

République Algérienne Démocratique et Populaire
UNIVERSITE ABDELHAMID BIN BADIS MOSTAGANEM



جامعة مستغانم
Université De Mostaganem

FACULTE DE SCIENCE ET TECHNOLOGIE
DEPARTEMENT GENIE ELECTRIQUE

Mémoire de fin d'études
De MASTER PROFESSIONNEL

Spécialité : *Électronique des Systèmes Embarqués*

Thème :

**Détection des arythmies cardiaques par
machine learning**

Réalisé par :

- Gana Habib Adel.
- Ghorma Mohamed.

Présenter devant le jury composé de :

- Président : Mr. Mordi Nasredine MCB
- Examineur : Mr. Mehidi Abdallah MCB
- Encadreur : Mr. Daoud Mohamed MCA

Promotion : 2025 - 2026.

Remerciement

Ce présent travail est le fruit de longues années d'études et d'efforts.

*Tout d'abord, nous remercions **Allah** qui nous a donné la force, la patience et la volonté pour mener à bien ce travail.*

*Nous adressons nos sincères remerciements à **nos enseignants** qui ont contribué à notre formation tout au long de notre parcours universitaire.*

*Nos remerciements les plus sincères s'adressent également à **notre encadreur Mr. Daoud Mohamed** pour ses encouragements, son soutien et ses orientations qui ont été d'une grande aide dans l'accomplissement de ce travail.*

*Nous remercions chaleureusement notre **membres du jury Mr. Mordi Nasredine et Mr. Mehidi Abdallah** qui ont accepté d'évaluer et de juger notre mémoire de fin d'études.*

*Nous adressons aussi nos sincères remerciements à **toute personne ayant contribué, de près, de loin**, à la réalisation de ce modeste travail.*

*Nous tenons à exprimer notre sincère gratitude au **Laboratoire de recherche Signaux et Systèmes (LSS)** de l'Université de Mostaganem et à tous ses membres pour l'aide qu'ils nous ont apportée en termes d'équipement, d'espaces de travail, de commodités, de conseils, d'orientation et de suivi jusqu'à l'achèvement de notre projet de fin d'études master.*

*Enfin, nous exprimons notre **profonde reconnaissance à nos chers parents**, pour leur soutien moral, leurs sacrifices, leurs encouragements et leur confiance tout au long de notre parcours académique.*

Adel & Mohamed

Table des matières :

Remerciements	2
Table des matières :	3
Résumé :	6
Abstract:	7
Liste des figures.....	9
Liste des tableaux.....	10
Liste des abréviations.....	11
Introduction générale :	12

Chapitre I: Eléments du base en cardiologie

I. Introduction :	14
I.1. Anatomie et fonctionnement du cœur :	14
I.1.1. Structure :	14
I.1.2. Fonctionnement :	15
I.2. Les arythmies cardiaques :	16
I.2.1. Définition de l'arythmie :	16
I.2.2. Causes et facteurs de risque :	16
I.2.3. Types d'arythmies :	17
I.3. Les signaux physiologiques :	17
I.4. L'électrocardiographie :	17
I.4.1. Électrocardiographie :	17
I.4.2. Électrocardiogramme (ECG) :	17
I.4.3. Types d'électrocardiogrammes :	18
I.4.4. Déroulement de ECG :	18
I.4.5. Les ondes du signal ECG :	19
I.4.6. Les segments et intervalles d'un ECG normal :	19
I.5. Conclusion :	20

Chapitre II: Intelligence artificielle:

II. Introduction :	22
II.1. IA (intelligence artificielle) :	22
II.1.1. Définition :	22
II.1.2. Éléments fondamentaux de IA :	22
II.2. Machine learning:	23
II.2.1. Fonctionnement du machine learning :	23
II.2.1.1. Acquisition des données :	23
II.2.1.2. Préparation des données pour l'analyse :	23
II.2.1.3. Evaluer les performances et la précision du modèle :	23
II.2.1.4. Réglage les paramètres de modèles :	23
II.2.1.5. Lancement du modèle :	24
II.2.2. Types de machine learning :	24
II.2.2.1. Apprentissage supervisé :	24
II.2.2.2. Apprentissage non supervisé :	25
II.2.2.3. Apprentissage par renforcement :	25
II.2.3. Algorithmes de machine learning :	26

II.2.3.1. Réseaux de neurones (RN) :	26
II.2.3.2. Vecteurs de support (SVM) :	27
II.2.3.3. K plus proches voisins (KNN) :	28
II.3. Deep learning :	29
II.3.1. Definition et fonctionemet :	29
II.3.2. Les différents algorithmes de Deep Learning :	29
II.3.1.1. Réseaux neuronaux convolutifs (CNN):	29
II.3.1.2. Réseaux neuronaux récurrents (RNN) :	30
II.3.1.3. Réseaux à fonction de base radiale (RBFN) :	31
II.4. Conclusion :	31

Chapitre III: Conception du système proposé:

III.Introduction :	33
III.1. Acquisition des signaux ECG:	33
III.1.1. Bases de données ECG :	33
III.2. Prétraitement du signal :	34
III.2.1. Suppression du bruit :	34
III.2.2. Filtrage du signal ECG :	35
III.2.3. Normalisation du signal :	36
III.2.4. Segmentation du signal ECG:	36
III.3. Extraction des caractéristiques morphologiques :	37
III.4. Classification des signaux :	37
III.5. Conclusion:	37

Chapitre IV: Résultats et discussion:

IV.Introduction :	39
IV.1. Environnement de travail:	39
IV.2. Architecture globale de systeme:	40
IV.3. Base de données ECG:	41
IV.3.1. Apprentissage et test :	41
IV.4. Prétraitement des signaux ECG :	41
IV.4.1. Signal ECG original :	41
IV.4.2. Filtrage du signal ECG :	42
IV.4.3. Normalisation du signal ECG :	42
IV.4.4. Segmentation des battements cardiaques :	43
IV.5. Méthodes d'évaluation :	43
IV.5.1. Matrice de confusion :	43
IV.5.2. La structure de la matrice de confusion :	43
IV.5.3. Accuracy :	44
IV.5.4. Précision :	44
IV.5.5. Recall :	44
IV.5.6. F1 score :	44
IV.6. Etude des performances :	45
IV.6.1. Réseaux de neurones (RN) :	45
Architecture du modèle :	45
IV.6.1.1. Etude de l'accuracy pour (N=200) :	46
IV.6.1.2. Etude de l'accuracy pour (N=400) :	48
IV.6.1.3. Etude de l'accuracy pour (N=600) :	48
IV.6.2. Vecteurs de support (SVM) :	49

IV.6.2.1. SVM Linéaire:.....	50
Pour N=200	50
Pour N=500:.....	51
Pour N=800.....	51
IV.6.2.2. SVM RBF:	52
Pour N=200:	52
Pour N=500:.....	53
Pour N=800.....	54
IV.6.3. CNN :	55
Scenario 1	55
Scenario 2.....	56
Scenario 3.....	56
IV.7. Conclusion :	57
Conclusion générale.....	58
Références	59

Résumé :

Les maladies cardiovasculaires sont l'une des principales causes de mortalité au monde. Parmi ceux-ci, les arythmies cardiaques, constituent un problème important pouvant causer des complications graves si elles ne sont pas détectées à temps. L'électrocardiogramme (ECG) est un signal physiologique largement utilisé pour l'analyse d'activité électrique du cœur et détecter les anomalies du rythme cardiaque.

Toutefois, l'analyse manuelle de ces signaux peut être complexe et nécessite une expertise médicale. Avec le développement des technologies de l'intelligence artificielle, les techniques de Machine Learning et deep learning offrent plusieurs solutions efficaces pour faciliter l'analyse des signaux biomédicaux (ECG).

Dans ce travail, nous proposons un système basé sur le Machine Learning et Deep Learning pour l'analyse et la classification des signaux ECG afin de détecter et classer les arythmies cardiaques. Le système comprend plusieurs étapes comme: prétraitement du signal pour réduire le bruit, l'extraction des caractéristiques morphologiques et l'utilisation d'un algorithme de classification pour identifier les différents types d'arythmies cardiaques.

Les résultats que nous obtenons montrent que l'utilisation de ces techniques permet d'améliorer l'Accuracy) de la détection automatique des arythmies cardiaques et peut constituer un outil d'aide au diagnostic médical.

Mots-clés: ECG, Arythmie cardiaque, Signal, Machine Learning, RN, Cœur, accuracy, Classification.

Abstract:

Cardiovascular diseases are one of the leading causes of death worldwide. Among them are cardiac arrhythmias, which pose a significant problem and can lead to serious complications if they are not detected early. The electrocardiogram (ECG) is a physiological signal widely used to analyze the heart's electrical activity and detect heart rhythm abnormalities.

However, manual analysis of these signals can be complex and requires medical expertise. With the development of artificial intelligence technologies, machine learning and deep learning techniques offer several effective solutions to facilitate the analysis of biomedical signals (ECGs).

In this work, we propose a system based on machine learning, including neural networks and support vector machines (SVMs), as well as deep learning methods, for the analysis and classification of ECG signals to detect and classify cardiac arrhythmia. The system comprises several steps, including signal preprocessing to reduce noise, extraction of morphological features, and the use of a classification algorithm to identify different types of cardiac arrhythmia.

Our results show that the use of these machine learning techniques improves the accuracy of automatic cardiac arrhythmia detection and can serve as a valuable tool for medical diagnosis.

Keywords: ECG, Cardiac arrhythmia, Signal, Machine Learning, RN, Heart, accuracy, Classification.

ملخص:

تُعدّ أمراض القلب والأوعية الدموية من بين الأسباب الرئيسية للوفيات على مستوى العالم. ومن بين هذه الأمراض، تُعتبر اضطرابات نظم القلب (Arrhythmias) من المشكلات الصحية المهمة التي قد تؤدي إلى مضاعفات خطيرة إذا لم يتم اكتشافها وتشخيصها في الوقت المناسب.

يُعدّ تخطيط القلب الكهربائي (ECG) إشارةً فسيولوجيةً تُستخدم على نطاق واسع لتحليل النشاط الكهربائي للقلب وتشخيص اضطرابات النظم القلبية. ومع ذلك، فإن التحليل اليدوي لإشارات تخطيط القلب قد يكون معقدًا ويتطلب خبرة طبية متخصصة.

ومع التطور المتسارع لتقنيات الذكاء الاصطناعي، أصبحت أساليب التعلم الآلي (Machine Learning) والتعلم العميق (Deep Learning) توفر حلولاً فعالة لأتمتة تحليل الإشارات الطبية الحيوية وتحسين دقة التشخيص.

في هذا العمل، نقترح نظامًا يعتمد على تقنيات التعلم الآلي والتعلم العميق لتحليل وتصنيف إشارات تخطيط القلب الكهربائي بهدف الكشف عن اضطرابات نظم القلب. ويتضمن النظام عدة مراحل أساسية، تتمثل في المعالجة المسبقة للإشارة للحد من الضوضاء والتشويش، واستخراج الخصائص المهمة، ثم استخدام خوارزميات التصنيف للتعرف على الأنواع المختلفة لاضطرابات النظم القلبية. أظهرت النتائج المتحصل عليها أن استخدام تقنيات التعلم الآلي والتعلم العميق يساهم في تحسين دقة الكشف الآلي عن الاضطرابات القلبية، مما يجعله أداةً واعدة لدعم التشخيص الطبي ومساعدة المختصين في اتخاذ القرارات السريري.

الكلمات المفتاحية: تخطيط كهربية القلب، اضطراب نظم القلب، الإشارة، التعلم الآلي، ممرض مسجل، القلب، الدقة، التصنيف.

Liste des figures :

Figure I.1 : Schéma général du cœur.	14
Figure I.2 : Circulation sanguine dans le cœur.	15
Figure I.3 : Arythmie cardiaque.	16
Figure I.4 : Signal ECG.	18
Figure I.5 : Les ondes du signal ECG.	19
Figure I.6 : Segments et intervalles d'un ECG normal.	20
Figure II.1: L'apprentissage supervisé.	24
Figure II.2: L'apprentissage non supervisé.	25
Figure II.3: L'apprentissage par renforcement.	26
Figure II.4: Réseaux de neurones.	27
Figure II.5: Support Vector Machines (SVM).	28
Figure II.6: k plus proches voisins (KNN).	28
Figure II.7: difference entre ML et DL.	29
Figure II.8: Fonctionnement du CNN.	30
Figure III.1: Les annotations des battements cardiaques de la base de donnée MIT-BIH.	34
Figure III.2 : Bruit d'un signal.	35
Figure III.3 : Exemple du filtrage.	35
Figure III.4 : Méthode de segmentation.	36
Figure IV.1 : Architecture globale de system.	40
Figure IV.2: Signal ECG originale.	41
Figure IV.3: Signal ECG filtré.	42
Figure IV.4: Signal ECG normalisé.	42
Figure IV.5: Segmentation des battements ECG.	43
Figure IV.6 : Matrice de confusion.	44
Figure IV.7 : Matrice de confusion et l'histogramme pour N=200 M=10.	46
Figure IV.8 : Matrice de confusion et l'histogramme pour N=200 M=20.	46
Figure IV.9 : Matrice de confusion et l'histogramme pour N=200 M=30.	47
Figure IV.10 : L'evolution de l'accuracy globale en fonction de N et M.	49
Figure IV.11 : Martice de confusion pour C= 0.1 ,1, 10.	50
Figure IV.12: Martice de confusion pour C1= 1 G=0.01, 0.1, 1.	52
Figure IV.13 : Rapport de classification Scenario 1.	55
Figure IV.14 : Rapport de classification Scenario 2.	56
Figure IV.15 : Rapport de classification Scenario 3.	56

Liste des tableaux:

Tableau 1 : nombre totale des battements.....	41
Tableau 2 : nombre des battements de TEST.....	45
Tableau 3 : Architecture de RN.....	45
Tableau 4 : L'accuracy, le taux d'erreur, de matrice global pour N=200.....	47
Tableau 5 : L'accuracy, le taux d'erreur, de matrice global pour N=400.....	48
Tableau 6 : L'accuracy, le taux d'erreur, de matrice global pour N=600.....	48
Tableau 7 : L'accuracy, le taux d'erreur pour N=200.....	50
Tableau 8 : L'accuracy, le taux d'erreur pour N=500.....	51
Tableau 9 : L'accuracy, le taux d'erreur pour N=800.....	51
Tableau 10 : L'accuracy, le taux d'erreur pour N=200.....	53
Tableau 11 : : L'accuracy, le taux d'erreur pour N=500.....	53
Tableau 12 : L'accuracy, le taux d'erreur pour N=800.....	54
Tableau 13: Architecture de CNN.....	55
Tableau 14 : L'accuracy pour chaque class(S1).....	55
Tableau 15 : L'accuracy pour chaque class(S2).....	56
Tableau 16 : L'accuracy pour chaque class(S3).....	56

Liste des abréviations:

- **ECG** : Electrocardiogramme
- **MCV** : maladies cardiovasculaires
- **ML** : Machine learning
- **DL** : Deep learning
- **IA** : intelligence artificielle
- **RN** : réseaux de neurons
- **SVM** : Support Vector Machines
- **KNN** : K plus proches voisins
- **CNN** : Réseaux neuronaux convolutifs
- **RNN** : Réseaux neuronaux récurrents
- **RBFN** : Réseaux à fonction de base radiale
- **TP** : True Positif
- **TN** : True Negatif
- **FP** : False Positif
- **FN** : False Negatif

Introduction générale :

Les maladies cardiovasculaires représentent l'une des principales causes de mortalité dans le monde selon **l'organisation mondiale de la santé** (OMS), des millions de personnes meurent chaque année à cause de troubles cardiaques. Parmi ces troubles, les arythmies cardiaques qui affectent le rythme normal du cœur et peuvent causer des complications graves si elles ne sont pas détectées à temps.

Le cœur est un organe essentiel du système circulatoire. Son fonctionnement repose sur une activité électrique qui permet la contraction et la relaxation des muscles cardiaques qui assurent la circulation du sang dans l'organisme. Cette activité électrique peut être enregistrée à l'aide d'un électrocardiogramme (ECG) qui est un outil médical largement utilisé pour analyser le fonctionnement du cœur et détecter différentes anomalies cardiaques.

Cependant, l'analyse manuelle des signaux ECG par les médecins peut être longue et complexe, notamment lorsque les données sont gros. Avec l'évolution des technologies et l'augmentation des données médicales, l'utilisation des techniques d'intelligence artificielle et de Machine Learning devient une solution efficace pour automatiser l'analyse des signaux biomédicaux.

Le ML permet aux systèmes informatiques d'apprendre à partir de données et d'identifier des modèles ou des arythmies dans ces données. Dans le domaine médical, ces techniques sont de plus en plus utilisées pour assister les médecins dans le diagnostic de différentes pathologies, notamment les maladies cardiovasculaires.

L'objectif principal de ce projet est de designer un système capable de traiter les signaux ECG et de réaliser un prétraitement pour réduire le bruit, d'extraire des caractéristiques importantes du signal, et finalement d'utiliser un algorithme de Machine Learning pour classifier les différents types d'arythmies.

Ce mémoire est organisé en quatre chapitres. Le premier chapitre présente les notions générales relatives au cœur, aux arythmies cardiaques et de signal Électrocardiographique. Le deuxième chapitre introduit les concepts fondamentaux du IA ainsi que les principaux algorithmes utilisés pour la classification. Le troisième chapitre décrit la conception du système proposé ainsi que les différentes étapes du traitement du signal ECG. Enfin, le quatrième chapitre présente les résultats obtenus ainsi que leur analyse et leur discussion.

Chapitre I:

Eléments du base en
cardiologie

I. Introduction :

Les maladies cardiovasculaires (MCV) représentent l'une des principales causes de mortalité dans le monde. Elles sont influencées par plusieurs facteurs de risque, tels que le diabète, l'obésité, le tabagisme et le manque d'activité physique. Le dépistage précoce joue un rôle crucial dans la réduction des complications et l'amélioration des soins aux patients [1].

Dans ce contexte, l'électrocardiogramme (ECG) constitue un outil fondamental pour l'analyse de l'activité électrique du cœur et la détection des arythmies cardiaques. L'utilisation des techniques de Machine Learning permet aujourd'hui de développer des systèmes intelligents capables d'assister les spécialistes dans la classification automatique des signaux ECG et l'identification précoce des arythmies cardiaques [2].

I.1. Anatomie et fonctionnement du cœur :

I.1.1. Structure :

Le cœur est un organe musculaire creux qui fonctionne comme une pompe et assure la circulation du sang à travers les artères et les veines. Il présente une forme proche de celle d'un cône inversé, avec une base orientée vers le haut et la droite, et un apex dirigé vers le bas et la gauche.

Situé dans le médiastin, qui correspond à la région centrale de la cage thoracique délimitée par les poumons, le sternum et la colonne vertébrale, le cœur est légèrement décalé vers la gauche du thorax. Chez l'adulte, son poids moyen est d'environ 300 grammes chez l'homme et 250 grammes chez la femme. Au repos, il est capable de propulser entre 4 et 5 litres de sang par minute.

Sur le plan anatomique, le cœur est constitué de quatre cavités: deux oreillettes et deux ventricules. Cette organisation permet d'assurer la propulsion du sang vers l'ensemble des cellules de l'organisme. Les cavités droites forment le cœur droit, tandis que les cavités gauches constituent le cœur gauche, chacun jouant un rôle spécifique dans la circulation sanguine [3].

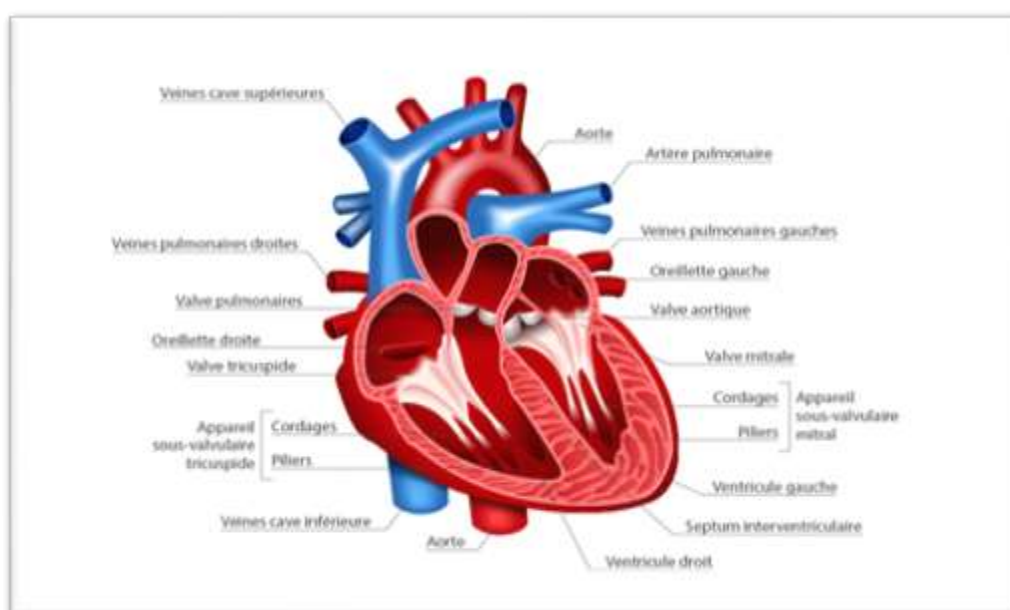


Figure I.1 : Schéma général du cœur.

I.1.2. Fonctionnement :

Pour faire circuler le sang dans le corps, le cœur se contracte et se dilate. Cette action de pompage s'illustre bien par l'alternance du serrement et du desserrement d'un poing. Chaque battement, le cœur expulse du sang dans les artères. C'est ce qui crée le pouls [4].

- D'abord, l'oreillette droite se remplit du sang appauvri en oxygène provenant du corps (des muscles, des organes, du cerveau et même du cœur). Lorsque l'oreillette est pleine, elle se contracte. À la contraction, la valvule tricuspide reliant l'oreillette droite et le ventricule droit s'ouvre. Le sang entre alors dans le ventricule droit.
- Lorsque le ventricule droit est plein, il se contracte à son tour pour pousser le sang dans les poumons par la valvule pulmonaire.
- Les poumons remplacent le dioxyde de carbone présent dans le sang par de l'oxygène. Le sang, maintenant oxygéné, est expulsé vers l'oreillette gauche.
- Lorsque l'oreillette gauche se contracte, la valvule mitrale reliant cette dernière au ventricule gauche s'ouvre. Ainsi, le sang pénètre dans ce ventricule.
- Le ventricule gauche expulse le sang oxygéné par la valvule aortique vers l'aorte, qui alimente le reste du corps cœur.
- Le sang riche en oxygène circule dans tout le corps. Ensuite, les veines ramènent le sang pauvre en oxygène à l'oreillette droite, qui s'en remplit, et le cycle recommence.

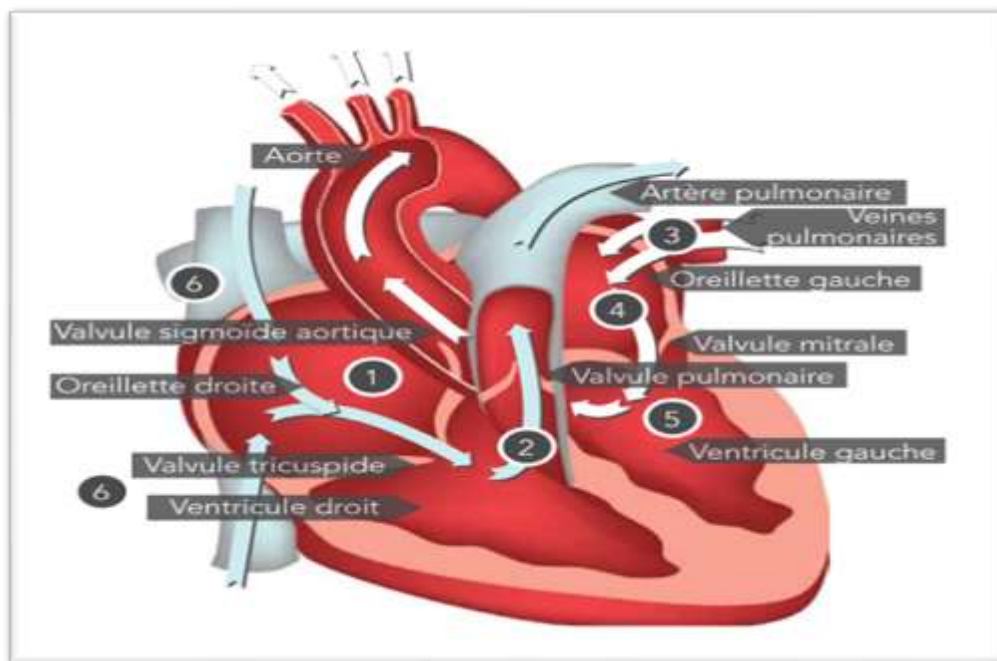


Figure I.2 : Circulation sanguine dans le cœur.

I.2. Les arythmies cardiaques :

I.2.1. Définition de l'arythmie :

Un changement du rythme habituel du muscle vital définit l'arythmie cardiaque. De nombreux observateurs disent que l'arythmie cardiaque se montre de diverses manières parce que le muscle vital pulse avec une rapidité excessive ou une lenteur démesurée. L'arythmie cardiaque est souvent identifiée comme un changement de mouvement.

Quand le corps se repose, le rythme normal du muscle vital reste entre 50 et 60 pulsations par minute. Une anomalie de la répétition des battements est provoquée par la présence de l'arythmie cardiaque, ce qui montre une disparition du mouvement normal. Les experts en santé affirment que l'arythmie cardiaque possède plusieurs catégories comme la bradycardie, la tachycardie ou la fibrillation ventriculaire. Cette maladie médicale devient plus courante aujourd'hui parce que le processus de l'âge touche un grand nombre de citoyens [5].

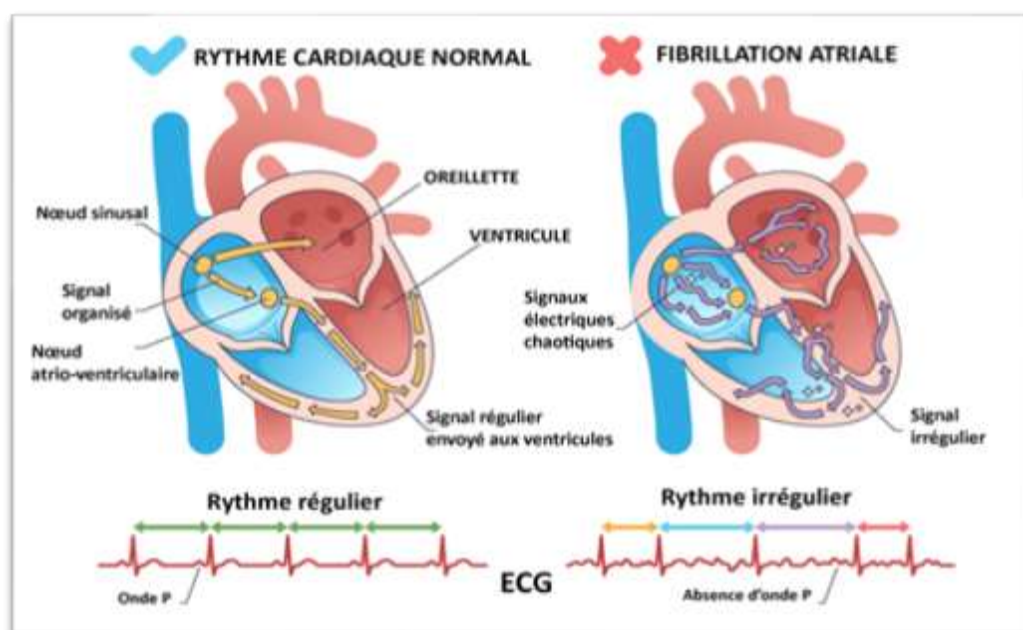


Figure I.3 : Arythmie cardiaque.

I.2.2. Causes et facteurs de risque :

Les causes de l'arythmie cardiaque sont nombreuses. Parmi ceux-ci, on trouve :

- L'âge de la personne concernée, car l'arythmie cardiaque touche des patients âgés dans la plupart des cas.
- Crise cardiaque antérieure, qui a fragilisé le cœur, et suite à laquelle des séquelles peuvent apparaître.
- Le diabète.
- Les troubles endocriniens, comme un dysfonctionnement des glandes thyroïdes.
- La consommation excessive de certaines substances, comme le tabac, la caféine, l'alcool ou les drogues.
- Le stress [5].

I.2.3. Types d'arythmies :

- **V-Battement Ventriculaire Prématuré** : C'est une extrasystole prenant naissance dans les ventricules, caractérisée sur l'ECG par un complexe QRS large, déformé et anormal. Elle traduit une activation électrique précoce et ectopique du muscle cardiaque inférieur, perturbant temporairement le rythme sinusal régulier.
- **L-Bloc de Branche Gauche** : C'est un trouble de conduction intraventriculaire où le signal électrique est ralenti ou bloqué au niveau de la branche gauche du faisceau de His. Sur l'ECG, cela se traduit par un élargissement du complexe QRS avec une morphologie typiquement élargie ou encochée.
- **R-Bloc de Branche Droit** : C'est une anomalie de conduction où la propagation de l'influx électrique vers le ventricule droit est retardée par un blocage de la branche droite.
- **A-Battement Auriculaire Prématuré** : C'est une contraction cardiaque précoce qui trouve son origine dans les oreillettes, en dehors du nœud sinusal normal. Elle se manifeste graphiquement par une onde P prématurée de morphologie altérée, généralement suivie d'un complexe QRS d'aspect normal.

I.3. Les signaux physiologiques :

On les appelle aussi bio-sinaux (en anglais bio-signals) et ils sont des enregistrements spatiaux (images), temporels (c'est-à-dire un signal évoluant dans le temps) ou bien spatio-temporels de phénomènes biologiques tels que les battements de cœur ou les contractions musculaires ou bien encore l'activité cérébrale [6].

I.4. L'électrocardiographie :

I.4.1. Électrocardiographie :

Un électrocardiographe est un appareil qui mesure et enregistre l'activité électrique du cœur :

- L'enregistrement ainsi obtenu s'appelle électrocardiogramme (souvent noté ECG).
- Willem Einthoven inventa cet appareil en 1895 (Prix Nobel en 1924).
- L'électrocardiogramme est aujourd'hui l'examen le plus fréquent pour étudier le fonctionnement du et détecter les troubles de coeur cardiaques.

I.4.2. Électrocardiogramme (ECG) :

L'électrocardiogramme (ECG) est un examen qui permet d'étudier le fonctionnement du cœur par mesurant l'activité électrique. A chaque battement du cœur, une impulsion électrique (onde) le traverse. Cette onde provoque la contraction du muscle cardiaque jusqu'à ce qu'il expulse le sang hors du cœur.

Un ECG est donc un appareil de mesure qui enregistre l'activité électrique d'un médecin peut dire si l'activité électrique ainsi observée est normale ou anormale [7].

L'ECG est un examen essentiel en cardiologie. Il est peut-être prescrit comme suit :

- À l'occasion d'un bilan médical avant une intervention chirurgicale.
- Lorsqu'un patient se plaint de palpitations ou de douleurs à la poitrine.
- En cas d'urgence (arrêt cardiaque, suspicion d'infarctus du myocarde).
- Pour détecter des arythmies cardiaques

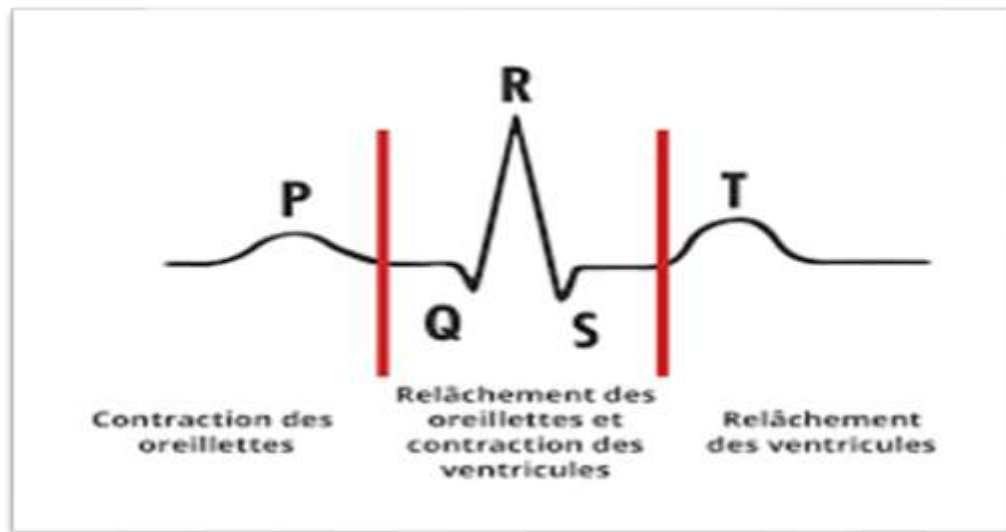


Figure I.4 : Signal ECG.

I.4.3. Types d'électrocardiogrammes :

- **L'électrocardiogramme de repos** est le cas le plus fréquent en pratique :
 - Il se pratique au cours d'une consultation médicale avec le patient allongé sur le dos.
 - Il peut aussi être réalisé en cas d'intervention d'urgence par les pompiers à l'aide d'appareils portatifs.
- **L'électrocardiogramme en continu pendant un effort** est enregistré pendant toute la durée d'un effort physique du patient comme un vélo ou sur un tapis roulant.
 - Cet ECG est souvent appelé le test à l'effort.
 - L'électrocardiogramme de stress est réalisé si le patient se plaint de palpitations ou de douleurs thoraciques alors que l'électrocardiogramme de repos est normal.

I.4.4. Déroulement de ECG :

L'examen peut être réalisé au cabinet du cardiologue, à l'hôpital et parfois à domicile. Pour effectuer l'électrocardiogramme, le médecin applique sur la peau une dizaine de petites électrodes disposées au niveau des bras, des jambes et de la poitrine. Pour une meilleure adhérence des électrodes, il peut être nécessaire de raser quelques zones de la peau. Puis l'examen se déroule de la manière suivante selon le type d'ECG :

- **L'enregistrement de l'électrocardiogramme de repos dure entre 5 et 10 minutes**
- **L'électrocardiogramme en continu pendant un effort dure entre 10 et 30 minutes**

I.4.5. Les ondes du signal ECG :

L'ECG est représenté par des lettres « P, Q, R, S et T », elles correspondent à l'activation électrique des différentes parties du myocarde qui est formé de plusieurs ondes, le signal représente la morphologie et l'amplitude de ces ondes mais la durée change selon les dérivations ECG utilisées :

- **Onde P** : c'est la première onde qui on peut détecter et elle correspond à la dépolarisation de l'oreillette. C'est une petite déviation positive avant le complexe QRS.
- **L'onde Q** : C'est la première déviation initiale vers le bas ou « négative ».
- **L'onde R** : Aussi appelé le Pic R. C'est ensuite la déviation suivante vers le haut (max onde)
- **L'onde S** : C'est ensuite la déviation suivante vers le bas, à condition qu'elle franchisse la ligne isoélectrique pour devenir brièvement négative avant de revenir à la ligne de base isoélectrique.
- **L'onde T** : Elle représente la repolarisation des ventricules. L'onde T normale a une amplitude plus faible que celle du complexe QRS [8].

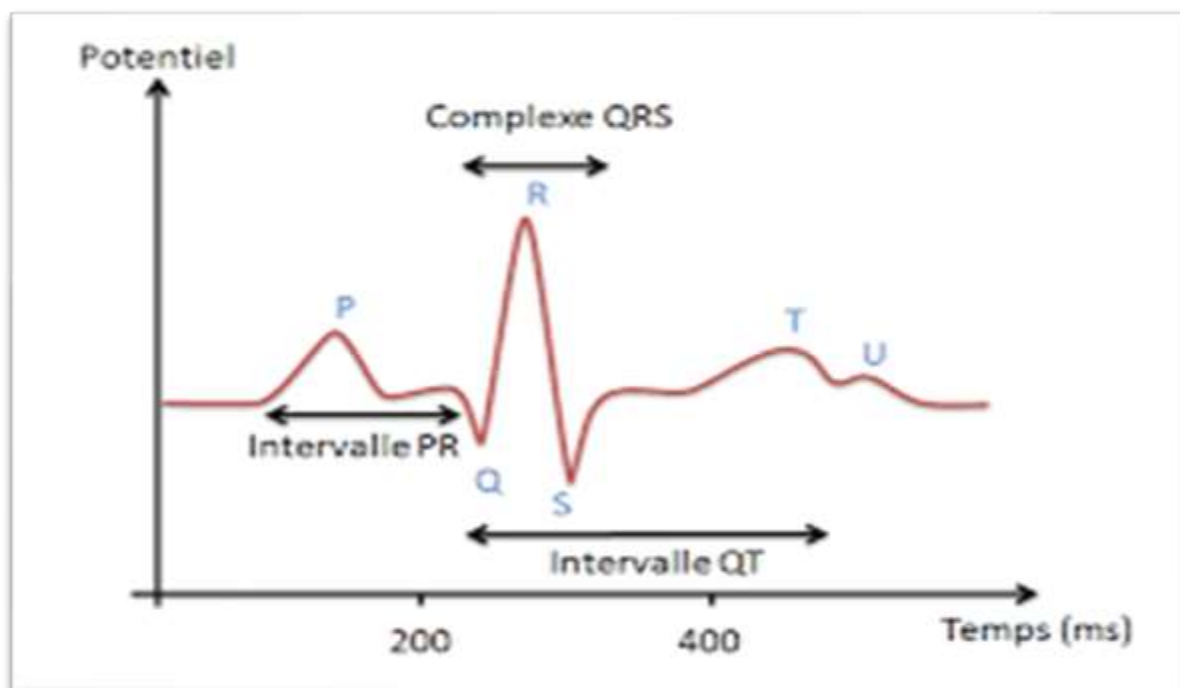


Figure I.5 : Les ondes du signal ECG.

I.4.6. Les segments et intervalles d'un ECG normal :

Outre les différentes ondes qui constituent les paramètres de base pour une bonne caractérisation d'un signal ECG, il existe tout un ensemble d'intervalles et de segments qui apportent des informations très utiles sur la vitesse de conduction de l'impulsion électrique dans les différentes parties du cœur.

Les intervalles et les segments les plus importants sont :

- **Intervalle RR** : C'est le temps qui s'écoule entre deux dépolarisations des ventricules. C'est grâce à cet intervalle que l'on peut calculer la fréquence cardiaque.
- **Segment PR** : (pause du nœud AV) Le segment PR correspond au délai entre la fin de la dépolarisation des oreillettes et le début de celle des ventricules. C'est le temps pendant lequel l'onde de dépolarisation est bloquée au niveau du nœud AV.
- **Intervalle PR** : (durée de la conduction auriculo-ventriculaire) L'intervalle PR correspond au temps que met l'onde de dépolarisation pour aller du nœud sinusal jusqu'aux cellules myocardiques ventriculaires.
- **Intervalle QT** : (durée de systole ventriculaire) Cet intervalle correspond au temps de systole ventriculaire, qui va du début de l'excitation des ventricules jusqu'à la fin de leur relaxation.
- **Segment ST** : (durée de stimulation complète des ventricules) Le segment ST correspond à la phase pendant laquelle les cellules ventriculaires sont toutes dépolarisées, le segment est alors isoélectrique [9].

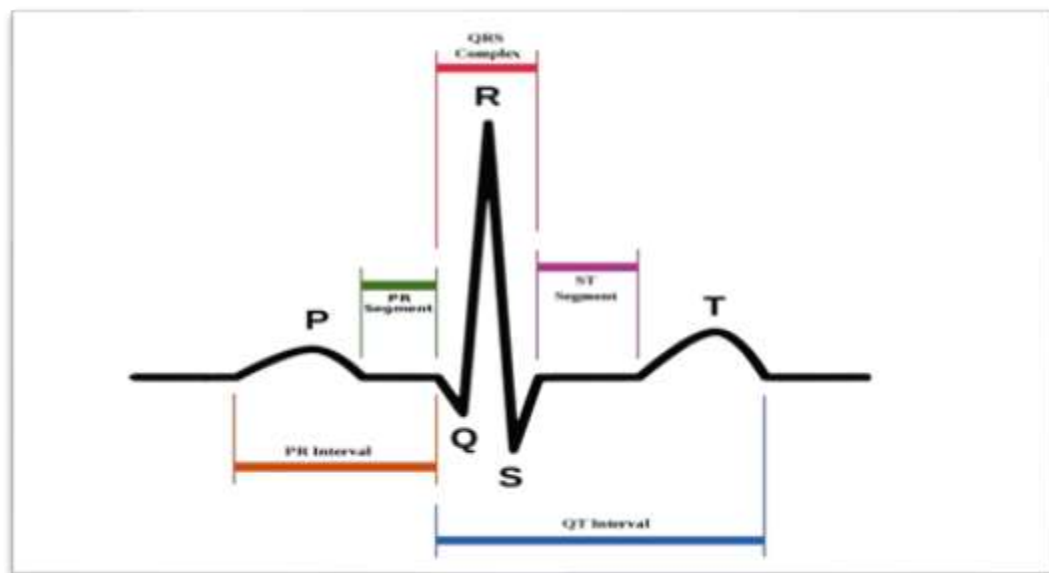


Figure I.6 : Segments et intervalles d'un ECG normal.

I.5. Conclusin :

Ce chapitre a présenté les notions fondamentales liées à l'électrocardiogramme (ECG) et à son importance dans l'analyse de l'activité électrique du cœur.

Les différentes composants(les ondes, les intervalles) du signal ECG ainsi que les principaux types de rythmes cardiaques et d'arythmies ont été étudiés et présente.

Cette analyse a permis de mieux comprendre les caractéristiques des signaux cardiaques et les anomalies pouvant affecter leur fonctionnement normal. Les connaissances présentées dans ce chapitre constituent une base essentielle pour le développement des méthodes automatiques de classification des signaux ECG dans les chapitres suivants.

Chapitre II:

Intelligence artificielle

II. Introduction :

Le recueil numérisé de l'activité électrique myocardique autorise son analyse par l'intelligence artificielle.

L'intelligence artificielle est utilisée dans le traitement des arythmies pour identifier les zones de fibrose ventriculaire ou la fibrose induite par l'arythmie, dans le but d'améliorer la qualité de l'ablation de la fibrillation auriculaire. Les technologies de l'information peuvent guider le processus d'ablation et évaluer le diagnostic en créant une chambre ventriculaire virtuelle personnalisée à partir de données d'imagerie.

Lors des morts subites, c'est moins l'analyse des données de l'activité électrique myocardique et de la délivrance du choc électrique salvateur que celle des données de santé de la population dans son ensemble, qui pourraient contribuer à identifier les meilleurs prédicteurs de risque. Les données recueillies lors d'exams électrocardiographiques banaux ou par les objets portables, dont les « montres connectées », permettent grâce à l'IA de prédire la survenue ultérieure ou la récurrence de troubles du rythme [10].

II.1. IA (intelligence artificielle) :

II.1.1. Définition :

L'intelligence artificielle est une discipline scientifique qui était concentrée sur la création de machines capables de réaliser des tâches nécessitant une forme d'intelligence humaine. Elle est définie par Marvin Minsky comme suit : « *la science qui consiste à faire aux machines des choses qui nécessiteraient de l'intelligence si elles étaient réalisées par des hommes* ».

Elle se divise principalement en deux branches: **le Machine Learning** et **le Deep Learning**. Le Machine Learning permet aux machines d'apprendre à partir de données et d'améliorer leurs performances. Mais le Deep Learning s'inspire du fonctionnement du cerveau humain à travers des réseaux de neurones artificiels pour traiter de grandes quantités d'informations [11].

II.1.2. Éléments fondamentaux de IA :

L'IA repose sur trois éléments fondamentaux:

- **Les données:** Les systèmes d'IA apprennent grâce aux données et ont besoin de grands volumes d'informations pour être entraînés efficacement. On répartit généralement les données en trois ensembles: entraînement, validation et test. L'utilisation de jeux de données variés facilite la généralisation de l'apprentissage du modèle, et lui permet d'avoir de meilleures performances sur de nouvelles données [12].
- **Les algorithmes:** Ce sont des ensembles de règles qui permettent aux systèmes d'IA de traiter les données et d'en déduire des décisions. Ils emportent le machine learning, qui apprend depuis les données sans un codage explicite, et le deep learning, fondé sur des réseaux neuronaux multicouches. Une autre méthode est l'apprentissage par renforcement, dans lequel un agent affine son comportement au fil du temps à partir d'un système de récompenses et de pénalités [13].
- **La puissance de calcul:** Les algorithmes d'intelligence artificielle nécessitent souvent d'importantes

ressources de calcul pour traiter d'énormes quantités de données et exécuter des algorithmes complexes, notamment dans le cas de l'apprentissage profond. De nombreuses organisations s'appuient sur du matériel spécialisé, comme les unités de traitement graphique (GPU), pour optimiser ces processus [14].

II.2. Machine learning:

Le Machine Learning est une forme d'intelligence artificielle qui enseigne aux ordinateurs à apprendre des données et à s'améliorer avec l'expérience et l'étude. Ce qui signifie que les ordinateurs s'améliorent dans leurs tâches en identifiant des modèles plutôt qu'en suivant des règles fixes. Un modèle de ML ne se base pas sur des instructions prédéfinies mais améliore ses performances grâce à l'exposition à de nouvelles données, de la même manière que les humains apprennent de leurs expériences.

Les entreprises contemporaines appliquent le ML afin d'identifier les fraudes, d'anticiper la demande et de personnaliser les suggestions. Ces systèmes adaptatifs progressent constamment grâce au feedback, ce qui renforce la précision (accuracy) et l'efficacité des processus dans tous les secteurs [15].

II.2.1. Fonctionnement du machine learning :

Le ML fonctionne par créant des modèles statistiques informatiques affinés pour un but donné en évaluant les données d'entraînement, plutôt que par l'approche classique où les programmeurs développent un algorithme statique qui tente de résoudre un problème. Les data specialists ajustent le modèle grâce à une série de variables établies, appelées hyperparamètres, et à des variables ajustées algorithmiquement, appelées paramètres d'apprentissage, au fur et à mesure que les ensembles de données passent par le modèle de ML et que la sortie obtenue est jugée sur la précision. Il déviser sur plusieurs étapes:

II.2.1.1. Acquisition des données :

L'entraînement des modèles ML repose sur la disponibilité de volumes importants des données. Ces données constituent le fondement essentiel sur lequel repose le système de machine learning.

II.2.1.2. Préparation des données pour l'analyse :

Dans la majorité des cas, les données brutes ne sont pas directement exploitables. Une phase de préparation est donc nécessaire pour assurer leur qualité et cohérence. Elle comprend le nettoyage des données, la suppression des doublons et le traitement des valeurs aberrantes.

II.2.1.3. Evaluer les performances et la précision du modèle :

Après la préparation, il est nécessaire d'évaluer le modèle à l'aide de données inédites. Ces données de test sont souvent extraites du jeu initial, mais conservées séparément afin de garantir une évaluation objective. Cette étape permet de mesurer la capacité du modèle à généraliser ses prédictions.

II.2.1.4. Reglage les paramètres de modèles :

À ce stade, le modèle est proche de sa phase de déploiement. Des ajustements supplémentaires peuvent être réalisés pour améliorer les performances, notamment à travers l'optimisation des hyperparamètres ou l'intégration de données plus spécifiques au contexte d'application.

II.2.1.5. Lancement du modèle :

Grâce à l'optimisation des résultats, le modèle est prêt pour traiter les données inédites dans une utilisation normale de la production. Lorsque le modèle est actif, les équipes de projet collectent des données sur le fonctionnement du modèle dans des scénarios réels. Cela peut être fait en surveillant des mesures de performances, telles que la précision des prédictions du modèle [16].

II.2.2. Types de machine learning :

Il existe trois principaux types de machine learning. Chacun a ses propres forces et limites, c'est pourquoi il est important de choisir la bonne approche pour la tâche spécifique à accomplir.

II.2.2.1. Apprentissage supervisé :

Dans cette apprentissage, l'algorithme est entraîné à partir d'un ensemble de données préalablement étiquetées, comprenant à la fois les variables d'entrée et les sorties attendues. L'objectif est d'établir une relation fonctionnelle entre ces entrées et les résultats correspondants. Pour ce faire, le modèle compare ses prédictions aux valeurs réelles, puis ajuste progressivement ses paramètres afin d'améliorer sa précision.

C'est l'une des méthodes les plus utilisées en Machine Learning. Il permet de prédire la réponse attendue à partir d'exemples annotés d'entrées correspondantes. Cette méthode est super populaire pour des tâches comme classification, détection et prédiction. Parmi les principales techniques on trouve la régression, les machines à vecteurs de support (SVM), et les réseaux de neurones profonds [17].

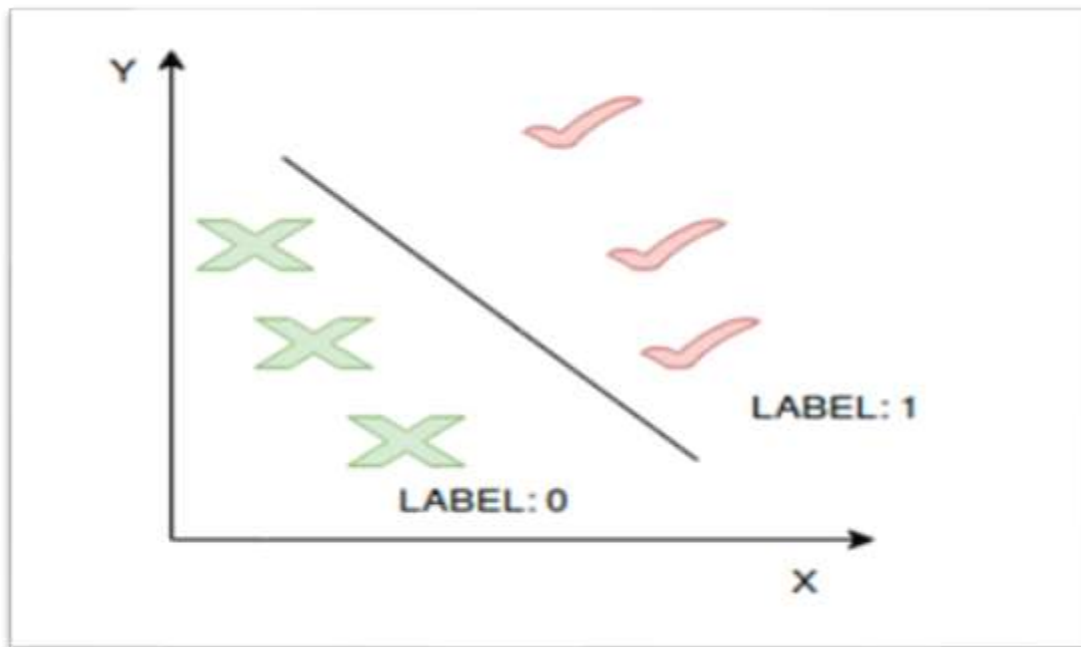


Figure II.1: L'apprentissage supervisé.

II.2.2.2. Apprentissage non supervisé :

Comme son nom l'indique, l'apprentissage non supervisé utilise des algorithmes d'auto-apprentissage, c'est-à-dire qu'il apprend sans étiquettes ni entraînement préalable. Le modèle reçoit des données brutes, et doit inférer ses propres règles et structurer les informations en fonction des similitudes, des différences et des modèles, sans instructions explicites sur la façon de travailler avec chaque élément de données.

Ce type d'apprentissage est particulièrement adapté aux tâches complexes impliquant de grands volumes de données. Il est couramment utilisé pour regrouper des observations en ensembles homogènes ou pour révéler des schémas cachés difficilement détectables par des autres méthodes. Il permet ainsi d'identifier des caractéristiques facilitant l'interprétation et la structuration des données.

À titre d'illustration, considérons un vaste ensemble de données. Un algorithme d'apprentissage non supervisé peut analyser ces informations et regrouper automatiquement les observations selon des critères implicites, tels que la température ou des conditions climatiques similaires. Ces regroupements émergent sans connaissance préalable des catégories [18].

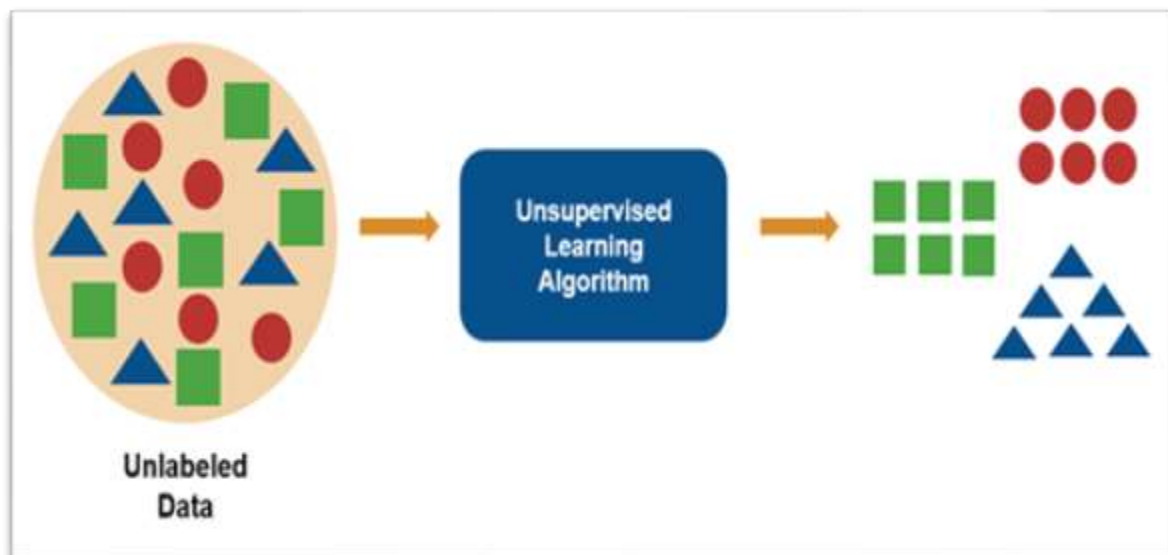


Figure II.2: L'apprentissage non supervisé.

II.2.2.3. Apprentissage par renforcement :

L'apprentissage par renforcement constitue une approche du ML dans laquelle un agent apprend à prendre des décisions à travers un processus itératif basé sur l'essai et l'erreur. Contrairement à l'apprentissage supervisé, cette méthode ne nécessite pas de données étiquetées ni d'intervention humaine directe durant la phase d'apprentissage. L'agent interagit avec son environnement, teste différentes actions et reçoit en retour un signal de récompense ou de pénalité, lui permettant d'évaluer la pertinence de ses choix.

Au fil des interactions, l'algorithme développe une stratégie optimale en maximisant les récompenses cumulées. Ce mécanisme permet aux systèmes intelligents, notamment les robots, de s'adapter à des situations nouvelles ou imprévisibles et d'identifier les actions les plus efficaces sans supervision explicite [19].

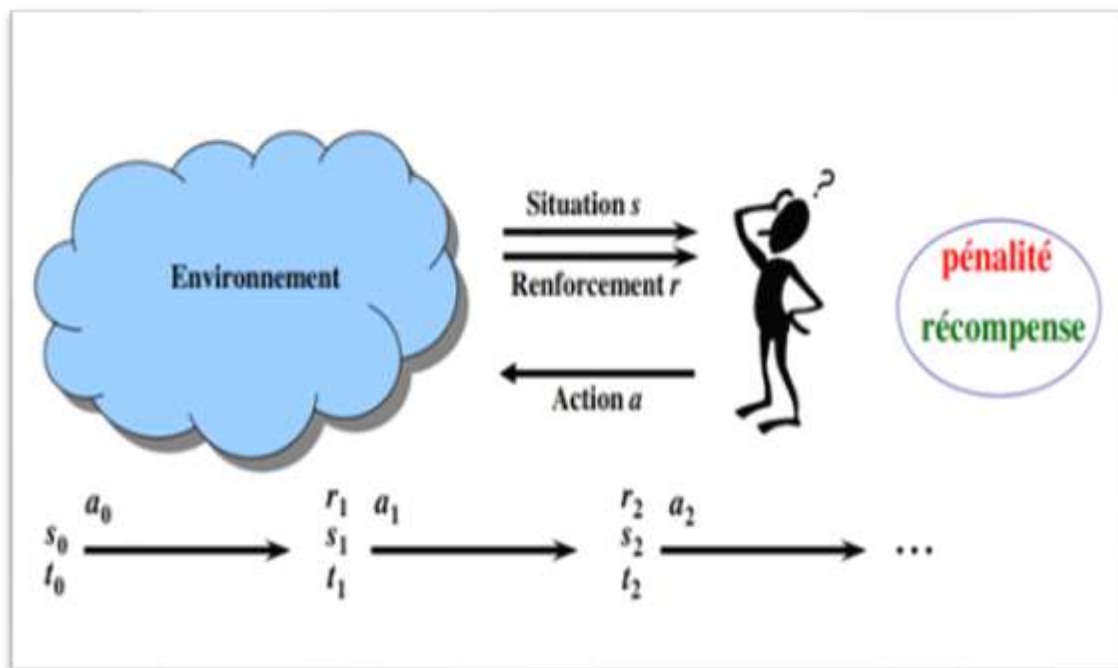


Figure II.3: L'apprentissage par renforcement.

II.2.3. Algorithmes de machine learning :

Les algorithmes constituent le cœur de tout projet de machine learning. Après la phase d'entraînement, ils permettent de générer des modèles capables de fournir des réponses ou de prendre des décisions basées sur des probabilités statistiques. Ces modèles peuvent être utilisés pour diverses finalités, telles que la reconnaissance d'objets dans des images (par exemple, l'identification de chats), la détection des arythmies cardiaques, ainsi que la réalisation de prédictions.

II.2.3.1. Réseaux de neurones (RN) :

Les réseaux de neurones artificiels représentent une approche inspirée du fonctionnement du cerveau humain. Ils se composent d'un ensemble d'unités de calcul interconnectées, appelées les neurones artificiels, capables de traiter et d'analyser et classifier l'information de manière distribuée.

Les réseaux de neurones, qui proviennent du Deep Learning, sont conçus pour traiter d'importants volumes de données et capturer des relations complexes. Grâce à leur capacité d'apprentissage, ils améliorent peu à peu la performance en ajustant leurs paramètres internes.

Ils font également preuve d'une bonne robustesse face à des données bruitées ou incomplètes. Leur utilisation dans de nombreux domaines, tels que la classification, la reconnaissance de formes et la détection de arythmies, s'explique par ces caractéristiques [20].

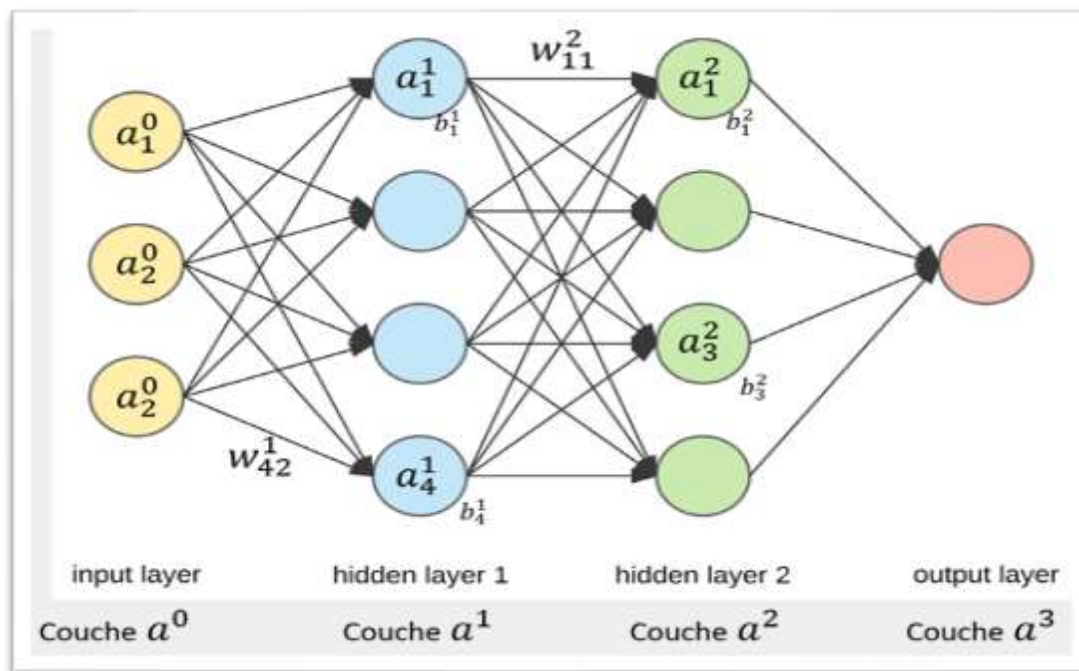


Figure II.4: Réseaux de neurones.

II.2.3.2. Vecteurs de support (SVM) :

Les machines à vecteurs de support (Support Vector Machines – SVM) représentent une méthode d'apprentissage supervisé largement utilisée pour les tâches de classification et de régression. Leur principe fondamental repose sur la détermination d'une frontière de décision optimale, appelée hyperplan, permettant de séparer les différentes classes dans un espace de dimension n .

L'objectif principal des SVM consiste à maximiser la marge de séparation entre les points de données les plus proches appartenant à des classes opposées, appelés vecteurs de support. Cette stratégie contribue à améliorer la capacité de généralisation du modèle face à de nouvelles données [21].

Il existe plusieurs types des SVM:

- **SVM linéaires :**

Les SVM linéaires conviennent particulièrement bien à la classification de données linéairement séparables. Dans ce cas, les différentes classes peuvent être séparées par une frontière de décision sans nécessiter qu'au préalable les données soient transformées.

L'idée centrale des SVM est de trouver l'hyperplan optimal qui va maximiser la marge de séparation entre les différentes classes. Cette marge correspond à la distance maximale entre l'hyperplan de décision et les points des différentes classes les plus proches de celui-ci, appelés vecteurs de support [22].

- **SVM RBF:**

Le noyau RBF (Radial Basis Function), également appelé noyau gaussien, est l'un des noyaux les plus utilisés dans les machines à vecteurs de support. Contrairement aux SVM linéaires, le noyau RBF permet de traiter des données non linéairement séparables en projetant les données d'entrée dans un espace de

dimension supérieure. Cette transformation permet de construire une frontière de décision non linéaire capable de mieux distinguer les différentes classes. Le principe du noyau RBF repose sur la mesure de similarité entre deux vecteurs de données. Plus deux points sont proches, plus la valeur du noyau est élevée [22].

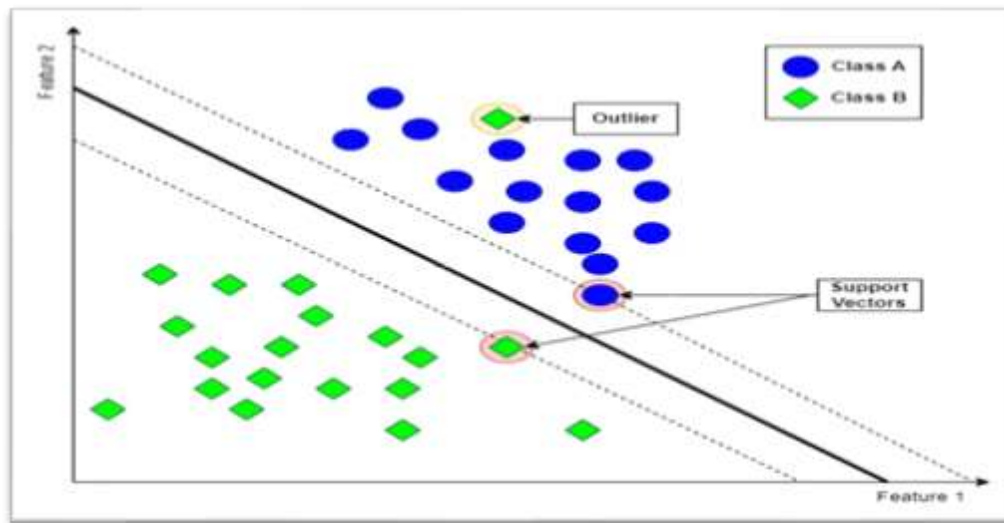


Figure II.5: Support Vector Machines (SVM).

II.2.3.3. K plus proches voisins (KNN) :

Cette algorithme est une technique d'apprentissage supervisé employée à des fins de classification et de régression. Il repose sur le principe que les observations similaires sont voisines dans l'espace des caractéristiques. En phase de prédiction, le modèle repère les k observations les plus proches de la nouvelle donnée et lui assigne la classe majoritaire (classification) ou la valeur moyenne (régression). Le choix du paramètre k est crucial, puisqu'il détermine directement la précision et la capacité du modèle à généraliser [23].

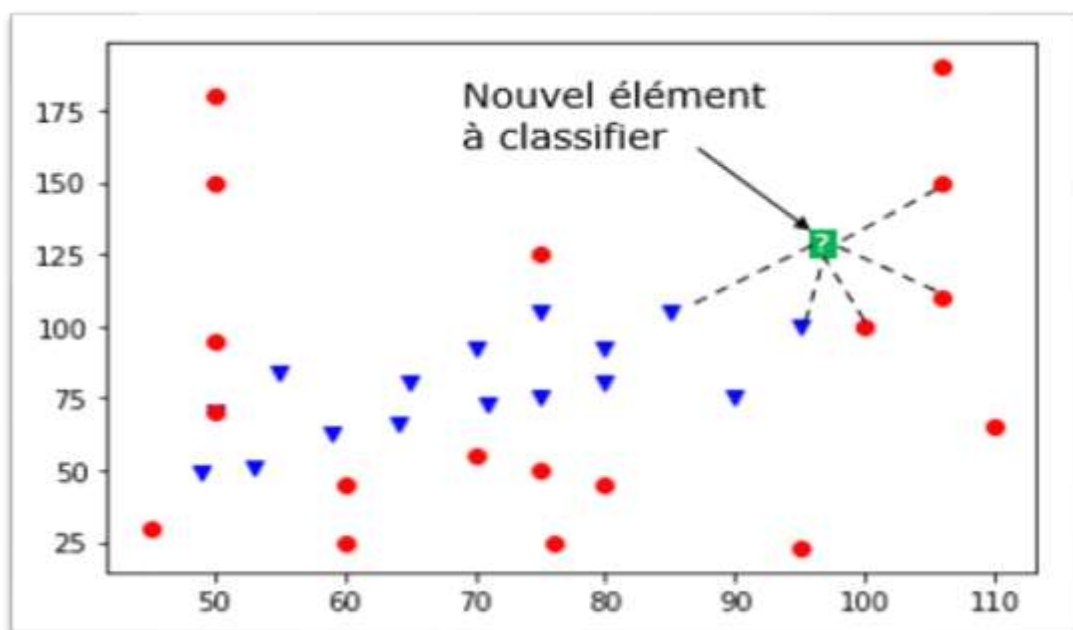


Figure II.6: k plus proches voisins (KNN).

II.3. Deep learning :

II.3.1. Definition et fonctionnement :

Le deep learning est une branche avancée de l'intelligence artificielle qui relève du domaine du machine learning. Il englobe un éventail de techniques d'apprentissage basé sur des modèles mathématiques, capables d'analyser et d'interpréter d'importants volumes de données. L'objectif du deep learning est alors d'entraîner les machines à apprendre par elles mêmes à partir des données sans avoir à expliciter toutes les règles de décision qu'elles doivent suivre

Le deep learning s'appuie principalement sur les réseaux de neurones artificiels, qui s'inspirent du fonctionnement du cerveau humain. Ces réseaux sont constitués de plusieurs couches de neurones interconnectés qui permettent de traiter, mémoriser et analyser l'information. Chaque couche extrait progressivement des caractéristiques de plus en plus complexes des données d'entrée, ce qui permet au système de réaliser des tâches avancées telles que la reconnaissance d'images, le traitement du signal, la classification et la prédiction [24].

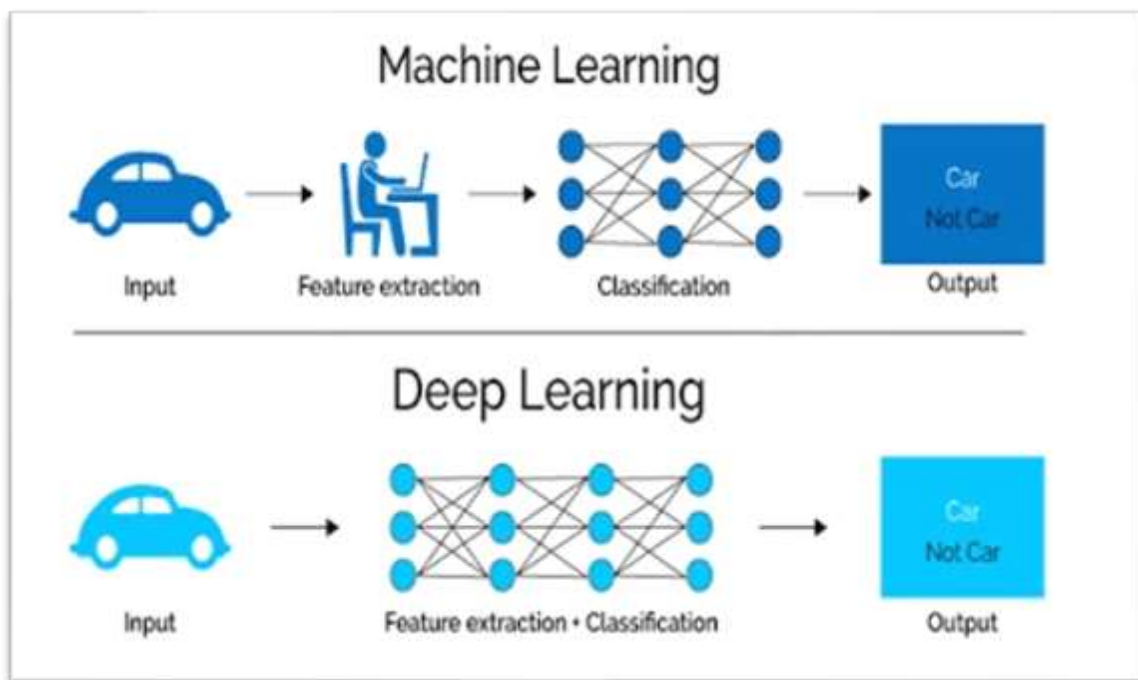


Figure II.7: difference entre ML et DL.

II.3.2. Les différents algorithmes de Deep Learning :

En deep learning, on utilise plusieurs types d'algorithmes. Chaque algorithme a ses propres caractéristiques et ses propres domaines d'application [25].

II.3.1.2. Réseaux neuronaux convolutifs (CNN):

Le CNN constitue une classe de réseaux de neurones profonds particulièrement efficace au traitement des données visuelles. Leur architecture comprend plusieurs couches successives qui permettent d'extraire automatiquement les caractéristiques importantes des données, comme les contours, les formes présentes dans une image, ou les ondes (P Q R S T) dans un signal ECG.

Le fonctionnement des CNN repose essentiellement sur des opérations de convolution. Ces opérations consistent à appliquer des filtres sur les données d'entrée pour détecter des caractéristiques. À mesure que les données progressent à travers les différentes couches du réseau, les caractéristiques extraites deviennent progressivement plus complexes et plus représentatives. Cette méthode permet de réduire l'intervention humaine dans l'extraction des caractéristiques et d'augmenter considérablement les performances de classification et de reconnaissance.

Les CNN sont largement utilisés dans plusieurs domaines, notamment la reconnaissance faciale, l'analyse d'images médicales, la détection d'objets, la vision par ordinateur ainsi que l'interprétation des images satellites. Dans le domaine biomédical, ils jouent un rôle important dans la détection d'anomalies et l'analyse automatique des signaux et images médicales [26].

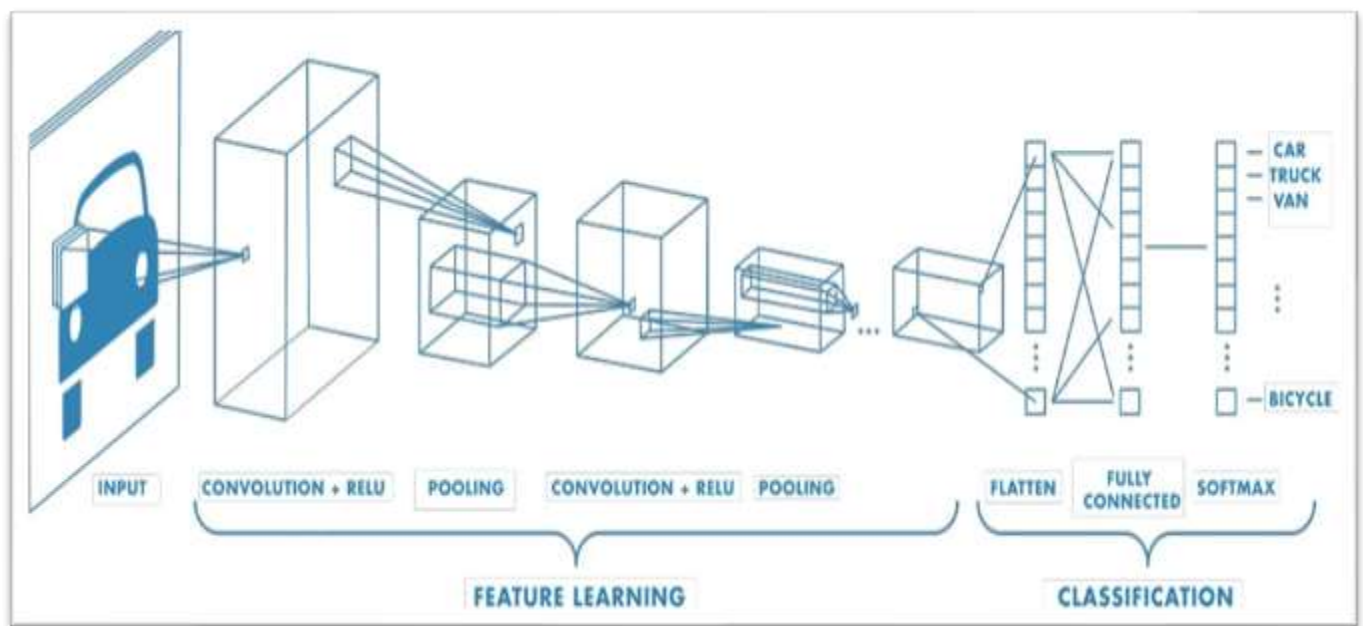


Figure II.8: Fonctionnement du CNN.

II.3.1.2. Réseaux neuronaux récurrents (RNN) :

Les réseaux neuronaux récurrents sont des modèles conçus pour traiter des données séquentielles ou temporelles. Contrairement aux réseaux neuronaux classiques, les RNN possèdent des connexions récurrentes permettant de conserver les informations des étapes précédentes grâce à une mémoire interne. Cette caractéristique leur permet de prendre en compte le contexte temporel lors du traitement des données.

Dans les RNN, la sortie obtenue à un instant donné peut être réutilisée comme entrée pour l'étape suivante. Ainsi, le réseau est capable d'apprendre les dépendances entre les différentes séquences de données. Toutefois, les RNN classiques présentent certaines limitations lorsqu'il s'agit de mémoriser des dépendances à long terme. Pour résoudre ce problème, des variantes plus avancées comme les réseaux LSTM (Long Short-Term Memory) ont été développées afin d'améliorer la capacité de mémorisation du modèle [26].

II.3.1.2. Réseaux à fonction de base radiale (RBFN) :

Les réseaux à fonction de base radiale (RBFN – Radial Basis Function Networks) constituent une architecture particulière de réseaux neuronaux feedforward. Ils se composent généralement de trois couches principales: une couche d'entrée, une couche cachée et une couche de sortie. La particularité de ces réseaux réside dans l'utilisation des fonctions de base radiales comme fonctions d'activation au niveau de la couche cachée.

Le principe des RBFN consiste à mesurer la similarité entre les données d'entrée et les centres des neurones cachés. Plus une donnée est proche du centre d'un neurone, plus l'activation produite est importante. Cette approche permet une modélisation efficace des relations non linéaires entre les données [27].

II.4. Conclusion :

Dans ce chapitre, nous avons présenté les concepts fondamentaux de l'intelligence artificielle (IA), du Machine Learning (ML) et du Deep Learning (DL).

Les différentes approches d'apprentissage ainsi que le rôle des réseaux de neurones artificiels dans l'analyse et la classification des données ont été étudiés.

Ces techniques offrent des outils performants pour l'extraction automatique de caractéristiques et la prise de décision à partir de grandes quantités de données. Elles constituent ainsi une base essentielle pour la conception de systèmes intelligents destinés à la classification des signaux ECG et à la détection automatique des arythmies cardiaques, qui seront développées dans les chapitres suivants.

Chapitre III:

Conception du système
proposé

III. Introduction :

Ce chapitre est consacré à la phase de conception du système proposé pour la classification des signaux électrocardiographiques (ECG). Après avoir présenté les aspects théoriques et l'étude bibliographique dans les chapitres précédents, nous abordons dans cette partie la mise en œuvre pratique de la solution retenue.

L'objectif principal de ce chapitre est de décrire l'architecture globale du système ainsi que les différentes étapes nécessaires à son développement. Cela inclut la préparation et le prétraitement des données, la conception du modèle basé sur les réseaux de neurones, ainsi que le choix des paramètres et des techniques d'apprentissage.

Nous détaillerons également les outils et les technologies utilisés pour la réalisation de ce projet, en mettant en évidence les différentes décisions prises lors de la conception afin d'optimiser les performances du modèle.

Enfin, ce chapitre permettra de poser les bases nécessaires pour l'évaluation et l'analyse des résultats qui seront présentées dans le chapitre suivant.

III.1. Acquisition des signaux ECG:

L'acquisition des signaux (ECG) constitue une étape essentielle dans l'analyse de l'activité électrique du cœur. Elle fonctionne en détectant les différences de tension créées par le battement du cœur, grâce à des électrodes posées sur la peau selon des positions standardisées. Le signal capté est généralement très faible, de l'ordre du millivolt, et doit être traité avant d'être utilisé. Ce traitement inclut une amplification et un filtrage pour réduire les bruits comme ceux du muscle ou les perturbations provenant du réseau électrique. Par la suite, le signal analogique est transformé en signal numérique grâce à un convertisseur analogique-numérique, afin qu'il puisse être traité par des systèmes informatiques. Une acquisition fiable est essentielle pour assurer la précision des analyses de diagnostic, surtout pour détecter les anomalies du rythme du cœur. Ainsi, la qualité du système qui capte les signaux influence directement l'efficacité des méthodes utilisées pour traiter et classer les signaux ECG [28].

III.1.1. Bases de données ECG :

Les bases de données ECG sont des outils importants pour faire de la recherche et du développement dans le domaine du traitement des signaux biomédicaux. Parmi les plus répandues, on trouve la base de données **MIT-BIH Arrhythmia**, qui comporte des enregistrements annotés pour étudier les arythmies cardiaques. Cette base contient des signaux issus de patients souffrant de différentes maladies, ce qui permet de tester et d'approfondir la validation des algorithmes de classification. Elles sont généralement accompagnées de commentaires exacts faits par des spécialistes, ce qui aide à former les modèles d'intelligence artificielle et d'apprentissage automatique.

L'utilisation de cette base standardisée assure que les travaux scientifiques peuvent être reproduits facilement et permettent de comparer efficacement différentes méthodes qui sont discutées dans les publications scientifiques [29].

La figure suivante représente les annotations des battements cardiaques :

Annotation	Description
N	Battement normal (Normal beat)
A	Contraction auriculaire prématuré (Atrial Premature Contraction)
V	Contraction ventriculaire prématuré (Premature Ventricular Contraction)
a	Contraction auriculaire prématuré aberré (aberrated Atrial Premature Contraction)
R	Bloc de branche droit (Right bundle branch block)
L	Bloc de branche gauche (Left bundle branch block)
P	Battement ectopique (Paced beat)
F	Fusion des battements V et N (fusion of ventricular and normal beat)
f	Fusion des battements P et N (fusion of paced and normal beat)
J	Battement nodal (ou jonctionnel) prématuré (Nodal (Junctional) premature beat)
j	Battement nodal (ou jonctionnel) échappé (nodal (junctional) escape beat)
E	Battement ventriculaire échappé (ventricular escape beat)
e	Battement auriculaire échappé (atrial escape beat)
S	Battement supraventriculaire prématuré (Supraventricular premature beat)
Q	Battement non classé (unclassified beat)

Figure III.1: Les annotations des battements cardiaques de la base de donnée MIT-BIH

III.2. Prétraitement du signal :

Le prétraitement du signal ECG constitue une étape cruciale améliorer la qualité des données avant leur analyse. Les signaux bruts sont souvent affectés par différents types de bruit qui peuvent altérer l'interprétation des caractéristiques cardiaques. Ainsi, plusieurs techniques de traitement sont appliquées, telles que la suppression du bruit, le filtrage et la normalisation. Ces étapes permettent d'obtenir un signal plus stable et exploitable, facilitant l'extraction des caractéristiques pertinentes et améliorant les performances des modèles de classification, notamment dans les approches basées sur l'apprentissage automatique et profond [30].

III.2.1. Suppression du bruit :

La réduction du bruit dans les signaux ECG a pour but de supprimer les désagréments comme le bruit des muscles, les interférences provenant du réseau électrique (50/60 Hz) et les fausses images causées par les mouvements.

Ces bruits peuvent cacher des parties essentielles du signal, comme les ondes P, Q, R, S et T. Des techniques comme les filtres adaptatifs (passe-bas, passe-bande), la transformée en ondelettes ou les méthodes basées sur la décomposition du signal sont souvent utilisées pour améliorer la qualité du signal tout en maintenant ses propriétés importantes [31].

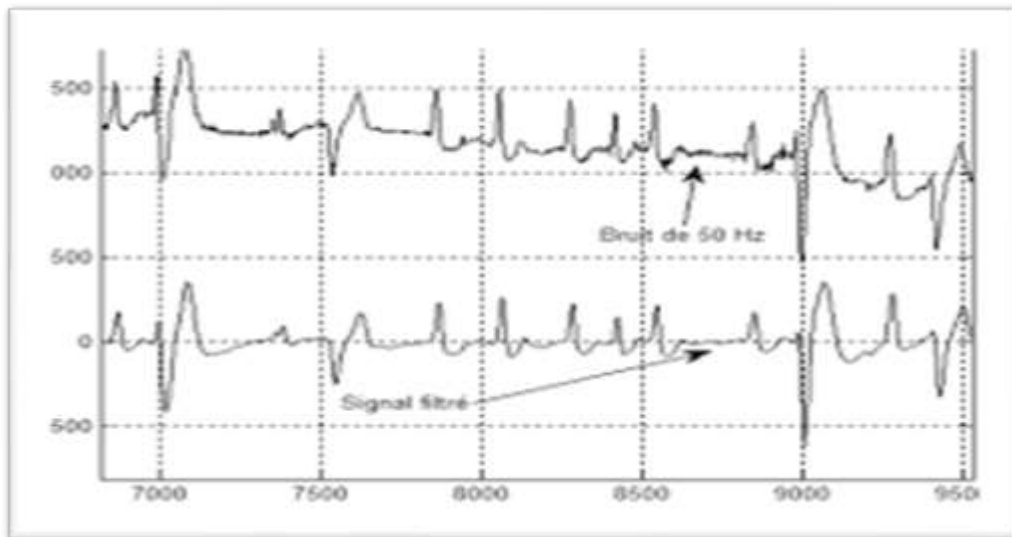


Figure III.2 : Bruit d'un signal.

III.2.2. Filtrage du signal ECG :

Le signal ECG est un signal multi-composantes et non stationnaire, souvent perturbé par différents types de bruit, notamment le bruit musculaire dont certaines composantes peuvent être corrélées au signal utile.

Cette approche permet principalement d'éliminer les composantes de haute fréquence. Cependant, elle ne supprime pas totalement les basses fréquences responsables des ondulations de la ligne de base du signal ECG.

Pour résoudre ce problème, un filtrage est appliqué, étant donné que ces ondulations se situent dans une très basse bande de fréquence, tandis que les complexes QRS appartiennent généralement à une bande de fréquence supérieure à 5 Hz. L'élimination des composantes de très basse fréquence permet ainsi de stabiliser la ligne de base du signal ECG [32].

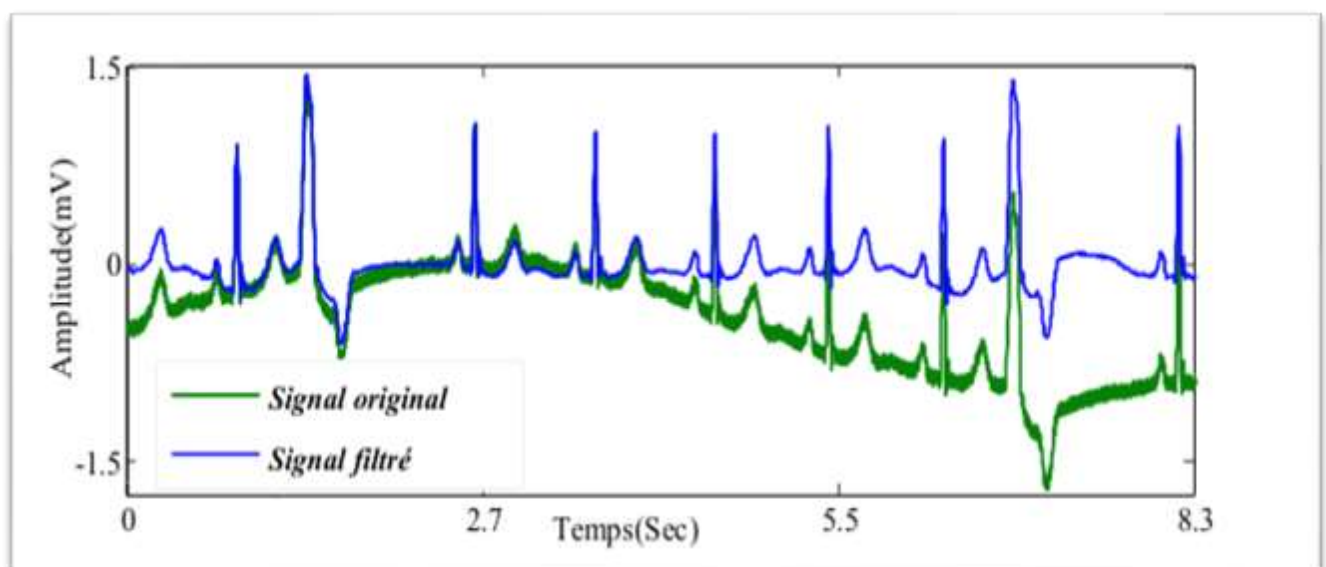


Figure III.3 : Exemple du filtrage.

III.2.3. Normalisation du signal :

La mise à l'échelle du signal ECG consiste à ajuster les amplitudes des données afin qu'elles soient comparables d'un enregistrement à l'autre. Cette étape est cruciale dans les systèmes d'apprentissage automatique, puisqu'elle permet d'accélérer l'apprentissage et d'améliorer l'accuracy des modèles. Les méthodes les plus courantes sont la normalisation Min-Max et la standardisation (Z-score).

Ces techniques permettent de réduire l'impact des variations individuelles et d'assurer une plus grande stabilité des algorithmes de classification [33].

III.2.4. Segmentation du signal ECG:

La segmentation du signal ECG consiste à diviser le signal en différentes parties représentatives d'un cycle cardiaque, notamment les ondes P, le complexe QRS et l'onde T. Cette étape est cruciale dans l'analyse automatique des signaux biomédicaux, parcequ'elle permet d'extraire les informations importantes nécessaires à l'extraction des caractéristiques et à la classification.

On réalise le plus souvent la segmentation par la détection des points caractéristiques du signal, appelés points fiduciaires (P, Q, R, S et T). Le pic R est le point de référence principal grâce à sa forte amplitude et sa facilité d'identification. Après localisation des pics R, le signal est découpé en segments correspondant à chaque battement cardiaque pour analyser la morphologie des différentes ondes ECG.

La littérature scientifique a proposé diverses méthodes de segmentation, notamment les approches fondées sur l'algorithme de Pan-Tompkins, les transformées en ondelettes (Wavelet Transform), ainsi que les méthodes plus récentes reposant sur l'apprentissage profond et les réseaux de neurones convolutifs (CNN). Ces techniques permettent d'améliorer la précision de détection des différentes composantes du signal ECG, même en présence de bruit [34].

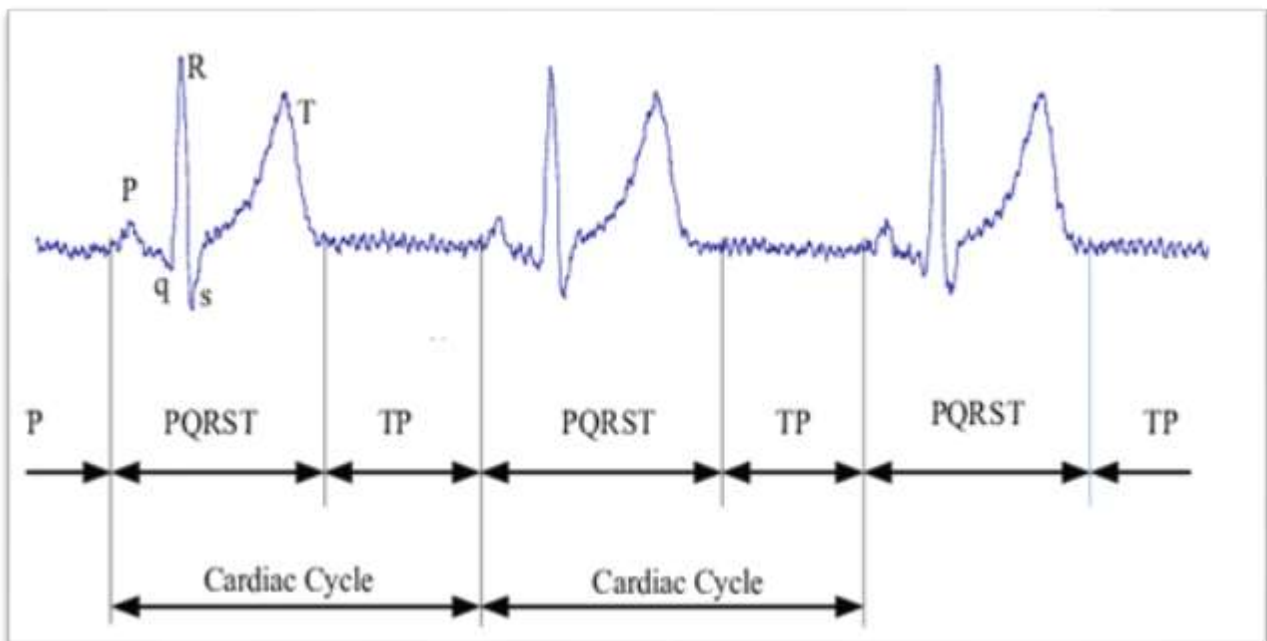


Figure III.4 : Méthode de segmentation.

III.3. Extraction des caractéristiques morphologiques :

Les caractéristiques temporelles sont les informations extraites directement du signal ECG dans le domaine temporel. Elles décrivent les variations d'amplitude et la morphologie du signal au cours du temps. Cette étude a utilisé plusieurs paramètres temporels pour caractériser de façon efficace l'activité cardiaque.

Ces paramètres incluent la moyenne du signal, l'écart-type, l'énergie et certaines mesures physiologiques associées aux ondes P, Q, R, S et T et des intervalles comme RR.

L'analyse temporelle est essentielle dans les systèmes de classification, puisqu'elle permet de détecter les anomalies cardiaques et d'améliorer les performances des modèles d'apprentissage automatique [35].

III.4. Classification des signaux :

La phase de classification représente une étape essentielle dans un système automatisé d'analyse des signaux ECG. Après les opérations de prétraitement et d'extraction des caractéristiques, les données obtenues sont utilisées afin d'identifier automatiquement les différentes classes cardiaques. Cette étape permet de distinguer les battements normaux des anomalies cardiaques.

Dans ce travail, les caractéristiques extraites du signal ECG, telles que les intervalles RR, sont introduites dans des modèles de classification supervisée. Ces modèles apprennent à partir d'un ensemble de données annotées afin de reconnaître les différentes catégories de signaux cardiaques.

Parmi les classifieurs les plus utilisés dans la littérature biomédicale, on retrouve les machines à vecteurs de support (SVM), ainsi que les méthodes basées sur les réseaux de neurones. Ces approches sont largement employées grâce leur capacité à traiter des données complexes et à améliorer la précision du diagnostic automatique.

L'évaluation des performances des modèles de classification est réalisée à l'aide de plusieurs métriques telles que l'**accuracy**, la précision, le rappel (recall) et le score F1. Une matrice de confusion est également utilisée afin d'analyser les erreurs de classification et de comparer les performances des différents modèles étudiés [36].

III.5. Conclusion:

Dans ce chapitre, nous avons présenté la méthodologie adoptée pour la mise en oeuvre du système de classification des signaux ECG, ainsi que les différentes étapes de prétraitement et de classification.

Le prétraitement a permis de nettoyer les signaux ECG en supprimant le bruit et les artefacts afin d'améliorer leur qualité et de faciliter leur analyse.

Nous avons également décrit la base de données utilisée ainsi que les techniques d'extraction des caractéristiques. La phase de classification a été réalisée à l'aide de modèles de Machine Learning permettant de distinguer les différents types de rythmes cardiaques.

L'architecture du modèle a été détaillée en mettant en évidence les choix effectués pour l'entraînement et l'évaluation.

Chapitre IV:

Résultats et discussion

IV. Introduction :

Ce travail porte sur la classification automatique des signaux ECG afin de détecter et d'identifier les différents types d'arythmies cardiaques. L'objectif principal est de concevoir un système intelligent capable d'analyser les signaux ECG avec une accuracy tout en réduisant l'intervention manuelle et le temps de diagnostic.

Le choix des algorithmes repose sur plusieurs critères tels que l'accuracy, la rapidité d'exécution, la capacité de généralisation et la complexité des données. Les modèles de Machine Learning sont privilégiés pour leur simplicité et efficacité sur des données structurées après extraction de caractéristiques. Les modèles de Deep Learning, quant à eux, sont choisis pour leur capacité à apprendre des représentations complexes et non linéaires des signaux ECG.

Enfin, une comparaison des performances des différents modèles est effectuée à l'aide de plusieurs métriques d'évaluation, notamment l'accuracy, la précision, le recall et le score F1.

Cette analyse permet d'identifier l'approche la plus performante pour la classification des signaux ECG et de mieux comprendre les avantages et limites des méthodes ML et DL dans le diagnostic automatique des maladies cardiaques.

IV.1. Environnement de travail:

Pour notre projet, nous avons choisi de mettre en oeuvre les algorithmes de classification d'ECG avec le langage Python, en raison de sa simplicité, flexibilité et de la disponibilité d'un grand nombre de bibliothèques dédiées au traitement du signal, au Machine Learning et au Deep Learning. L'environnement de développement utilisé facilite l'analyse des données, l'entraînement des modèles et la visualisation des résultats.

Les principales bibliothèques utilisées dans ce travail sont les suivantes:

- **Python 3.14**: langage de programmation utilisé pour le développement des modèles de classification et le traitement des données ECG [37].
- **TensorFlow**: bibliothèque de Deep Learning utilisée pour la création et l'entraînement des réseaux de neurones profonds tels que les CNN [38].
- **Scikit-learn**: bibliothèque de Machine Learning utilisée pour l'implémentation des algorithmes comme SVM, Random Forest et MLP ainsi que l'évaluation des performances [39].
- **NumPy**: bibliothèque utilisée pour les calculs numériques et la manipulation des tableaux de données ECG [40].
- **Matplotlib**: bibliothèque utilisée pour la visualisation graphique des signaux ECG, des résultats et des performances des modèles [41].
- **WFDB (WaveForm DataBase)**: bibliothèque utilisée pour charger et manipuler les bases de données ECG comme MIT-BIH Arrhythmia Database [42].

```
import wfdb
import numpy as np
import matplotlib.pyplot as plt
```

IV.2. Architecture globale de systeme:

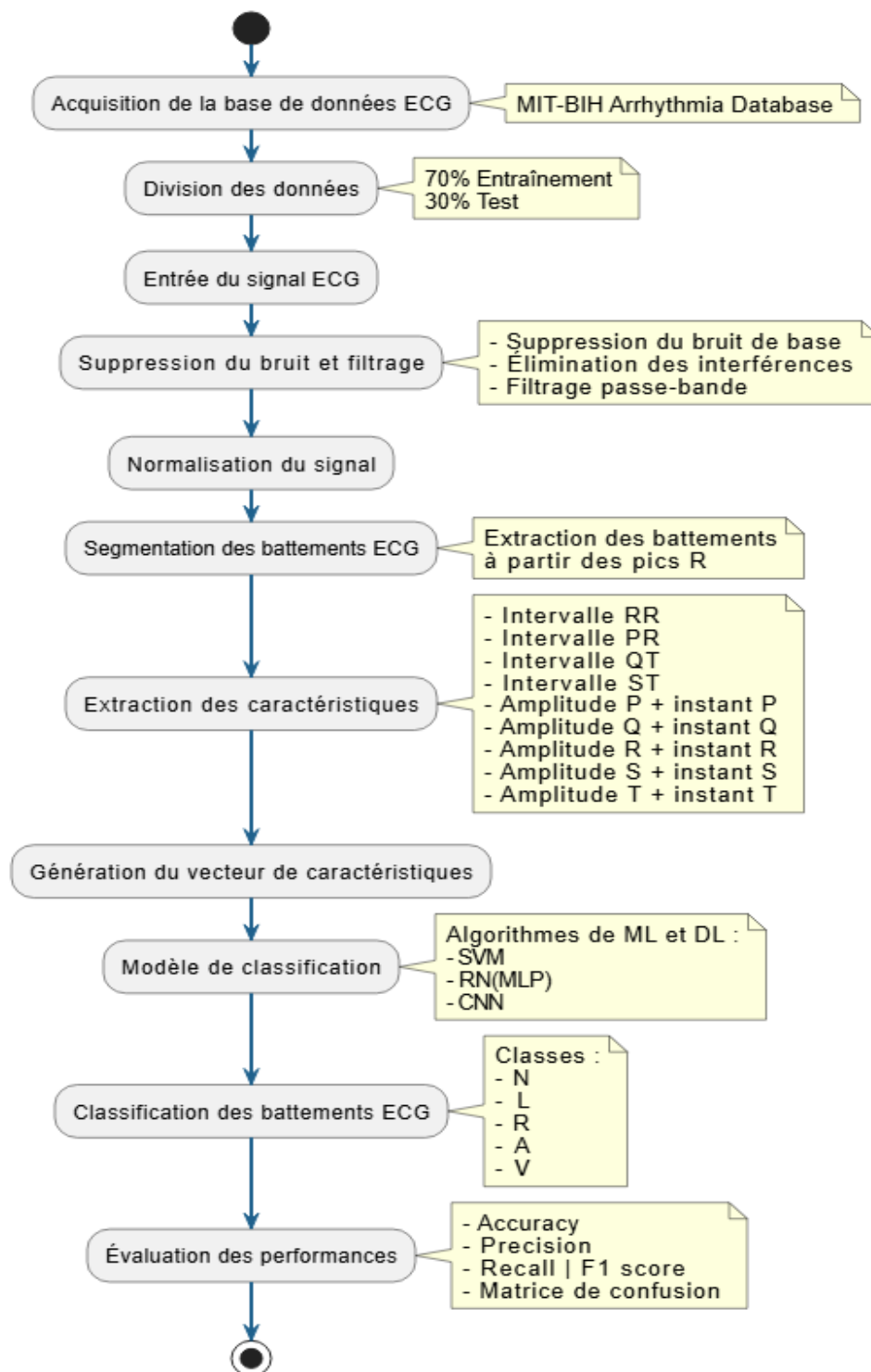


Figure IV.1 : Architecture globale de system

IV.3. Base de données ECG:

Dans ce projet, nous avons utilisé la base de données **MIT-BIH**, qui est l'une des bases les plus utilisées dans le domaine de la classification des signaux ECG et de la détection des arythmies cardiaques.

Cette base contient plusieurs enregistrements ECG provenant de différents patients, avec des annotations médicales permettant d'identifier les types de battements cardiaques.

Class	N	L	V	A	R
Nombre de battements	90589	2777	7239	803	8026

Tableaux 1 : nombre totale des battements.

IV.3.1. Apprentissage et test :

Pour évaluer correctement les performances des modèles, les données ont été séparées en 2 ensembles:

- **70%** des données pour l'apprentissage (**Training Set**)
- **30%** des données pour le test (**Testing Set**)

```
x_train, x_test, y_train, y_test = train_test_split(
    segments, labels, test_size=0.30, random_state=42, stratify=labels
)
```

IV.4. Prétraitement des signaux ECG :

Dans cette étude, toutes les opérations de prétraitement ont été appliquées à l'ensemble des enregistrements de la base de données MIT-BIH. Cependant, afin d'assurer la lisibilité et la clarté du document, les figures présentées dans ce chapitre représentent uniquement un exemple issu d'un seul enregistrement ECG. Tous les autres patients de la base de données subissent les mêmes étapes de traitement (filtrage, normalisation et segmentation).

IV.4.1. Signal ECG originale :

Le signal ECG brut représente l'activité électrique du cœur avant toute opération de traitement. Ce signal contient généralement du bruit, des interférences et des variations dues aux mouvements du patient ou aux équipements d'acquisition. L'observation du signal original permet de visualiser la forme initiale des battements cardiaques avant le prétraitement.

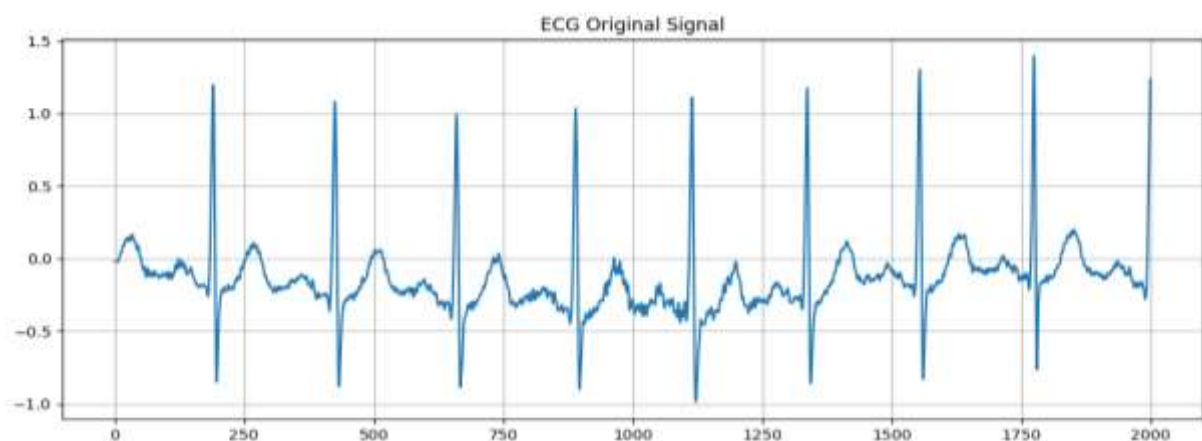


Figure IV.2: Signal ECG originale.

IV.4.2. Filtrage du signal ECG :

Afin d'améliorer la qualité du signal ECG, un filtrage passe-bande de 4eme ordre qui éliminer tout fréquence sauf [0.5 a 40 Hz] a été appliqué.

```
def butter_bandpass(sig, fs=360, low=0.5, high=40):  
    nyq = fs / 2  
    b, a = butter(4, [low/nyq, high/nyq], btype='band')  
    return filtfilt(b, a, sig)
```

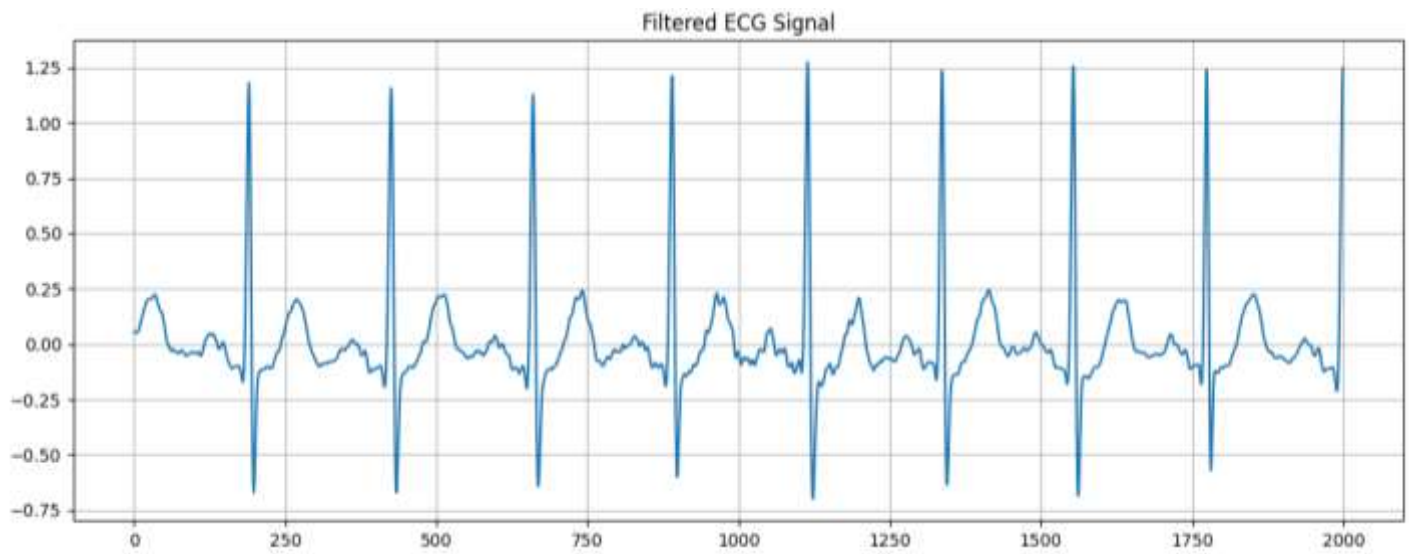


Figure IV.3: Signal ECG filtré.

IV.4.3. Normalisation du signal ECG :

Après le filtrage, une normalisation du signal a été réalisée afin de standardiser les amplitudes des signaux ECG. Cette opération améliore les performances du modèle d'apprentissage automatique en réduisant les différences d'échelle entre les données.

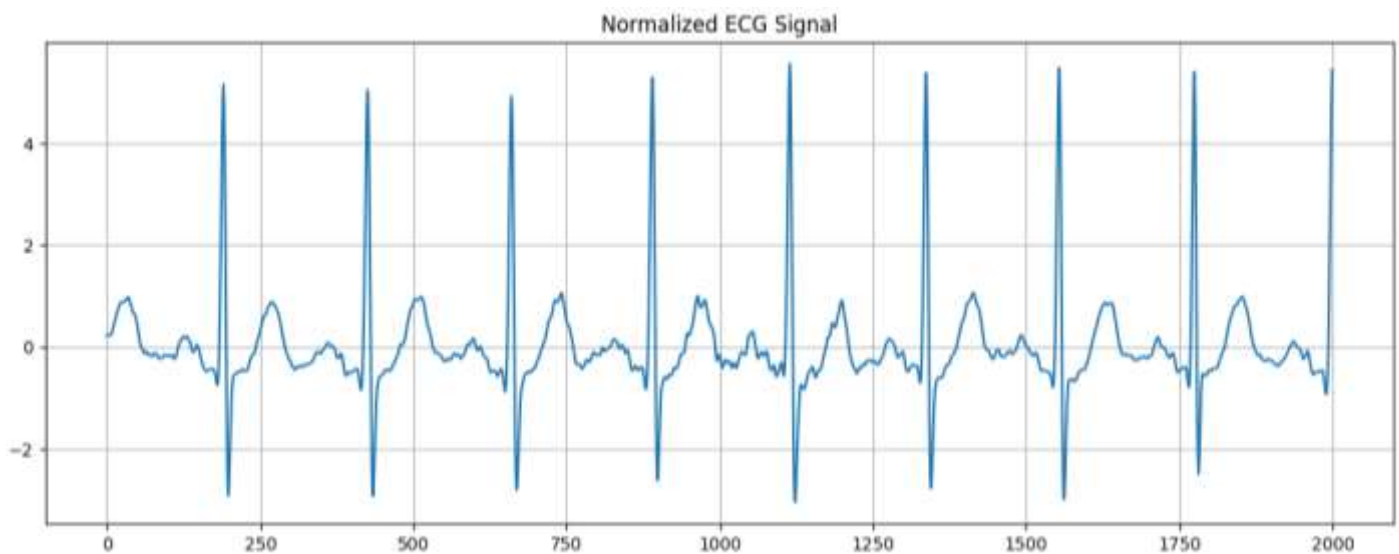


Figure IV.4: Signal ECG normalisé.

IV.4.4. Segmentation des battements cardiaques :

La segmentation consiste à extraire individuellement les battements cardiaques à partir du signal ECG. Chaque segment est centré autour du pic R afin de faciliter l'analyse des ondes P, QRS et T. Cette étape est essentielle pour l'extraction des caractéristiques utilisées dans la classification. En vue que le pic R est centralisé à 100 entre 0 et 250 (échantillons).

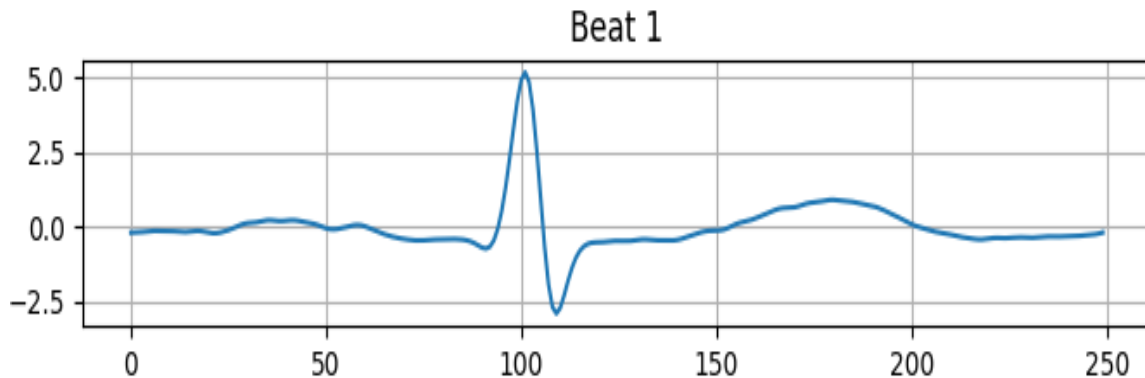


Figure IV.5: Segmentation des battements ECG.

IV.5. Méthodes d'évaluation :

IV.5.1. Matrice de confusion :

La matrice de confusion $N \times M$ (aussi appelée matrice d'erreur) est un outil fondamental pour l'évaluation des performances d'un modèle de classification. Elle compare les classes réelles aux classes prédites et permet d'identifier les prédictions correctes et incorrectes. Les résultats sont répartis en quatre catégories principales : vrais positifs (TP), vrais négatifs (TN), faux positifs (FP) et faux négatifs (FN). Cette matrice constitue la base du calcul de plusieurs métriques d'évaluation telles que l'accuracy, la précision, le rappel et le score F1 [43].

IV.5.2. La structure de la matrice de confusion :

➤ **Vrais positifs (True Positif : TP)**

Le modèle prédit correctement la classe positive. Par exemple, dans une tâche de détection d'objets, si une personne est réellement présente dans l'image, le modèle parvient correctement à dessiner une boîte englobante autour de cette personne.

➤ **Vrais négatifs (True Négatif : TN)**

Le modèle prédit correctement la classe négative. Cela s'avère déterminant dans des situations telles que la détection d'anomalies, lorsque le système parvient à identifier avec précision qu'une pièce produite est sans défaut.

➤ **Faux positif (False positif : FP)**

Le modèle prédit de façon erronée la classe positive. Cette erreur, souvent appelée « erreur de type I », survient lorsque le système détecte un objet qui n'existe pas, comme une caméra de sécurité qui prend une ombre pour un intrus.

➤ Faux négatif (False Négatif : FN)

Le modèle prédit de façon erronée la classe négative. Cette dernière, appelée (erreur de type II), survient quand le modèle ne parvient pas à détecter un objet présent, c'est-à-dire qu'il « rate » la cible [44].

Confusion matrix		Reality	
		Negative : 0	Positive : 1
Prediction	Negative : 0	True Negative : TN	False Negative : FN
	Positive : 1	False Positive : FP	True Positive : TP

Figure IV.6 : Matrice de confusion.

IV.5.3. Accuracy :

L'accuracy mesure la proportion totale des prédictions correctes effectuées par le modèle par rapport à l'ensemble des observations. Elle fournit une vue globale des performances du classificateur et est particulièrement pertinente lorsque les classes sont relativement équilibrées. Une valeur élevée d'accuracy indique que le modèle classe correctement la majorité des échantillons [45].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4.1)$$

IV.5.4. Précision :

La précision évalue la capacité du modèle à produire des prédictions positives correctes. Elle correspond à la proportion des échantillons prédits comme positifs qui sont effectivement positifs. Cette métrique est particulièrement importante lorsque le coût des faux positifs est élevé [45].

$$Precision = \frac{TP}{TP+FP} \quad (4.2)$$

IV.5.5. Recall :

Le rappel (ou Sensibilité) évalue la capacité du modèle à correctement identifier les échantillons qui appartiennent effectivement à une classe spécifique. Il correspond à la proportion de cas positifs bien détectés parmi l'ensemble des cas positifs réels. [45].

$$Recall = \frac{TP}{TP+FN} \quad (4.3)$$

IV.5.6. F1 score :

Le score F1 est la moyenne harmonique de la précision et du rappel. Il permet d'avoir une mesure équilibrée de la performance du modèle en tenant compte à la fois des faux positifs et des faux négatifs. [45].

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4.4)$$

IV.6. Etude des performances :

Nous avons utilisé 3 algorithmes différents (RN, SVM, CNN) dans notre projet afin d'observer les différences et de déterminer lequel est le plus précis (Accuracy).

- Le nombre total des battements de test (30%) est de 29786

Class	N	L	V	R	A
Nombre de battement	22354	2421	2071	2176	764

Tableaux 2 : nombre des battements de TEST.

IV.6.1. Réseaux de neurones (RN) :

Architecture du modèle :

L'architecture du Perceptron Multicouche (MLP) adoptée est structurée de manière optimisée avec deux couches cachées de 64 et 32 neurones. Cette disposition permet une réduction progressive de la dimensionnalité pour extraire les caractéristiques clés des signaux ECG. L'utilisation de la fonction d'activation ReLU assure une modélisation non linéaire efficace tout en accélérant l'apprentissage. Enfin, la couche de sortie comprend 5 neurones avec la fonction Softmax pour classifier les battements, offrant un compromis idéal entre la précision du diagnostic et la légèreté requise pour les systèmes embarqués.

```
model = Sequential([
    Dense(64, activation='relu', input_shape=(X_train.shape[1],)),
    BatchNormalization(), Dropout(0.3),

    Dense(32, activation='relu'), BatchNormalization(), Dropout(0.2),

    Dense(n_classes, activation='softmax')
```

Type de couche	Nombre de neurones	Fonction d'activation
Couche d'entrée	14 neurones	/
Couche cachée 1	64 neurones	ReLU
Couche cachée 2	32 neurones	ReLU
Couche de sortie	5 neurones	Softmax

Tableau 3 : Architecture de RN.

- On a fait plusieurs tests pour étudier séparément les performances de notre classifieur en fonction du nombre d'échantillons de données (N) et le nombre de nœuds dans la couche cachée (M).

M = [10, 20, 30] et N = [200, 400, 600]

IV.6.1.1. Etude de l’accuracy pour (N=200) :

- Pour M=10 :

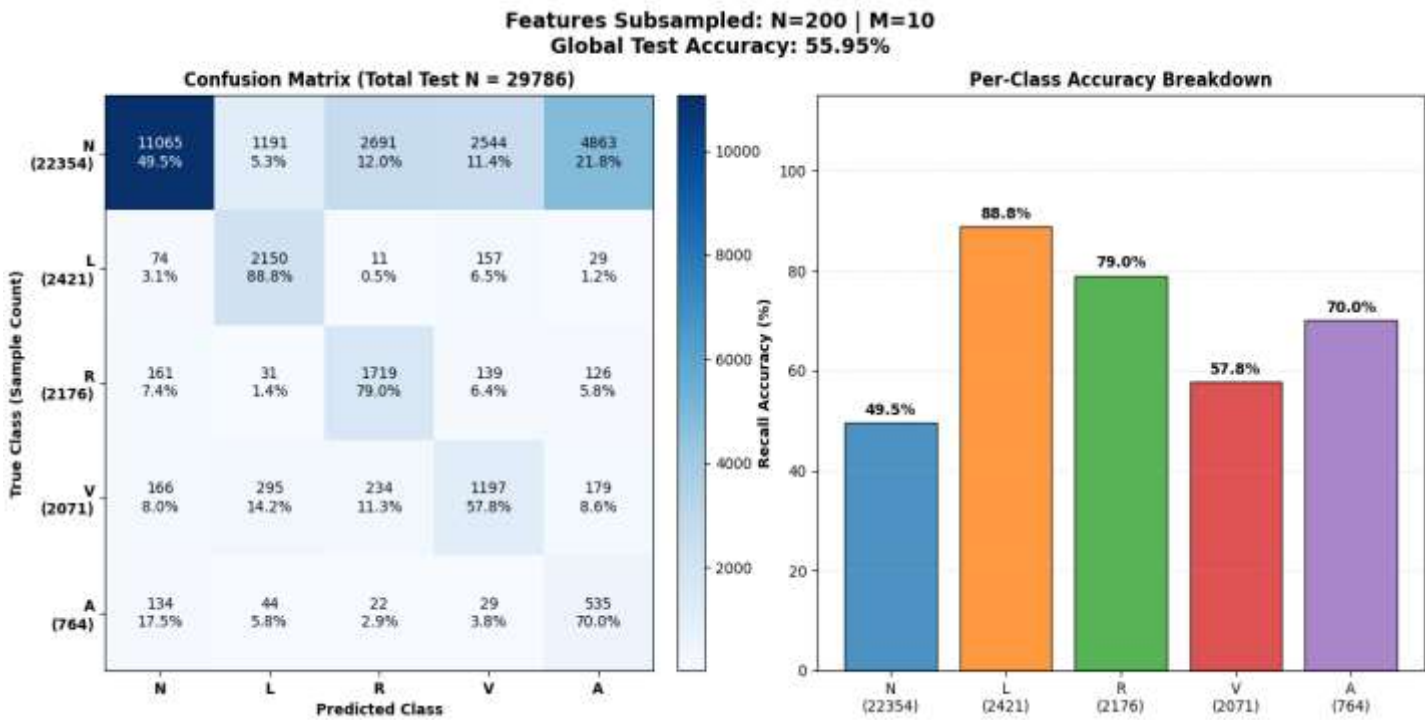


Figure IV.7 : Matrice de confusion et l’histogramme pour N=200 | M=10

- Pour M=20

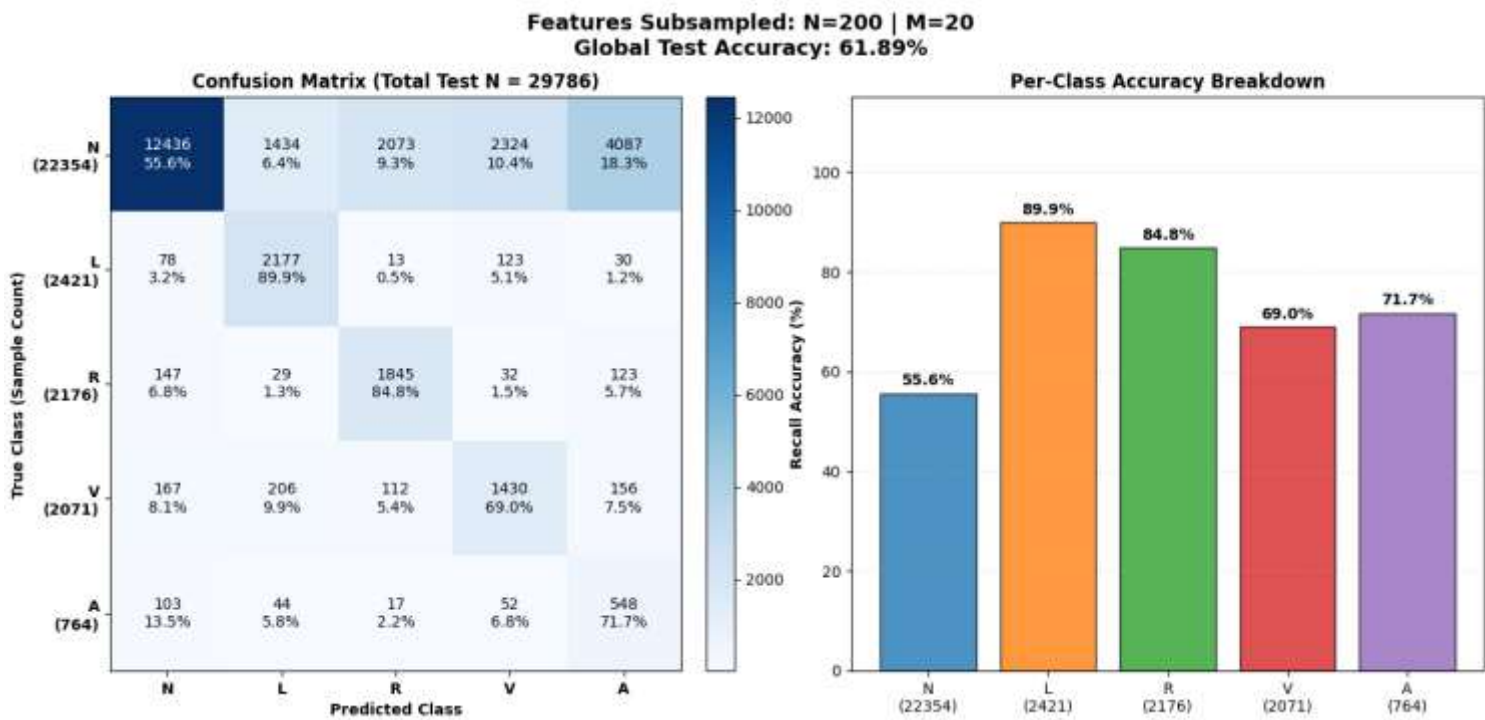


Figure IV.8 : Matrice de confusion et l’histogramme pour N=200 | M=20

• Pour M=30

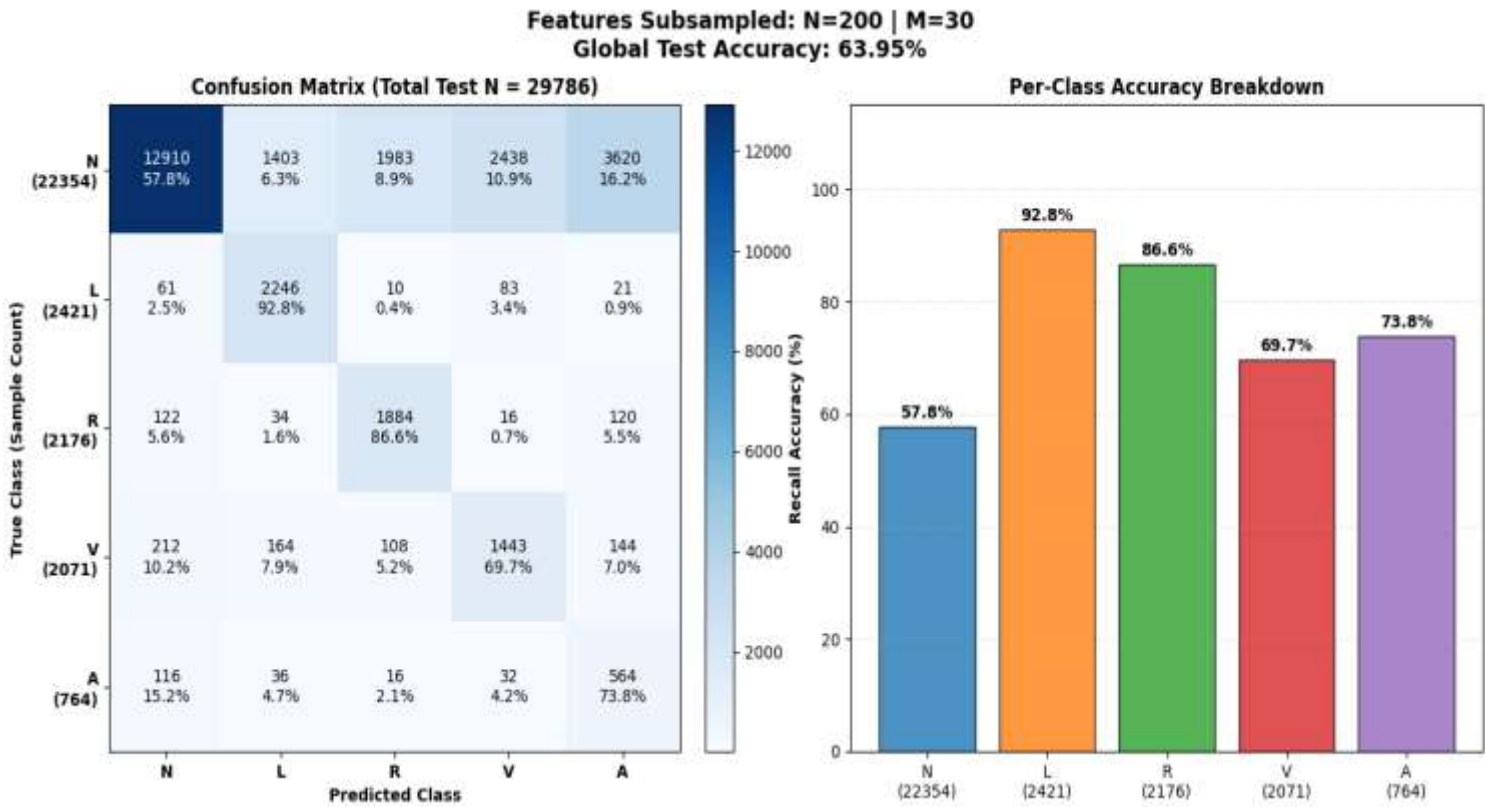


Figure IV.9 : Matrice de confusion et l’histogramme pour N=200 | M=30

M= 10	global	A	R	N	L	V
Accuracy	0.560	0.700	0.790	0.495	0.888	0.578
Erreur rate	0.440	0.300	0.210	0.505	0.112	0.422
M= 20	global	A	R	N	L	V
Accuracy	0.619	0.717	0.848	0.556	0.899	0.690
Erreur rate	0.381	0.283	0.152	0.444	0.101	0.310
M= 30	global	A	R	N	L	V
Accuracy	0.639	0.738	0.866	0.578	0.928	0.697
Erreur rate	0.361	0.262	0.134	0.422	0.072	0.303

Tableau 4 : L’accuracy et le taux d’erreur globale et de chaque classe pour N=200

IV.6.1.2. Etude de l'accuracy pour (N=400) :

M= 10	global	A	R	N	L	V
Accuracy	0.649	0.696	0.858	0.593	0.942	0.683
Erreur rate	0.351	0.304	0.142	0.407	0.058	0.317
M= 20	global	A	R	N	L	V
Accuracy	0.680	0.797	0.903	0.616	0.941	0.783
Erreur rate	0.320	0.203	0.097	0.384	0.059	0.217
M= 30	global	A	R	N	L	V
Accuracy	0.763	0.779	0.904	0.725	0.952	0.793
Erreur rate	0.237	0.221	0.096	0.275	0.048	0.207

Tableau 5 : L'accuracy et le taux d'erreur globale et de chaque classe pour N=400

IV.6.1.3. Etude de l'accuracy pour (N=600) :

M= 10	global	A	R	N	L	V
Accuracy	0.725	0.724	0.896	0.682	0.949	0.748
Erreur rate	0.275	0.276	0.104	0.318	0.051	0.252
M= 20	global	A	R	N	L	V
Accuracy	0.769	0.801	0.915	0.726	0.956	0.854
Erreur rate	0.231	0.199	0.085	0.274	0.044	0.146
M= 30	global	A	R	N	L	V
Accuracy	0.840	0.815	0.911	0.817	0.955	0.878
Erreur rate	0.160	0.185	0.089	0.183	0.045	0.122

Tableau 6 : L'accuracy et le taux d'erreur globale et de chaque classe pour N=600

Remarque :

Comme on peut le voir dans les tableaux 4, 5, 6, chaque fois que l'on augmente la valeur de N et M, l'accuracy augmente également. On peut donc affirmer que les modifications ou les changements directs apportées à N et M ont un impact sur l'accuracy.

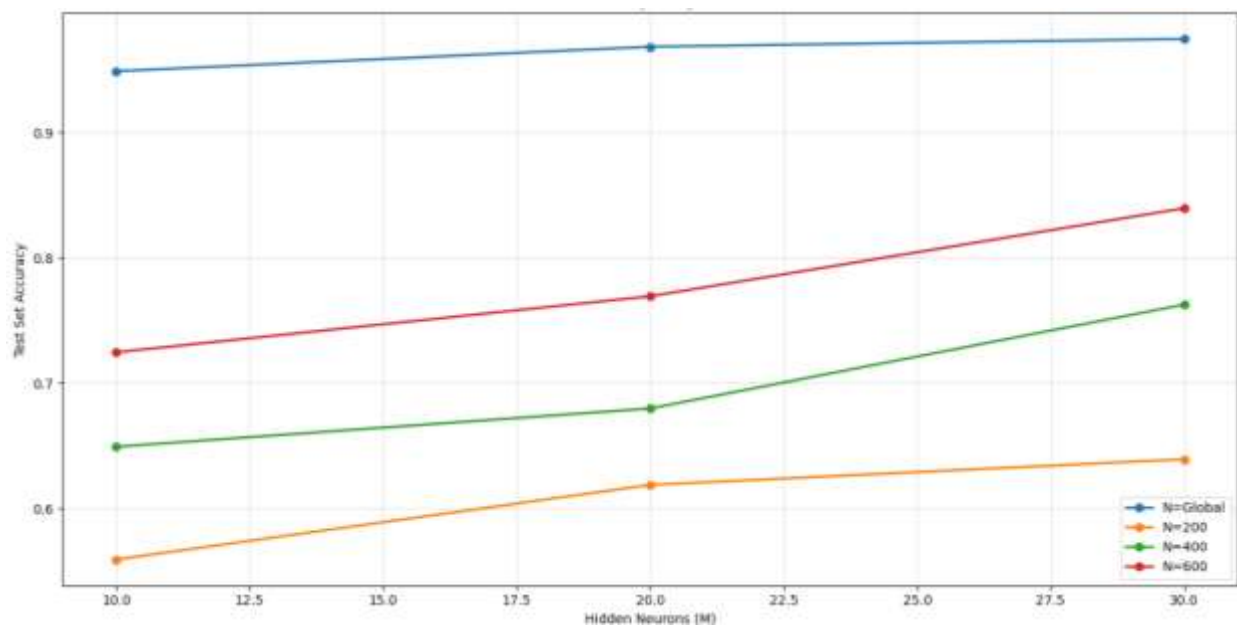


Figure IV.10 : L'évolution de l'accuracy globale en fonction de N et M

✓ Interprétation des Résultats et Discussion :

Cette étude comparative démontre l'impact direct de la taille de l'échantillon (N) et de la variation du nombre de noyaux (M) dans la 1^{ère} couche sur l'accuracy globale.

Les expérimentations menées sur N = 200 et N = 400 révèlent une grand accuracy à mesure que le volume de données et la profondeur du réseau augmentent. Le palier de N = 600 enregistre les performances maximales, offrant une excellente stabilité et accuracy par classe d'arythmie.

Il en résulte que la performance du modèle est une fonction strictement croissante de la quantité de données combinée à un réglage architectural précis. En conclusion, nous observons que plus le nombre d'échantillons est élevé, plus l'accuracy obtenue est grande. la configuration N = 600 associée au réseau de neurones le plus optimisé s'impose comme la solution idéale pour cette étude.

IV.6.2. Vecteurs de support (SVM) :

Dans cette section, nous abordons la phase d'évaluation expérimentale de l'algorithme des SVM. L'objectif principal est d'analyser l'influence du choix du parametre et d'étudier l'accuracy du modèle face à la variation de ses principaux hyperparamètres.

Pour ce faire, notre stratégie d'optimisation repose sur une recherche sur grille (*Grid Search*) appliquée à deux configurations distinctes :

- **Le noyau Linéaire:** où nous évaluons l'impact du paramètre de régularisation **C** [0.1, 1, 10, 100].
- **Le noyau RBF:** où nous ajustons conjointement le paramètre de régularisation **C1** [1, 10, 20] et le paramètre de forme **gamma** [0.01, 0.1, 1].

Ces deux approches sont systématiquement testées à travers différentes dimensions de sous-ensembles de données N [200, 500, 800] afin de retenir la configuration la plus performante pour la classification des arythmies.

IV.6.2.1. SVM Linéaire:

- Pour N=200 et changement de C:

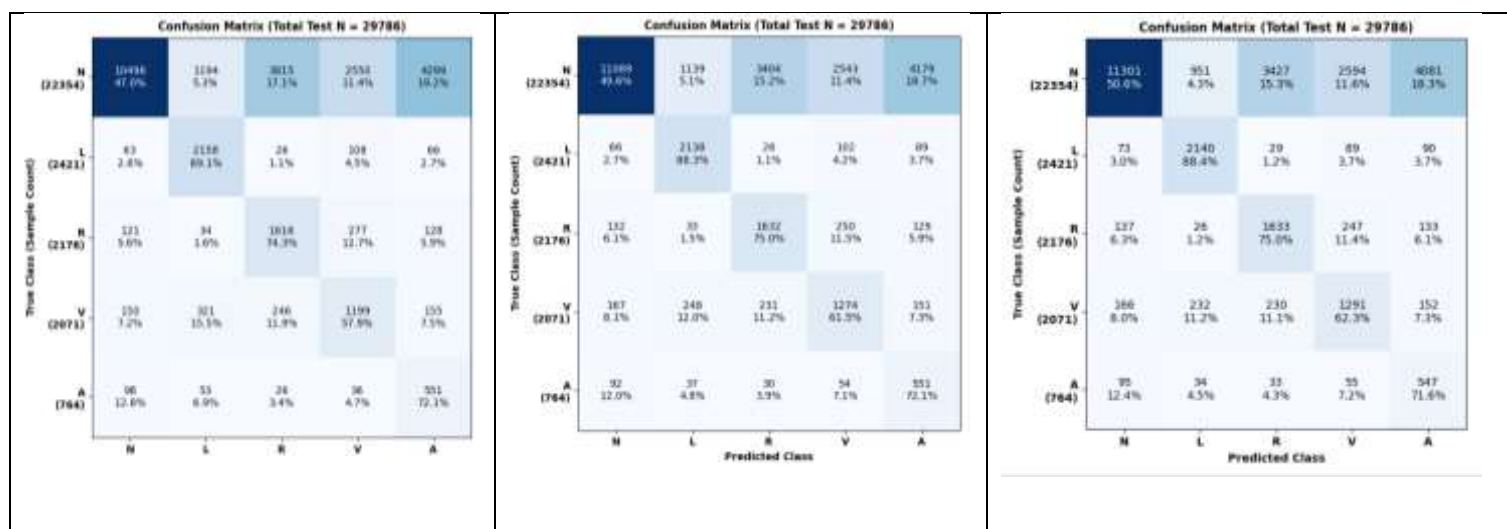


Figure IV.11 : Martice de confusion pour C= 0.1 ,1, 10

Linear C=0.1	global	A	R	N	L	V
Accuracy	0.538	0.721	0.743	0.470	0.891	0.579
Erreur rate	0.462	0.279	0.257	0.530	0.109	0.421
Linear C=1	global	A	R	N	L	V
Accuracy	0.560	0.721	0.750	0.496	0.883	0.615
Erreur rate	0.440	0.279	0.250	0.504	0.117	0.385
Linear C=10	global	A	R	N	L	V
Accuracy	0.568	0.716	0.750	0.506	0.884	0.623
Erreur rate	0.432	0.284	0.250	0.494	0.116	0.377
Linear C=100	global	A	R	N	L	V
Accuracy	0.569	0.713	0.750	0.507	0.883	0.623
Erreur rate	0.431	0.287	0.250	0.493	0.117	0.377

Tableau 7 : L'accuracy, le taux d'erreur pour N=200

- Pour N=500 et changement de C:

Linear C=0.1	global	A	R	N	L	V
Accuracy	0.533	0.738	0.746	0.462	0.891	0.578
Erreur rate	0.467	0.262	0.254	0.538	0.109	0.422
Linear C=1	global	A	R	N	L	V
Accuracy	0.543	0.740	0.743	0.476	0.886	0.588
Erreur rate	0.457	0.260	0.257	0.524	0.114	0.412
Linear C=10	global	A	R	N	L	V
Accuracy	0.544	0.745	0.744	0.477	0.883	0.590
Erreur rate	0.456	0.255	0.256	0.523	0.117	0.410
Linear C=100	global	A	R	N	L	V
Accuracy	0.544	0.745	0.744	0.477	0.882	0.591
Erreur rate	0.456	0.255	0.256	0.523	0.118	0.409

Tableau 8 : L'accuracy, le taux d'erreur pour N=500

- Pour N=800 et changement de C:

Linear C=0.1	global	A	R	N	L	V
Accuracy	0.549	0.740	0.707	0.488	0.891	0.574
Erreur rate	0.451	0.260	0.293	0.512	0.109	0.426
Linear C=1	global	A	R	N	L	V
Accuracy	0.553	0.742	0.714	0.490	0.892	0.589
Erreur rate	0.447	0.258	0.286	0.510	0.108	0.411
Linear C=10	global	A	R	N	L	V
Accuracy	0.554	0.742	0.716	0.492	0.891	0.591
Erreur rate	0.446	0.258	0.284	0.508	0.109	0.409
Linear C=100	global	A	R	N	L	V
Accuracy	0.554	0.742	0.715	0.492	0.892	0.591
Erreur rate	0.446	0.258	0.285	0.508	0.108	0.409

Tableau 9 : L'accuracy, le taux d'erreur pour N=800

Remarques :

Comme on peut le voir dans les **tableaux 7, 8, 9**, et contrairement aux résultats de RN, l'impact direct et les modifications des règles C et N n'améliorent pas significativement l'accuracy, ce qui explique les légers changements observés.

✓ **Interprétation des Résultats et Discussion :**

L'analyse des rapports de performance du SVM à noyau linéaire met en évidence un phénomène de stabilisation systématique : pour toutes les tailles d'échantillons **N** et malgré la variation du paramètre de régularisation **C**, l'accuracy globale demeure invariante, se stabilisant autour de **0.555 (55.5%)**. Cette stabilisation structurelle s'explique par les facteurs scientifiques suivants :

- **Non-séparabilité linéaire des descripteurs :** Les caractéristiques morphologiques extraites des signaux ECG (telles que les amplitudes des ondes P-Q-R-S-T) présentent une distribution hautement complexe et un chevauchement sévère entre les 5 classes d'arythmies.
- Inefficacité théorique de **C** en espace rigide, Par conséquent, l'augmentation du nombre d'échantillons **N** n'apporte aucune amélioration, car le modèle souffre d'un sous-apprentissage structurel lié à sa faible capacité de représentation.

IV.6.2.2. SVM RBF:

On a utilise:

```
C_values = [1, 10, 20, 50]
gamma_values = [0.001, 0.01, 0.1, 1]
```

Pour N= 200:

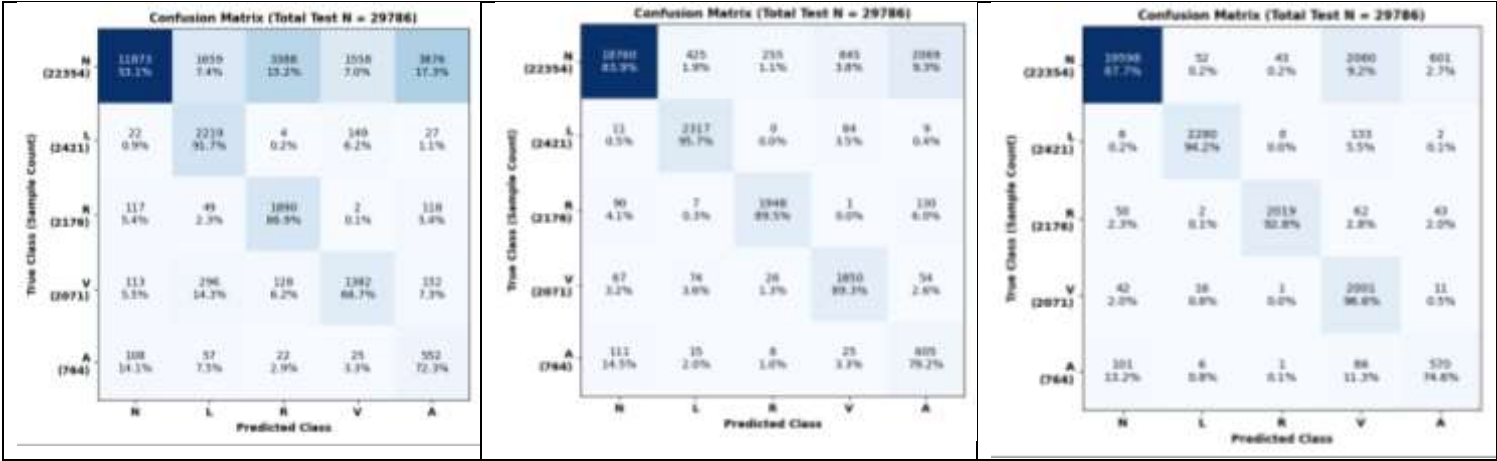


Figure IV.12: Martice de confusion pour C1= 1 | G=0.01, 0.1, 1

RBF C=1 G=0.01	global	A	R	N	L	V
Accuracy	0.601	0.723	0.869	0.531	0.917	0.667
Erreur rate	0.399	0.277	0.131	0.469	0.083	0.333
RBF C=1 G=0.1	global	A	R	N	L	V
Accuracy	0.855	0.792	0.895	0.839	0.957	0.893
Erreur rate	0.145	0.208	0.105	0.161	0.043	0.107
RBF C=1 G=1	global	A	R	N	L	V
Accuracy	0.889	0.746	0.928	0.877	0.942	0.966
Erreur rate	0.111	0.254	0.072	0.123	0.058	0.034
RBF C=10 G=0.01	global	A	R	N	L	V
Accuracy	0.750	0.771	0.893	0.710	0.943	0.801
Erreur rate	0.250	0.229	0.107	0.290	0.057	0.199
RBF C=10 G=0.1	global	A	R	N	L	V
Accuracy	0.873	0.829	0.956	0.853	0.975	0.907
Erreur rate	0.127	0.171	0.044	0.147	0.025	0.093
RBF C=10 G=1	global	A	R	N	L	V
Accuracy	0.889	0.766	0.941	0.874	0.951	0.963
Erreur rate	0.111	0.234	0.059	0.126	0.049	0.037
RBF C=20 G=0.01	global	A	R	N	L	V
Accuracy	0.795	0.784	0.896	0.765	0.950	0.833
Erreur rate	0.205	0.216	0.104	0.235	0.050	0.167
RBF C=20 G=0.1	global	A	R	N	L	V
Accuracy	0.876	0.821	0.969	0.856	0.976	0.900
Erreur rate	0.124	0.179	0.031	0.144	0.024	0.100
RBF C=20 G=1	global	A	R	N	L	V
Accuracy	0.889	0.766	0.941	0.874	0.951	0.963
Erreur rate	0.111	0.234	0.059	0.126	0.049	0.037

Tableau 10 : L'accuracy, le taux d'erreur pour N=200

Pour N= 500:

RBF C=1 G=0.01	global	A	R	N	L	V
Accuracy	0.641	0.776	0.880	0.574	0.945	0.706
Erreur rate	0.359	0.224	0.120	0.426	0.055	0.294
RBF C=1 G=0.1	global	A	R	N	L	V
Accuracy	0.902	0.827	0.934	0.893	0.965	0.921
Erreur rate	0.098	0.173	0.066	0.107	0.035	0.079
RBF C=1 G=1	global	A	R	N	L	V
Accuracy	0.915	0.843	0.958	0.903	0.959	0.972
Erreur rate	0.085	0.157	0.042	0.097	0.041	0.028
RBF C=10 G=0.01	global	A	R	N	L	V
Accuracy	0.833	0.780	0.907	0.814	0.942	0.845
Erreur rate	0.167	0.220	0.093	0.186	0.058	0.155
RBF C=10 G=0.1	global	A	R	N	L	V
Accuracy	0.916	0.878	0.972	0.904	0.975	0.925
Erreur rate	0.084	0.122	0.028	0.096	0.025	0.075
RBF C=10 G=1	global	A	R	N	L	V
Accuracy	0.909	0.866	0.965	0.893	0.961	0.968
Erreur rate	0.091	0.134	0.035	0.107	0.039	0.032
RBF C=20 G=0.01	global	A	R	N	L	V
Accuracy	0.862	0.791	0.914	0.848	0.955	0.869
Erreur rate	0.138	0.209	0.086	0.152	0.045	0.131
RBF C=20 G=0.1	global	A	R	N	L	V
Accuracy	0.919	0.887	0.975	0.908	0.973	0.922
Erreur rate	0.081	0.113	0.025	0.092	0.027	0.078
RBF C=20 G=1	global	A	R	N	L	V
Accuracy	0.906	0.868	0.965	0.890	0.961	0.968
Erreur rate	0.094	0.132	0.035	0.110	0.039	0.032

Tableau 11 : L'accuracy, le taux d'erreur pour N=500

Pour N= 800:

RBF C=1 G=0.01	global	A	R	N	L	V
Accuracy	0.682	0.770	0.880	0.624	0.937	0.771
Erreur rate	0.318	0.230	0.120	0.376	0.063	0.229
RBF C=1 G=0.1	global	A	R	N	L	V
Accuracy	0.905	0.861	0.943	0.893	0.973	0.933
Erreur rate	0.095	0.139	0.057	0.107	0.027	0.067
RBF C=1 G=1	global	A	R	N	L	V
Accuracy	0.927	0.876	0.966	0.917	0.965	0.975
Erreur rate	0.073	0.124	0.034	0.083	0.035	0.025
RBF C=10 G=0.01	global	A	R	N	L	V
Accuracy	0.850	0.801	0.919	0.832	0.945	0.873
Erreur rate	0.150	0.199	0.081	0.168	0.055	0.127
RBF C=10 G=0.1	global	A	R	N	L	V
Accuracy	0.928	0.884	0.981	0.918	0.981	0.938
Erreur rate	0.072	0.116	0.019	0.082	0.019	0.062
RBF C=10 G=1	global	A	R	N	L	V
Accuracy	0.922	0.893	0.968	0.910	0.968	0.967
Erreur rate	0.078	0.107	0.032	0.090	0.032	0.033
RBF C=20 G=0.01	global	A	R	N	L	V
Accuracy	0.874	0.817	0.921	0.861	0.963	0.884
Erreur rate	0.126	0.183	0.079	0.139	0.037	0.116
RBF C=20 G=0.1	global	A	R	N	L	V
Accuracy	0.927	0.886	0.979	0.917	0.980	0.935
Erreur rate	0.073	0.114	0.021	0.083	0.020	0.065
RBF C=20 G=1	global	A	R	N	L	V
Accuracy	0.921	0.894	0.968	0.908	0.968	0.967
Erreur rate	0.079	0.106	0.032	0.092	0.032	0.033

Tableau 12 : L'accuracy, le taux d'erreur pour N=800

➤ **Remarque :** Nous pouvons voir dans **les tableaux 10, 11, 12**, qu'après les modifications que nous avons apportées au C1, G, N, l'accuracy augmente à chaque fois, mais configuration de tableau 12 offre les performances maximales et la plus grande stabilité, où la richesse morphologie de la base de données est pleinement exploitée par le modèle RBF, ce qui se traduit par une grand accuracy et une réduction optimale des taux d'erreur.

➤ Interprétation des Résultats et Discussion :

Contrairement au noyau linéaire qui stagne structurellement, le SVM à noyau RBF démontre de manière empirique que l'accuracy globale est une fonction strictement croissante du volume de données N.

Les expérimentations prouvent que plus la taille de l'échantillon augmente (N = 200 , 500 , 800), plus le modèle RBF parvient à enrichir sa frontière de décision et à élever son accuracy globale. En combinant cet apport massif de données quantitatives à un ajustement optimal du couple d'hyperparamètres (C1, G), le réseau surmonte définitivement la non-séparabilité des signaux ECG, s'imposant ainsi comme le modèle le plus robuste et le plus performant pour la classification des arythmies.

IV.6.3. CNN :

Afin d'évaluer notre modèle CNN et de quantifier l'impact de la masse des données d'apprentissage sur ses performances, nous avons implémenté une approche de validation multi-scénarios.

Ce protocole consiste à entraîner le modèle sur trois sous-ensembles distincts (Scénarios 1, 2 et 3) caractérisés par une réduction homothétique progressive des effectifs de chaque classe d'arythmie (N, L, R, A, V), tout en préservant le déséquilibre intrinsèque de la base de données. L'architecture du Perceptron Multicouche (MLP) adoptée est structurée de manière optimisée avec trois couches cachées de 128, 64 et 32 neurones.

Type de couche	Nombre de neurones	Fonction d'activation
Couche d'entrée	14 neurones	/
Couche cachée 1	128 neurones	ReLU
Couche cachée 2	64 neurones	ReLU
Couche cachée 3	32 neurones	ReLU
Couche de sortie	5 neurones	Softmax

Tableau 13 : Architecture de CNN.

Scenario 1:

- ✓ Nombre total de : 35380 échantillons
- ✓ Entraînement distribution: {'N': 26659, 'L': 2857, 'R': 2719, 'A': 567, 'V': 2578}

Classification Report:					
	precision	recall	f1-score	support	
N	0.99	0.99	0.99	11426	
L	0.96	0.96	0.96	1224	
R	0.98	0.99	0.99	1166	
V	0.93	0.95	0.94	1105	
A	0.87	0.77	0.82	243	
accuracy			0.98	15164	
macro avg	0.95	0.93	0.94	15164	
weighted avg	0.98	0.98	0.98	15164	

Figure IV.13 : Rapport de classification Scenario 1.

Class	N	L	R	V	A
Accuracy	0.9914	0.9649	0.9940	0.9493	0.7737

Tableau 4.14 : L'accuracy pour chaque classe de S1.

Scenario 2:

- ✓ Nombre total de: 30183 échantillons
- ✓ Entrainement distribution: {'N': 22851, 'L': 2368, 'R': 2268, 'A': 486, 'V': 2210}

	precision	recall	f1-score	support
N	0.99	0.99	0.99	11426
L	0.97	0.97	0.97	1224
R	0.97	1.00	0.98	1166
V	0.94	0.94	0.94	1105
A	0.92	0.81	0.86	243
accuracy			0.99	15164
macro avg	0.96	0.94	0.95	15164
weighted avg	0.99	0.99	0.99	15164

Figure IV.14 : Rapport de classification Scenario 2.

Class	N	L	R	V	A
Accuracy	0.9944	0.9665	0.9957	0.9412	0.8107

Tableau 4.15 : L'accuracy pour chaque classe de S2.

Scenario 3:

- ✓ Nombre total de : 20122 échantillons
- ✓ Entrainement distribution: {'N': 215234, 'L': 1579, 'R': 1512, 'A': 324, 'V': 1473}

	precision	recall	f1-score	support
N	0.99	0.99	0.99	11426
L	0.93	0.97	0.95	1224
R	0.94	1.00	0.97	1166
V	0.92	0.93	0.92	1105
A	0.91	0.67	0.78	243
accuracy			0.98	15164
macro avg	0.94	0.91	0.92	15164
weighted avg	0.98	0.98	0.98	15164

Figure IV.15 : Rapport de classification Scenario 3.

Class	N	L	R	V	A
Accuracy	0.9852	0.9681	0.9983	0.9276	0.6749

Tableau 4.16 : L'accuracy pour chaque classe de S3.

➤ Interprétation des Résultats et Discussion :

L'analyse des résultats montre que les performances du modèle sont influencées par la quantité et l'équilibre des données d'apprentissage. Le Scénario 2 obtient les meilleures performances globales (98,54 %), confirmant l'intérêt d'un meilleur équilibre entre les classes.

Les classes L et R conservent une précision élevée, même avec un nombre réduit d'échantillons, grâce à leurs caractéristiques morphologiques ECG distinctives. En revanche, la classe A présente une baisse notable du rappel dans le Scénario 3, traduisant une plus grande sensibilité au manque de données. Ces observations soulignent l'importance d'un ensemble d'apprentissage suffisamment représentatif pour garantir une classification fiable des différents types d'arythmies cardiaques.

IV.7. Conclusion :

Ce chapitre a été entièrement dédié à la phase expérimentale, à la présentation des résultats et à l'analyse comparative des différents modèles développés. Après avoir mis en place notre environnement de simulation et configuré le pipeline de traitement sur la base de données MIT-BIH Arrhythmia Database, nous avons évalué et confronté les performances des approches classiques de Machine Learning (SVM, RN) à celles du Deep Learning (CNN).

En résumé, les contributions et les constatations clés de ce chapitre se déclinent comme suit :

- **Mise en place de scénarios d'évaluation:** Nous avons validé nos modèles à travers différents scénarios expérimentaux (Scénarios, C1, G) afin d'analyser rigoureusement l'impact de la taille des échantillons (N) et de la distribution des données sur la qualité de l'apprentissage.
- **Limites de la séparabilité linéaire :** Les tests ont mis en évidence l'échec du SVM linéaire, qui a plafonné à **55,5 %** d'exactitude, prouvant que les descripteurs ECG ne sont pas linéairement séparables. L'introduction de la non-linéarité (via le SVM à noyau RBF et le réseau MLP) a permis de corriger ce biais et d'élever significativement les performances.
- **Suprématie du Deep Learning :** Le modèle CNN s'est imposé comme la solution la plus performante et la plus robuste de notre étude. En configuration optimale, il a atteint une exactitude globale (Accuracy) remarquable de **98,54 %**, confirmant l'immense avantage de l'extraction automatisée des caractéristiques directement à partir du signal brut.

Conclusion générale :

Dans ce travail, nous avons proposé une approche complète pour la classification automatique des arythmies cardiaques à partir de signaux ECG. Ce système repose sur un pipeline rigoureux comprenant le prétraitement des signaux bruts (filtrage, normalisation et segmentation), l'extraction de caractéristiques morphologie, et enfin la classification. Afin d'identifier la solution la plus performante pour cette problématique médicale critique, nous avons implémenté, testé et comparé différents algorithmes de Machine Learning (RN, SVM) et de Deep Learning (Réseaux de neurones convolutifs CNN).

En réponse directe à la problématique centrale de ce mémoire nos travaux apportent des conclusions claires :

- L'utilisation concrète de l'IA sur les signaux ECG nécessite impérativement de passer par une chaîne de traitement numérique (nettoyage du bruit, filtrage, segmentaion....etc) pour que les algorithmes puissent "lire" le signal sans biais.
- La différence fondamentale entre les algorithmes testés réside dans leur capacité géométrique à gérer la complexité et le chevauchement des anomalies cardiaques. Les modèles de Machine Learning classiques (comme le SVM Linéaire) échouent face à la non-séparabilité des descripteurs ECG, se heurtant à un plafond de verre structurel à **55,5 %** d'accuracy. À l'inverse, les algorithmes non linéaires (comme le SVM RBF et le RN) s'adaptent beaucoup mieux à la distribution des données, montrant une performance strictement croissante avec l'augmentation de la taille de l'échantillon (N). Enfin, le Deep Learning (CNN) se distingue radicalement des autres en éliminant le besoin d'extraire manuellement les caractéristiques, il apprend à extraire ses propres filtres discriminants directement à partir du signal brut.

Dans notre cas, les expérimentations ont clairement démontré que l'architecture CNN offre les meilleurs résultats par rapport aux autres approches. En configuration optimale (Scénario 2), le CNN a atteint une Accuracy globale remarquable de **98,54 %** et un F1-score de **0,95**, surpassant les réseaux de neurones classiques (MLP) et les SVM.

L'étude a notamment mis en lumière les limites du SVM linéaire, qui a stagné à **55,5 %** d'exactitude en raison de la non-séparabilité linéaire des descripteurs ECG, tandis que le noyau RBF et le CNN ont su capturer avec brio la complexité non linéaire et les déformations morphologiques des complexes QRS.

D'après les expériences répertoriées dans notre propre pratique expérimentale, nous constatons qu'il est indispensable de tester plusieurs paradigmes algorithmiques pour répondre efficacement à une problématique donnée. Il n'existe pas de règle générale stipulant qu'un algorithme est universellement meilleur qu'un autre; si le SVM RBF s'avère robuste et stable sur des effectifs réduits, le Deep Learning (CNN) excelle dans l'extraction automatisée de caractéristiques de haut niveau, bien qu'il reste sensible au fort déséquilibre des classes (*Imbalance Data*) lorsque le volume de données chute.

Grâce aux avancées fulgurantes de l'intelligence artificielle, le Machine Learning et le Deep Learning nous permettent aujourd'hui de modéliser des phénomènes physiologiques hautement complexes. Dans le domaine de la santé, où une fraction de seconde et une précision absolue peuvent sauver des vies humaines, l'automatisation du diagnostic de l'ECG s'impose comme un outil d'aide à la décision indispensable pour seconder les cardiologues et optimiser la prise en charge des pathologies cardiaques.

Pour finir, ce PFE a été une opportunité exceptionnelle d'élargir notre horizon scientifique, de confronter nos connaissances théoriques aux exigences de la pratique et de nous initier de manière rigoureuse à la recherche scientifique.

Références :

- [1]. <https://world-heart-federation.org/what-is-cvd>
- [2]. <https://www.biosample.fr/blog/intelligence-artificielle-dans-le-secteur-de-la-sante>
- [3]. Obraska P. Perlemuter L. Quevauvilliers J Médecine, appareil cardiovasculaire, Edition Masson Tome II 1968.
- [4]. Goldberger, Ary L ; clinical electrocardiography : a simplified approach.2013.
- [5]. <https://www.elsan.care/fr/pathologie-et-traitement/maladies-cardiovasculaires/arythmie-cardiaque-definition-causes-traitement>
- [6]. Enderle, Joseph Bronzino, ‘ Introduction to biomedical signal processing’, Third Edition, Academic Press Series in Biomedical Engineering, 2012.
- [7]. Goldberger, Ary L ; clinical electrocardiography : a simplified approach.2013.
- [8]. L. Breiman, ‘Random forests’, Machine learning, vol. 45, no. 1, pp. 5–32, 2001
- [9]. R. Benali, ‘ Analyse du signal ECG par réseau adaptif d’ondelettes en vue de la reconnaissance de pathologies cardiaques ‘, Doctorat en Sciences, Université Abou Bekr Belkaid, Avril 2013.
- [10]. <https://www.sciencedirect.com/science/article>
- [11]. <https://www.netapp.com/fr/artificial-intelligence/what-is-artificial-intelligence/>
- [12]. <https://binaintelligence.com>
- [13]. <https://iartificial.blog>
- [14]. <https://www.thats-ai.org/fr-CH/units/l-ia-peut-apprendre>
- [15]. <https://www.sap.com/africa/resources/what-is-machine-learning>
- [16]. <https://www.oracle.com/artificial-intelligence/machine-learning/what-is-machine-learning/#how-machine-learning-works>
- [17]. <https://www.databricks.com/fr/blog/supervised-vs-unsupervised-learning>
- [18]. <https://cloud.google.com/discover/what-is-unsupervised-learning?hl=fr>
- [19]. <https://www.coursera.org/learn/unsupervised-learning-recommenders-reinforcement-learning?>
- [20]. <https://www.cognizant.com/us/en/glossary/neural-network>
- [21]. <https://www.ibm.com/docs/en/spss-modeler/saas?topic=models-about-svm>
- [22]. <https://www.geeksforgeeks.org/machine-learning/support-vector-machine-algorithm>
- [23]. <https://www.elastic.co/what-is/knn>
- [24]. <https://www.futura-sciences.com/tech/definitions/intelligence-artificielle-deep-learning-17262/>
- [25]. <https://www.geeksforgeeks.org/deep-learning/convolutional-neural-network-cnn-in-machine-learning/>
- [26]. <https://www.lemagit.fr/conseil/Reseaux-de-neurones-RNN-et-CNN-comment-les-differencier>
- [27]. <https://deeptai.org/machine-learning-glossary-and-terms/radial-basis-function-network>
- [28]. John G. Webster, ‘ Medical Instrumentation: Application and Design’, Wiley, 2010.
- [29]. George B. Moody and Roger G. Mark, “The impact of the MIT-BIH Arrhythmia Database,” IEEE Engineering z in Medicine and Biology Magazine, vol. 20, no. 3, pp. 45–50, 2001.
- [30]. Rangayyan Rangaraj M., ‘Analyse des signaux biomédicaux : une approche par étude de cas’, Wiley-IEEE Press, 2002.
- [31]. S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, 1999.
- [32]. Zidelmal Zahia. ‘Reconnaissance d’arythmies cardiaque par Support Vector Machines’, Thèse de Doctorat, Université Mouloud Mammeri-Tizi Ouzou. 2012.
- [33]. Rangayyan Rangaraj M., ‘Biomedical Signal Analysis: A Case-Study Approach’. Christopher M. Bishop, ‘Pattern Recognition and Machine Learning’, Springer, 2006.
- [34]. Wenhao Li and Yuan Zhou, ‘Data Interpretation Methods and Biomedical Applications, Department of

Biomedical Informatics' , Beijing 100191, China.

[35]. Universite de Constantine, Doctorat en Sciences, Par : Mohamed Lamine TALBI, 2011.

[36]. Article de Institut de Recherche en Informatique, Mathématiques, Automatique et Signal (IRIMAS) UR 7499 12 rue des Frères Lumière, "Classification des signaux ECG" par Yacine BELGUERMI.

[37]. <https://www.python.org>

[38]. <https://www.tensorflow.org>

[39]. <https://scikit-learn.org>

[40]. <https://numpy.org>

[41]. <https://matplotlib.org>

[42]. <https://wfdb.readthedocs.io>

[43]. <https://www.mygreatlearning.com/blog/confusion-matrix-an-overview-with-python-and-r/>

[44]. <https://www.ultralytics.com/fr/glossary/confusion-matrix>

[45]. <https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall?>