

N° d'ordre :
N° de série :

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



UNIVERSITE ECHAHID HAMMA LAKHDAR - EL OUED
FACULTÉ DES SCIENCES EXACTES
Département D'Informatique



Mémoire de Fin D'étude
Présenté pour l'obtention du Diplôme de

MASTER ACADEMIQUE

Domaine : **Mathématique et Informatique**
Filière : **Informatique**
Spécialité : **Systèmes Distribués et Intelligence Artificielle**

Présenté par :

- **BERHOUM Adel**
- **BELHADI Mohamed Tahar**

Thème

**Apprentissage automatique des maladies
cardiaques dans le système Big data**

Soutenue le : 22 Juin 2019 devant le jury:

Dr. Lejdel Brahim	MCA	Président
M. Khelaifa Abdennacer	MAA	Rapporteur
M. Naoui Med Anouar	MAA	Encadreur
Dr. Mahmoudi Fayçal	Cardiologue	Co-Encadreur

Année Universitaire: 2018/2019

Remerciement

*Tout d'abord nous remercions notre Dieu qui nous a donné
la force et la volonté pour élaborer ce travail.*

Nous adressons nos vifs remerciements à notre encadreur

M. « NAOUI Med Anouar »

et le Co-encadreur

Dr . « MAHMOUDI Fayçal » Cardiologue ,

*qui n'ont cessé de fournir des conseils et des orientations et ne nous ont
pas privés de temps.*

*Nous adressons nos sincères remerciements à tous les enseignants et à
toutes les personnes qui, par leurs paroles, leurs écrits, leurs conseils
et leurs critiques, ont fait part de leurs réflexions et ont accepté
de nous rencontrer et de répondre à nos questions
lors de nos recherches.*

*Notre reconnaissance va aussi à tous ceux qui ont collaboré
à notre formation en particulier les enseignants du
département d'Informatique,
Universitaire Echahid Hamma Lakhdar
El-Oued.*

*Aussi à nos collègues de la promotion 2018-2019 On
remercie également tous ceux qui ont participé de près ou
de loin à élaborer ce travail.*



Résumé

Les maladies cardiaques en général et la Fibrillation ventriculaire (FV) en particulier sont des troubles du fréquence rythme cardiaque associés à des conséquences fatales et à la mort subite. En conséquence, les outils et les méthodes jouent un rôle essentiel dans la classification, la prédiction et l'analyse des maladies cardiaques.

Nous proposons un système d'apprentissage automatique qui classe différents types d'arythmie et encourage les cardiologues à prendre les mesures préventives appropriées. Le système détecte et classe automatiquement les arythmies dans les enregistrements ECG. Le modèle à préparer est lié par des systèmes distribués pour attirer l'ECG et l'envoyer aux médecins.

L'ECG, formé de fréquence de rythme cardiaque, sera classé et prédit par l'algorithme d'apprentissage automatique dans le système Big data selon la description de la maladie dans les normes internationales AAMI pour lire l'ECG [49]. L'objectif de système est d'aider les cardiologues à prendre la décision appropriée en bon moment.

Mots Clés

Les maladies cardiaques , Fibrillation ventriculaire, ECG, L'apprentissage automatique, Big Data



Abstract

Heart disease in general and Ventricular fibrillation (VF) in particular are cardiac rhythm disorders associated with fatal consequences and sudden death. As a result, tools and methods play a vital role in the classification, prediction and analysis of heart disease.

We propose a Machine learning system that diagnoses different types of arrhythmias and encourages cardiologists to take appropriate precautions. The system automatically detects and classifies arrhythmias in ECG recordings. The model to be prepared is linked by distributed systems to attract the ECG sent by the patients to the doctors.

The ECG training heart rate frequency will be class and predicted by the Machine learning algorithm in the Big Data system and will meet the description of the disease according to international AAMI standards for reading the ECG [49]. The goal of the system is to help cardiologists to make the right decision at the right time.

Keywords

Heart Disease, Ventricular fibrillation, Electrocardiogram, Machine learning, Big Data .

ملخص

أمراض القلب بشكل عام والرجفان البطيني على وجه الخصوص هي اضطرابات ضربات القلب المرتبطة بالنتائج القاتلة والموت المفاجئ. نتيجة لذلك ، تلعب الأدوات والأساليب دورًا حيويًا في التنبؤ بأمراض القلب وإدارتها.

نقترح نظامًا للتعلم الآلي يقوم بتشخيص أنواع مختلفة من عدم انتظام ضربات القلب ويشجع أطباء القلب على اتخاذ التدابير الوقائية المناسبة ، يقوم هذا النظام باكتشاف وتصنيف عدم انتظام ضربات القلب تلقائيًا من خلال تسجيلات ECG وقراءتها. ويرتبط النموذج المنجز بالأنظمة الموزعة على شبكة الإنترنت. لتلقي ECG المرسل من قبل المرضى إلى الأطباء.

يتم التنبؤ بالمرض من خلال خوارزمية التعلم الآلي في نظام البيانات الكبيرة، وسوف يستجيب لوصف المرض بطريقة معينة لقراءة تخطيط القلب وفقا للمعايير الدولية AAMI [49]. الهدف من العملية هو مساعدة أطباء القلب على اتخاذ القرار المناسب في الوقت المناسب.

الكلمات المفتاحية :

أمراض القلب ، الرجفان الأذيني ، تخطيط القلب ، التعلم الآلة ، البيانات الضخمة

Sommaire

Sommaire	I
Liste des figures	IV
Liste des tableaux	VI
Liste des abréviations	VI
Introduction général	01
Chapitre I : Généralité sur électrocardiogramme (ECG)	
1. Introduction	03
1.1. Définition de cœur	03
1.2. Histoire de l'ECG	04
1.3. But de l'ECG	05
2. Principe de fonctionnement de L'ECG	06
2.1. L'électricité cardiaque	06
2.2. Cycle d'électrophysiologie cardiaque	06
2.3. Calibrage de l'enregistrement ECG	08
2.4. Calcul la fréquence cardiaque (Fc)	08
2.5. La méthode de 300	08
2.6. Comment trouver et calculer l'axe électrique de cœur ?	09
2.6.1. Méthode de calcul des axes.	09
2.6.2. Résultats attendu.	09
3. Les dérivations de L'ECG	10
3.1. Les six dérivations frontales	10
3.2. Les dérivations précordiales (standards)	11
3.3. Étapes d'analyse de tracé	11
4. Identifier l'anomalie dans l'ECG (le trouble du rythme)	12
4.1. Anomalies des ondes P, des complexes QRS et des ondes T	13
4.2. D'autres anomalies du segment ST et de l'onde T	13
4.3. ECG pour certaines maladies	14
5. Conclusion	16
Chapitre II : L'apprentissage automatique et le système Big data	
1. Introduction	18
1.1. Historique	18
1.2. Définition d'apprentissage automatique	19
1.3. Surapprentissage	19
2. Types des problèmes de l'apprentissage automatique	20
2.1. Apprentissage supervisé	20
2.2. Apprentissage non-supervisé	20
2.3. Régression vs classification	21
3. Différents algorithmes de l'apprentissage automatique	21
3.1. Régression linéaire	21
3.2. Machine à vecteurs de support (SVM)	22
3.3. Arbre de décision	22
3.4. Réseaux de neurones	23

3.5. Le bagging et le boosting	23
3.5.1. Le bagging	23
3.5.2. Le boosting	23
3.6. L'algorithme Random Forest	24
3.6.1. Avantages et inconvénients	24
3.6.2. Principe de L'algorithme	24
3.6.3. Applications de l'algorithme	24
4. Deep learning	25
5. Big data et Caractéristiques	25
6. MongoDB	26
6.1. Les différentes terminologies MongoDB	26
6.2. Les concepts de base MongoDB	27
7. Internet des objets	27
7.1. Les couches de l'internet des objets	27
8. Travaux connexes	28
9. Conclusion	29

Chapitre III : Conception

1. Introduction	31
2. Objectif	31
3. Description des données	31
3.1. Contenu des données	31
3.2. Les annotations des catégorie des classes	32
3.3. Signification des battements de chaque classe	32
3.4. Les arythmies MIT-BIH	34
3.5. Prétraitement des signaux ECG	34
4. Architecture générale du système	36
4.1. Couche IoT	36
4.2. Couche Big data	36
4.2.1. Pour quoi Big data ?	36
4.3. Couche ML	37
1. Architecture détaillée du système	37
1.1. Couche IoT	37
1.2. Couche Big data	38
1.3. Couche ML	39
1.3.1. L'algorithme Random Forest distribuée	39
1.3.2. Description de l'algorithme Random_Forest distribuée	39
2. Fonctions du système	40
3. Evaluation des algorithmes	41
3.1. L'ensemble des algorithmes du système	41
3.2. Les métriques utilisées au système	42
3.3. Description du résultat des métriques	43
4. Conclusion	46

Chapitre IV : Implémentation

1. Introduction	48
2. Environnement et Outils du développement	48
2.1. Langage python	48
2.1.1. Pour quoi python ?	48
2.1.2. Scikit-Learn	48
2.1.3. Pandas	49
2.1.4. NumPy	49
2.2. Outils Big Data	49
2.2.1. MongoDB	49
2.2.2. Fichiers de données	49
3. Matériel de configuration	49
4. Etapes d'implémentation	50
5. Résultats obtenues	50
5.1. Apprentissage automatique du modèle	50
5.1.1. Apprentissage automatique	50
5.1.2. Construction de modèle	54
5.2. Prédire les maladies	56
6. Comparaison entre les travaux connexes et notre travail	58
7. Conclusion	59
Conclusion général et perspectives	61
Bibliographié	63

Liste des figures

Chapitre I : Généralité sur électrocardiogramme (ECG)

Figure I.1: Schéma fonctionnel du cœur.	02
Figure I.2: Willem Einthofen, prix Nobel pour l'invention d'un dispositif ECG	04
Figure I.3: Electrocardiographe ECG Contec100G	04
Figure I.4: Electrocardiographe ECG Edan PADECG Numérique sans fil avec interprétation et tablette tactile androïde	04
Figure I.5: L'électrophysiologie cardiaque du cœur	05
Figure I.6: Intervalles de l'ECG normal, Incluant l'onde U	06
Figure I.7: Calibrage de l'enregistrement ECG	07
Figure I.8: Relation entre les carreaux du papier ECG et le temps	07
Figure I.9: Bloc de 5 petits carrés	07
Figure I.10: Les vecteurs d'activation cardiaque	08
Figure I.11: L'axe électrique d'activation cardiaque	08
Figure I.12: Placement des dérivations frontales	09
Figure I.13: Le tri axe des dérivations frontales	09
Figure I.14: Placement des dérivations sur le corps précordiales	10
Figure I.15: Sus-décalage du segment ST. b. Sous-décalage du segment ST	12
Figure I.16: ECG enregistré chez une jeune étudiante de 20 ans ; aucune anomalie décelée à l'examen clinique	13
Figure I.17: ECG d'un homme hospitalisé en unité de soins coronariens pour un infarctus du myocarde, qui brutalement	13
Figure I.18: ECG enregistré chez une femme de 70 ans de petite taille, en insuffisance cardiaque, dont les symptômes	14

Chapitre II : L'apprentissage automatique et le système Big data

Figure II.1: Architecture du machine learning	18
Figure II.2: Le concept de compromis biais-variance	18
Figure II.3: Apprentissage supervisé	19
Figure II.4: Apprentissage non supervisé	19
Figure II.5: Illustration de la différence entre régression linéaire et classification linéaire	20
Figure II.6: Régression linéaire	20
Figure II.7: Choix du séparation des deux classes	21
Figure II.8: Un modèle de graphe d'une arbre de décision	21
Figure II.9: Architecture réseaux de neurones	22
Figure II.10: Architecture bagging	22
Figure II.11: Architecture boosting	22
Figure II.12: Architecture de l'algorithme Random Forest	23
Figure II.13: La relation entre l'intelligence artificielle, le ML et le Deep learning	24
Figure II.14: Architecture MongoDB	25
Figure II.15: Architecture de l'Internet des objets, en bleu ; les contributions de l'IMS	26

Chapitre III : Conception

Figure III.1 : Battement d'ECG de la classe 0	31
Figure III.2 : Battement d'ECG de la classe 1	31
Figure III.3 : Battement d'ECG de la classe 2	32
Figure III.4 : Battement d'ECG de la classe 3	32
Figure III.5 : Battement d'ECG de la classe 4	32
Figure III.6 : Un exemple d'extraire un battement ECG d'une signal ECG de 10 ms	34
Figure III.7 : Architecture générale du système en	35
Figure III.8 : Architecture détaillée du système en couches	36
Figure III.9 : Architecture Clients / Serveur	37
Figure III.10 : Algorithme Random_Forest distribuée pour j nœuds de base de données	38
Figure III.12 : Prédire la classe du état du santé de cœur	39
Figure III.13 : Formule de prédictions de la classification	41
Figure III.14 : Formule de Score F1	41
Figure III.15 : Formule de Erreur quadratique moyenne	41
Figure III.16 : Précision de teste (Validation) du classification	42
Figure III.17 : Score F1 du test du classification	43
Figure III.18 : L'erreur quadratique moyenne de l'apprentissage.	44

Chapitre IV : Implémentation

Figure IV.1 : Architecture de l'algorithme Machine Learning	50
Figure IV.2 : Requêtes MongoDB pour importée le document d'entraînement	49
Figure IV.3 : Requêtes MongoDB pour importée le document de test	49
Figure IV.4 : Code Python pour connecté MongoClient avec MongoDB.	52
Figure IV.5 : Résultat de code Python Après la connexion de MongoClient avec MongoDB.	52
Figure IV.6 : Importation des bibliothèques du système	53
Figure IV.7 : Lecture et conversion de la base de données	53
Figure IV.8 : Résultat de lecture et conversion de la base de données.	53
Figure IV.9 : Configuration de l'ensemble des algorithmes d'apprentissage automatique.	54
Figure IV.10 : Appel des algorithmes pour l'apprentissage automatique avec les métriques pour crée les modèles.	54
Figure IV.11 : Affichage de numéros de deux premier modèles avec leurs métriques.	55
Figure IV.12 : Code Python du résultat final de vote.	55
Figure IV.13 : Résultat final du vote.	55
Figure IV.12 : Requêtes MongoDB pour importée des collections rythme cardiaque des patients	56
Figure IV.13 : Code Python pour prédit les classes de rythme cardiaque des patients	56
Figure IV.14 : Résultats de prédiction des classes de rythme cardiaque des patients	57
Figure IV.15 : Inverser la méthode d'extraction de battement pour le patient_1	57

Liste des tableaux

Chapitre I : Généralité sur électrocardiogramme (ECG)

Tableau I.1: Placez les dérivations bipolaires sur le corps	09
Tableau I.2: Placez les dérivations unipolaires sur le corps	09
Tableau I.3: Reconnaître les anomalies de l'ECG	11

Chapitre II : L'apprentissage automatique et système le Big data

Tableau II.1: Quelques exemples de Machine Learning proposés par Mitchell	17
Tableau II.2: Concepts de MongoDB vs concepts BD Relationnel	25

Chapitre III : Conception

Tableau III.1 : L'ensemble de données sur l'arythmie	30
Tableau III.2 : Regroupement des types de battement de cœur marqués MIT-BIH, selon les normes AAMI	31
Tableau III.3 : Les algorithmes d'apprentissage automatique du système	40
Tableau III.4 : Les résultats de métriques	42

Chapitre IV : Implémentation

Tableau VI.1: Comparaison du précision de la classification de ECG.	58
--	----

List des abréviations

FV	: Fibrillation Ventriculaire.
ECG	: Electrocardiogramme Ou Electrocardiographie.
BAV	: Bloc Arythmie Ventriculaire.
ML	: Machine Learning.
SVM	: Support Vectors Machine
NN	: Neural Networks.
CNN	: Convolutional Neural Network
DWT	Discrete Wavelet Transform
AAMI	Association for the Advancement of Medical Instrumentation
MIT- BIH	Laboratories at the Boston Hospital Institute and the Massachusetts Institute of Technology

Introduction général

Introduction général

La révolution des technologies de l'information et de la communication permet aux médecins de communiquer avec leurs patients pour diagnostiquer leurs cas sans avoir besoin de mobilité.

Au cours des dernières années, un nouveau terme appelé Big Data [18], est apparu dans la scène. Des nombreuses sociétés et des laboratoires scientifiques ont été créés pour le stockage des fichiers très volumineux, allant jusqu'à des téra d'octets et plus, et pour exploiter cette énorme quantité de données sous forme de fichiers, images, textes ou sons, par des scientifiques pour l'intérêt scientifique.

Les sociétés médicales et les grands hôpitaux ont établi leurs bases de données pendant des nombreuses années pour les exploiter en science et diagnostiquer la maladie à travers de nombreux exemples de patients et tenter de trouver la cause commune entre eux; cette énorme quantité de données est stockée et traitée par les outils Big Data [42] et les algorithmes intelligents appelés algorithmes d'apprentissage automatique [8]. Par conséquent, nous présentons les deux problématiques suivantes :

- Comment stocker des données sur les maladies cardiaques dans le système Big data?
- Comment construire un bon modèle d'apprentissage automatique pour prédire les classes des maladies cardiaques ?

Pour résoudre ces problèmes, nous avons procédé comme suit:

Nous commençons par le premier chapitre : nous présentons le cœur humain et des généralités de l'appareil ECG, de son enregistrement du rythme cardiaque et sa méthode de lecture, menant au diagnostic de la maladie dans certains cas.

Ensuite le deuxième chapitre : nous présentons un résumé de l'apprentissage automatique avec une explication de l'algorithme Random Forest et le Big Data système.

Dans le troisième chapitre : nous présentons la base de données sur laquelle nous avons travaillé et l'architecture proposée pour la résolution du problème et l'évaluation des algorithmes.

Enfin le dernier chapitre : nous mettons en place des techniques pour mettre en œuvre l'architecture proposée à l'aide des outils du système Big Data et l'apprentissage automatique.

Chapitre I

Généralité sur électrocardiogramme (ECG)

1. Introduction

Au cours de ce chapitre, et après avoir donné un aperçu de l'organe principal du corps humain, à savoir le cœur. Nous présentons la fonctionnement général du système cardiovasculaire, de manière plus détaillée, le principe de travail et d'identifier le trouble du rythme de l'électrocardiogramme (ECG). Normal ou anormal pour le rythme de cœur et la manière de diagnostiquer la majorité des cas inhabituels qui causent ces maladies.

1.1. Définition de cœur

Le cœur est un muscle qui travaille volontairement, qui se contractent et se relâchent en cadence d'une manière rythmique pour pomper le sang vers le corps, avec un rythme de 60 à 100 fois par minute pour une personne normal, déterminé par un groupe de cellules musculaires (**Figure I.1**). Ce rythme initialisé par le nœud sinusal qui produit un courant provoque des impulsions du cœur. Ces impulsions sont suivies et analysées par un électrocardiogramme l'ECG sans intervention chirurgicale, mais uniquement par des électrodes, qui adhèrent à la surface de la peau à des endroits spécifiques, pour diagnostic des nombreuses maladies. Les troubles du rythme cardiaque peuvent également affecter la fonctionnement de la pompe cardiaque. Un diagnostique qui détecte les arythmies et ses causes permet au médecins de choisir le diagnostique approprié [1].

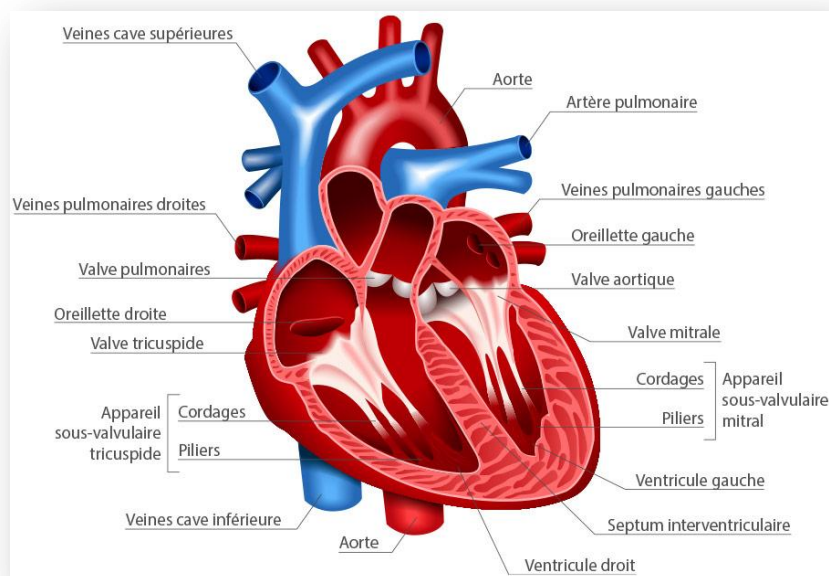


Figure I.1: Schéma fonctionnel du cœur.

1.2. Histoire de l'ECG

L'ECG a traversé plusieurs étapes historiques pour atteindre sa forme actuelle [2] :

- 1842 : Le physicien italien Carlo Matteucci montre le phénomène physiologique produit par le muscle de contraction, chaque contraction du cœur étant accompagnée d'un courant électrique.
- 1843 : Le physiologiste allemand Emil Dubois Reymond démontre une possibilité réelle de contraction musculaire, confirmant les recherches de Carlo Matteucci.
- 1870 : Alexander Moorhead, de l'hôpital St. Bartholomew de Londres, aurait enregistré un plan cardiaque électrique.
- 1878 : Les physiciens britanniques John Borden Sanderson et Frederick Big ont enregistré l'énergie électrique du cœur de la grenouille avec une échelle poétique à deux niveaux (appelée plus tard P, QRS et T).
- 1884 : John Burden Sanderson et Frederick Page publient certains de ces enregistrements sur les phénomènes d'excitation électrique au cœur de la tortue.
- 1887 : Le physicien britannique Augustus de. Walter, du Collège de médecine St. Mary à Londres, a publié le premier schéma électrique du cœur humain. Il a été enregistré sur Thomas Joswell, un technicien de laboratoire.
- 1891 : Les physiciens britanniques William Baillis et Edward Starling de College College London améliorent le niveau poétique. Reliez la pointe au bras droit et au cuir chevelu et visualisez le "contraste triphasé associé à chaque contraction du cœur". Ces écarts seront appelés plus tard P, QRS, T. Baillis, WM et Starling EH. Sur les différences électriques du cœur chez l'homme.
- 1895 : En utilisant une formule de correction électrique améliorée développée indépendamment par Porsche, Ainhoffen met en évidence cinq écarts décrits par P, Q, R, S et T.
- 1906 : Antofen publie le premier classement du GEC. Ventricule hypertrophique gauche et droit, hypertrophie auriculaire gauche et droite, éléments d'onde U sur QRS, contractions ventriculaires prématurées, ventricule jumeau ventriculaire, flutter auriculaire et masse auriculaire ventrale complète: normales et anormales.
- 1908 : Edward Schaffer, de l'Université d'Édimbourg, est le premier à acheter le galvanomètre à usage médical.
- 1910 : Le premier magazine américain sur les électrocardiogrammes, Walter James de l'Université Columbia et Harthy Williams de la Faculté de médecine de l'Université Cornell à New York. Ils décrivent l'hypertrophie ventriculaire, les spectres ventriculaires et ventriculaires et la fibrillation auriculaire ventriculaire.
- 1911 : Thomas Lewis publie The Heartbeat Machine à Londres: Shaw and Sons, dédié à Einhoven.
- 1920 : Harold Barty, New York, publie le premier infarctus du myocarde avec électrocardiogramme chez l'homme, décrivant l'onde T comme élargie. Barty Hipp. Signe ECG de blocage des artères coronaires.
- 1924 : Willem Einthoven, **prix Nobel** pour l'invention d'un dispositif ECG.
- 1928 : Le changement ECG de Frank Sanborn dans la première version portable de 50 livres fonctionne sur une batterie de voiture de 6 volts.
- 1938 : L'American Heart Association et la Heart Association de Grande-Bretagne déterminent la position standard des branches V1-V6 précondiales. AR Barnes, Pardee HEB, PD blanche. Et d'autres Unification des indices
- 1942 : Emmanuel Goldberger ajoute aux dérivés introverts de aRr, aVl, aVf. Cela lui permet, avec 6 V1-V6 pré-filetés, de réaliser le premier planificateur de cœur à 12 canaux, qui est encore utilisé aujourd'hui (**Figure I.3** et **Figure I.4** [5]).

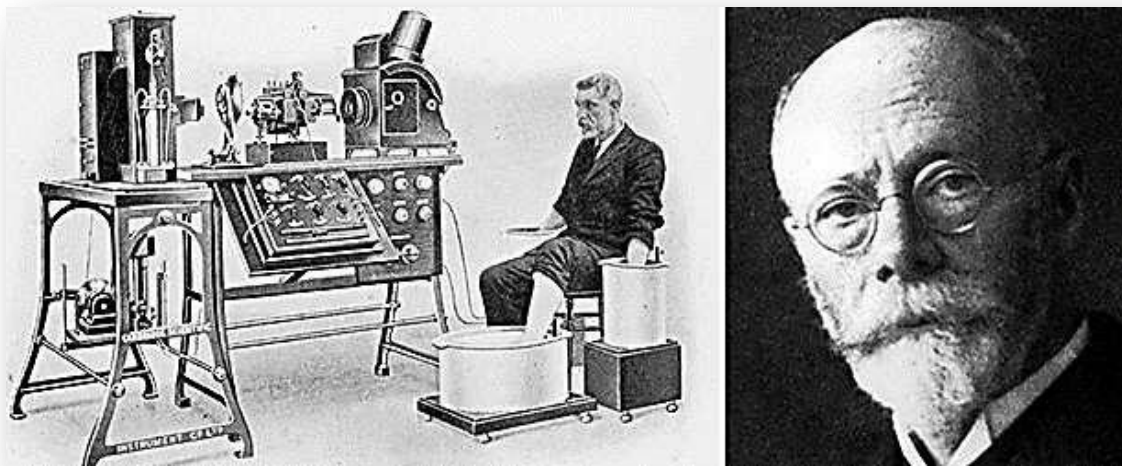


Figure I.2: Willem Einthoven, prix Nobel pour l'invention d'un dispositif ECG.

Willem Einthoven : (21 mai 1860- 29 septembre 1927, Pays-Bas) est un physiologiste néerlandais, prix Nobel de physiologie ou médecine de 1924, « pour sa découverte du mécanisme de l'électrocardiogramme » (**Figure I.2**) [3]. Son galvanomètre à cordes a rendu plus simple et pratique l'utilisation de l'électrocardiographie. [4]



Figure I.3: Electrocardiographe ECG Contec100G.



Figure I.4: Electrocardiographe ECG Edan PADECG Numérique sans fil avec interprétation et tablette tactile androïde.

1.3. But de l'ECG

Le but de l'électrocardiogramme, permet de mesurer et enregistrer l'activité électrique du cœur est de clarifier la régularité du rythme cardiaque. Dans des cas irréguliers, nous en déduisons, les défauts de conduction et le diagnostic précoce de nombreuses maladies cardiaques telles que le manque d'infarctus, ainsi que de déterminer l'ampleur du déséquilibre muscle cardiaque et électrolytique et la toxicité par des certains médicaments.

En termes généraux, l'électrocardiogramme permet de diagnostiquer une maladie cardiaque telle que l'angine de poitrine, les crises cardiaques et la cardiomyopathie, ce qui garantit un enregistrement à long terme de l'activité électrique du cœur [6].

2. Principe de fonctionnement de L'ECG

Nous devons être conscients que le diagnostic ECG est un outil et non une fin en soi. Son analyse dépend de quelques principes de base et de règles simples [2].

2.1. L'électricité cardiaque

La contraction des muscles du corps est provoquée par des modifications électriques appelées «dépoléarisation» et par des électrodes connectées au corps pouvant détecter ces modifications. Et le muscle cardiaque similaire aux muscles du corps peut détecter leurs modifications électriques à condition que le corps du patient soit complètement relaxé afin d'éviter les modifications électriques des muscles squelettiques [2].

2.2. Cycle d'électrophysiologie cardiaque

L'impulsion initiale de l'électrocardiogramme pour chaque cycle cardiaque provient des ondes de conduction rapide (faisceau du support), sont ensuite multipliées pour provoquer une contraction ventriculaire, suivie d'une repolarisation (figure : 1.4) [2].

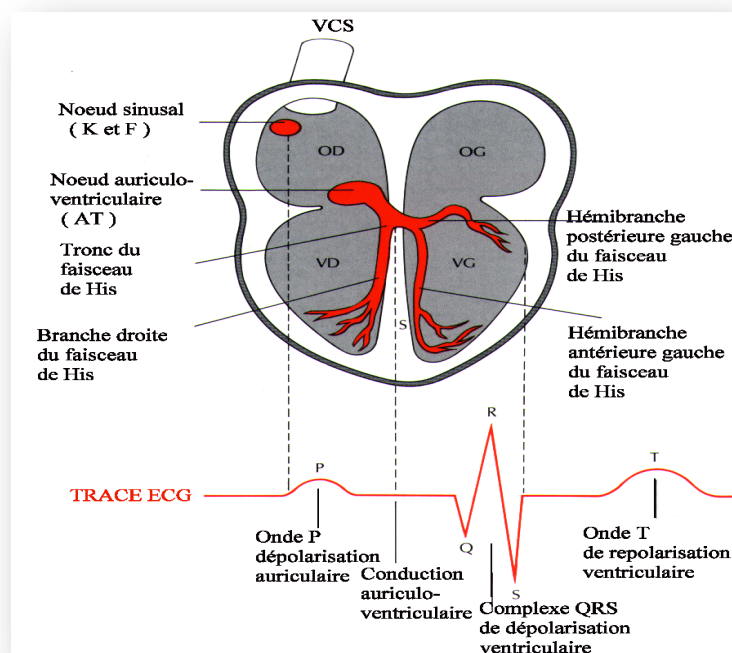


Figure I.5: L'électrophysiologie cardiaque du cœur.

Par conséquent, les rythme du cœur produit des ondes, des segments et des intervalles suivantes:

- **L'onde P** : l'onde de dépolarisation auriculaire, souvent positive, de faible amplitude (1 à 3 m volts) et $< 0,12$ seconde en D2 (**Figure I.5**).
- **Le complexe QRS** : c'est l'activation et à la dépolarisation des ventricules (**Figure I.6** [7]).
 - L'onde Q : première déflexion négative : activation septale.
 - L'onde R : première déflexion positive : activation pariétale du VG.
 - L'onde S : déflexion négative qui suit l'onde R : activation basale du VG.

L'intervalle de La durée QRS varie de 0,06 à 0,1 seconde (3 à 5 petits carreaux) et se mesure du début jusqu'à la fin du QRS.

- **Le segment PR** : sa durée d'1/10e de seconde entre l'activation auriculaire et l'activation ventriculaire, par le passage de l'influx du Nœud auriculo - ventriculaire au faisceau de His.

L'intervalle de l'onde P de fin du QRS et correspond à 0,03 à 0,04 seconde (moins de 2 petits carreaux).

- **L'onde T** : la durée de repolarisation ventriculaire. Son amplitude est inférieure à 2 mm.
- **Le segment ST** : La durée uniforme de d'excitation des ventricules jusqu'à la phase de récupération des ventricules. Sont intervalle de la fin de l'onde S ou R jusqu'au début de l'onde T (**Figure I.6**). si sa largeur plus d'1 mm par rapport à la ligne isoélectrique est anormal.
- **L'onde U** : repolarisation tardive de zones myocardiques d'amplitude inscrite entre de l'onde P et de l'onde T. $<$ à quartier de l'amplitude de l'onde T.
- **L'intervalle PR** : La durée de conduction auriculo-ventriculaire. C'est l'intervalle du début de l'onde P en allant jusqu'au début du QRS, il est de 0,12 à 0,23 seconde.

L'intervalle QT : C'est l'intervalle de dépolarisation (QRS) (**Figure I.6**), d'excitation (ST) et de repolarisation (T) des ventricules, du QRS jusqu'à la fin de l'onde T, le QT est fonction de la fréquence cardiaque; qui se calcule avec la formule de **Bazett** ($QTc = QT / Fc \leq 0,44$ seconde) QT corrigé c'est QTc [2].

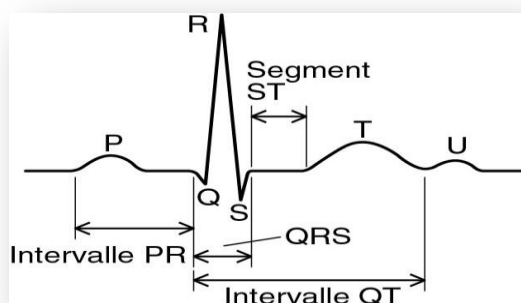


Figure I.6: Intervalles de l'ECG normal, Incluant l'onde U

2.3. Calibrage de l'enregistrement ECG

Avec la hauteur d'onde **P**, des complexes **QRS** et d'onde **T**, nous pouvons fournir une quantité limitée d'informations, mais nous devons nous concentrer sur l'étalonnage des performances. La tête d'enregistrement de 1 cm (2 grandes carreaux) doit être remplie d'un signal standard de 1 mV (**Figure I.7**), le signal de « calibrage » devant être inclus dans chaque enregistrement [7].

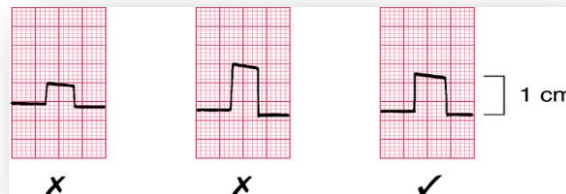


Figure I.7: Calibrage de l'enregistrement ECG.

2.4. Calcul de la fréquence cardiaque(F_c)

Les changements d'activité électrique sont enregistrés en plaçant une ligne sur une bande de papier millimétrique. Le taux d'enregistrement ECG standard est de 25 mm / s. Chaque grand carreau (5 mm) représente 0,2 seconde ou 200 millièmes de seconde (200 ms) (**Figure I.8**). En conséquence, il y a 5 grandes carreaux par seconde et 300 par minute [7].

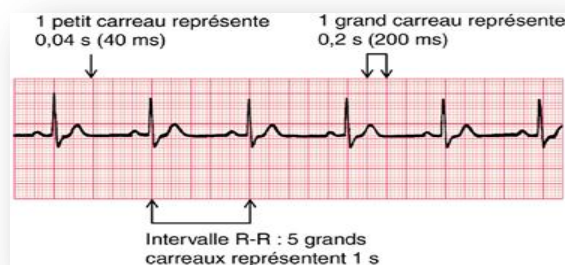


Figure I.8: Relation entre les carreaux du papier ECG et le temps.

2.5. La méthode des 300

Un petit carré correspond à 0,04 seconde et 60 secondes font 1500 petits carrés. Chaque bloc de 5 petits carrés est marqué en trait gras (**Figure I.9**). On repère une onde R coïncidant avec un trait gras qui deviendra un multiple des autres traits gras suivants. Entre 2 traits gras il y a 5 petits carreaux. [2]

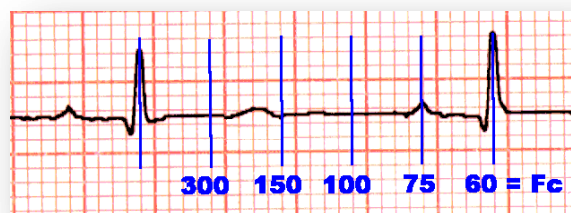


Figure I.9: Bloc de 5 petits carrés.

Donc l'onde R qui coïncide au trait gras, départ des 1500, le prochain, on aura 1500/5 soit 300, au suivant 1500/10 soit 150 puis 100, 75, 60, L'onde R suivante permet d'indiquer la fréquence cardiaque en relation avec le trait gras correspondant. C'est la méthode la plus rapide et la plus utilisée [2].

2.6. Comment trouver et calculer l'axe électrique de cœur ?

Chaque flèche dirigée représente l'activité électrique (**Figure I.10**), les vecteurs dans la barrière et les ventricules représentent les activités de la barrière et du ventricule. Lors de l'ajout de tous les vecteurs au bloc et aux ventricules par rapport à l'arrangement, cela permet de tracer un vecteur moyen appelé axe moyen du QRS (**Figure I.11**) [2].

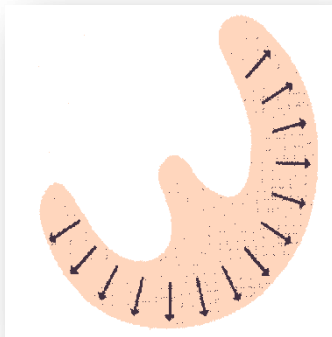


Figure I.10: Les vecteurs d'activation cardiaque.

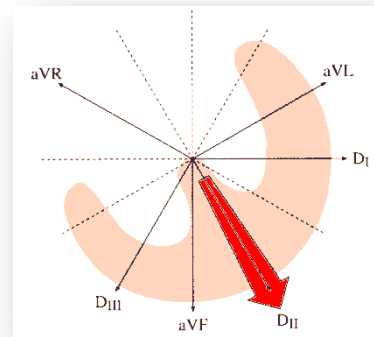


Figure I.11: L'axe électrique d'activation cardiaque.

2.6.1. Méthode de calcul des axes :

- Mesurer l'amplitude de QRS dans deux isotopes.
- Retardez ces valeurs sur les axes correspondant au circuit trigonométrique pour obtenir le résultat vectoriel.
- Ajouter la valeur positive de R aux valeurs négatives de Q et S pour trouver l'amplitude du QRS.

2.6.2. Résultats attendu :

- L'axe moyen du QRS normal est compris entre 0 et + 90°.
- L'axe est dit gauche quand l'angle est compris entre 0 et -90°.
- L'axe est dit droit quand l'angle est compris entre +90° et -90°.

3. Les dérivations de L'ECG

Le cœur est caractérisé de 12 dérivations [2] :

3.1. Les six dérivations frontales : DI, DII, DIII, VR, VL, VF.

○ D1, D2, D3 sont des dérivations bipolaires qui traduisent la différence de potentiel entre deux membres (Tableau I.1).

DI	: Entre bras droit (pôle -) et bras gauche (pôle +)
DII	: Entre bras droit (pôle -) et jambe gauche (pôle +)
DIII	: Entre bras gauche (pôle -) et jambe gauche (pôle +)

Tableau I.1: Placement des dérivations bipolaires sur le corps.

○ aVR, aVL, et aVF, sont les dérivations unipolaires chacun correspond au membre auquel (Tableau I.2), selon la théorie de Wilson et Golberger, le l'électrode exploratoire positif correspond à la pointe appliquée (Figure I.12 et Figure I.13), et le voltage est amplifiée pour obtenir la même amplitude que DI, DII et DIII.

aVR	: Respectivement le bras droit
aVL	: Respectivement le bras gauche
aVF	: Respectivement la jambe gauche

Tableau I.2: Placement des dérivations unipolaires sur le corps.

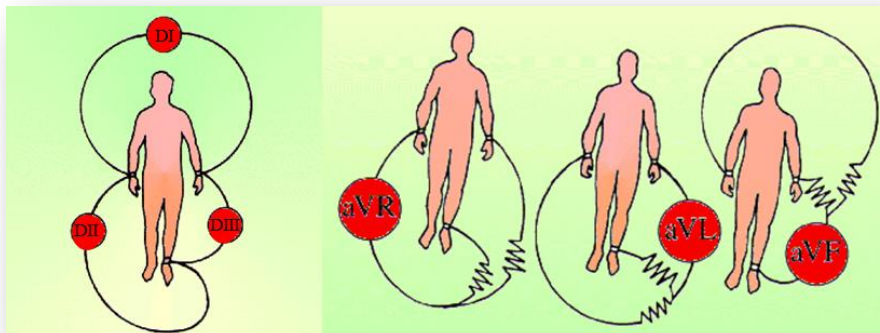


Figure I.12: Placement des dérivations frontales.

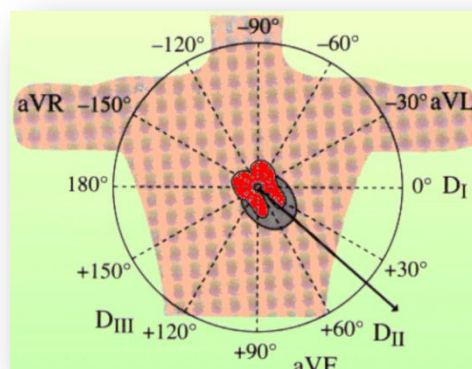


Figure I.13: Le tri axe des dérivations frontales .

3.2. Les dérivations précordiales (standards) : V1 à V6, sont unipolaires fixées en des points définis sur la thoracique (de 1 à 6) par la méthode qui désignés par Wilson (Figure I.14) [2].

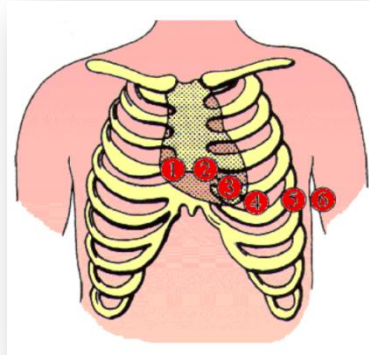


Figure I.14: Placement des dérivations sur le corps

3.3. Étapes d'analyse de tracé :

- Allumez l'appareil et ajustez sa vitesse: 25 mm / s.
- Positionnement correct des électrodes.
- Eviter les parasites dans les tracés électriques
- Méthodisme et lecture stricte:
- Lire les dérivations de D1 à V6.
- Lire de droite à gauche.
- Les assemblages de réalité virtuelle doivent être négatifs
- En cas de doute, l'ECG doit être redessiné et comparé au précédent.

Les dérivations DII et V1 sont les plus utilisées sur l'électrocardiogramme (ECG) pour l'étude du rythme cardiaque [2].

4. Identifier l'anomalie dans l'ECG (le trouble du rythme)

L'anomalie électro-cardiographique, on le voit une fois, et on ne l'oublie jamais (**Tableau I.3**) [7]. Cependant, dans les cas difficiles, il est utile de se poser les questions suivantes :

- L'anomalie est-elle occasionnelle ou soutenue ?
- Voit-on des ondes P ?
- Existe-t-il autant de complexes QRS que d'ondes P ?
- Les ventricules se contractent-ils régulièrement ou irrégulièrement ?
- La morphologie des complexes ventriculaires est-elle normale ?
- Quelle est la fréquence ventriculaire ?

Anomalies	Ondes P	Rapport P/QRS	QRS			Rythme
			Régularité	Morphologie	Fréquence	
Occasionnelles (ex : extrasystoles)				Normale		Supra ventriculaire
				Anormale		Ventriculaire
Soutenues	Présentes	P/QRS = 1/1	Régulier	Normale	Normale	Rythme sinusale
			Légèrement irrégulier	Normale	$\geq 150/\text{min}$	Tachycardie atriale
					Normale	Arythmie sinusale
		Plus d'ondes P que de complexes QRS	Régulier	Normale	Lente	Echappement auriculaire
					Rapide	Tachycardie atriale avec bloc
					Lente	BAV deuxième degré
	Absentes		Régulier	Anormale	Lente	BAV complet
				Normale	Rapide	Tachycardie jonctionnelle
					Lente	Echappement jonctionnel
			Irrégulier	Anormale	Rapide	Tachycardie jonctionnelle avec bloc de branche ou tachycardie ventriculaire
				Normale	Quelle que soit-la fréquence	Fibrillation auriculaire
			Irrégulier	Anormale	Quelle que soit-la fréquence	Fibrillation auriculaire avec bloc de branche
						Fibrillation ventriculaire ou arrêt cardiaque
		Absence de complexes QRS				

Tableau I.3: Reconnaître les anomalies de l'ECG.

4.1. Anomalies des ondes P des complexes QRS et des ondes T

Une fois que les capteurs ECG du patient sont connectés et utilisés, une charge cardiaque est générée. Pour interpréter ce diagramme, trouvez les réponses aux questions suivantes:

- L'anomalie de l'onde P ?

Peuvent être :

- Normales ;
- Une hauteur inhabituelle ;
- Une largeur inhabituelle.
- Trouver l'axe cardiaque ? (Regardez les complexes **QRS** en **DI**, **DII** et **DIII**) ?
- La durée du complexe **QRS** si il normale ?
- L'anomalie des complexes **QRS** – exactement l'anormales des ondes **Q** ?

Peuvent être :

- Trop larges;
- Trop hauts;
- Présenter une onde **Q** anormale.
- Le segment **ST** (Figure I.15) est-il « sus » ou « sous »- décalé ?

Peuvent être :

- Normal;
- Sur-élevé;
- Sous-décalé.
- L'onde **T** elle normale ?

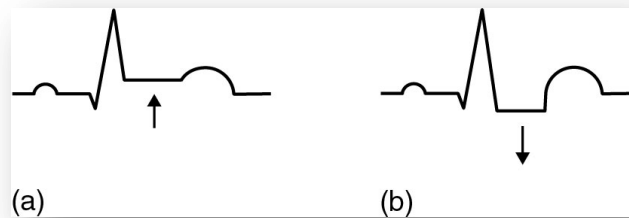


Figure I.15: Sus-décalage du segment ST. b. Sous-décalage du segment ST.

4.2. D'autres anomalies du segment ST et de l'onde T

L'ECG est affecté par les niveaux de potassium, de calcium et de magnésium dans le segment **ST** et de l'onde **T** :

- Aplatissement de l'onde **T** et apparition d'une bosse à la fin d'une onde **T** appelée "onde U" en raison d'un manque de potassium.
- Influencé hyperkaliémie donner un aspect de grandes ondes **T** pointues et la disparition du segment **ST**, et les complexes **QRS** peuvent être élargis.
- un allongement de l'intervalle **QT** provoque par Une hypocalcémie, et raccourcit ce dernier provoque par une calcémie élevée.

4.3. ECG pour certaines maladies

Avant de commencer, vous devez d'abord connaître l'état normal de l'ECG [7].

○ ECG 01 :

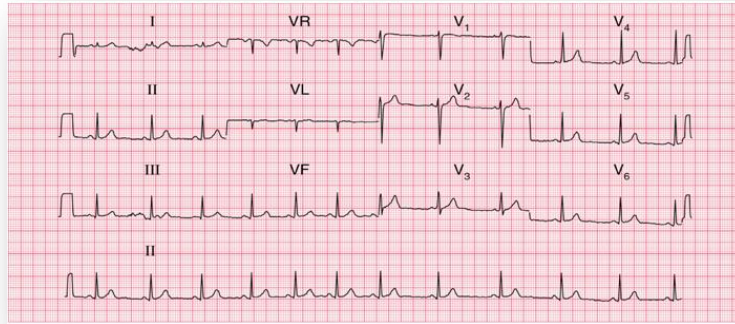


Figure I.16: ECG enregistré chez une jeune étudiante de 20 ans ; aucune anomalie décelée à l'examen clinique .

En déduira de cette ECG (Figure I.16):

- Arythmie sinusale bien visible sur le « long DII » ;
- Perturbation de la fréquence cardiaque (bien visibles en VF et V3) sont dues à l'arythmie sinusale ;
- Espace PR normal, 120 ms ;
- Axe de QRS normal ;
- Durée des complexes QRS : 80 ms ; leur amplitude est normale ;
- Segment ST isoélectrique dans toutes les dérivations ;
- Inversion de l'onde T uniquement en VR.

Conclusion: Ce tracé est en tous points parfaitement normaux.

○ ECG 02 :

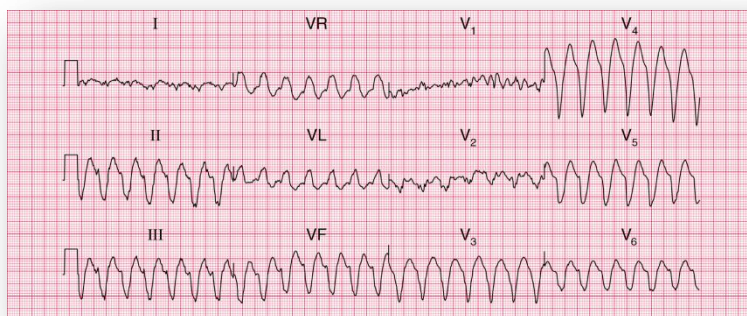


Figure I.17: ECG d'un homme hospitalisé en unité de soins coronariens pour un infarctus du myocarde, qui brutalement.

En déduira de l'ECG 02 (Figure I.17) :

- Tachycardie à complexes larges à 160/min ;
- Aucune onde p visible ;
- Déviation axiale gauche ;
- Durée des complexes QRS de 200 ms ;
- Tous les complexes QRS sont orientés vers le bas dans les dérivations précordiales ;
- Artéfacts en DI et V1-V2.

Description de l'ECG 02 : Ce patient est en œdème pulmonaire et demande un traitement d'urgence avec une tachycardie ventriculaire et demande un traitement d'urgence.

Soins au patient : prépare à la car diversion par choc électrique, après lui injecte administrer de la lidocaïne et du furosémide par voie intraveineuse [7].

○ **ECG 03 :**

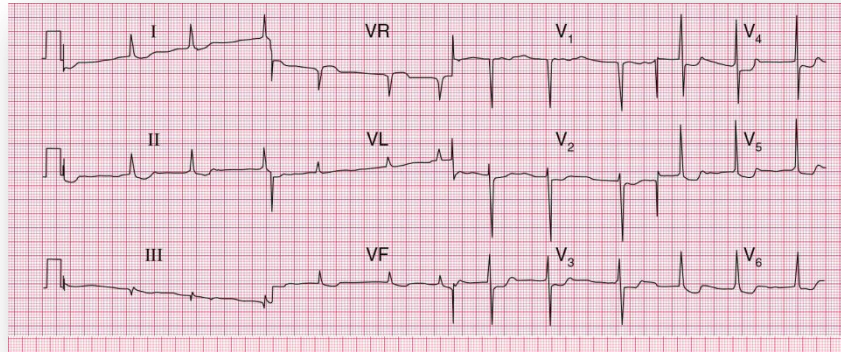


Figure I.18: ECG enregistré chez une femme de 70 ans de petite taille, en insuffisance cardiaque, dont les symptômes.

En déduira de l'ECG 03 (Figure I.18) :

- Fibrillation auriculaire ;
- Axe cardiaque normal ;
- Complexes QRS normaux ;
- Segment st incliné vers le bas, bien visible en v4-v6 ;
- Onde u bien visible en V2.

Description de l'ECG 03 : Le rythme totalement irrégulier avec complexes QRS fins est nécessairement dû à la fibrillation auriculaire. Les ondes U suggèrent une hypokaliémie

Soins au patient : La kaliémie doit être dosée en urgence. Un grand nombre de situations peuvent être associées à la fibrillation auriculaire mais, chez les sujets âgés en particulier, la maladie importante dont il faut se souvenir est la thyrotoxicose car, dans ce cas, la fibrillation auriculaire peut être la seule manifestation clinique de cette affection [7].

7 - Conclusion

À la fin de ce chapitre, il est clair que cet organe de base du corps humain, est un muscle, activé par le nœud sinusal au sommet de l'oreillette droite pour activer la circulation sanguine dans tout le corps, et entraîner des contractions cardiaques et une vie successives. Ces impulsions sont suivies et analysées par un électrocardiogramme sans intervention chirurgicale, au moyen d'électrodes qui adhèrent à la surface de la peau à des endroits spécifiques.

L'ECG assure un enregistrement à long terme de l'activité électrique du cœur, ce qui nous amène à conclure que ces entiers sont des séries chronologiques qui peuvent être analysées, classées et stockées pour nous fournir des bases de données utilisables pour diverses études de rythme cardiaque.

Chapitre II

L'apprentissage automatique

&

le système Big data

1. Introduction

Dans ce chapitre, nous devons être en mesure de formuler notre problème de santé sous la forme d'un problème humain en science des données. Ensuite, nous traduisons les éléments de notre problème humain qui peuvent être analysés par des algorithmes d'apprentissage automatique. En général, après la collection des données, nous devons savoir ce que nous voulons. Et comment ça se faire?

Des algorithmes d'apprentissage automatique peuvent être dirigé la présentation d'un problème de science des données selon un vocabulaire précis [8].

1.1. Historique

Au début de la seconde moitié du XXe siècle, les premières définitions de l'apprentissage automatique, attribuées à l'une des pionniers de l'intelligence artificielle, **Arthur Samuel**: « *définissent l'apprentissage de la machine comme un domaine d'étude permettant à l'appareil d'apprendre sans être programmé explicitement* » [9].

Ensuite en 1997 une définition plus précise, à **Tom Mitchell** de l'université de Carnegie Mellon: « *Un programme informatique permettant de tirer des leçons de l'expérience acquise en ce qui concerne certaines classes de mesure de la performance de T et P, si la performance de la P, avec E expérience améliorée* ». Cette définition est jointe à un tableau illustratif (**Tableau II.1**).

Cas d'application	Apprentissage des dames	Reconnaissance de l'écriture	Apprentissage de la conduite de robot
Tâches T	Jouer aux dames	Reconnaître et classer mots manuscrits à l'intérieure image	Conduite sur voie publique à quatre voies autoroutes utilisant la vision des capteurs
Performance mesure P	Pourcentage de jeux gagnés contre des adversaires	Pourcentage de mots correctement classifié	Distance moyenne parcourue avant une erreur (à en juger par surveillant humain)
Entraînement expérience E	Jouer à des jeux de pratique contre lui-même	Une base de données manuscrite mots avec donnés classifications	Une séquence d'images et commandes de direction enregistrées tout en observant un humain chauffeur

Tableau II.1: Quelques exemples de Machine Learning proposés par Mitchell.

1.2. Définition d'apprentissage automatique

L'apprentissage automatique (Machine learning) est une technologie d'intelligence artificielle (Figure II.1). C'est grâce à des programmes spécifiques qui permettent à la machine ou à l'ordinateur un apprentissage automatique selon des processus systématiques et simple pour pouvoir effectuer des opérations complexes [10].

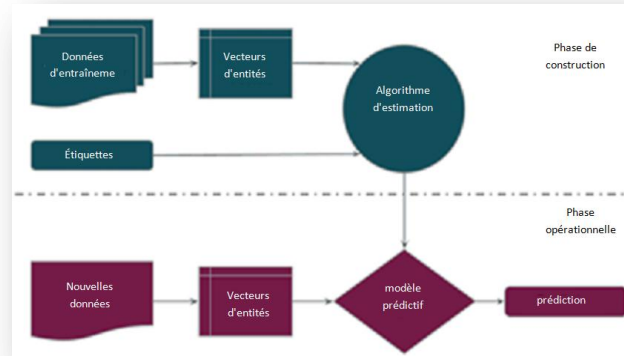


Figure II.1: Architecture du l'apprentissage automatique.

1.3. Surapprentissage

Il faux de croire que les performances d'un modèle dépendent des prédictions correctes établies sur le groupe de contrôle utilisé pour l'apprentissage, plus le résultat obtenu son bon. Nous cherchons donc à prédire ceux qui n'ont pas encore été observés et à juger de la qualité des algorithmes apprentissage automatique par sa capacité à faire de nouvelles prédictions après l'apprentissage (Figure II.2). Evitez donc un apprentissage excessif afin que le modèle ne soit pas complexe [15].

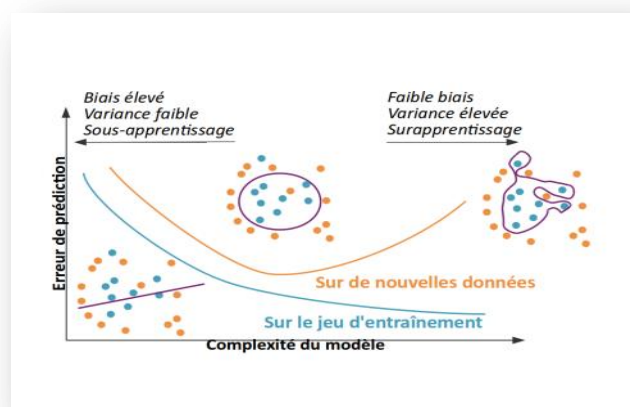


Figure II.2: Le concept de compromis biais-variance.

Pour éviter ce la, on divise les données disponibles en deux groupes distincts. Le premier sera l'ensemble d'entraînement, et le deuxième sera l'ensemble de test [15].

2. Types des problèmes de l'apprentissage automatique

L'apprentissage automatique utilisée dans des nombreux domaines d'application et chaque domaine contient une grande liste des classes de problèmes, de sorte qu'il existe une diversité dans la méthode d'apprentissage automatique.

2.1. Apprentissage supervisé

Nous avons une liste d'exemples étiquetés, c'est-à-dire accompagnés de la valeur à laquelle on peut s'attendre. Les étiquettes agissent en tant que "enseignant" et supervisent l'algorithme d'apprentissage pour savoir comment faire des prédictions. Il s'agit d'un apprentissage supervisé (Figure II.3). Il est le type le plus simple des problèmes de L'apprentissage automatique [8].

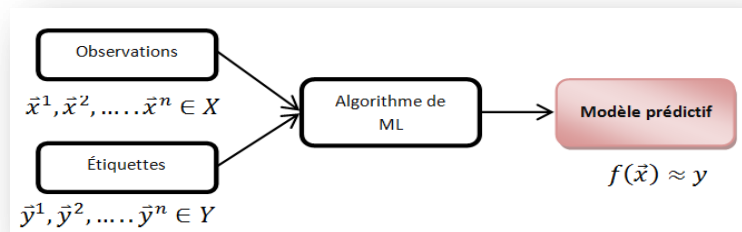


Figure II.3: Apprentissage supervisé.

Les données sont souvent définies dans les espaces $X = \mathbb{R}^p$. Il y a beaucoup de représentations, telles que classification et régression.

2.2. Apprentissage non-supervisé

Si un problème pouvant être formalisés comme suit : n observations : $\{X^i\}_{i=1, \dots, n}$ sont données à l'avance pour décrites dans un espace X, qui vérifie certaines propriétés en suivant une méthode d'apprendre une fonction c'est un branche du machine learning appelle l'apprentissage non supervisé (Figure II.4) [8] ; Comme (Partitionnement (Clustering), Estimation de la densité).

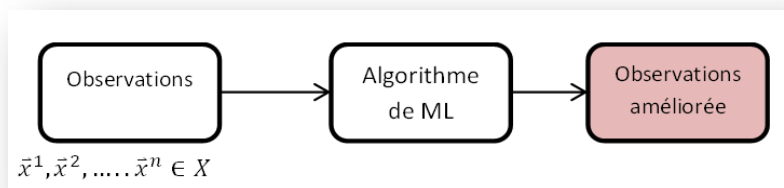


Figure II.4: Apprentissage non supervisé.

2.3. Régression vs classification

L'identifier de type de sortie attendu de notre programme nous aideons à choisir l'algorithme d'apprentissage automatique. Cela répond également aux questions suivantes: est-ce une valeur continue (un nombre)? Est-ce une valeur discrète (une catégorie)? Le premier cas s'appelle la régression et le second est la classification (**Figure II.5** [12]) [10].

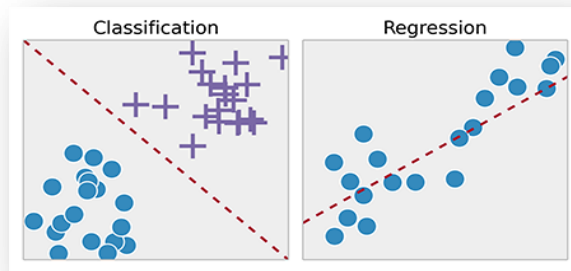


Figure II.5: Illustration de la différence entre régression linéaire et classification linéaire.

3. Les différent algorithmes de l'apprentissage automatique

Des algorithmes d'apprentissage automatique que nous estimons liés à notre problème de classification.

3.1. Régression linéaire

Fonction mathématique des algorithmes prédiction permettant de déterminer la relation entre les variables attendues et la variable cible [13].

- Régression linéaire (**Figure II.6**).
- Régression linéaire multivariée.

Une régression polynomiale peut entraîner des relations complexes qui ne sont pas nécessairement linéaires.

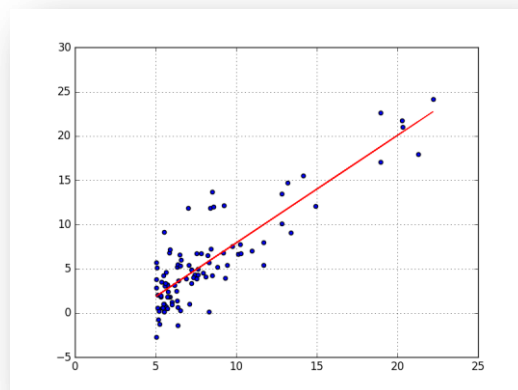


Figure II.6: Régression linéaire.

3.2. Machine à vecteurs de support (SVM)

Machine à vecteurs de support est un algorithme de classification binaire. Similaire à la régression logistique. Nous avons deux catégories dans notre figure:

- La régression logistique choisira la séparation des deux classes par le trait rouge.
- SVM choisira de séparer les deux classes par le trait vert.

Le SVM choisira la séparation la plus nette possible entre les deux classes (classifier aux marges larges) (**Figure II.7**) [13].

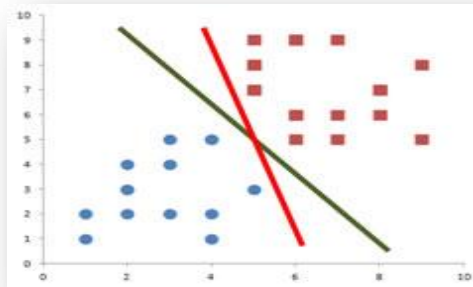


Figure II.7: Choix du séparation des deux classes.

3.3. Arbres de décision

L'arbre de décision est utilisé pour les problèmes de classification et de régression. Il permet de voir la logique d'interprétation des données (contrairement aux algorithmes de type boîte noire tels que SVM, NN,... etc.).

Un arbre de décision est un arbre dans lequel chaque nœud représente une caractéristique (attribut), chaque lien (branche) représente une décision (règle) et chaque feuille représente un résultat (**Figure II.8**) (valeur catégorique ou continue) [13].

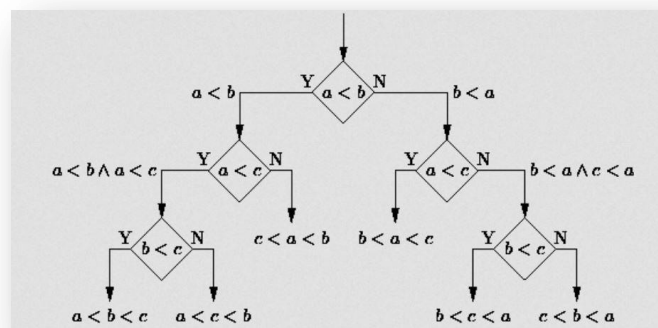


Figure II.8: Un modèle de graphe d'une arbre de décision.

3.4. Réseaux de neurones

Les réseaux de neurones sont des algorithmes avec des patterns complexes se trouvent dans les données. Après avoir apprennent sur les données de formation pour accomplir une tâche spécifique (Figure II.9 [6]). [13]

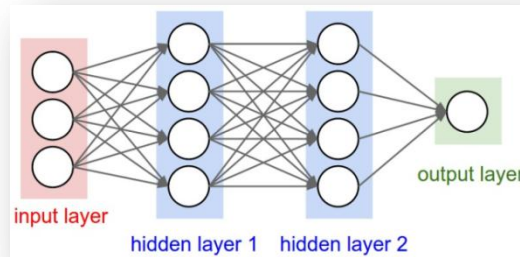


Figure II.9: Architecture réseaux de neurones.

3.5. Le bagging et le boosting

3.5.1. Bagging : Est une méthode permettant de générer plusieurs versions d'un prédicteur et les utilisées pour obtenir un prédicteur agrégé (Figure II.10). Bagging aide à réduire la variance par rapport les modèles qui peuvent être très précis, mais uniquement sur les données qui ont été formés [14].

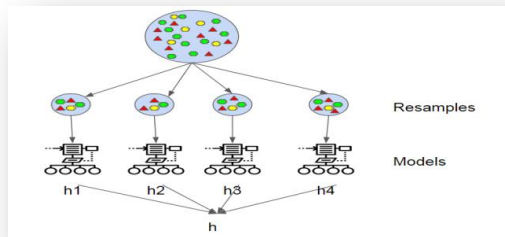


Figure II.10: Architecture bagging.

3.5.2. Boosting : Est un travail d'équipe. Chaque modèle exécuté détermine les fonctionnalités sur lesquelles le modèle suivant se concentrera (Figure II.11). Qui fait référence à un groupe d'algorithmes qui utilisent des moyennes pondérées pour transformer les apprenants faibles en apprenants plus forts [14].

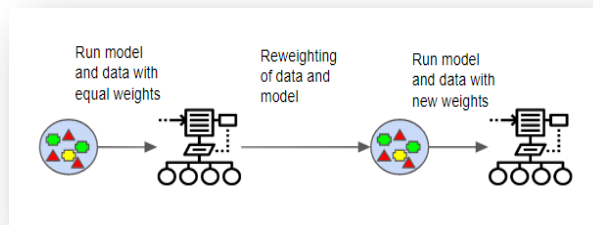


Figure II.11: Architecture boosting. [14]

Le bagging et le boosting sont deux techniques formidables pour réduire la variance.

3.6. L'algorithme Random Forest

L'algorithme Random Forest est un algorithme de classification supervisé. Ce qui consiste à créer une forêt de quelque soit la manière et la rendre aléatoire (**Figure II.12**). Il existe une relation directe entre le nombre d'arbres de la forêt et les résultats qu'elle peut obtenir: « quand le nombre d'arbres est grand, le résultat est précis » [24].

3.6.1. Avantages et inconvénients

- Pour les applications dans les problèmes de classification, l'algorithme Random Forest évitera le problème de sur-ajustement ;
- Le même algorithme peut être utilisé pour les tâches de classification et de régression ;
- Peut être utilisé pour identifier les entités les plus importantes à partir de l'ensemble de données d'apprentissage.
- Problème si nombre de variables pertinentes très faibles ;
- Déploiement d'un tel modèle reste compliqué.

3.6.2. Principe de L'algorithme

L'algorithme Random Forest généralement supervisé. Comporte de deux étapes: l'une est la création d'une forêt aléatoire, la deuxième consiste à effectuer une prévision à partir du classifieur de forêt aléatoire créé dans la première étape. il utilise un système de vote puis décide de la classe, il prend la moyenne de toutes les sorties de chacun des arbres de décision. Cela fonctionne bien avec les grands ensembles de données à haute dimensionnalité [24].

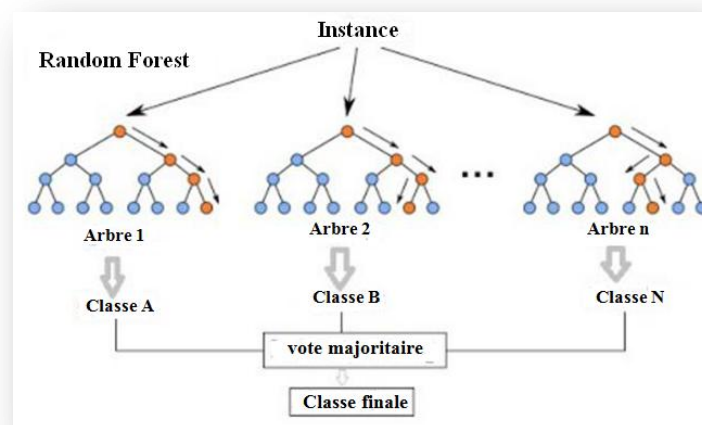


Figure II.12: Architecture de l'algorithme Random Forest.

3.6.3. Applications de l'algorithme: banque, médecine, bourse et commerce électronique.

4. Deep Learning

Deep Learning est un nouveau domaine de recherche en Machine Learning. Qui vise à rapprocher l'apprentissage automatique de l'un de ses objectifs initiaux « l'intelligence artificielle ». Cela dépend de la manipulation de grandes quantités de données en ajoutant des couches au réseau de neurone artificiel pour extraire les propriétés des données brutes à travers plusieurs couches de traitement composées de transformations linéaires et non linéaires [17].

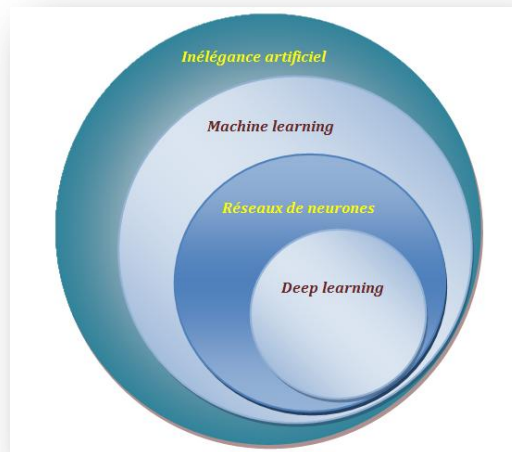


Figure II.13: La relation entre l'intelligence artificielle, le ML et le Deep learning.

5. Big data

Le Big data est un ensemble de données volumineux pouvant être analysés par l'ordinateur qui utilise des outils permettant de traiter ces données avec une manière totalement différente que les systèmes de gestion classique pour révéler des tendances et des associations [18].

5.1. Caractéristiques Big data

Règle des 5V [19] :

- Le volume : indique la masse d'informations produite chaque seconde.
- La vitesse : indique la rapidité de l'élaboration et du déploiement des nouvelles données.
- La variété : permet de faire l'analyse, la comparaison, la reconnaissance, le classement des données de différents types sur différents sites etc.
- La véracité : indique la fiabilité et la crédibilité des informations collectées.
- La valeur : indique le profit que l'on peut tirer de l'utilisation du Big Data.

6. MongoDB

Le système distribué MongoDB (Figure II.14), permet de positionner la donnée nécessaire avec disponibilité, la sociabilité et la conformité selon les besoins de client qui a la liberté d'utiliser MongoDB n'importe où, quel que soit sur un site ou d'une utilisation de MongoDB en tant que service dans le « Cloud » public. Le schéma suivant montre l'architecture MongoDB [20].

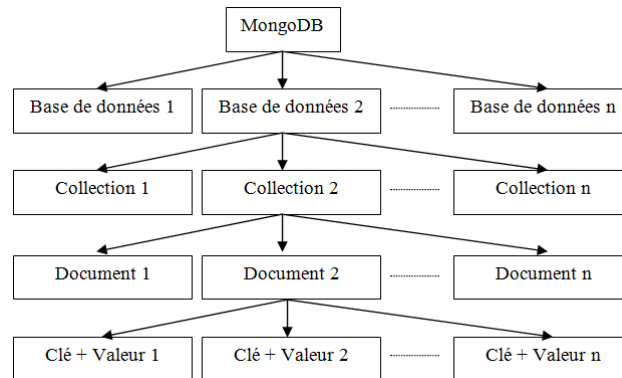


Figure II.14: Architecture MongoDB.

Les équivalentes terminologies de MongoDB par rapport base de données Relationnel :

base de données Relationnel	MongoDB
Base de données	Database
Tables	Collections
Lignes	Documents
Index	Index

Tableau II.2: Concepts de MongoDB vs concepts BD Relationnel.

6.1. Les différentes terminologies MongoDB

Les différentes terminologies nécessaires pour MongoDB (Tableau II.2) [20] sont les suivantes :

- **Base de données:** Un seul serveur MongoDB est constitué de diverses bases de données, chacune d'entre elles constituant un conteneur physique de collections.
- **Collection:** Une collection dans MongoDB est équivalente à une table de base de données et elle existe dans une base de données unique. Il comprend un groupe de documents MongoDB.
- **Document:** Le document peut être défini en tant qu'instance d'une collection MongoDB. Il comprend un ensemble de paires clé-valeur. Tous les documents incluent un

schéma dynamique, ce qui signifie que les documents comprenant la même collection n'ont pas besoin d'avoir le même ensemble de champs et la même structure.

6.2. Les concepts de base MongoDB

Pour apprendre MongoDB il faut comprendre 3 concepts de base de MongoDB :

- **Modèle de données:** MongoDB stocke les données sous forme de "documents", structures de données riches dans des bases de données.
- **Réplication:** MongoDB est conçu pour être exécuté sur plusieurs centres de données, avec un mécanisme de tolérance aux pannes intégré basé sur la réplication de bases de données.
- **Eclatement:** MongoDB est une technologie informatique distribuée. Cela signifie que plusieurs serveurs connectés via un réseau TCP / IP peuvent se présenter à une application comme une base de données "unique" "logique". Les applications client émettent les mêmes requêtes et commandes pour un système fonctionnant sur un seul serveur ou sur un cluster MongoDB à 1 000 nœuds.

7. Internet des objets (IoT)

Internet des objets (Internet of things : IoT) est la connexion d'un réseau Internet pour des divers objets et de leur combine . Pour ce faire, suivre les protocoles qui garantissent l'échange des informations et de données et assurent la communication entre les divers dispositifs [10].

7.1. Les couches de l'internet des objets

L'Internet des Objets organisés en trois couches principales:

- La couche perception : le matériel nécessaire pour parvenir à la collection de données contextuelles des objets connectés.
- La couche réseau : l'assurance de la connectivité inter-objets connectés et entre objets intelligents et les autres hôtes de l'Internet
- La couche application : définit les profils des services intelligents et les mécanismes de gestion de données de différents types

8. Travaux connexes

Voici quelques résultats des recherches pour la littérature sur l'ECG :

- En 2013, T. Lee et M. Chu ont achevé leurs recherches sur l'ECG sous le titre "*Application of higher order cumulant features for cardiac health diagnosis using ecg signals*". En utilisant les bases de données de l'entropie de paquets d'ondelettes. En utilisant l'algorithme de Machine à vecteurs de support (SVM) et ils ont obtenu une précision de 93.8 % [27].
- En 2016, R. J. Martis, U. R. Acharya, C. M. Lim, K. Mandana, A. K. Ray, et C. Chakraborty, ont achevé leurs recherches sur l'ECG sous le titre "*Étiquetage ECG utilisant l'entropie de paquets d'ondelettes et forêts aléatoires*". En utilisant les bases de données de l'entropie de paquets d'ondelettes. En utilisant l'algorithme de forêt aléatoire et ils ont obtenu une précision de 94,5% [26].
- En 2017, U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan et M. Adam, ont achevé leurs recherches sur l'ECG sous le titre "*Application of deep convolutional neural network for automated detection of myocardial infarction using ecg signals*". En utilisant les bases de données des signaux EEG. En utilisant l'algorithme Convolutional Neural Network (CNN) avec Augmentation et ils ont obtenu une précision de 93.5% [28].
- En 2018, Mohammad Kachuee, Shayan Fazeli, Majid Sarrafzadeh, Université de Californie à Los Angeles (UCLA), Los Angeles, États-Unis, ont achevé leurs recherches sur l'ECG sous le titre "*ECG Heartbeat Classification*". En utilisant les bases de données des arythmies MIT-BIH. En utilisant l'algorithme Convolutional Neural Network (CNN) avec Deep résiduel et ils ont obtenu une précision de 93.4% [29].

9. Conclusion

Les algorithmes d'apprentissage automatique tirent leur inspiration à partir des sources diverses telles que la théorie des probabilités aux intuitions géométriques par le biais des approches heuristiques. Faisant de l'apprentissage automatique est un outil puissant pour prédire les analyses et décrire les données dans leur divers types des problèmes.

L'apprentissage de données est une application d'apprentissage automatique sur le Big data à travers des environnements distribués comme MongoDB et à l'aide de IOT. Tel que Big data est une ensemble des données volumineux caractérisées par 5V [19].

Ces étapes sert comme une base permettant d'entamer la prochaine phase de l'étude conceptuelle.

Conclusion général & perspectives



Conclusion général

L'assurance de la santé est une condition nécessaire pour tout le monde. Mais chaque personne est exposé aux nombreuses maladies. Certaines maladies nécessitent un diagnostic immédiat par le médecin. Avec le développement de la science, les soins de santé sont devenus plus répandus dans les hôpitaux et les cliniques.

La grande quantité de données de soins de santé sur le diagnostic des maladies cardiaques dans le monde est produites par les hôpitaux et de nombreux laboratoires médicaux [33], pour être exploitées dans le domaine de la recherche scientifique, du diagnostic correct et de la classification des maladies ou pour prévoir les maladies.

Dans ce travail, nous avons abordé les moyens les plus importants d'exploiter les grandes données (Big data) et leurs sources, méthodes d'analyse et d'utilisation [18]. Nous avons créé un système de classification des maladies cardiaques à l'aide de dix (10) algorithmes d'apprentissage automatique utilisées pour nous donner le bon modèle qui permettent au médecin d'intervenir immédiatement en cas d'urgence. Les résultats de l'apprentissage automatique ont donné une bonne valeur de précision (97.4 %) et de moins d'erreur (12.9%) grâce à l'utilisation de l'algorithme Random Forest distribuée.

Nous avons testé notre modèle pour prédire de nombreux cas. Ce modèle est capable de faire des prédictions précises et de distinguer des différentes classes.

La révolution des technologies de l'information et de la communication doit être exploitée avec l'intelligence artificielle pour intégrer tels travaux scientifiques dans la vie humaine.

Bibliographie

Bibliographies

- [1]: @miscauthor = Ecole Suprieur Polytechnique de Dakar, title = traitement numrique du signalcardiaque (ECG), howpublished = "<https://fr.slideshare.net/cheikhdiabang/traitement-umrique-du-signal-cardiaque-ecg/>", year = 2015, note = "[Online; accessed 07 /04/2019]"
- [2]: @miscauthor = A.L.I.E.2, title = Anatomie et lectrophysiologie du coeur, ralisation, langage et anomalies des lectrocardiogrammes., howpublished = "<http://foulon.chez-alice.fr/Alie/>", note = "[Online; accessed 07 /04/2019]"
- [3]: @miscauthor = Sivakumaran, Nivethika , Nanthakumar, Rohini, title = Role of Biomedical Engineering for Diagnose and Treatment., howpublished = "<https://www.researchgate.net/publication/>", note = "[Online; accessed 20/04/2019]"
- [4]: @misc author = Copyright Nobel Media AB 2019, title = The Nobel Prize in Physiology or Medicine 1924., howpublished = "<https://www.nobelprize.org/prizes/medicine/1924/einthoven/facts/>", note = "[Online; accessed 10/04/2019]"
- [5]: @misc author = Cardiorestore, title = Le matriel de cardiologie, plus pro, moinscher., howpublished = "<https://www.cardiorestore.fr/>", note = "[Online; accessed 10/04/2019]"
- [6]: @misc author = Anatomiecardiaque, title = Anatomiecardiaque., howpublished = "<http://www.cardiodiac.net/ActiviteElectriqueCoeur.htm>", note = "[Online; accessed 10/04/2019]"
- [7]: @book author=John R. Hampton, title=The ECG Made Easy, publisher=Elsevier Masson SAS, year=2013
- [8]: @book author=Chlo-AgatheAzencott,Charge de recherche au CBIO, title=Introduction au Machine Learning, publisher=Elsevier Masson SAS, year=2013
- [9]: @book author=Eric Biernat , Michel Lutz, title=Data science : fondamentaux et tudes de cas Machine learning avec Python et R, publisher=Eyrolles, year=2015
- [10]: @article author = "M. Han and H. Zhang", title = "Business intelligence architecture based on internet of things", journal = "Journal of Theoretical and Applied Information Technology", volume = "50", number = "01", pages = "90{95}", year = "2013",
- [11]: @article author ="Luigi Atzori, Antonio Iera , GiacomoMorabito ", title = "The Internet of Things: A survey", journal = "Computer Networks", volume = "54", number = "15", pages = "1{19}", year = "2010", DOI = " doi: /10.1060/j.comnet.2010.05.010"
- [12]: @onlineknuthwebsite, author = "Bastien L", title = "Machine Learning et Big Data : d_nition et explications ", url = "<https://www.lebigdata.fr/machine-learning-et-big-data>", addendum = "(accessed: 04 /02/2019)",
- [13]: @onlineknuthwebsite, author = "YounesBenzaki", title = "9 Algorithmes de Machine Learning que chaque Data Scientist doit connaitre", url = "<https://mrmint.fr/9-algorithmes-de-machine-learning-que-chaque-data-scientist-doit-connaître>", addendum = "(accessed: 07 /02/2019)",
- [14]: @articlescikit-learn, title=Scikit-learn: Machine Learning in Python, author=Pedregosa, F. and Varoquaux, G. and Gramfort, A. and Michel, V. and Thirion, B. and Grisel, O. and Blondel, M. and Prettenhofer, P. and Weiss, R. and Dubourg, V. and Vanderplas, J. and Passos, A. And Cournapeau, D. and Brucher, M. and Perrot, M. and Duchesnay, E., journal=Journal of Machine Learning Research, volume=12, pages=2825{2830, year=2011
- [15]: @onlineknuthwebsite, author = "Ben Rogojan", title = "Boosting and Bag- ging: How To Develop A Robust Machine Learning Algorithm", url = "<https://medium.com/@SeattleDataGuy/how-to-develop-a-robust-algorithm-c38e08f32201>", addendum = "(accessed: 17 /03/2019)",

- [16]: @onlineknuthwebsite, author = "Chlo-Agathe Azencott", title = "Comprenez ce qui fait un bon modle d'apprentissage ", url = "https://openclassrooms.com/fr/courses/4297211-evaluez-et-ameliorez-les-performances-dun-modele-de-machine-learning/4297218-comprenez-ce-qui-fait-un-bon-modele-d-apprentissage", addendum = "(accessed: 09 /02/2019)",
- [17]: @onlineknuthwebsite, author = "Alexis ANZIEU", title = "L'introduction au Machine Learning la plus simple au monde", url = "https://medium.com/@alexis.anzieu/samuser-avec-le-machine-learning-a40159bb0546", addendum = "(accessed: 21 /03/2019)",
- [18]: @onlineknuthwebsite, author = "George Seif", title = "7 Practical Deep Learning Tips", url = "https://towardsdatascience.com/7-practical-deep-learning-tips-97a9f514100e", addendum = "(accessed: 25 /03/2019)",
- [19]: @book author= Pirmin Lemberger, Marc Batty, Mdric Morel, Jean-Luc Raalli, title=Big Data et Machine Learning, Les concepts et les outils de la data science Broch, publisher=Eyrolles, year=2016
- [20]: @onlineknuthwebsite, author = "Stefan Schmidt", title = "Les 3 V du Big Data : Volume, Vitesse et Varit", url = "https://www.journaldunet.com/solutions/expert/51696/les-3-v-du-big-data/volume{vitesse-et-variete.shtml", addendum = "(accessed: 07 /05/2019)",
- [21]: @onlineknuthwebsite, author = "Margaret Rouse", title = "Guide to NoSQL databases: How they can help users meet big data needs", url = "https://searchdatamanagement.techtarget.com/de_nition/MongoDB", addendum = "(accessed: 25 /03/2019)",
- [22]: @onlineknuthwebsite, author = "Alexis ANZIEU", title = "L'introduction au Machine Learning la plus simple au monde", url = "https://medium.com/@SeattleDataGuy/how-to-develop-arobust-algorithm-38e08f32201", addendum = "(accessed: 21 /03/2019)",
- [23]: @onlineknuthwebsite, Cris Ewing, Nicholas Hunt-Walker author = "Yannis Chaouche", title = "Identi_ez les di_erents types d'apprentissage automatiques", url = "https://openclassrooms.com/fr/courses/4011851-initiez-vous-au-machine-learning/4020611-identi_ez-les-di_erents-types-dapprentissage-automatiques", addendum = "(accessed: 21/03/2019)",
- [24]: @onlineknuthwebsite, author = "Cris Ewing, Nicholas Hunt-Walker", title = "The Decision Tree Classi_er", url = "https://codefollows.github.io/sea-python-401d5/lectures/decision-tree.html", addendum = "(accessed: 25 /03/2019)",
- [25]: @onlineknuthwebsite, author = "Saimadhu Polamuri ", title = "How The Random Forest Algorithm Works In Machine Learning", url = "https://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learning", addendum = "(accessed: 04/04/2019)",
- [26]: @onlineknuthwebsite, author = "Yann Deval ", title = "GPU IOT", url = "https://www.imsbordeaux.fr/fr/gpu/gpu-iot", addendum = "(accessed: 07/04/2019)",
- [27]: @article author = "Li, T.; Zhou, M.", title = "Ecgc_lassi_cation using wavelet packet entropy and random forests", journal = "Entropy", volume = "18", number = "8", pages = "1-16", year = "2016", DOI = "https://doi.org/10.3390/e18080285"
- [28]: @article article, author = Martis, Roshan and Acharya, U Rajendra and Lim, Choo and Mandana, Kayapanda and Ray, Ajoy and Chakraborty, Chandan, year = 2013, month = 08, pages = 1350014, title = Application of higher order cumulant features for cardiac health diagnosis using ECG signals, volume = 23, journal = International Journal of Neural Systems, doi = 10.1142/S0129065713500147
- [29]: @article author = "U. Rajendra Acharya, Hamido Fujita, ShuLih Oh, Yuki Hagiwara, Jen Hong Tan, Muhammad Adam", title = "Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals", journal = "Information

- Sciences", volume = "415", number = "416", pages = "190-198", year = "2017", DOI = "<https://doi.org/10.1016/j.ins.2017.06.027>"
- [30]: @inproceedingsinproceedings, author = Kachuee, Mohammad and Fazeli, Shayan and Sarrafzadeh, Majid, year = 2018, month = 06, pages = 443-444, title = ECG Heartbeat Classification: A Deep Transferable Representation, doi = 10.1109/ICHI.2018.00092
- [31]: @articlearticle, author = Ramalingam, V V and Dandapath, Ayantan and Karthik Raja, M, year = 2018, month = 03, pages = 684, title = Heart disease prediction using machine learning techniques: A survey, volume = 7, journal = International Journal of Engineering and Technology, doi = "10.14419/ijet.v7i2.8.10557"
- [32]: @articlearticle, author = Jean-Nol Montagne, year = 2018, month = 01, pages = 61, title = L'internet des objets : Big Data, Internet des Objets, Intelligence Artificielle,
- [33]: @book author= Association for the Advancement of Medical Instrumentation, title=A. Pour l'avancement de l'instrumentation médicale et autre., Testing and reporting performance results of cardiac rhythm and ST segment measurement, algorithms, publisher=Association for the Advancement of Medical Instrumentation, year=2013
- [34]: @articlearticle, author = Mohammad Kachuee, ShayanFazeli, MajidSarrafzadeh, year = 2018, month = 04, pages = 61, title = ECG Heartbeat Classification: A Deep Transferable Representation, doi = "10.1109/ICHI.2018.00092"
- [35]: @onlineknuthwebsite, author = "David Bowie ", title = "Fichier CSV - Qu'est-ce qu'un fichier CSV ?", url = "https://debitoor.fr/termes-comptables/_chier-csv", addendum = "(accessed: 09/03/2019)",
- [36]: @onlineknuthwebsite, author = "George B. Moody ", title = "MIT-BIH Arrhythmia Database Directory", url = "<https://physionet.org/physiobank/database/html/mitdbdir/intro.htm>", addendum = "(accessed: 05/03/2019)",
- [37]: @onlineknuthwebsite, author = "Pedregosa, F. and Varoquaux, G. and Gramfort, A. and Michel, V. and Thirion, B. and Grisel, O. and Blondel, M. and Prettenhofer, P. and Weiss, R. and Dubourg, V. and Vanderplas, J. and Passos, A. and Cournapeau, D. and Brucher, M. and Perrot, M. and Duchesnay, E. ", title = "scikit-learn Machine Learning in Python", url = "<https://scikit-learn.org/stable/>", addendum = "(accessed: 15/03/2019)",
- [38]: @onlineknuthwebsite, author = "Aditya Mishra", title = "Metrics to Evaluate your Machine Learning Algorithm", url = "<https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234>", addendum = "(accessed: 20/03/2019)",
- [39]: @onlineknuthwebsite, author = "Scala Dream Team,", title = "WiFi, Bluetooth, Zigbee, Z-Wave : quel réseau choisir pour ses objets connectés ?", url = "https://www.groupe-scala.com/wi_-bluetooth-zigbee-z-wave-quel-reseau-choisir-pour-ses-objets-connectes/", addendum = "(accessed: 25/03/2019)",
- [40]: @onlineknuthwebsite, author = "TutorialToUs", title = "MongoDB Tutorial ", url = "<http://tutorialtous.com/mongodb/index.php>", addendum = "(accessed: 25/03/2019)",
- [41]: @onlineknuthwebsite, author = "Amazon", title = "BerryKing Heartbeat Bluetooth and ANT ou Sportbeat pour Garmin Wahoo Polar RUNTASTIC STRAVA ENDOMONDO", url = "<https://www.amazon.fr/BerryKing-Heartbeat-frquence-cardiaque-lquipement/dp/B0182GISA2>", addendum = "(accessed: 25/03/2019)",
- [42]: @onlineknuthwebsite, author = "George B. Moody ", title = "MIT-BIH Arrhythmia Database Directory", url = "<https://www.physionet.org/physiobank/database/mitdb/>", addendum = "(accessed: 10/04/2019)",
- [43]: @articlearticle, author = Estes, Edward, year = 2014, month = 10, pages = , title = Big data and the electrocardiogram, volume = 48, journal = Journal of electrocardiology, doi = 10.1016/j.jelectrocard.2014.10.008
- [44]: @onlineknuthwebsite, author = "Python Software Foundation", title = "Functions Defined", url = "<https://www.python.org/>", addendum = "(accessed: 15/04/2019)",

- 
-
- [45]: @book author= Sarah Guido, Andreas C. Muller, title=Le machine learning avec Python, publisher=O'Reilly, First, year=2018
- [46]: @onlineknuthwebsite, author = "pandas community", title = "Python Data Analysis Library", url = "<https://pandas.pydata.org/>", addendum = "(accessed: 15/04/2019)",
- [47]: @onlineknuthwebsite, author = "NumPy Developers", title = "NumPy", url = "<https://www.numpy.org/>", addendum = "(accessed: 15/04/2019)",
- [48]: @onlineknuthwebsite, author = "Real Python", title = "Introduction to MongoDB and Python", url = "<https://realpython.com/introduction-to-mongodb-and-python/>", addendum = "(accessed: 15/04/2019)",
- [49]: @onlineknuthwebsite, author = "W3Schools", title = "Python MongoDB", url = "https://www.w3schools.com/python/python_mongodb_getstarted.asp", addendum = "(accessed: 20/04/2019)",
- [50]: @onlineknuthwebsite, author = "AAMI", title = "Standards Development", url = "<https://www.aami.org/Standards/>", addendum = "(accessed: 25/04/2019)",