^n \hat{\alpha}_i \hat{\alpha}_j y_i y_j \sum_{m=1}^M \omega_m K_m(x_i, x_j) + \sum_{i=1}^n \hat{\alpha}_i \quad (\text{IV.9})$$

où $\hat{\alpha}$ sont les multiplicateurs de Lagrange que maximise IV.9 et qui peuvent être établi en utilisant n'importe quel algorithme de solveur SVM.

IV.4.5 MKL appliqué à la sélection des variables

IV.4.5.1 Formulation mathématique du problème

Dans ce contexte, une nouvelle approche MKL est développée pour effectuer une sélection de variables. Donc, l'idée est de constituer un système convexe composé de P noyaux de base.

$$\begin{cases} K(x_i, x_j) = \sum_{m=1}^P \omega_m K_m(x_i^m, x_j^m) \\ \text{sous contraintes } \omega_m \geq 0, \quad \sum_{m=1}^P \omega_m = 1 \end{cases} \quad (\text{IV.10})$$

où P est le nombre total de noyaux, défini ici par le nombre total de variables considérées et ω_m sont les poids associés à chaque variable.

Or, chaque matrice de Gram K_m est calculée sur un ensemble de données d'entraînement de couples-variables, et non de couples-sujets, tels que :

$$K_m = \begin{bmatrix} K_m(x_1^m, x_1^m) & K_m(x_1^m, x_2^m) & \cdots & K_m(x_1^m, x_n^m) \\ K_m(x_2^m, x_1^m) & K_m(x_2^m, x_2^m) & \cdots & K_m(x_2^m, x_n^m) \\ \vdots & \vdots & \ddots & \vdots \\ K_m(x_n^m, x_1^m) & K_m(x_n^m, x_2^m) & \cdots & K_m(x_n^m, x_n^m) \end{bmatrix} \quad (\text{IV.11})$$

Dans l'approche originale, le nombre de noyaux combinés donné en IV.2 est un choix personnel en fonction de l'application, donc l'utilisateur définit un nombre de noyaux de son choix. Cependant, dans cette nouvelle approche, le nombre de noyaux nécessaires est fixé par le nombre de variables étudiées. Chaque matrice de Gram K_m sélectionne une seule caractéristique d'indice "m", où $\{m = 1, \dots, P\}$. Tout en appliquant le même noyau sur toutes les variables pour standardiser leurs combinaisons respectives. On réécrit donc l'équation IV.11 comme suit :

$$K_m = \begin{bmatrix} K(x_1^m, x_1^m) & K(x_1^m, x_2^m) & \cdots & K(x_1^m, x_n^m) \\ K(x_2^m, x_1^m) & K(x_2^m, x_2^m) & \cdots & K(x_2^m, x_n^m) \\ \vdots & \vdots & \ddots & \vdots \\ K(x_n^m, x_1^m) & K(x_n^m, x_2^m) & \cdots & K(x_n^m, x_n^m) \end{bmatrix} \quad \forall m \in \{1, \dots, P\} \quad (\text{IV.12})$$

D'où,

$$\min_{\omega} J(\omega) \quad \text{tel que} \quad \sum_{m=1}^P \omega_m = 1 \quad \omega_m \geq 0 \quad \forall m \quad (\text{IV.13})$$

$$\text{Où } J(\omega) = \begin{cases} \max_{\alpha} & -\frac{1}{2} \sum_{m=1}^P \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j \left(\omega_m K_m(x_i^m, x_j^m) \right) + \sum_{i=1}^n \alpha_i \\ \text{sous contraintes} & \begin{cases} 0 \leq \alpha_i \leq C \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases} \quad \forall i \end{cases} \quad (\text{IV.14})$$

IV.4.5.2 Méthode de résolution

IV.4.5.2.1 Formulation de Lagrange : Initialement, toutes les composantes ω_m sont initialisées à $1/P$, puis la fonction objectif $J(\omega)$ sera calculée à l'aide d'un solveur SVM.

$$J(\omega) = -\frac{1}{2} \sum_{m=1}^P \sum_{i,j=1}^n \hat{\alpha}_i \hat{\alpha}_j y_i y_j \left(\omega_m K_m(x_i^m, x_j^m) \right) + \sum_{i=1}^n \hat{\alpha}_i \quad (\text{IV.15})$$

où $\hat{\alpha}$ sont les multiplicateurs de Lagrange que maximise IV.14, obtenu après résolution du problème dual SVM.

IV.4.5.2.2 Calcul de gradient : Après avoir calculé $J(\omega)$, nous devons résoudre l'équation IV.13. Pour commencer, nous devons calculer la dérivée de $J(\omega)$ en supposant qu'elle ne dépend pas de $\hat{\alpha}$.

$$\frac{\partial J(\omega)}{\partial \omega_m} = -\frac{1}{2} \sum_{i,j=1}^n \hat{\alpha}_i \hat{\alpha}_j y_i y_j \left(K_m(x_i^m, x_j^m) \right) \quad (\text{IV.16})$$

Ensuite, le gradient réduit de $J(\omega)$ est calculé par :

$$[\nabla_{red} J]_m = \begin{cases} \frac{\partial J(\omega)}{\partial \omega_m} - \frac{\partial J(\omega)}{\partial \omega_\nu}, & \text{si } m \neq \nu \\ \sum_{m \neq \nu} \left(-\frac{\partial J(\omega)}{\partial \omega_m} + \frac{\partial J(\omega)}{\partial \omega_\nu} \right) & \text{si } m = \nu \end{cases} \quad (\text{IV.17})$$

où ω_ν est la plus grande composante de ω à l'indice ν .

IV.4.5.2.3 Calcul de la direction de descente : La direction de descente utilisée pour mettre à jour les poids ω_m est donnée par la formule suivante :

$$D_m = \begin{cases} 0 & \text{si } \omega_m = 0 \text{ et } \left(-\frac{\partial J(\omega)}{\partial \omega_m} + \frac{\partial J(\omega)}{\partial \omega_\nu} \right) > 0 \\ -\frac{\partial J(\omega)}{\partial \omega_m} + \frac{\partial J(\omega)}{\partial \omega_\nu}, & \text{si } m \neq \nu \text{ et } \omega_m > 0 \\ \sum_{m \neq \nu} \left(\frac{\partial J(\omega)}{\partial \omega_m} - \frac{\partial J(\omega)}{\partial \omega_\nu} \right) & \text{si } m = \nu \text{ et } \omega_m > 0 \end{cases} \quad (\text{IV.18})$$

IV.4.5.2.4 Règle d'ajustement des poids : Une fois le gradient de $J(\omega)$ et la direction de la descente D sont calculés, nous commençons à mettre à jour les poids ω selon la formule $\omega \leftarrow \omega + \delta D$ où δ est le pas de descente.

Le pas maximal admissible δ_{max} est déterminé par la valeur en-tête de D . Nous recherchons donc l'indice de la composante minimale de l'argument calculée par :

$$v = \underset{\{m/D_m < 0\}}{\text{argmin}} \frac{-\omega_m}{D_m} \Rightarrow \delta_{max} = \frac{-\omega_v}{D_v} \quad (\text{IV.19})$$

IV.4.5.2.5 Minimisation de la fonction objectif : Après une première mise à jour, $J^*(\omega)$ est calculé à l'aide d'un solveur SVM comme suite :

$$J^*(\omega) = -\frac{1}{2} \sum_{m=1}^P \sum_{i,j=1}^n \hat{\alpha}_i \hat{\alpha}_j y_i y_j \left(\omega_m^* K_m(x_i^m, x_j^m) \right) + \sum_{i=1}^n \hat{\alpha}_i \quad (\text{IV.20})$$

où chaque $\omega_m^* = \omega_m + \delta_{max} D_m$, et on vérifie si la valeur objective diminue ou non $J^*(\omega) < J(\omega)$. À la première itération, nous fixons $J^*(\omega) = J(\omega)$. Si la valeur objective diminue, ω est ajusté, ce qui

implique donc ($\omega_v = 0$), et nous fixons $J(\omega) = J^*(\omega)$. Cette démarche est répétée jusqu'à ce que $J^*(\omega)$ cesse de diminuer.

IV.4.5.2.6 Condition d'optimalités : Le processus se termine lorsque le critère d'arrêt est satisfait. Ce critère d'arrêt peut être fonction de l'écart de dualité, des contraintes KKT, de la variation de ω entre deux étapes consécutives, ou d'un nombre maximal d'itérations. Dans la présente étude, on a choisi l'écart de dualité MKL.

$$\text{DyalGap} = C_1 - C_2 \leq \varepsilon \quad (\text{IV.21})$$

Où

$$C_1 = \max_m \sum_{i,j=1}^n \hat{\alpha}_i \hat{\alpha}_j y_i y_j \left(K_m(x_i^m, x_j^m) \right) \quad (\text{IV.22})$$

$$C_2 = \sum_{m=1}^P \sum_{i,j=1}^n \hat{\alpha}_i \hat{\alpha}_j y_i y_j \left(\omega_m^* K_m(x_i^m, x_j^m) \right) \quad (\text{IV.23})$$

Par conséquent, le processus s'arrête lorsque $\text{DyalGap} \leq \varepsilon$, où ε est un seuil de tolérance. Puis, nous cherchons toutes les pondérations $\omega_m = 0$, à savoir que leurs matrices Gram K_m associées sont éliminées. Ainsi, toutes les pondérations $\omega_m > 0$ strictement supérieurs à zéro correspondent aux variables retenues pour construire la combinaison optimale de noyaux K_m . Dans ce sens, ces caractéristiques seront choisies comme la combinaison la plus appropriée pour résoudre le problème de la MKL.

IV.4.6 Modèle idéal

Une fois le processus de sélection arrêté, on choisit les variables pertinentes à travers les valeurs des poids ω_m . Selon le choix des paramètres du noyau σ et celui du paramètre de régularisation C , divers sous-ensembles de caractéristiques peuvent être obtenus. Chaque sous-ensemble retenu est relié à un classificateur SVM afin d'évaluer leurs performances. Conformément au cadre actuel, nous sommes intéressés par un sous-ensemble qui combine un nombre minimum de noyaux. Bien entendu, il revient de sélectionner les combinaisons qui présentent un nombre minimum de variables. Le meilleur résultat a été obtenu en combinant 13 variables (figure IV.11).

TABLE IV.11: Détails du modèle optimal sur la base de validation et test.

	Métriques d'évaluation			
	Sensibilité	Spécificité	Justesse	échantillons
Validation	99.66	99.58	99.61	2 590
Test	99.57	99.97	99.77	128 063
Test FA			99.97	62 163
Test RSN			99.60	65 900

L'objectif principal de ce travail est de construire une méthode de diagnostic rapide et efficace permettant d'exploiter des segments de battements cardiaques courts pour un suivi médical automatique précis. Nous avons ciblé six dynamiques du signal du rythme cardiaque, chacune a été segmentée en une fenêtre de 30 secondes et transférée à travers 12 formes de méthodes linéaires, géométriques et non linéaires, ceci a abouti à un ensemble de 72 prédicteurs. Une approche MKL a été conçue pour supprimer les caractéristiques non informatives et non pertinentes.

Une fois le processus de sélection arrêté, on choisit les variables pertinentes à travers les valeurs des poids ω_m . Selon le choix des paramètres du noyau σ et celui du paramètre de régularisation C , divers sous-ensembles de caractéristiques peuvent être obtenus. Chaque sous-ensemble retenu est relié à un classificateur SVM afin d'évaluer leurs performances. Conformément au cadre actuel, nous sommes intéressés par un sous-ensemble qui combine un nombre minimum de noyaux. Bien entendu, il revient de sélectionner les combinaisons qui présentent un nombre minimum de variables. Le meilleur résultat a été obtenu en combinant 13 variables.

Ainsi, seules 13 variables ont conduit à la constitution d'un classificateur plus précis. De plus, nous souhaitons démontrer que MKL peut être appliqué avec succès à des tâches de sélection de variables pour les pipelines d'apprentissage machine. La performance rapportée a atteint 99,77% pour un modèle SVM utilisant le processus de sélection de variables MKL.

Ces résultats révèlent la réussite du concept de sélection de variables à l'aide de l'approche MKL. Deuxièmement, nous avons pu détecter les épisodes FA avec un taux de reconnaissance très important de 99,97% sur une base de données contenant 62163 segments, appliquant 13 caractéristiques. Ce qui permet d'atteindre un double objectif. Le tableau IV.12 résume les performances des mesures statistiques données sous la forme d'une matrice de confusion. Celle-ci indique le nombre d'épisodes correctement classifiés et le nombre d'épisodes mal classifiés.

TABLE IV.12: Matrice de confusion

		Valeur prédite		
		RSN	FA	
Valeur réelle	RSN	VN = 65 633	FP = 267	65 900
	FA	FN = 19	VP = 62 144	62 163
		65 652	62 411	

IV.5 Conclusion

Dans ce chapitre, nous avons présenté trois systèmes différents de diagnostic de l'arythmie. L'une des principales distinctions entre chaque système est l'entrée d'apprentissage construite. Nous avons introduit diverses formulations de la dynamique pour alimenter les algorithmes d'apprentissage automatique. Nous avons d'abord combiné trois formes de dynamiques pour détecter la présence de fibrillation auriculaire. Ensuite, nous avons ajouté deux dynamiques supplémentaires pour détecter à la fois la fibrillation auriculaire et trois formes d'arythmie courantes, qui sont le flutter auriculaire, la bradycardie sinusale et la tachycardie supraventriculaire. Dans les deux systèmes, nous avons tenté de construire un ensemble réduit de variables pertinentes en utilisant des stratégies factorielles et sélectives. L'objectif est de concevoir des systèmes de diagnostic rapides et fiables, néanmoins, nous avons toujours été confrontés au problème de dimensionnalité. En effet, le sous-ensemble de variables sélectionnées pour alimenter l'algorithme d'apprentissage n'est pas assez pertinent pour atteindre les performances souhaitées. Autrement dit, certaines variables pertinentes ne font pas partie du sous-ensemble final ce qui a pour conséquence la dégradation des performances de prédiction. De ce fait, nous avons été naturellement poussés à concevoir et développer des méthodes de sélection de variables pour atteindre les objectifs souhaités. Un système de diagnostic est présenté dans la troisième partie du chapitre, introduisant un nouvel algorithme de sélection de variables, basé sur l'apprentissage multinoyaux. Les résultats d'application obtenus pour ce système démontrent la pertinence du concept de sélection de variables utilisant l'apprentissage multinoyaux. De plus, nous avons pu détecter les épisodes de la fibrillation auriculaire avec un taux de reconnaissance très élevé de 99,97% sur une base de données contenant 62163 segments.

CONCLUSION GÉNÉRALE

Au cours de la dernière décennie, la santé numérique a connu une immense révolution avec les progrès technologiques et le développement d'outils de diagnostic intelligents et automatisés. À ce propos, l'équipe de recherche du laboratoire LIS a été investie dans la mise au point de nouvelles stratégies automatisées d'analyse, de détection et de diagnostic liées à la santé.

Dans cette thèse, nous présentons un système automatisé d'aide au diagnostic de la fibrillation auriculaire. Depuis toujours, de nombreux critères ont été imposés par les experts, comme l'efficacité du diagnostic tout en minimisant les fausses alarmes (minimisation des faux positifs), et la rapidité de calcul tout en minimisant les opérations de traitement. La conception d'un modèle de diagnostic d'arythmie fiable généralement nécessite la prise en compte de connaissances étendues mettant en évidence la présence d'une anomalie, exprimée sous forme de marqueurs ou d'indicateurs. Tandis que la combinaison de toutes les connaissances et caractéristiques essentielles nécessite un calcul numérique massif et compliqué. Pour répondre aux exigences des experts et pour éviter les modèles complexes, l'analyse des intervalles R-R a été privilégiée, vu que l'extraction des ondes R est beaucoup plus facile que l'extraction des autres composantes morphologique de l'ECG. Les méthodes de description des intervalles R-R sont très étendues, en fonction du nombre de variables extraites et du nombre de combinaisons de dynamiques utilisées.

Le premier travail de recherche proposé dans cette thèse consiste à extraire les séries chronologiques des ondes R. Un algorithme de détection des ondes R a été conçu à cette fin et se compose de deux parties principales. La première partie implique une détection simple basée sur des seuils d'amplitude et de temps fixes. Cependant, cette technique présente quelques limitations liées à l'altération morphologique des signaux ECG. Dans le cas où la valeur de l'amplitude du signal ECG passe en dessous du seuil, l'algorithme détecte plus de pics. Ceci nous amène à introduire un processus de correction dans la deuxième partie, visant à corriger les surdétectations induites par la première partie.

Considérant les travaux de recherche sur le diagnostic automatique de la FA, la majorité s'est principalement portée sur l'analyse des variables extraites des intervalles R-R. L'intervalle R-R, défini comme l'oscillation entre deux battements cardiaques successifs, est calculé comme l'intervalle de temps séparant deux ondes R consécutives du complexe QRS. Ici, nous proposons des paradigmes alternatifs approfondis, à savoir des formes de dérivées temporelles des intervalles R-R. Elles sont présentées par la dynamique de base des intervalles R-R et ses dérivées successives, normales et absolues. De telles dynamiques peuvent étendre le champ de représentation du comportement cardiaque en intégrant de nouveaux indicateurs vitaux et fournir un meilleur système de diagnostic.

En raison des investissements excessifs dans l'analyse manuelle de données ECG volumineuses, la mise en œuvre d'un système de l'IA automatisé pour le diagnostic de l'arythmie est désormais indispensable pour améliorer la rapidité et la qualité de diagnostic des enregistrements ECG à long terme. Dans le cadre de cette thèse, nous avons développé différents systèmes de diagnostic de l'arythmie, plus particulièrement de la fibrillation auriculaire.

Un système de diagnostic combinant trois dynamiques, chacun décrit par 11 variables descriptives, est construit en premier lieu. La structure du modèle est déterminée en évaluant les composantes de chaque sous-ensemble fournies par la technique d'analyse factorielle (AF) via des approches d'apprentissage machine, telle que SVM (support vector machine), WKNN (weighted K-nearest neighbors) et LR (logistic regression). Le modèle optimal atteint une $Sen = 97.14$, $Spe = 99.27$, $VPP = 97.57$, $TFP = 99.14$ et $ACC = 98.78$. Ensuite, nous présentons un système de diagnostic conçu pour distinguer les épisodes de NSR de quatre formes différentes d'arythmie. Ce système est composé de cinq dynamiques, chacune d'entre elles étant décrite à l'aide de 13 variables. L'approche proposée consiste principalement à évaluer l'ensemble initial de variables à l'aide d'une méthode d'analyse de la variance (ANOVA). Ensuite, les caractéristiques sélectionnées à partir de cette dernière sont soumises à des analyses multivariées avancées. Dans cette étude, nous avons développé un pipeline de méthodes de réduction de la dimensionnalité et d'algorithmes d'apprentissage automatique. Les méthodes de réduction de la dimensionnalité comprennent des méthodes non supervisées, telles que l'analyse en composantes principales (ACP) et l'analyse factorielle (AF), et une méthode supervisée, à savoir l'élimination des caractéristiques récursives (RFE). En outre, pour les algorithmes d'apprentissage, les méthodes SVM, ETC (extra trees classifier), et WKNN ont été utilisées pour évaluer la qualité des sous-ensembles induits. Le pipeline construit vise à sélectionner un modèle optimal composé d'une stratégie de réduction et d'un algorithme de classification, permettant d'obtenir les meilleures métriques de discrimination. Le modèle optimal atteint une $Sen = 98.42$, $Spe = 98.08$, $VPP = 98.08$, $TFP = 98.42$ et $ACC = 98.25$ en utilisant un classificateur ETC avec neuf variables sélectionnées par la méthode RFE. Enfin, la troisième partie est consacrée à la présentation du modèle de diagnostic ultime, construit à l'aide d'un nouvel algorithme de sélection de variables. Nous avons adopté le principe de l'apprentissage multinoyaux pour concevoir une stratégie de sélection de caractéristiques pertinente. L'algorithme conçu permettant d'extraire l'ensemble de variables consistantes adaptées au processus de diagnostic. Dans cette structure, nous avons ciblé six dynamiques, chacune d'entre elles étant décrite par 12 variables, aboutissant à un ensemble de 72 variables. Celles-ci sont ensuite étudiées à l'aide de l'algorithme MKL (multiple kernel learning) pour la sélection de variables. Le modèle obtenu atteint une $Sen = 99.57$, $Spe = 99.97$, $VPP = 99.97$, $TFP = 99.59$ et $ACC = 99.77$. L'objectif principal de ce travail est de démontrer que la méthode MKL peut s'appliquer à des tâches de sélection variable pour les pipelines d'apprentissage automatique. Tout d'abord, les résultats obtenus révèlent le succès de la stratégie MKL. Deuxièmement, nous avons pu détecter des épisodes de FA avec un taux de reconnaissance très élevé de 99,97% sur une base de données contenant 62163 segments.

Nous pensons que de nombreux travaux de recherche sont encore nécessaires en matière de discrimination et de classification des irrégularités cardiaques :

1. L'optimisation de l'ensemble du processus de prétraitement (filtrage et détection des ondes R) en réduisant le temps de calcul et en établissant des techniques de filtrage appropriées et adaptées aux différents signaux fournis par le projet. L'idée est de construire un algorithme de traitement adapté à différents types de signaux ECG, y compris une adaptation dynamique du seuil pour sélectionner les composantes morphologiques du tracé ECG. Nous tentons d'étudier des outils de seuillage dynamique, ainsi que des algorithmes de filtrage adaptatif.
2. La mise en place d'un système doté de plusieurs singularités électrocardiographiques. Au lieu de se concentrer uniquement sur la détection des ondes R, nous cherchons à développer des outils de diagnostic utilisant toutes les caractéristiques essentielles nécessaires pour diagnostiquer l'arythmie.
3. Etudes multidiagnostics pour diverses pathologies cardiaques. Jusqu'à présent, nous nous sommes principalement concentrés sur la détection de la fibrillation auriculaire, car c'est la forme la plus fréquente d'arythmie. Toutefois, il existe d'autres arythmies pour lesquelles le diagnostic revêt une importance considérable. C'est pourquoi nous poursuivons nos recherches sur le diagnostic de l'arythmie en vue d'un système qui puisse identifier l'existence de l'arythmie et en préciser la forme exacte.
4. Pronostic de la fibrillation auriculaire.

ÉTUDE ET PRONOSTIC DE LA DYNAMIQUE DE PROPAGATION DU SRAS-COV-2

IV.6	Introduction	116
IV.7	Contexte et objectifs	117
IV.8	Modélisation de la propagation de l'épidémie de COVID-19	118
IV.8.1	Intérêt des modèles mathématiques en épidémiologie	118
IV.8.2	Modélisation mathématique des maladies infectieuses	119
IV.8.2.1	Modèle SEIR généralisé	119
IV.8.2.2	Modèle proposé	119
IV.8.3	Données COVID-19	122
IV.8.4	Implantation du modèle	122
IV.8.4.1	Phase de mise en route	122
IV.8.4.2	Phase d'apprentissage	123
IV.9	Résultats de prédiction	124
IV.9.1	Les effets de la relaxation du confinement	127
IV.10	Conclusion	129

IV.6 Introduction

Les agents pathogènes émergents et réémergent sont des défis mondiaux pour la santé publique [131]. En janvier 2020, un nouveau coronavirus (SARS-CoV-2) a été identifié [132] comme l'agent pathogène qui a infecté plusieurs groupes de patients avec une pneumonie de cause inconnue dans la région de Wuhan, province de Hubei, en Chine. La progression massive de cette nouvelle maladie infectieuse à travers le monde conduit à la nouvelle pandémie de coronavirus (COVID-19) [133, 10]. Plus que la moitié de la population mondiale a été confinée dans des conditions restrictives de distanciation sociale afin de réduire les risques d'infections et de décès. Plus de 90 pays ont connu jusqu'à présent des formes de confinement social extrêmement sévère et dramatique. En France, la durée du confinement a été de huit semaines, du 17 mars 2020 au 11 mai 2020. Cette mesure restrictive s'efforce de limiter la propagation du COVID-19 en réduisant les interactions sociales. Avant la France, l'Asie et certains pays européens tels que l'Italie et l'Espagne ont appliqué également des mesures restrictives pour réduire le nombre de nouveaux cas et les charges hospitalières et de soins intensifs correspondants. Le taux de mortalité intrinsèque du SRAS-CoV-2 semble faible ($\sim 1\%$) [134], mais la dynamique de propagation conduit à une sollicitation excessive et une saturation progressive des ressources hospitalières.

L'incapacité à prendre en charge le flux massif de patients atteints de COVID-19 grave concomitant contribue à un taux de mortalité supérieur à celui observé [135]. Pour réduire ces conséquences dramatiques, des mesures contraignantes sont mises en place avant (pour retarder le pic épidémique) ou après (pour éviter un second pic épidémique) le confinement associant restrictions de mobilité, interdiction de rassemblement, suspension des activités scolaires et professionnelles, et quarantaine curative ou préventive.

Pour soutenir les autorités sanitaires dans la gestion de toutes ces mesures coercitives, certes très dommageables économiquement, mais nécessaires, nous sommes impliqués dans la mise au point d'un modèle de prédiction, piloté par les données observées de COVID-19, pour estimer le flux de cas susceptibles d'être hospitalisés dans les établissements nationaux et locaux accueillant des infections à COVID-19. Ceci est destiné à évaluer la propagation de l'infection dans le temps, et à alerter les autorités sanitaires sur les risques de débordements inattendus.

Dans ce chapitre, nous présentons une méthodologie améliorée pour étudier la dynamique de transmission de COVID-19, en considérant deux compartiments essentiels déduits des cas confirmés de COVID-19, à savoir le compartiment des cas hospitalisés et le compartiment des cas admis aux unités de soins intensifs. Nous avons conçu un modèle mathématique composé de neuf équations différentielles, chacune décrivant la variabilité dynamique d'un compartiment donné, y compris la dynamique des cas insusceptibles (P), susceptibles (S), exposés (E), infectés (I), confirmés (C), hospitalisés (H), en soins intensifs (R - Réanimation), guéris (G) et décédés (D). La répartition des flux entre les compartiments est calculée à l'aide d'un algorithme de moindre carré non linéaire. Pour construire le modèle, nous avons utilisé les statistiques observées du premier avril au 30 juin

pour l'apprentissage des hyperparamètres et les données du premier au 21 juillet pour évaluer les performances de l'estimation. En outre, les estimations de tendances statistiques des compartiments C , H , R , G et D ont été étendues jusqu'au 31 août.

Dans cette recherche, deux expériences sont réalisées : la première utilise tous les compartiments du modèle (c'est-à-dire les neuf équations différentielles). En outre, nous avons mené une deuxième analyse particulière, dans laquelle nous avons supprimé la dynamique P des cas insusceptibles. Ceci a pour objectif d'évaluer le comportement de propagation dans le cas où aucune mesure restrictive n'est adoptée pour empêcher la propagation du virus (contexte de déconfinement). Le comportement de propagation de l'épidémie de COVID-19 après le confinement a été également analysé dans ce chapitre sous différents scénarios.

IV.7 Contexte et objectifs

La propagation des maladies infectieuses au sein de la population est un phénomène dynamique caractérisé par un nombre variable d'individus sains et infectés. Cette variation est fonction de la fréquence de contacts établis entre les individus infectés et les individus sains non immunisés qui deviennent infectés en retour. La dynamique de la propagation de l'infection peut être étudiée à l'aide de modèles compartimentaux mathématiques en épidémiologie. Ces modèles sont appliqués pour analyser la dynamique de propagation de l'infection en fonction du nombre de cas infectés, des cas non affectés (sensibles) et des cas soustraits du cercle de transmission (soit guéris ou décès).

Au cours des dernières décennies, de nombreux modèles mathématiques ont été proposés pour modéliser la propagation de pandémies infectieuses. L'objectif principal de ces modèles est de décrire les caractéristiques de transmission des agents pathogènes entre les compartiments sensibles, infectés et guéris. Il consiste également à comprendre le processus de transmission du virus en se basant sur un taux de mutation défini et adapter à la transition entre ces compartiments [136]. Les taux de mutation sont des hyperparamètres qui doivent être bien définis, appelés aussi taux de mutation inter-compartiments [137].

Les méthodes proposées dans ce contexte ont été largement axées sur la prévision des tendances COVID-19, que ce soit au niveau des pays ou au niveau mondial. Peipei Wang et al. [138] ont présenté une méthode de prédiction de la tendance épidémique de COVID-19 basée sur les modèles Logistic et Prophet. Ils ont mené des expériences sur la pandémie mondiale et sur certains pays afin d'anticiper le pic de la vague épidémique. Ils se sont basés sur les données épidémiologiques recueillies entre le 22 janvier 2020 et le 16 juin 2020. Xie Yingkang et al. [139] ont utilisé un modèle SIS pour étudier la stabilité de l'épidémie tenant compte des taux d'infection de plusieurs périmètres interférant les uns avec les autres à travers des réseaux complexes. Chaeyoung Lee et al. [140] ont proposé un modèle de type SUC (Susceptible-Unidentified infected-Confirmed) pour estimer le taux de population infectée et non identifiée en Chine. Di Domenico et al. [141] se sont intéressés aux évolutions des

taux d'hospitalisation et de soins intensifs, en utilisant un modèle stochastique discret à structure d'âge basé sur la matrice des contacts sociaux.

Les personnes vulnérables infectées sollicitent fortement les ressources hospitalières en général et celles des unités de soins intensifs en particulier. Ces ressources sont indispensables pour contrôler les infections prises en charge dans les établissements sanitaires. Considérant les travaux actuels, ce qui semble manquer de préoccupation est la prédiction correcte des cas susceptibles de nécessiter des équipements thérapeutiques, tels que des lits et des ventilateurs ambulatoires. Comme il s'agit de ressources limitées, la projection future des cas prévus nécessitant une hospitalisation et occupant au moins une place est évidemment d'une importance cruciale pour anticiper les risques de dépassement de la capacité des ressources hospitalières disponibles, connus sous le nom de saturation.

Le confinement de la population a été adopté pour réduire les risques de propagation du virus en restreignant la liberté de déplacement. Le confinement sanitaire permet notamment de limiter les contacts entre les personnes pour éviter la propagation de l'infection. De plus, le confinement est un moyen aussi pour réduire les formes graves d'infection et préserver le bon fonctionnement des services de santé. En termes épidémiologiques, les efforts sont requis pour limiter le taux de reproduction du virus, noté R_0 . Le paramètre R_0 quantifie le nombre moyen d'individus infectés à travers un agent contagieux. Le taux de reproduction doit être réduit à moins de 1 pour empêcher l'épidémie de se propager.

IV.8 Modélisation de la propagation de l'épidémie de COVID-19

IV.8.1 Intérêt des modèles mathématiques en épidémiologie

Les modèles compartimentaux sont principalement appliqués pour la modélisation de la dynamique des maladies infectieuses. Il s'agit des approches mathématiques permettent de décrire le mécanisme de l'épidémie émergent. La formulation mathématique du modèle compartimental est basée sur deux concepts fondamentaux : les compartiments et les règles. Selon l'épidémie concernée, les compartiments classent la population en différents états de santé en fonction du statut infectieux de chaque individu. Tandis que, les règles (qui décrivent les taux de mutation inter-compartiments) déterminent les proportions d'individus qui transitent d'un compartiment à l'autre.

Les dynamiques de l'hospitalisation (H) et des soins intensifs (R) décrivent la situation épidémiologique instantanée t pour une population donnée. Ces deux dynamiques sont caractérisées par des fluctuations importantes et des brusques augmentations et diminutions, ce qui complique leur pronostic. Pour répondre à ce défi, nous avons construit deux équations différentielles adaptées à la dynamique de chaque compartiment, à savoir H et R (A.16, A.17). Ces équations différentielles comprennent trois facteurs temporels, qui sont donnés par $\omega(t)$, $\rho(t)$ et $\rho_d(t)$. Le terme $\rho_d(t)$ est utilisé pour estimer le taux de décharge des admissions en USI et pour préserver le statut exact du compartiment R .

Les épidémies atteignent leur pic lorsque le nombre de places d'hospitalisation et de soins intensifs est épuisé. Une fois le sommet atteint, le nombre d'infections graves commence à diminuer. En pratique, tant que le taux d'admission dans l'USI est inférieur aux seuils de capacité, les patients peuvent être accueillis - c'est ce qu'on appelle la phase ascendante. En revanche, si le nombre de cas dans l'USI dépasse le seuil de capacité, la situation devient plus pressante. De ce fait, la modélisation des maladies infectieuses a toujours suscité une attention soutenue des autorités sanitaires, vu qu'elle permet de comprendre le comportement dynamique de virus, sa précipitation de propagation et son impact sur les structures sanitaires.

IV.8.2 Modélisation mathématique des maladies infectieuses

IV.8.2.1 Modèle SEIR généralisé

L'approche développée dans ce chapitre adopte un concept similaire à celui de la modélisation SEIR généralisée [142]. Le modèle SEIR généralisé proposé par L. Peng et al. [142] divise la population en sept compartiments P (insusceptibles), S (susceptibles), E (exposés), I (infectés), Q (quarantaine), R (guéris) et D (décédés). Le compartiment P concerne les individus qui sont protégés contre la contamination, immunisés ou isolés. Cependant, le compartiment S comprend les individus non immunisés qui n'ont pas encore atteint l'infection mais qui sont potentiellement ciblés. Le compartiment E désigne les individus susceptibles avoir croisé un individu infecté par le virus. Un individu exposé est considéré comme étant infecté mais pas encore infectieux - en période de latence. Une période de latence se produit entre l'infection et la contagiosité, c'est-à-dire entre le compartiment des individus exposés et celui des individus infectés. Il s'agit d'une période de latence asymptomatique suivie d'une période d'incubation. Ainsi, le compartiment I désigne les individus symptomatiques ayant une capacité de transmission mais qui ne sont pas encore isolés ("quarantaine"). Une fois qu'un individu est infecté et que sa contamination est confirmée, il passe dans le compartiment Q . L'individu reste infectieux mais isolé par la politique de quarantaine.

Du compartiment Q , deux scénarios sont décrits via le modèle SEIR généralisé. Premièrement, l'individu décède, dans ce cas il est classé dans le compartiment D . Deuxièmement, l'infection s'arrête d'elle-même et procure une immunité contre la réinfection, dans ce cas, l'individu est affecté au compartiment G . Les flux inter-compartimentaux sont pondérés comme suit : taux de protection α , taux d'infection β , période d'incubation γ (temps de latence moyen), temps de quarantaine moyen δ , taux de guérison $\lambda(t)$ et taux de mortalité $\kappa(t)$. Les taux de guérison $\lambda(t)$ et de mortalité $\kappa(t)$ sont tous deux des fonctions dépendantes du temps.

IV.8.2.2 Modèle proposé

La plupart des modèles épidémiologiques connus se limitent à des compartiments standard : les individus susceptibles, exposés, infectés, guéris et décédés [143, 142, 144–148]. Cependant, dans le contexte de la pandémie actuelle de COVID-19, les compartiments les plus critiques et essentiels

qui devraient être abordés sont les patients hospitalisés et ceux des unités de soins intensifs. Pour cibler et mettre en évidence ces deux compartiments, nous avons proposé un modèle épidémiologique modifié, inspiré du modèle généralisé SEIR [142]. Le diagramme synoptique des compartiments est présenté dans la figure IV.5. Les flèches noires indiquent la structure arborescente entre les équations différentielles des compartiments. Cependant, les flèches rouges indiquent les mutations des individus entre les compartiments après confirmation de la COVID-19.

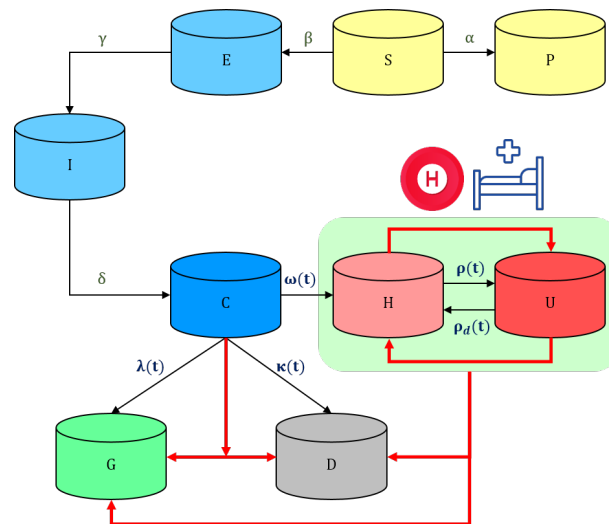


FIGURE IV.5: Le diagramme synoptique du modèle proposé.

IV.8.2.2.1 Formulation mathématique Pour généraliser les fonctionnalités du pronostic, nous incluons deux compartiments épidémiques supplémentaires : les cas hospitalisés et les cas admis en unités de soins intensifs. En outre, nous remplaçons le compartiment de quarantaine Q dans le modèle SEIR généralisé par le nombre cumulé de cas confirmés, ceci permet de mieux comprendre l'ampleur de l'épidémie au fil du temps. Trois nouveaux compartiments sont donc introduits sous la forme C (confirmé), indiquant le nombre cumulé de cas d'infection confirmés. H (hospitalisé) fait référence au nombre de patients hospitalisés après une infection. Enfin, le compartiment R (soins intensifs) désigne le nombre de patients admis dans les unités de soins intensifs.

La description de chaque compartiment proposé est caractérisée par une équation différentielle, comme résumé dans A.2. De même, les flux inter-compartimentaux sont évalués selon le taux de protection α , le taux d'infection β , la période d'incubation γ , le temps de quarantaine moyen δ , le taux de guérison $\lambda(t)$ et le taux de mortalité $\kappa(t)$. En outre, les paramètres supplémentaires ajoutés sont : le taux d'admission hospitalière $\omega(t)$, le taux d'admission aux unités de soins intensifs $\rho(t)$, le taux de décharge moyen des admissions aux unités de soins intensifs $\rho_d(t)$, et le taux de mortalité en dehors des hôpitaux $\kappa_d(t)$.

IV.8.2.2.2 Interprétation des équations La modélisation suppose une décroissance stable de la population susceptibles, ce qui introduit un taux de protection positif α pour refléter les mutations intervenues sur le plan sanitaire. Le taux de protection α dépend principalement des mesures préventives mises en place pour réduire les risques d'infection, telles que la sensibilisation au port de masques, les gestes barrière, la traçabilité des contacts croisé des cas infectés, etc. Le taux de guérison $\lambda(t)$ augmente progressivement au cours du temps, tandis que le taux de mortalité $\kappa(t)$ diminue progressivement jusqu'à ce que le virus atteigne la phase de stabilisation puis d'éradication. Chacune des fonctions $\lambda(t)$ et $\kappa(t)$ doit être adéquate et adaptée aux dynamiques du contexte épidémiologique. En effet, E. Cheynet [143] a suggéré de nombreuses formules pour les fonctions $\lambda(t)$ et $\kappa(t)$.

La dérivée $\frac{d}{dt}$ peut être utilisée pour mesurer la variation (soit croissante ou décroissante) des compartiments $P(t)$, $S(t)$, $E(t)$, $I(t)$, $C(t)$, $H(t)$, $U(t)$, $R(t)$ et $D(t)$ dans le temps t . Par exemple, $dS(t)/dt$ définit le flux des individus entre deux instantes t_i et t_{i+1} dans $S(t)$. Une variation négative indique que des individus sortent du compartiment. Tandis qu'une variation positive signifie que des individus arrivent dans le compartiment. La somme de tous les compartiments représente le nombre total de la population $N = P(t) + S(t) + E(t) + I(t) + Q(t)$.

Pour un instant t donné, le terme $\alpha S(t)$ représente le nombre de personnes protégées contre l'infection. Ce terme est donc introduit dans le compartiment P (A.9) et retiré du compartiment S (A.10). Le terme $(S(t) \times I(t))/N$ désigne le nombre de contacts effectués entre des individus sains et des individus infectés. $(\beta/N) \times (S(t) \times I(t))$ indique le nombre de cas nouvellement infectés. Le signe "-" dans le compartiment $S(t)$ signifie une diminution progressive de ce compartiment dans le temps. Les cas nouvellement infectés sont donc retirés de S et introduits dans E (A.11). En outre, les cas exposés seront introduits au compartiment I par $\gamma E(t)$ (A.12). Puis, on additionne $\delta I(t)$ au compartiment Q (A.13) et on le retire de I (A.12).

De cas de quarantaine, trois compartiments d'individus sont considérés : Le premier concerne les cas exclus du cycle de transmission du virus (c'est-à-dire les individus guéris G (A.14) ou décédés D (A.15)). Le deuxième concerne les personnes confinées à domicile conformément à une politique de contrôle domestique. Le troisième est constitué de personnes hospitalisées. Les personnes hospitalisées se répartissent en deux parties : Unité hospitalière conventionnelle, mentionnée comme H (A.16) et Unité de soins intensifs, mentionnée comme R (A.17). Un individu peut commuter entre H et R plusieurs fois dans les deux sens, avant qu'il ne soit affecté au compartiment R ou D .

IV.8.2.2.3 Motivations Le compartiment d'hospitalisation est déterminé en fonction de la fréquence quotidienne des admissions et décharges. Pour un temps donné t (en pratique jour d), le nombre de personnes signalées comme étant hospitalisées est composé à la fois des nouvelles admissions enregistrées au moment t et des personnes demeurant hospitalisées depuis. Pour modéliser la variabilité dynamique de ces admissions, nous avons introduit un coefficient $\omega(t)$ adapté à cet effet. Le terme $\omega(t)C(t)$ désigne la proportion des cas hospitalisés calculée sur le total des personnes confirmées à la date t . L'état du compartiment H est défini par la soustraction de trois termes : (1)

Les cas signalés de guérison après hospitalisation $\lambda(t)H(t)$; (2) Les cas signalés de décès après hospitalisation $\kappa(t)H(t)$; (3) Les cas de décès en dehors de l'hôpital $\kappa_d(t)(C(t) - H(t))$.

Le terme $C(t) - H(t)$ représente les cas infectés, placés en quarantaine et ne nécessitant pas une hospitalisation. Le taux de mortalité mesuré sur les patients hospitalisés s'exprime par κ_0 et κ_1 selon la formule dépendant du temps $\kappa(t) = \kappa_0 \times \exp(-\kappa_1 t)$. Les décès COVID-19 non comptabilisés survenus dans les Ehpad ou à domicile sont intégrés via une fonction de taux de mortalité réduite, à savoir $\kappa_d(t)$, similaire à $\kappa(t)$. La fonction $\kappa_d(t)$ est exprimée à l'aide des paramètres ν_a , et ν_b , selon la formule dépendant du temps $\kappa_d(t) = \nu_a \times \exp(-\nu_b(t + \tau_D))$. Le compartiment R est défini en fonction des cas provenant du compartiment H à chaque instant t , ainsi que par les patients sortis après une période donnée. Le terme $\rho_0 \times \exp(-\frac{1}{2}(\rho_1 t))$ est utilisé pour estimer la proportion des cas traités en USI par rapport à la taille effective des patients hospitalisés au moment t . Toutefois, le terme $\frac{\rho_0}{\eta_a} \times \exp(-\frac{1}{2} \frac{\rho_1}{\eta_b}(t + \tau_U))$ est appliqué pour adapter l'état du compartiment R en supprimant la proportion approximative de décharge.

IV.8.3 Données COVID-19

Dans la présente étude, nous avons extrait cinq ensembles de données historiques provenant de [149] et [150], à savoir, les données C , R , D , H et R . Les données du premier avril au 30 juin ont été utilisées pour l'apprentissage des hyperparamètres et les données du premier au 21 juillet pour évaluer les performances de l'estimation.

IV.8.4 Implantation du modèle

IV.8.4.1 Phase de mise en route

La propagation d'un agent infectieux dans la population est un phénomène dynamique : le nombre d'individus sains et infectés varie dans le temps, en fonction des contacts établis entre les individus infectés et non immunisés. Cette dynamique peut être adressée par la modélisation d'équations différentielles et la détermination du processus de propagation de l'épidémie par la résolution numérique de ces équations. Une première étape dans le processus de modélisation consiste à collecter des données et à définir les paramètres initiaux pour alimenter le modèle. La conduite des compartiments $P(t)$, $S(t)$, $E(t)$, $I(t)$, $C(t)$, $H(t)$, $U(t)$, $R(t)$ et $D(t)$ peut être obtenue par la résolution numérique des équations différentielles associées. Pour procéder au calcul, il est nécessaire de configurer certains paramètres : les valeurs des états initiaux, c'est-à-dire les données de P , S , E , I , C , H , R , R et D à l'instant $t = 0$, c'est-à-dire, P_0 , S_0 , E_0 , I_0 , C_0 , H_0 , U_0 , R_0 , et D_0 .

Les états initiaux de C_0 , H_0 , U_0 , R_0 , et D_0 sont définis en fonction des cas déclarés dans chaque compartiment respectif à $t = 0$. Pour les autres compartiments, nous définissons $I_0 = q_1 \times C_0$, $E_0 = q_2 \times C_0$, $P_0 = 0$, et S_0 comme : $S_0 = N - P_0 - E_0 - I_0 - C_0$, où N est le nombre total de la population. Le terme t_0 ou $t = 0$ définit la première date que nous avons fixée pour lancer la construction du modèle, dans ce cas, t_0 correspond aux observations du premier avril. Les termes q_1

et q_2 sont utilisés pour initier le nombre de cas exposés et infectés à t_0 . Généralement, q_1 est égal à un, et q_2 est égal à zéro. Pour une estimation fiable de la tendance épidémiologique de chaque compartiment, tous les hyperparamètres du système doivent être adaptés correctement. Dans le cadre actuel, nous avons adopté 12 hyperparamètres.

IV.8.4.2 Phase d'apprentissage

Après l'initialisation des compartiments, l'étape suivante consiste à déterminer les valeurs optimales des hyperparamètres, minimisant l'erreur quadratique entre les données estimées et les données de distribution souhaitées. Pour cela, nous avons implémenté un algorithme de moindre carré non linéaire (gradient-descent et Levenberg-Marquardt) pour adapter notre modèle aux données observées. Généralement, la construction d'un tel modèle implique plusieurs hyperparamètres, dont chaque observation peut être exprimée en fonction de ces hyperparamètres. L'objectif de l'approche de moindre carrés non linéaire est d'estimer les meilleures solutions pour ces hyperparamètres en fonction de l'ensemble des données observées. Le critère d'erreur utilisé pour optimiser les hyperparamètres du modèle est donné par :

$$Err = \frac{|X_o - X_e|}{X_o} \quad (IV.24)$$

Pour évaluer les performances du modèle, nous avons utilisé l'erreur absolue moyenne en pourcentage (EAMP), également appelée moyenne absolue des écarts en pourcentage (MAPD). Il s'agit d'une mesure permettant d'estimer la précision d'une méthode de pronostic basée sur des données statistiques. Le MAPD est appliqué pour évaluer l'estimation des tendances, servant également comme fonction objective pour les problèmes de régression dans l'apprentissage automatique. Il exprime généralement la précision sous la forme d'un rapport défini par la formule :

$$MAPD = \varepsilon_X = \frac{100}{N_X} \times \sum_{i=1}^{N_X} \left| \frac{X_o - X_e}{X_e} \right|, \quad (IV.25)$$

où X_o est la série de données observées, X_e les données estimées et N_X la durée de la série.

— Calcule de la bande de confiance

La bande de confiance a été calculé à l'aide des formules ci-dessous. L'exemple est donné pour le compartiment H , puis le même principe a été appliqué pour tous les autres bandes.

$$H_{max} = \hat{H} + \sigma(\%) \times \omega(t) \times \hat{H} \quad (IV.26)$$

$$H_{min} = \hat{H} - \sigma(\%) \times \omega(t) \times \hat{H} \quad (IV.27)$$

où \hat{H} représente les estimations, et $\sigma(\%)$ contrôle la déviation appliquée aux estimations.

Ces bandes de confiance sont conçues pour couvrir les incidents ou les évolutions imprévues survenant au fil du temps. Il s'agit des incidents qui se produisent au cours de la phase ascendante de

l'épidémie et qui peuvent dévier les tendances des compartiments. Dans ce cas, il est nécessaire de mettre à jour le modèle pour intégrer un éventuel changement de dynamique de l'épidémie.

IV.9 Résultats de prédiction

Dans cette analyse, deux expériences sont effectuées : la première fait appel à tous les compartiments du modèle (les neuf équations différentielles). Or, dans la seconde, nous avons retiré la dynamique P des cas insusceptibles. Autrement dit, nous avons supprimé le taux de protection pour évaluer les tendances de transmission du virus dans un cadre où aucune mesure restrictive n'est adoptée pour empêcher la propagation du virus (contexte de déconfinement). Ainsi, les estimations de tendances statistiques des compartiments C , H , R , G et D ont été étendues jusqu'au 31 août.

TABLE IV.13: 1er expérience : Pourcentage d'erreurs d'estimation atteint en phase d'apprentissage et test en utilisant les données COVID-19 pour la France.

L'erreur absolue moyenne en pourcentage (EAMP)					
	ε_C	ε_H	ε_R	ε_G	ε_D
Apprentissage	0.9382	1.1997	1.8097	1.1211	1.0070
Test	1.0703	1.2368	2.1532	1.3715	1.0672

TABLE IV.14: 1er expérience : Pourcentage d'erreurs d'estimation atteint en phase d'apprentissage et test en utilisant les données COVID-19 pour la région PACA.

L'erreur absolue moyenne en pourcentage (EAMP)					
	ε_C	ε_H	ε_R	ε_G	ε_D
Apprentissage	0.5019	2.3349	5.3540	1.1083	2.1034
Test	0.4914	2.9725	7.9978	1.3372	2.0582

TABLE IV.15: 2e expérience : Pourcentage d'erreurs d'estimation atteint en phase d'apprentissage et test en utilisant les données COVID-19 pour la France.

L'erreur absolue moyenne en pourcentage (EAMP)					
	ε_C	ε_H	ε_R	ε_G	ε_D
Apprentissage	1.0461	1.2127	1.8088	1.1452	1.0538
Test	0.9047	1.2100	2.1353	1.1672	1.0188

TABLE IV.16: 2e expérience : Pourcentage d'erreurs d'estimation atteint en phase d'apprentissage et test en utilisant les données COVID-19 pour la région PACA.

L'erreur absolue moyenne en pourcentage (EAMP)					
	ε_C	ε_H	ε_R	ε_G	ε_D
Apprentissage	0.4959	2.4260	5.3300	1.3117	2.0728
Test	0.4854	2.9240	8.0887	1.3033	1.8873

— Écart absolue entre les valeurs prédites et observées

Dans le cadre actuel, l'attention a été principalement consacrée à l'amélioration des capacités de prédiction des cas susceptibles d'être admis dans les services hospitaliers en général, et dans les unités de soins intensifs en particulier. Compte tenu des courbes obtenues pour les compartiments H et R , les résultats sont très intéressants. Dans les figures IV.6, IV.7, IV.8, et IV.9, nous avons illustré les écarts entre les données H et R observées et prédites dans le contexte national et régional.

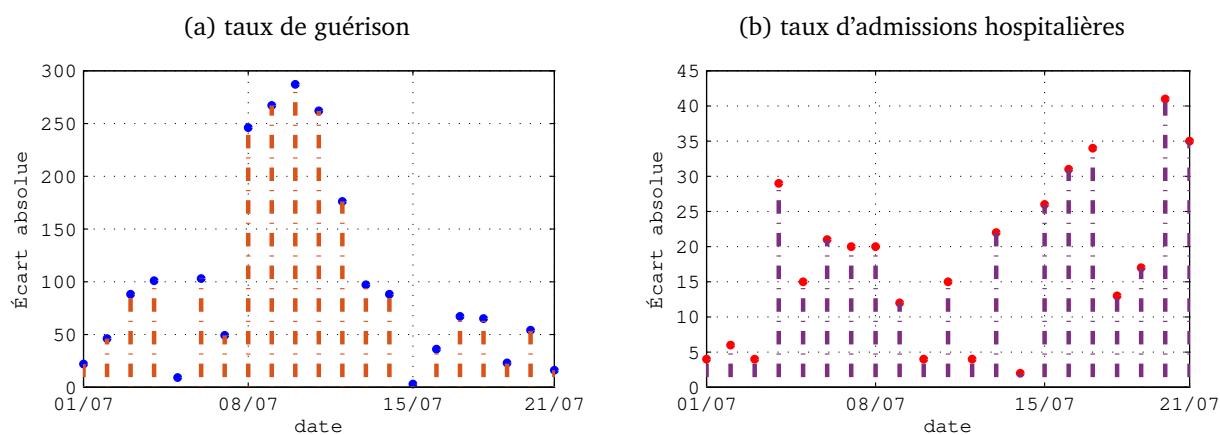


FIGURE IV.6: 1er cas : Écart absolu entre les estimations et les valeurs observées du modèle national.

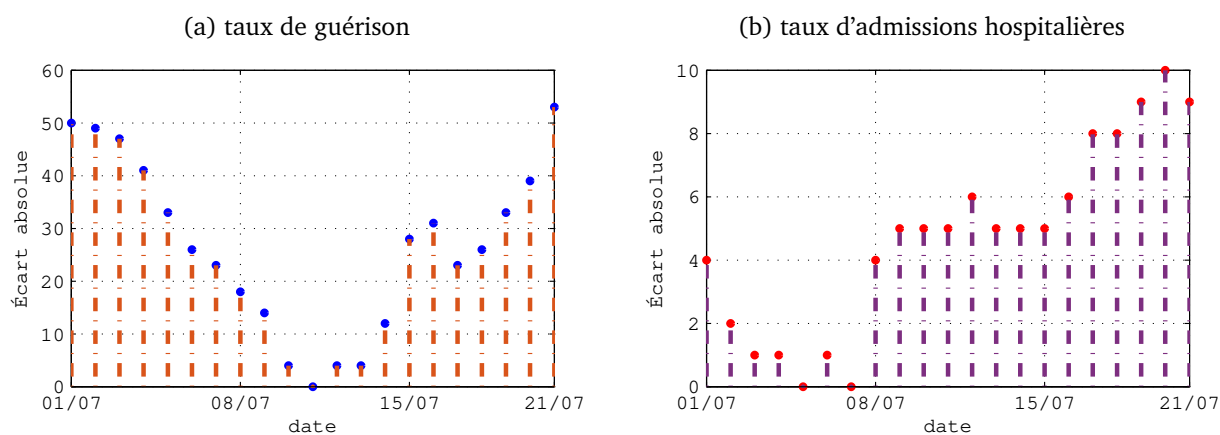


FIGURE IV.7: 1er cas : Écart absolu entre les estimations et les valeurs observées du modèle régional.

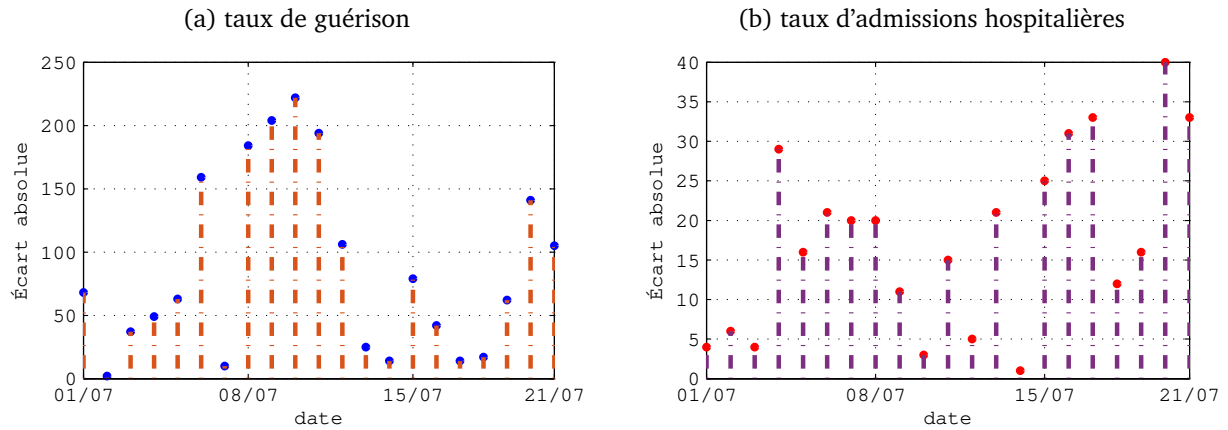


FIGURE IV.8: 2e cas : Écart absolu entre les estimations et les valeurs observées du modèle national.

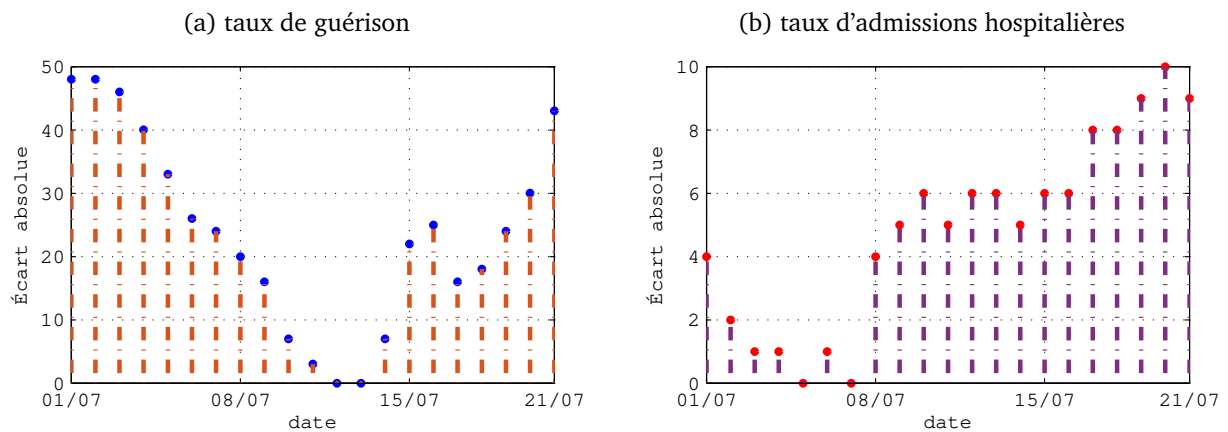


FIGURE IV.9: 2e cas : Écart absolu entre les estimations et les valeurs observées du modèle régional.

Selon les diagrammes illustrés ci-dessus, les écarts identifiés pour les cas d'hospitalisation en utilisant le modèle national, est limité à 5 estimations supérieures à un écart de 150, et le reste des estimations au-dessus du 150. Pour les cas de soins intensifs, la plupart des estimations sont inférieures à l'écart de 30. Les résultats obtenus pour le modèle national démontrent sa capacité effective à prédire les charges futures au sein des compartiments H et R . Le modèle régional a également démontré son efficacité pour prédire la probabilité d'admission dans les compartiments H et R avec un écart de moins de 50 pour les patients hospitalisés et de moins de 10 pour les soins intensifs.

IV.9.1 Les effets de la relaxation du confinement

Le comportement de propagation de l'épidémie de COVID-19 après le confinement a été également analysé sous différents scénarios. Tout d'abord, la période de projection entre le 11 mai et le 31 août est répartie en huit intervalles distincts. Pour évaluer les effets du relâchement du confinement, nous utilisons un taux de protection nul et un taux d'infection variant. Dans le contexte actuel, nous avons abordé deux cas d'étude : La première décrit le mécanisme de propagation de COVID-19 en fonction du nombre de cas signalés le 11 mai dans chaque compartiment, tout en déplaçant trois quarts du total des cas de compartiment P vers le compartiment S (où le 11 mai est le 40ème jour de la date estimée t_{40}). Donc les valeurs initiales du modèle simulé sont données comme :

$$P_{t_0}^{dec} = \frac{1}{4}P_{t_{40}} \quad (IV.28)$$

$$S_{t_0}^{dec} = S_{t_{40}} + \frac{3}{4}P_{t_{40}} \quad (IV.29)$$

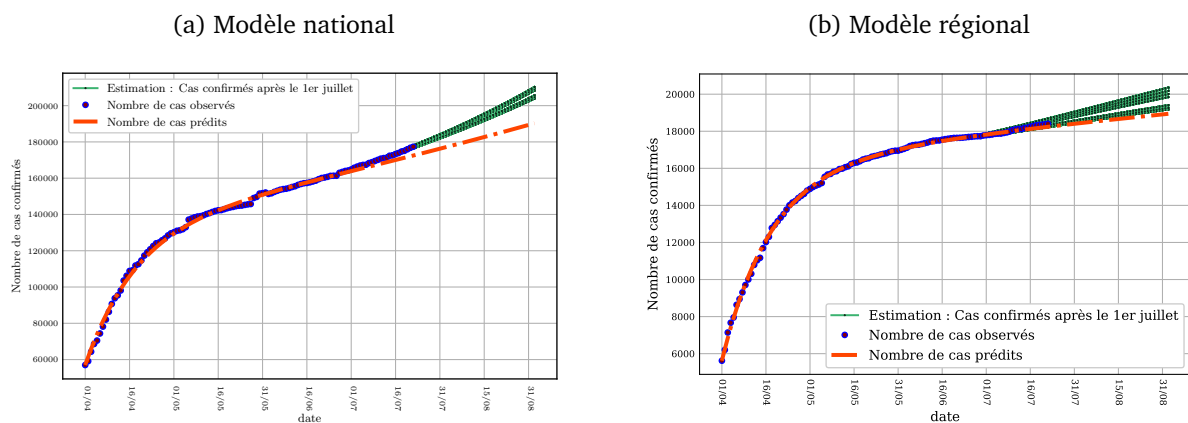


FIGURE IV.10: 1er cas : Impact de la relaxation du confinement sous différents scénarios possibles pour le modèle national et régional.

La seconde étude décrit le mécanisme de propagation de COVID-19 en affectant deux tiers des cas de compartiment P au compartiment S . Donc les valeurs initiales du modèle simulé sont données comme :

$$P_{t_0}^{dec} = \frac{1}{3}P_{t_{40}} \quad (IV.30)$$

$$S_{t_0}^{dec} = S_{t_{40}} + \frac{2}{3}P_{t_{40}} \quad (IV.31)$$

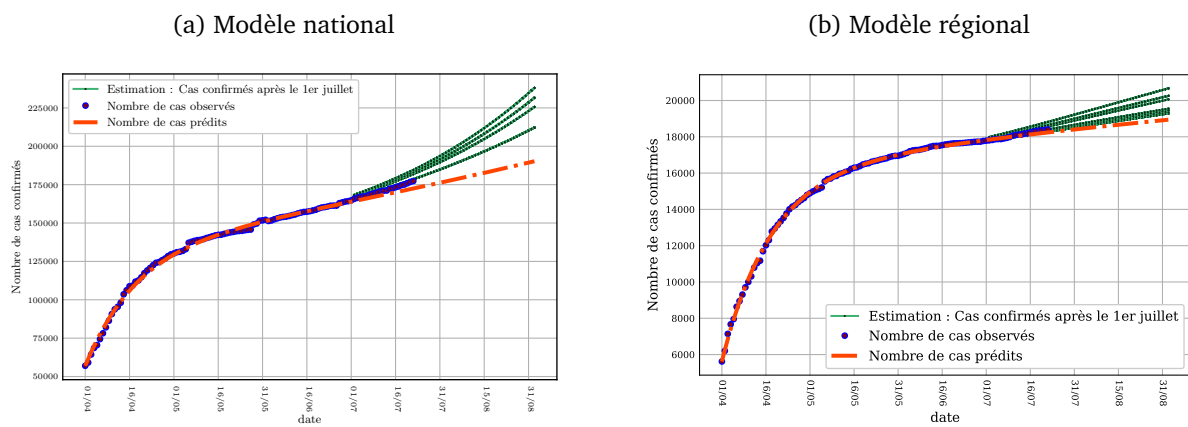


FIGURE IV.11: 2e cas : Impact de la relaxation du confinement sous différents scénarios possibles pour le modèle national et régional.

Les figures IV.10 et IV.11 illustrent l'impact de la relaxation du confinement sur l'évolution des cas infectés. Toutefois, de nombreux facteurs inconnus restent problématiques après le déconfinement. En effet, combien de personnes ont été immunisées après la première vague ? En supposant que 10% de la population a déjà été immunisée, c'est-à-dire que $S(0) = 9/10$ de la population restante est susceptible d'être infectée. Le modèle prévoit environ 220 000 cas d'infection en France et environ 25 000 cas d'infection dans la région PACA jusqu'à la fin du mois d'août.

IV.10 Conclusion

Dans ce chapitre, nous avons proposé un modèle épidémiologique amélioré, comprenant deux nouveaux compartiments, à savoir les hospitalisations et les unités de soins intensifs. Ces compartiments permettent d'évaluer les contraintes liées à l'épidémie de COVID-19 sur le plan sanitaire national et local. Les résultats démontrent la capacité du modèle à prédire les flux de cas estimés aux admissions hospitalières ou aux soins intensifs. Nous avons étudié la dynamique de propagation de COVID-19 selon différents scénarios en fonction de la situation de déconfinement. Nous avons également estimé les effets du confinement et de sa relaxation aux niveaux national et régional. La précision du modèle est principalement attestée par les compartiments H et U, qui sont l'objet principal de notre étude. Sur la base des tendances estimées de ces compartiments, une évaluation rigoureuse des tendances est réalisée. Ceci permet aux autorités sanitaires de surveiller les risques de la propagation du virus COVID-19 et à gérer les ressources hospitalières.

Limitations : Les tendances prédites dans l'analyse actuelle sont susceptibles de s'écarter des données réelles enregistrées. Plusieurs raisons peuvent expliquer ces écarts : au niveau de la lutte contre la propagation de l'infection, la réorientation des stratégies de dépistage permettant d'augmenter le nombre de tests quotidiens, la dégradation de la pratique de distanciation sociale, la poursuite des activités sociales ou la relaxation des mesures restrictives. Ces facteurs peuvent influencer le comportement de transmission de l'infection, entraînant des fluctuations soudaines dans le temps. En effet, une mise à jour des données COVID-19 dans le modèle est nécessaire pour fournir des performances précises.

A.1 Modèle SEIR généralisé

Les équations différentielles décrivant le modèle SEIR généralisé sont données sous la forme suivante :

$$\frac{dP}{dt} = \alpha S(t), \quad (\text{A.1})$$

$$\frac{dS}{dt} = -\beta \frac{S(t) \times I(t)}{N} - \alpha S(t), \quad (\text{A.2})$$

$$\frac{dE}{dt} = \beta \frac{S(t) \times I(t)}{N} - \gamma E(t), \quad (\text{A.3})$$

$$\frac{dI}{dt} = \gamma E(t) - \delta I(t), \quad (\text{A.4})$$

$$\frac{dQ}{dt} = \delta I(t) - \kappa(t)Q(t) - \lambda(t)Q(t), \quad (\text{A.5})$$

$$\frac{dG}{dt} = \lambda(t)Q(t), \quad (\text{A.6})$$

$$\frac{dD}{dt} = \kappa(t)Q(t). \quad (\text{A.7})$$

La population totale respecte l'équation suivante :

$$N = P(t) + S(t) + E(t) + I(t) + Q(t) + G(t) + D(t). \quad (\text{A.8})$$

En outre, les hyperparamètres sont définis comme suit :

- β : le taux d'infection.
- α : le taux de protection.
- γ : le temps de latence moyen.
- δ : le temps de quarantaine moyen.
- $\lambda(t)$: le taux de guérison.
- $\kappa(t)$: le taux de mortalité.

A.2 Modèle proposé

Les équations différentielles décrivant le modèle proposé sont données sous la forme suivante :

$$\frac{dP}{dt} = \alpha S(t), \quad (\text{A.9})$$

$$\frac{dS}{dt} = -\beta \frac{S(t) \times I(t)}{N} - \alpha S(t), \quad (\text{A.10})$$

$$\frac{dE}{dt} = \beta \frac{S(t) \times I(t)}{N} - \gamma E(t), \quad (\text{A.11})$$

$$\frac{dI}{dt} = \gamma E(t) - \delta I(t), \quad (\text{A.12})$$

$$\frac{dC}{dt} = \delta I(t), \quad (\text{A.13})$$

$$\frac{dG}{dt} = \lambda(t)C(t), \quad (\text{A.14})$$

$$\frac{dD}{dt} = \kappa(t)C(t) \quad (\text{A.15})$$

$$\frac{dH}{dt} = \omega(t)C(t) - \kappa(t)H(t) - \lambda(t)H(t) - \kappa_d(t)(C(t) - H(t)), \quad (\text{A.16})$$

$$\frac{dR}{dt} = \rho(t)H(t) - \rho_d(t)H(t), \quad (\text{A.17})$$

Les paramètres $\lambda(t)$, $\kappa(t)$, $\kappa_d(t)$, $\omega(t)$, $\rho(t)$ et $\rho_d(t)$ sont fonction du temps et peuvent être exprimés comme suit :

$$\lambda(t) = \lambda_0 e^{-\lambda_1 t} \quad (\text{A.18})$$

$$\kappa(t) = \kappa_0 e^{-\kappa_1 t} \quad (\text{A.19})$$

$$\kappa_d(t) = \nu_a e^{-\nu_b(t+\tau_D)} \quad (\text{A.20})$$

$$\omega(t) = \omega_0 e^{-\omega_1 t} \quad (\text{A.21})$$

$$\rho(t) = \rho_0 e^{-\frac{1}{2}(\rho_1 t)} \quad (\text{A.22})$$

$$\rho_d(t) = \frac{\rho_0}{\eta_a} e^{-\frac{1}{2} \frac{\rho_1}{\eta_b} (t+\tau_U)} \quad (\text{A.23})$$

Ainsi, les paramètres β , α , γ , δ , $\lambda(t)$, et $\kappa(t)$ sont définis comme dans le modèle précédent. En outre, les nouveaux paramètres $\kappa_d(t)$, $\omega(t)$, $\rho(t)$, et $\rho_d(t)$ sont définis comme suit :

- $\omega(t)$: le taux des admissions hospitalières, c'est-à-dire la proportion de personnes infectées hospitalisées
- $\rho(t)$: le taux d'admission aux unités de soins intensifs
- $\rho_d(t)$: le taux moyen de décharge des admissions dans les unités de soins intensifs après une période donnée.
- $\kappa_d(t)$: taux de mortalité réduit, ajusté pour estimer les décès non comptabilisés en dehors des structures hospitalières.

BIBLIOGRAPHIE

- [1] Y. Trardi, B. Ananou, and M. Ouladsine. An advanced arrhythmia recognition methodology based on r-waves time-series derivatives and benchmarking machine-learning algorithms. In *2020 European Control Conference (ECC)*, pages 1377–1382, 2020. doi : 10.23919/ECC51009.2020.9143678.
- [2] T. Youssef, Bouchra Ananou, and Mustapha Ouladsine. An Effective Data-Driven Diagnostic Strategy for Cardiac Pathology Screening. In *15th European Workshop on Advanced Control and Diagnosis, ACD 2019*, Bologna, Italy, November 2019. URL <https://hal-amu.archives-ouvertes.fr/hal-02486832>.
- [3] Y. Trardi, B. Ananou, Z. Haddi, and M. Ouladsine. A novel method to identify relevant features for automatic detection of atrial fibrillation. In *2018 26th Mediterranean Conference on Control and Automation (MED)*, pages 613–618, 2018. doi : 10.1109/MED.2018.8442479.
- [4] Y. Trardi, B. Ananou, Z. Haddi, and M. Ouladsine. Multi-dynamics analysis of qrs complex for atrial fibrillation diagnosis. In *2018 5th International Conference on Control, Decision and Information Technologies (CoDIT)*, pages 1067–1072, 2018. doi : 10.1109/CoDIT.2018.8394935.
- [5] Z. Haddi, B. Ananou, Y. Trardi, J.-F. Pons, S. Delliaux, J.-C. Deharo, and M. Ouladsine. Advanced machine learning coupled with heart-inter-beat derivatives for cardiac arrhythmia detection. In *2020 American Control Conference (ACC)*, pages 5433–5438, 2020. doi : 10.23919/ACC45564.2020.9147991.
- [6] Z. Haddi, B. Ananou, Y. Trardi, M. Ouladsine, J. Pons, S. Delliaux, and J. Deharo. Relevance vector machine as data-driven method for medical decision making. In *2019 18th European Control Conference (ECC)*, pages 1011–1016, 2019. doi : 10.23919/ECC.2019.8796141.
- [7] Z. Haddi, B. Ananou, Y. Trardi, J. Pons, S. Delliaux, M. Ouladsine, and J. Deharo. An efficient pattern recognition kernel-based method for atrial fibrillation diagnosis. In *2018 Computing in Cardiology Conference (CinC)*, volume 45, pages 1–4, 2018. doi : 10.22489/CinC.2018.090.
- [8] Zouhair Haddi, Bouchra Ananou, Trardi Youssef, and Mustapha Ouladsine. Fusion of multiple univariate data analysis-based detectors to build a specific fingerprint of atrial fibrillation. 12 2020. doi : 10.22489/CinC.2020.449.
- [9] J. Pan and W. J. Tompkins. A real-time qrs detection algorithm. *IEEE Transactions on Biomedical Engineering*, BME-32(3) :230–236, 1985.
- [10] Organisation Mondiale de la Santé (OMS). Allocution liminaire du directeur général de l’oms lors du point presse sur la covid-19 - 11 mars 2020, 2020. URL <https://www.who.int>.
- [11] Institut Pasteur de Lille. maladies cardiovasculaires, 03 janvier 2019. URL <https://www.pasteur-lille.fr/recherche-medicale/les-thematiques-de-recherche/la-recherche-sur-les-maladies-cardiovasculaires/>.
- [12] Ary L. Goldberger, Zachary D. Goldberger, and Alexei Shvilkin. *Chapter 14 - Supraventricular Arrhythmias, Part I : Premature Beats and Paroxysmal Supraventricular Tachycardias*, pages 130–143. Elsevier, Jan 2018. ISBN 978-0-323-40169-2.
- [13] Ary L. Goldberger, Zachary D. Goldberger, and Alexei Shvilkin. *Chapter 15 - Supraventricular Arrhythmias, Part II : Atrial Flutter and Atrial Fibrillation*, pages 144–155. Elsevier, Jan 2018. ISBN 978-0-323-40169-2.

- [14] Ary L. Goldberger, Zachary D. Goldberger, and Alexei Shvilkin. *Chapter 16 - Ventricular Arrhythmias*, pages 156–171. Elsevier, Jan 2018. ISBN 978-0-323-40169-2.
- [15] Cees B. de Vos, Ron Pisters, Robby Nieuwlaat, Martin H. Prins, Robert G. Tieleman, Robert-Jan S. Coelen, Antonius C. van den Heijkant, Maurits A. Allesie, and Harry J.G.M. Crijns. Progression from paroxysmal to persistent atrial fibrillation. *Journal of the American College of Cardiology*, 2010.
- [16] Massimo Zoni-Berisso, Fabrizio Lercari, Tiziana Carazza, and Stefano Domenicucci. Epidemiology of atrial fibrillation : European perspective. *Clinical epidemiology*, 6 :213–220, Jun 2014. ISSN 1179-1349. doi : 10.2147/CLEPS47385.
- [17] Jan Heeringa, Deirdre A.M. van der Kuip, Albert Hofman, Jan A. Kors, Gerard van Herpen, Bruno H.Ch. Stricker, Theo Stijnen, Gregory Y.H. Lip, and Jacqueline C.M. Witteman. Prevalence, incidence and lifetime risk of atrial fibrillation : the Rotterdam study. *European Heart Journal*, 27(8) :949–953, 03 2006. ISSN 0195-668X. doi : 10.1093/eurheartj/ehi825. URL <https://doi.org/10.1093/eurheartj/ehi825>.
- [18] Renate B. Schnabel, Xiaoyan Yin, Philimon Gona, Martin G. Larson, Alexa S. Beiser, David D. McManus, Christopher Newton-Cheh, Steven A. Lubitz, Jared W. Magnani, Patrick T. Ellinor, Sudha Seshadri, Philip A. Wolf, Ramachandran S. Vasam, Emelia J. Benjamin, and Daniel Levy. 50 year trends in atrial fibrillation prevalence, incidence, risk factors, and mortality in the framingham heart study : a cohort study. *Lancet (London, England)*, 386(9989) :154–162, Jul 2015. ISSN 1474-547X. doi : 10.1016/S0140-6736(14)61774-8.
- [19] Alan S. Go, Elaine M. Hylek, Kathleen A. Phillips, YuChiao Chang, Lori E. Henault, Joe V. Selby, and Daniel E. Singer. Prevalence of Diagnosed Atrial Fibrillation in Adults National Implications for Rhythm Management and Stroke Prevention : the Anticoagulation and Risk Factors In Atrial Fibrillation (ATRIA) Study. *JAMA*, 285(18) :2370–2375, 05 2001. ISSN 0098-7484. doi : 10.1001/jama.285.18.2370. URL <https://doi.org/10.1001/jama.285.18.2370>.
- [20] Kim E. Barrett, Susan M. Barman, Heddwen L. Brooks, and Jason X.J. Yuan. *Chapter 29 : Origin of the Heartbeat & the Electrical Activity of the Heart*. McGraw-Hill Education, 2019.
- [21] U Rajendra Acharya, Jasjit Suri, Jos Spaan, and S. Krishnan. *Advances in Cardiac Signal Processing*. 10 2014. doi : 10.1007/978-3-540-36675-1.
- [22] Healthengine. cardiovascular system (heart) anatomy, 2006. URL <https://healthengine.com.au/info/cardiovascular-system-heart>.
- [23] Cardiosecur. notions de base sur le coeur, 2020. URL <https://www.cardiosecur.com/fr/magazine/articles-specialises/notions-de-base-sur-le-coeur/coeur-en-bonne-sante>.
- [24] J. Gordon Betts Kelly A. Young, James A. Wise, Eddie Johnson, Brandon Poe, Dean H. Kruse, Oksana Korol, Jody E. Johnson, Mark Womble, and Peter DeSaix. *Cardiac Muscle and Electrical Activity*. OpenStax, 2013.
- [25] Richard Klabunde. *Cardiovascular physiology concepts*. Lippincott Williams & Wilkins, 2011.
- [26] S. N. Ramli, R. Ahmad, M. F. Abdollah, and E. Dutkiewicz. A biometric-based security for data authentication in wireless body area network (wban). In *2013 15th International Conference on Advanced Communications Technology (ICACT)*, pages 998–1001, 2013.
- [27] Ary L. Goldberger, Zachary D. Goldberger, and Alexei Shvilkin. *Chapter 7 - Atrial and Ventricular Enlargement*, pages 50–60. Elsevier, Jan 2018. ISBN 978-0-323-40169-2.
- [28] Maria Rasmussen, Preman Kumarathurai, Andreas Fabricius-Bjerre, Bjørn Larsen, Helena Domínguez, Ulla Davidsen, Thomas Gerds, Jorgen Kanters, and Ahmad Sajadieh. P-wave indices as predictors of atrial fibrillation. *Annals of Noninvasive Electrocardiology*, 04 2020. doi : 10.1111/anec.12751.
- [29] Ary L. Goldberger, Zachary D. Goldberger, and Alexei Shvilkin. *Chapter 8 - Ventricular Conduction Disturbances : Bundle Branch Blocks and Related Abnormalities*, pages 61–72. Elsevier, Jan 2018. ISBN 978-0-323-40169-2.

- [30] Ary L. Goldberger, Zachary D. Goldberger, and Alexei Shvilkin. *Chapter 2 - ECG Basics : Waves, Intervals, and Segments*, pages 6–10. Elsevier, Jan 2018. ISBN 978-0-323-40169-2.
- [31] Ary L. Goldberger, Zachary D. Goldberger, and Alexei Shvilkin. *Chapter 3 - How to Make Basic ECG Measurements*, pages 11–20. Elsevier, Jan 2018. ISBN 978-0-323-40169-2.
- [32] Ary L. Goldberger, Zachary D. Goldberger, and Alexei Shvilkin. *Chapter 17 - Atrioventricular (AV) Conduction Abnormalities, Part I : Delays, Blocks, and Dissociation Syndromes*, pages 172–182. Elsevier, Jan 2018. ISBN 978-0-323-40169-2.
- [33] Kyuhyun Wang, Richard W. Asinger, and Henry J.L. Marriott. St-segment elevation in conditions other than acute myocardial infarction. *New England Journal of Medicine*, 349(22) :2128–2135, 2003. doi : 10.1056/NEJMra022580.
- [34] Ary L. Goldberger, Zachary D. Goldberger, and Alexei Shvilkin. *Chapter 5 - The Normal ECG*, pages 32–40. Elsevier, Jan 2018. ISBN 978-0-323-40169-2.
- [35] Goldberger Ary L., Amaral Luis A. N., Glass Leon, Hausdorff Jeffrey M., Ivanov Plamen Ch., Mark Roger G., Mietus Joseph E., Moody George B., Peng Chung-Kang, and Stanley H. Eugene. Physiobank, physiotoolkit, and physionet. *Circulation*, 101(23) :e215–e220, Jun 2000.
- [36] Simona Petrutiu, Alan V. Sahakian, and Steven Swiryn. Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans. *EP Europace*, 9(7) :466–470, 05 2007. ISSN 1099-5129. doi : 10.1093/europace/eum096. URL <https://doi.org/10.1093/europace/eum096>.
- [37] G. MOODY. A new method for detecting atrial fibrillation using r-r intervals. *Computers in Cardiology*, pages 227–230, 1983.
- [38] George B. Moody and Roger G. Mark. Mit-bih arrhythmia database, 1992. URL <https://physionet.org/physiobank/database/mitdb/>.
- [39] G. E. Moody. Spontaneous termination of atrial fibrillation : a challenge from physionet and computers in cardiology 2004. In *Computers in Cardiology, 2004*, pages 101–104, 2004.
- [40] The Arrhythmia Laboratory At Boston's Beth Israel Hospital (Now The Beth Israel Deaconess Medical Center). The mit-bih normal sinus rhythm database, 1990. URL <https://physionet.org/physiobank/database/nsrdb/>.
- [41] Phyllis Stein and Rochelle Goldsmith. Normal sinus rhythm rr interval database, 2003. URL <https://physionet.org/physiobank/database/nsr2db/>.
- [42] Manab Kumar Das and Samit Ari. Ecg beats classification using mixture of features. *International Scholarly Research Notices*, 2014 :178436, Sep 2014. doi : 10.1155/2014/178436. URL <https://doi.org/10.1155/2014/178436>.
- [43] Eduardo José da S. Luz, William Robson Schwartz, Guillermo Cámara-Chávez, and David Menotti. Ecg-based heartbeat classification for arrhythmia detection : A survey. *Computer Methods and Programs in Biomedicine*, 2016.
- [44] Saeed Babaeizadeh, Richard E. Gregg, Eric D. Helfenbein, James M. Lindauer, and Sophia H. Zhou. Improvements in atrial fibrillation detection for real-time monitoring. *Journal of Electrocardiology*, 42(6) :522 – 526, 2009. ISSN 0022-0736. doi : <https://doi.org/10.1016/j.jelectrocard.2009.06.006>. URL <http://www.sciencedirect.com/science/article/pii/S0022073609002507>.
- [45] S. Hargittai. Is it possible to detect atrial fibrillation by simply using rr intervals? In *Computing in Cardiology 2014*, 2014.
- [46] P. Langley, M. Dewhurst, L. Y. Di Marco, P. Adams, F. Dewhurst, J. C. Mwita, R. Walker, and A. Murray. Accuracy of algorithms for detection of atrial fibrillation from short duration beat interval recordings. *Medical Engineering & Physics*, 2012.
- [47] Douglas E. Lake and J. Randall Moorman. Accurate estimation of entropy in very short physiological time series : the problem of atrial fibrillation detection in implanted ventricular devices. *American Journal of Physiology-Heart and Circulatory Physiology*, 2011.

- [48] S. Cerutti, L. T. Mainardi, A. Porta, and A. M. Bianchi. Analysis of the dynamics of rr interval series for the detection of atrial fibrillation episodes. In *Computers in Cardiology 1997*, pages 77–80, 1997.
- [49] Andrius Petrenas, Vaidotas Marozas, and Leif Sörnmo. Low-complexity detection of atrial fibrillation in continuous long-term monitoring. *Computers in Biology and Medicine*, 2015.
- [50] Lu Hong-wei, S. Ying, Lin Min, Li Pi-Ding, and Z. Zheng. A probability density function method for detecting atrial fibrillation using r-r intervals. *Medical engineering & physics*, 31 1 :116–23, 2009.
- [51] K. Tateno and L. Glass. A method for detection of atrial fibrillation using rr intervals. In *Computers in Cardiology 2000. Vol.27 (Cat. 00CH37163)*, pages 391–394, 2000.
- [52] B. Young, D. Brodnick, and R. Spaulding. A comparative study of a hidden markov model detector for atrial fibrillation. In *Neural Networks for Signal Processing IX : Proceedings of the 1999 IEEE Signal Processing Society Workshop (Cat. No.98TH8468)*, pages 468–476, 1999.
- [53] David Duverney, Jean-Michel Gaspoz, Vincent Pichot, Frédéric Roche, Richard Brion, Anestis Antoniadis, and Jean-Claude Barthélémy. High accuracy of automatic detection of atrial fibrillation using wavelet transform of heart rate intervals. *Pacing and Clinical Electrophysiology*, 25(4) :457–462, 2002. doi : 10.1046/j.1460-9592.2002.00457.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1460-9592.2002.00457.x>.
- [54] G. Hindricks, M. Taborsky, P. Wohlgemuth, G. Rieger, F. Beckers, and B. Albers. Atrial fibrillation detection by a subcutaneous monitoring device. In *2008 Computers in Cardiology*, pages 413–416, 2008.
- [55] A. Ghodrati, B. Murray, and S. Marinello. Rr interval analysis for detection of atrial fibrillation in ecg monitors. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 601–604, 2008.
- [56] Xiuhua Ruan, Changchun Liu, Chengyu Liu, Xinpei Wang, and Peng Li. Automatic detection of atrial fibrillation using r-r interval signal. In *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*, 2011.
- [57] R. Colloca, A. E. Johnson, L. Mainardi, and G. D. Clifford. A support vector machine approach for reliable detection of atrial fibrillation events. In *Computing in Cardiology 2013*, pages 1047–1050, 2013.
- [58] M. Mohebbi and H. Ghassemian. Detection of atrial fibrillation episodes using svm. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 177–180, 2008.
- [59] Babak Mohammadzadeh Asl, Seyed Kamaledin Setarehdan, and Maryam Mohebbi. Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal. *Artificial Intelligence in Medicine*, 44(1) :51 – 64, 2008. ISSN 0933-3657. doi : <https://doi.org/10.1016/j.artmed.2008.04.007>. URL <http://www.sciencedirect.com/science/article/pii/S0933365708000559>.
- [60] Maryam Mohebbi and Hassan Ghassemian. Prediction of paroxysmal atrial fibrillation based on non-linear analysis and spectrum and bispectrum features of the heart rate variability signal. *Computer methods and programs in biomedicine*, 105 :40–9, 01 2012. doi : 10.1016/j.cmpb.2010.07.011.
- [61] S. G. Artis, R. G. Mark, and G. B. Moody. Detection of atrial fibrillation using artificial neural networks. In *[1991] Proceedings Computers in Cardiology*, pages 173–176, 1991.
- [62] U. Rajendra Acharya, Meena Sankaranarayanan, Jagadish Nayak, Chen Xiang, and Toshiyo Tamura. Automatic identification of cardiac health using modeling techniques : A comparative study. *Information Sciences*, 2008.
- [63] Farid Yaghoubi and Ahmad Ayatollahi. Robust genetic programming-based detection of atrial fibrillation using rr intervals. *Expert Systems*, 29, 03 2010. doi : 10.1111/j.1468-0394,.

- [64] Marta Carrara, L. Carozzi, Travis J Moss, Marco de Pasquale, Sergio Cerutti, Manuela Ferrario, Douglas E. Lake, and J. Randall Moorman. Heart rate dynamics distinguish among atrial fibrillation, normal sinus rhythm and sinus rhythm with frequent ectopy. *Physiological measurement*, 2015.
- [65] Alan Kennedy, Dewar D. Finlay, Daniel Guldenring, Raymond R. Bond, Kieran Moran, and James McLaughlin. Automated detection of atrial fibrillation using r-r intervals and multivariate-based classification. *Journal of Electrocardiology*, 2016.
- [66] Elias Ebrahimzadeh, Maede Kalantari, Mohammadamin Joulani, Reza Shahrokhi Shahraki, Farahnaz Fayaz, and Fereshteh Ahmadi. Prediction of paroxysmal atrial fibrillation : A machine learning based approach using combined feature vector and mixture of expert classification on hrv signal. *Computer Methods and Programs in Biomedicine*, 2018.
- [67] Jibin Wang. A deep learning approach for atrial fibrillation signals classification based on convolutional and modified elman neural network. *Future Generation Computer Systems*, 2019.
- [68] Oliver Faust, Alex Shenfield, Murtadha Kareem, Tan Ru San, Hamido Fujita, and U. Rajendra Acharya. Automated detection of atrial fibrillation using long short-term memory network with rr interval signals. *Computers in Biology and Medicine*, 2018.
- [69] Rasmus S. Andersen, Abdolrahman Peimankar, and Puthusserypady Sadasivan. A deep learning approach for real-time detection of atrial fibrillation. *Expert Systems with Applications*, 2019.
- [70] H. H. So and K. L. Chan. Development of qrs detection method for real-time ambulatory cardiac monitor. In *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 'Magnificent Milestones and Emerging Opportunities in Medical Engineering' (Cat. No.97CH36136)*, volume 1, pages 289–292, 1997.
- [71] Leif Sörnmo and Pablo Laguna. *Chapter 7 - ECG Signal Processing*, pages 453–566. Biomedical Engineering. Academic Press, Burlington, Jan 2005. ISBN 978-0-12-437552-9. URL <http://www.sciencedirect.com/science/article/pii/B9780124375529500076>.
- [72] N. M. Arzeno, Z. Deng, and C. Poon. Analysis of first-derivative based qrs detection algorithms. *IEEE Transactions on Biomedical Engineering*, 55(2) :478–484, 2008.
- [73] Leif Sörnmo and Pablo Laguna. *Chapter 8 - ECG Signal Processing : Heart Rate Variability*, pages 567–631. Biomedical Engineering. Academic Press, Burlington, Jan 2005. ISBN 978-0-12-437552-9. URL <http://www.sciencedirect.com/science/article/pii/B9780124375529500088>.
- [74] Gustavo Lenis, Nicolas Pilia, Axel Loewe, Walther H. W. Schulze, and Olaf Dössel. Comparison of baseline wander removal techniques considering the preservation of st changes in the ischemic ecg : A simulation study. *Computational and Mathematical Methods in Medicine*, 2017 :9295029, Mar 2017. ISSN 1748-670X. doi : 10.1155/2017/9295029. URL <https://doi.org/10.1155/2017/9295029>.
- [75] Nikita R Pai, Chaithra S, Rekha J, and Sowmya R Bangari. A comparative study of digital fir and iir band-pass filter. *International Research Journal of Engineering and Technology (IRJET)*, 06, May 2019. URL <https://www.irjet.net/archives/V6/i5/IRJET-V6I5465.pdf>.
- [76] Bhaskar Mishra and Rajesh Mehra. Design and simulation of low pass iir filter for ecg interference reduction. *International Journal of Engineering Research & Technology*, 03, Mars 2014. ISSN 2278-0181. URL <https://www.ijert.org/design-and-simulation-of-low-pass-iir-filter-for-ecg-interference-reduction>.
- [77] Hooman Sedghamiz. Complete pan tompkins implementation ecg qrs detector., August 2020. URL <https://www.mathworks.com/matlabcentral/fileexchange/45840-complete-pan-tompkins-implementation-ecg-qrs-detector>.
- [78] B. Logan and J. Healey. Robust detection of atrial fibrillation for a long term telemonitoring system. In *Computers in Cardiology, 2005*, pages 619–622, 2005.
- [79] Julien Oster and Gari Clifford. Impact of the presence of noise on rr intervals-based atrial fibrillation detection. *Journal of electrocardiology*, 48, 09 2015. doi : 10.1016/j.jelectrocard.2015.08.013.

- [80] João Paulo do Vale Madeiro, José Maria da Silva Monteiro Filho, and Priscila Rocha Ferreira Rodrigues. *Chapter 4 - Techniques for QRS Complex Detection*, pages 89–118. Academic Press, Jan 2019. ISBN 978-0-12-814035-2. URL <http://www.sciencedirect.com/science/article/pii/B9780128140352000104>.
- [81] Tanushree Sharma and Kamallesh Kumar Sharma. Qrs complex detection in ecg signals using locally adaptive weighted total variation denoising. *Computers in Biology and Medicine*, 87 : 187 – 199, 2017. ISSN 0010-4825. doi : <https://doi.org/10.1016/j.compbiomed.2017.05.027>. URL <http://www.sciencedirect.com/science/article/pii/S001048251730149X>.
- [82] M.Sabarimalai Manikandan and K.P. Soman. A novel method for detecting r-peaks in electrocardiogram (ecg) signal. *Biomedical Signal Processing and Control*, 7(2) :118 – 128, 2012. ISSN 1746-8094. doi : <https://doi.org/10.1016/j.bspc.2011.03.004>. URL <http://www.sciencedirect.com/science/article/pii/S1746809411000292>.
- [83] Yongli Chen and Huilong Duan. A qrs complex detection algorithm based on mathematical morphology and envelope. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pages 4654–4657, 2005.
- [84] R. Rodríguez, A. Mexicano, J. Bila, S. Cervantes, and R. Ponce. Feature extraction of electrocardiogram signals by applying adaptive threshold and principal component analysis. *Journal of Applied Research and Technology*, 13(2) :261 – 269, 2015. ISSN 1665-6423. doi : <https://doi.org/10.1016/j.jart.2015.06.008>. URL <http://www.sciencedirect.com/science/article/pii/S1665642315000103>.
- [85] Cuiwei Li, Chongxun Zheng, and Changfeng Tai. Detection of ecg characteristic points using wavelet transforms. *IEEE Transactions on Biomedical Engineering*, 42(1) :21–28, 1995.
- [86] V. X. Afonso, W. J. Tompkins, T. Q. Nguyen, and Shen Luo. Ecg beat detection using filter banks. *IEEE Transactions on Biomedical Engineering*, (2) :192–202, 1999.
- [87] S. Kadambe, R. Murray, and G. F. Boudreaux-Bartels. Wavelet transform-based qrs complex detector. *IEEE Transactions on Biomedical Engineering*, 46(7) :838–848, 1999.
- [88] Laurence Keselbrenner, Michel Keselbrenner, and Solange Akselrod. Nonlinear high pass filter for r-wave detection in ecg signal. *Medical Engineering & Physics*, 19(5) :481 – 484, 1997. ISSN 1350-4533. doi : [https://doi.org/10.1016/S1350-4533\(97\)00013-1](https://doi.org/10.1016/S1350-4533(97)00013-1). URL <http://www.sciencedirect.com/science/article/pii/S1350453397000131>.
- [89] D. Benitez, P.A. Gaydecki, A. Zaidi, and A.P. Fitzpatrick. The use of the hilbert transform in ecg signal analysis. *Computers in Biology and Medicine*, 31(5) :399 – 406, 2001. ISSN 0010-4825. doi : [https://doi.org/10.1016/S0010-4825\(01\)00009-9](https://doi.org/10.1016/S0010-4825(01)00009-9). URL <http://www.sciencedirect.com/science/article/pii/S0010482501000099>.
- [90] K. V. Suarez, J. C. Silva, Y. Berthoumieu, P. Gomis, and M. Najim. Ecg beat detection using a geometrical matching approach. *IEEE Transactions on Biomedical Engineering*, 54(4) :641–650, 2007.
- [91] Xiaomin Xu and Ying Liu. Ecg qrs complex detection using slope vector waveform (svw) algorithm. In *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 2, pages 3597–3600, 2004.
- [92] M. L. Ahlstrom and W. J. Tompkins. Automated high-speed analysis of holter tapes with microcomputers. *IEEE Transactions on Biomedical Engineering*, BME-30(10) :651–657, 1983.
- [93] H. A. N. Dinh, D. K. Kumar, N. D. Pah, and P. Burton. Wavelets for qrs detection. In *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 2, pages 1883–1887 vol.2, 2001.
- [94] S. K. Bashar, Y. Noh, A. J. Walkey, D. D. McManus, and K. H. Chon. Verb : Vfdm-based electrocardiogram reconstruction and beat detection algorithm. *IEEE Access*, 7 :13856–13866, 2019.

- [95] Chandan Nayak, Suman Kumar Saha, Rajib Kar, and Durbadal Mandal. An optimally designed digital differentiator based preprocessor for r-peak detection in electrocardiogram signal. *Biomedical Signal Processing and Control*, 49 :440 – 464, 2019. ISSN 1746-8094. doi : <https://doi.org/10.1016/j.bspc.2018.09.005>. URL <http://www.sciencedirect.com/science/article/pii/S174680941830243X>.
- [96] Ashish Kumar, Ramana Ranganatham, Rama Komaragiri, and Manjeet Kumar. Efficient qrs complex detection algorithm based on fast fourier transform. *Biomedical Engineering Letters*, 9(1) :145–151, Feb 2019. ISSN 2093-985X. doi : 10.1007/s13534-018-0087-y. URL <https://doi.org/10.1007/s13534-018-0087-y>.
- [97] X. Tang, Q. Hu, and W. Tang. A real-time qrs detection system with pr/rt interval and st segment measurements for wearable ecg sensors using parallel delta modulators. *IEEE Transactions on Biomedical Circuits and Systems*, 12(4) :751–761, 2018.
- [98] Mahsa Akhbari, Nasim Montazeri Ghahjaverestan, Mohammad B. Shamsollahi, and Christian Jutten. Ecg fiducial point extraction using switching kalman filter. *Computer Methods and Programs in Biomedicine*, 157 :129 – 136, 2018. ISSN 0169-2607. doi : <https://doi.org/10.1016/j.cmpb.2018.01.018>. URL <http://www.sciencedirect.com/science/article/pii/S0169260716308975>.
- [99] Ricardo Rodriguez Jorge, Edgar Martínez García, Rafael Torres Córdoba, Jiri Bila, and J. Mizera-Pietraszko. Adaptive threshold, wavelet and hilbert transform for qrs detection in electrocardiogram signals. In Fatos Xhafa, Santi Caballé, and Leonard Barolli, editors, *Advances on P2P, Parallel, Grid, Cloud and Internet Computing*, pages 777–786, Cham, 2018. Springer International Publishing. ISBN 978-3-319-69835-9.
- [100] Xuanyu Lu, Maolin Pan, and Yang Yu. Qrs detection based on improved adaptive threshold. *Journal of Healthcare Engineering*, 2018 :1–8, 03 2018. doi : 10.1155/2018/5694595.
- [101] Shweta Jain, M.K. Ahirwal, Anil Kumar, V. Bajaj, and G.K. Singh. Qrs detection using adaptive filters : A comparative study. *ISA Transactions*, 66 :362 – 375, 2017. ISSN 0019-0578. doi : <https://doi.org/10.1016/j.isatra.2016.09.023>. URL <http://www.sciencedirect.com/science/article/pii/S001905781630386X>.
- [102] Jie Lian, Lian Wang, and Dirk Muessig. A simple method to detect atrial fibrillation using rr intervals. *The American Journal of Cardiology*, 107(10) :1494 – 1497, 2011. ISSN 0002-9149. doi : <https://doi.org/10.1016/j.amjcard.2011.01.028>. URL <http://www.sciencedirect.com/science/article/pii/S0002914911003407>.
- [103] Xingran Cui, Emily Chang, Wen-Hung Yang, Bernard Jiang, Albert C Yang, and Chung-Kang Peng. Automated detection of paroxysmal atrial fibrillation using an information-based similarity approach. *Entropy*, 19 :677, 12 2017. doi : 10.3390/e19120677.
- [104] Yan Wang, Shoushui Wei, Shuai Zhang, Yatao Zhang, Lina Zhao, Chengyu Liu, and Alan Murray. Comparison of time-domain, frequency-domain and non-linear analysis for distinguishing congestive heart failure patients from normal sinus rhythm subjects. *Biomedical Signal Processing and Control*, 42 :30 – 36, 2018. ISSN 1746-8094. doi : <https://doi.org/10.1016/j.bspc.2018.01.001>. URL <http://www.sciencedirect.com/science/article/pii/S1746809418300016>.
- [105] Md Saiful Islam, Nassim Ammour, Naif Alajlan, and Hatim Aboalsamh. Rhythm-based heartbeat duration normalization for atrial fibrillation detection. *Computers in Biology and Medicine*, 72 : 160 – 169, 2016. ISSN 0010-4825. doi : <https://doi.org/10.1016/j.compbiomed.2016.03.015>. URL <http://www.sciencedirect.com/science/article/pii/S0010482516300683>.
- [106] K. Tateno and L. Glass. Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of rr and δrr intervals. *Medical and Biological Engineering and Computing*, 39(6) :664–671, Nov 2001. ISSN 1741-0444. doi : 10.1007/BF02345439. URL <https://doi.org/10.1007/BF02345439>.
- [107] U Rajendra Acharya, Paul Joseph, N Kannathal, Choo Lim, and Jasjit Suri. Heart rate variability : A review. *Medical & biological engineering & computing*, 44 :1031–51, 01 2007. doi : 10.1007/s11517-006-0119-0.

- [108] Marek Malik. Heart rate variability : Standards of measurement, physiological interpretation, and clinical use. *Circulation*, 93 :1043–1065, 03 1996.
- [109] Jean-François Pons, Zouhair Haddi, Jean-Claude Deharo, Ahmed Charai, Rachid Bouchakour, Mustapha Ouladsine, and Stéphane Delliaux. Heart rhythm characterization through induced physiological variables. *Scientific Reports*, 7(1) :5059, 2017.
- [110] Yeong Pong Meau, Fatimah Ibrahim, Selvanathan A.L. Narainasamy, and Razali Omar. Intelligent classification of electrocardiogram (ecg) signal using extended kalman filter (ekf) based neuro fuzzy system. *Computer Methods and Programs in Biomedicine*, 82(2) :157 – 168, 2006. ISSN 0169-2607. doi : <https://doi.org/10.1016/j.cmpb.2006.03.003>. URL <http://www.sciencedirect.com/science/article/pii/S0169260706000654>.
- [111] Jeen-Shing Wang, Wei-Chun Chiang, Yu-Liang Hsu, and Ya-Ting C. Yang. Ecg arrhythmia classification using a probabilistic neural network with a feature reduction method. *Neurocomputing*, 116 :38 – 45, 2013. ISSN 0925-2312. doi : <https://doi.org/10.1016/j.neucom.2011.10.045>. URL <http://www.sciencedirect.com/science/article/pii/S0925231212006960>. Advanced Theory and Methodology in Intelligent Computing.
- [112] Wikipedia contributors. Bessel’s correction, 2020. URL https://en.wikipedia.org/w/index.php?title=Bessel%27s_correction&oldid=980227409.
- [113] S. Dash, K. H. Chon, S. Lu, and E. A. Raeder. Automatic real time detection of atrial fibrillation. *Annals of Biomedical Engineering*, 37(9) :1701–1709, Sep 2009. ISSN 1573-9686. doi : 10.1007/s10439-009-9740-z. URL <https://doi.org/10.1007/s10439-009-9740-z>.
- [114] Marek Malik. Heart rate variability. *Annals of Noninvasive Electrocardiology*, 1(2) :151–181, 1996. doi : 10.1111/j.1542-474X.1996.tb00275.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1542-474X.1996.tb00275.x>.
- [115] Rosangela Akemi Hoshi, Carlos Marcelo Pastre, Luiz Carlos Marques Vanderlei, and Moacir Fernandes Godoy. Poincaré plot indexes of heart rate variability : Relationships with other nonlinear variables. *Autonomic Neuroscience*, 177(2) :271 – 274, 2013. ISSN 1566-0702. doi : <https://doi.org/10.1016/j.autneu.2013.05.004>. URL <http://www.sciencedirect.com/science/article/pii/S1566070213001264>.
- [116] M. Brennan, M. Palaniswami, and P. Kamen. Do existing measures of poincare plot geometry reflect nonlinear features of heart rate variability? *IEEE Transactions on Biomedical Engineering*, 48(11) :1342–1347, 2001.
- [117] Chandan K. Karmakar, Jayavardhana Gubbi, Ahsan H. Khandoker, and Marimuthu Palaniswami. Analyzing temporal variability of standard descriptors of poincaré plots. *Journal of Electrocardiology*, 43(6) :719 – 724, 2010. ISSN 0022-0736. doi : <https://doi.org/10.1016/j.jelectrocard.2010.09.001>. URL <http://www.sciencedirect.com/science/article/pii/S0022073610003821>.
- [118] S M Pincus. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, 88(6) :2297–2301, 1991. ISSN 0027-8424. doi : 10.1073/pnas.88.6.2297. URL <https://www.pnas.org/content/88/6/2297>.
- [119] Joshua S. Richman and J. Randall Moorman. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, 278(6) :H2039–H2049, 2000. doi : 10.1152/ajpheart.2000.278.6.H2039. URL <https://doi.org/10.1152/ajpheart.2000.278.6.H2039>.
- [120] W. Chen, Z. Wang, H. Xie, and W. Yu. Characterization of surface emg signal based on fuzzy entropy. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 15(2) :266–272, June 2007.
- [121] Yatao Zhang, Shoushui Wei, Yutao Long, and Chengyu Liu. Performance analysis of multiscale entropy for the assessment of ecg signal quality. *Journal of Electrical and Computer Engineering*, 2015 :563915, Apr 2015. ISSN 2090-0147. doi : 10.1155/2015/563915. URL <https://doi.org/10.1155/2015/563915>.

- [122] Weiting Chen, Jun Zhuang, Wangxin Yu, and Zhizhong Wang. Measuring complexity using fuzzyen, apen, and sampen. *Medical Engineering & Physics*, 31(1) :61 – 68, 2009. ISSN 1350-4533. doi : <https://doi.org/10.1016/j.medengphy.2008.04.005>. URL <http://www.sciencedirect.com/science/article/pii/S1350453308000726>.
- [123] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1) :389–422, Jan 2002. ISSN 1573-0565. doi : 10.1023/A:1012487302797. URL <https://doi.org/10.1023/A:1012487302797>.
- [124] Zena M. Hira and Duncan F. Gillies. A review of feature selection and feature extraction methods applied on microarray data. *Advances in Bioinformatics*, 2015 :198363, Jun 2015. ISSN 1687-8027. doi : 10.1155/2015/198363. URL <https://doi.org/10.1155/2015/198363>.
- [125] Gert R. G. Lanckriet, Nello Cristianini, Peter Bartlett, Laurent El Ghaoui, and Michael I. Jordan. Learning the kernel matrix with semidefinite programming. *J. Mach. Learn. Res.*, 5 :27–72, December 2004. ISSN 1532-4435.
- [126] Sören Sonnenburg, Gunnar Rätsch, and Christin Schäfer. A general and efficient multiple kernel learning algorithm. 01 2006.
- [127] Marius Kloft, Ulf Brefeld, and Alexander Zien. Lp-norm multiple kernel learning. *Journal of Machine Learning Research*, 12 :663–663, 03 2011.
- [128] Zenglin Xu, Rong Jin, Shenghuo Zhu, Michael Lyu, and Irwin King. Smooth optimization for effective multiple kernel learning. 01 2010.
- [129] Matthieu Kowalski, Marie Szafranski, and Liva Ralaivola. Multiple indefinite kernel learning with mixed norm regularization. In *Proceedings of the 26th Annual International Conference on Machine Learning, ICML '09*, page 545–552, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605585161. doi : 10.1145/1553374.1553445. URL <https://doi.org/10.1145/1553374.1553445>.
- [130] Alain Rakotomamonjy, Francis Bach, Stephane Canu, and Yves Grandvalet. Simplemkl. *Journal of Machine Learning Research*, 9, 11 2008.
- [131] George F. Gao. From “a”iv to “z”ikv : Attacks from emerging and re-emerging pathogens. *Cell*, 172(6) :1157 – 1159, 2018. ISSN 0092-8674. doi : <https://doi.org/10.1016/j.cell.2018.02.025>.
- [132] Zhu Na, Zhang Dingyu, Wang Wenling, Li Xingwang, Yang Bo, Song Jingdong, Zhao Xiang, Huang Baoying, Shi Weifeng, Lu Roujian, Niu Peihua, Zhan Faxian, Ma Xuejun, Wang Dayan, Xu Wenbo, Wu Guizhen, Gao George F., and Tan Wenjie. A novel coronavirus from patients with pneumonia in china, 2019. *New England Journal of Medicine*, 382(8) :727–733, 2020. doi : 10.1056/NEJMoa2001017.
- [133] European Centre for Disease Prevention and Control. Cluster of pneumonia cases caused by a novel coronavirus, wuhan, china, 2020, 2020. URL <https://www.ecdc.europa.eu>.
- [134] World Health Organization (WHO). Coronavirus disease 2019 (covid-19) : situation report, 2020.
- [135] Long-quan Li, Tian Huang, Yong-qing Wang, Zheng-ping Wang, Yuan Liang, Tao-bi Huang, Hui-yun Zhang, Wei-ming Sun, and Yu-ping Wang. 2019 novel coronavirus patients’ clinical characteristics, discharge rate and fatality rate of meta-analysis. *Journal of Medical Virology*, 03 2020. doi : 10.1002/jmv.25757.
- [136] Gustavo Barbosa Libotte, Fran Sérgio Lobato, Gustavo Mendes Platt, and Antônio J. Silva Neto. Determination of an optimal control strategy for vaccine administration in covid-19 pandemic treatment. *Computer Methods and Programs in Biomedicine*, 196 :105664, Nov 2020. ISSN 0169-2607.
- [137] Shumaila Azam, Jorge E. Macías-Díaz, Nauman Ahmed, Ilyas Khan, Muhammad S. Iqbal, Muhammad Rafiq, Kottakkaran S. Nisar, and Muhammad O. Ahmad. Numerical modeling and theoretical analysis of a nonlinear advection-reaction epidemic system. *Computer Methods and Programs in Biomedicine*, 193 :105429, Sep 2020. ISSN 0169-2607.

- [138] Peipei Wang, Xinqi Zheng, Jiayang Li, and Bangren Zhu. Prediction of epidemic trends in covid-19 with logistic model and machine learning technics. *Chaos, Solitons & Fractals*, 139 : 110058, Oct 2020. ISSN 0960-0779.
- [139] Yingkang Xie, Zhen Wang, Junwei Lu, and Yuxia Li. Stability analysis and control strategies for a new sis epidemic model in heterogeneous networks. *Applied Mathematics and Computation*, 383 :125381, Oct 2020. ISSN 0096-3003.
- [140] Chaeyoung Lee, Yibao Li, and Junseok Kim. The susceptible-unidentified infected-confirmed (suc) epidemic model for estimating unidentified infected population for covid-19. *Chaos, Solitons & Fractals*, 139 :110090, Oct 2020. ISSN 0960-0779.
- [141] Laura Di Domenico, Giulia Pullano, Chiara Sabbatini, Pierre-Yves Boelle, and Vittoria Colizza. Expected impact of lockdown in ile-de-france and possible exit strategies. 2020. doi : 10.1101/2020.04.13.20063933.
- [142] Liangrong Peng, Wuyue Yang, Dongyan Zhang, Changjing Zhuge, and Liu Hong. Epidemic analysis of covid-19 in china by dynamical modeling. *medRxiv*, 2020. doi : 10.1101/2020.02.16.20023465.
- [143] E. Cheynet. Generalized SEIR Epidemic Model (fitting and computation), 2020.
- [144] Meng Wang and Jingtao Qi. A deterministic epidemic model for the emergence of covid-19 in china. *medRxiv*, 2020. doi : 10.1101/2020.03.08.20032854.
- [145] Zhilan Feng and Horst R. Thieme. Recurrent outbreaks of childhood diseases revisited : The impact of isolation. *Mathematical Biosciences*, 1995. doi : [https://doi.org/10.1016/0025-5564\(94\)00069-C](https://doi.org/10.1016/0025-5564(94)00069-C).
- [146] Howard Weiss. The sir model and the foundations of public health. 2013.
- [147] Yang Yang, Qingbin Lu, Mingjin Liu, Yixing Wang, Anran Zhang, Neda Jalali, Natalie Dean, Ira Longini, M. Elizabeth Halloran, Bo Xu, Xiaoi Zhang, Liping Wang, Wei Liu, and Liqun Fang. Epidemiological and clinical features of the 2019 novel coronavirus outbreak in china. *medRxiv*, 2020. doi : 10.1101/2020.02.10.20021675.
- [148] Henrik Salje, Cécile Tran Kiem, Noémie Lefrancq, Noémie Courtejoie, Paolo Bosetti, Juliette Paireau, Alessio Andronico, Nathanaël Hozé, Jehanne Richet, Claire-Lise Dubost, Yann Le Strat, Justin Lessler, Daniel Levy-Bruhl, Arnaud Fontanet, Lulla Opatowski, Pierre-Yves Boelle, and Simon Cauchemez. Estimating the burden of sars-cov-2 in france. *Science*, 2020. doi : 10.1126/science.abc3517.
- [149] SantePubliqueFrance. Infection au nouveau coronavirus (sars-cov-2), covid-19, 2020. URL <https://www.santepubliquefrance.fr>.
- [150] ARS-Paca. Agence régionale de santé paca, 2020. URL <https://www.paca.ars.sante.fr/>.

Abstract

Cardiovascular disease (CVD) is one of the most pressing health concerns and the leading cause of death worldwide. According to the World Health Organization (WHO), CVD is responsible for 17.9 million deaths worldwide each year, or 31% of all deaths. In France, CVD is the second leading cause of death after cancer, with approximately 150,000 deaths per year. Myocardial infarction also called a heart attack, is the most deadly form of CVD in the world. It causes 18,000 deaths per year in France, i.e., 10% of total mortality. In this thesis, we are interested in CVD, and more precisely in one of its leading causes, namely cardiac arrhythmias. Academic research and industry rely on technological advances to develop IT tools for arrhythmia detection. In this work, we address this issue by proposing a new diagnostic strategy to distinguish healthy subjects in the presence of ectopic beats from subjects with atrial fibrillation. This strategy is based on the analysis of additional derivatives extracted from the R-R interval time series. This approach is composed of a process of derivative calculation and feature extraction. We applied different classification algorithms to build the diagnostic model, including support vector machine and multi-kernel learning. Also, we have developed a high-performance variable selection algorithm based on multi-kernel programming. The developed approach has been validated on different cardiac arrhythmia databases. The results obtained demonstrate the efficiency and robustness of the developed method.

Resumé

Les maladies cardiovasculaires (MCV) sont parties des préoccupations sanitaires les plus pressantes. Les MCV présentent la première cause de décès dans le monde. Selon l'Organisation mondiale de la santé (OMS), les MCV sont à l'origine de 17,9 millions de décès dans le monde chaque année, soit 31% de l'ensemble des décès. En France, les MCV sont la deuxième cause de décès après le cancer, avec environ 150 000 décès par an. L'infarctus du myocarde, également appelé crise cardiaque, est la forme de MCV la plus meurtrière au monde. Il provoque 18 000 décès par an en France, soit 10% de la mortalité totale. Dans cette thèse, nous nous intéressons aux MCV, et plus précisément à l'une de ses principales causes, à savoir les arythmies cardiaques. Les recherches académiques et les industriels s'appuient sur les avancées technologiques pour développer des outils informatiques pour la détection de l'arythmie. Dans ce travail, nous discutons cette problématique en proposant une nouvelle stratégie de diagnostic qui permet de distinguer les sujets sains en présence de battements ectopiques des sujets atteints de la fibrillation auriculaire. Cette stratégie est basée sur l'analyse de dérivés complémentaires extraits de la série chronologique des intervalles R-R. Pour construire le modèle de diagnostic, nous avons appliqué différents algorithmes de classification, notamment les séparateurs à vaste marge et l'apprentissage multinoyaux. En outre, nous avons développé un algorithme de sélection de variables très performant, basé sur la programmation multinoyaux. L'approche développée a été validée sur différentes bases de données d'arythmies cardiaques. Les résultats obtenus démontrent l'efficacité et la robustesse de la méthode développée.