

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

UNIVERSITE MOHAMED BOUDIAF - M'SILA

FACULTE DE Technologie
DEPARTEMENT Electronique
N° :



FILIÈRE : Electronique
OPTION : Instrumentation

Mémoire présenté pour l'obtention
Du diplôme de Master Académique

Par :

CHERIFI Wassim

BOUDILMI Fateh

THÈME

Prédiction des maladies cardiaques à l'aide des
techniques d'apprentissage statistiques

Soutenu devant le jury composé de :

Dr.	Université M ^{ed} Boudiaf –M'sila	Président
Dr.	Université M ^{ed} Boudiaf –M'sila	Rapporteur
Dr.	Université M ^{ed} Boudiaf –M'sila	Co-Rapporteur
Dr.....	Université M ^{ed} Boudiaf –M'sila	Examineur

Année universitaire : 2021 /2022

Remerciements

Tout d'abord, Nous remercions ALLAH miséricorde de nous avoir donné la volonté, le courage et la patience qui nous ont permis de réaliser ce travail.

Nous tenons à exprimer nos vifs remerciements à notre encadreur Dr. Djerioui Mohamed d'avoir proposé ce sujet, et dirigé notre travail durant la période d'élaboration de ce mémoire. Nous exprimons notre profonde gratitude pour nous avoir fait profiter de ces connaissances, mais aussi de sa méthode de travail, et surtout de sa rigueur scientifique.

Nos remerciements s'adressent également à tous les membres de Jury, qui ont accepté de juger notre travail.

Nos remerciements vont aussi à tous les enseignants du département d'électronique.

Enfin, nos remerciements A tous ceux qui ont contribué de près ou de loin à la réalisation de ce projet de fin d'études.

Dédicaces

On voudrait dédier ce modeste travail à tous ceux qui sont chers à nos cœurs, qui ont pris soin de nous et qui ont tout sacrifié pour nous, nos chères mamans.

A nos papas qui nous ont soutenus tout le long de nos cursus scolaire et universitaire et qui ont su donner tant d'efforts pour notre bien et réussite.

A nos frères et sœurs qui sont chers à nos cœurs.

À tous nos amis qui nous ont accompagnés Le long de nos cursus, spécialement : Mohy Eddin, Karim, Fares, Haroune, Ayoub, Aboubaker, Abdelhake.

À tous nos professeurs de l'école primaire à l'université.

À tous ceux qui sont chers à nos cœurs.

Table de matière

LIST DES FIGURES	6
LIST DES TABLEAUX	7
INTRODUCTION GENERALE	1
CHAPITRE I : LES MALADIES CARDIAQUES	3
1 INTRODUCTION	4
2 TYPES DES MALADIES CARDIAQUES	4
2.1 Arythmie.....	4
2.2 Maladie cardiaque congénitale.....	5
2.3 Maladie vasculaire.....	5
2.4 Valvulopathie cardiaque.....	6
2.5 Maladie du muscle cardiaque (cardiomyopathie)	6
2.6 Insuffisance cardiaque.....	7
3 SYMPTOMES.....	8
3.1 Symptômes de MC dans les vaisseaux sanguins.....	8
3.2 Symptômes de MC dus à un rythme cardiaque irrégulier.....	8
3.3 Symptômes de MC causée par une cardiopathie congénitale	8
3.4 Symptômes de MC causés par la cardiomyopathie.....	9
3.5 Symptômes de MC causés par des problèmes de valves cardiaques	9
4 CAUSES ET FACTEURS DE RISQUE.....	9
5 DIAGNOSTIC.....	11
5.1 Des analyses de sang	11
5.2 ECG (électrocardiogramme)	12
5.3 Tests de stress.....	12
5.4 Échocardiographie.....	12
5.5 Angiographie coronarienne et cathétérisme cardiaque	13
5.6 Radiographie pulmonaire	13
5.7 Tomographie par faisceau d'électrons ou EBCT.....	13
5.8 IRM cardiaque.....	13
6 CONCLUSION	13
CHAPITRE II : METHODES D'APPRENTISSAGE STATISTIQUES.....	15
1 INTRODUCTION	16
2 L'INTELLIGENCE ARTIFICIELLE	16

2.1	Définition de l'Intelligence artificielle	16
2.2	Historique	17
3	L'APPRENTISSAGE STATISTIQUE	18
3.1	Types d'apprentissage	19
3.1.1	L'apprentissage supervisé (Supervised Learning)	19
3.1.2	L'apprentissage non supervisé (Unsupervised Learning)	20
3.1.3	Apprentissage semi-supervisé (Semi-Supervised Learning)	20
3.1.4	L'apprentissage par renforcement (Reinforcement Learning)	21
3.2	Algorithmes d'apprentissage statistique utilisées	21
3.2.1	Decision Tree	21
3.2.2	Support Vector Machine (SVM)	24
3.2.3	Random Forest (RF)	25
3.2.4	Gradient Boosting	26
4	APPLICATIONS D'APPRENTISSAGE STATISTIQUES	27
4.1	Reconnaissance d'images	28
4.2	Reconnaissance de la parole	28
4.3	Prévision de trafic	28
4.4	Recommandations de produits	28
4.5	Voitures autonomes	28
4.6	Filtrage des courriers indésirables et des logiciels malveillants	28
4.7	Diagnostic médical	28
5	CONCLUSION	29
	CHAPITRE III : SIMULATION	30
1	INTRODUCTION	31
2	PROBLÉMATIQUE	31
3	ARCHITECTURE DU SYSTEME	32
4	BASE DES DONNÉES UTILISÉE	33
5	PRETRAITEMENT DES DONNEES	34
5.1	Normalisation	35
6	PROTOCOLE EXPÉRIMENTAL	35
6.1	Matrice de confusion	36
6.2	Accuracy	36
6.3	Précision	37
6.4	Sensitivity	37
6.5	F1-score	37

7	ALGORITHMES DE PREDICTION	38
7.1	Support Vector machine.....	38
7.2	Décision Tree	39
7.3	Random Forest	40
7.4	Gradiant Boosting	41
8	EVALUATION DES PERFORMANCES	42
8.1	Accuracy.....	42
8.2	Précision.....	43
8.3	Sensitivity.....	43
8.4	F1 Score.....	44
9	COMPARAISON AVEC L'ETAT DE L'ART	44
10	CONCLUSION.....	45
	CONCLUSION GENERALE.....	46
	Reference	47

LIST DES FIGURES

Fig. I. 1.	Arythmie cardiaque.....	5
Fig. I. 2.	Maladie vasculaire.....	5
Fig. I. 3.	Valvulopathie cardiaque.....	6
Fig. I. 4.	Maladie du muscle cardiaque.....	7
Fig. I. 5.	Insuffisance cardiaque.....	7
Fig. II. 1.	Processus d'apprentissage automatique.....	18
Fig. II. 2.	Subdivisions Supervised Learning.....	20
Fig. II. 3.	Exemple de Decision Tree.....	22
Fig. II. 4.	Hyperplan optimal pour Support Vector Machine avec deux classes.....	24
Fig. II. 5.	Algorithme de Random Forest.....	25
Fig. II. 6.	Algorithme de Gradient Boosting.....	27
Fig. III. 1.	Architecture du système de prédiction des maladies cardiaques.....	32
Fig. III. 2.	Résultat obtenu selon l'accuracy.....	42
Fig. III. 3.	Résultat obtenu selon la précision.....	43
Fig. III. 4.	Résultat obtenu selon la sensitivity.....	43
Fig. III. 5.	Résultat obtenu selon F1 Score.....	44

LIST DES TABLEAUX

Tableau. III. 1. Description des attributs de base des données utilisée.....	33
Tableau. III. 2. Statistiques descriptives des paramètres utilisés.	34
Tableau. III. 3. Matrice de confusion.....	36
Tableau. III. 4. Terme associés à la matrice de confusion.	36
Tableau. III. 5. Résultats obtenus avec le model SVM.....	38
Tableau. III. 6. La matrice de confusion des patients avec le model SVM.	39
Tableau. III. 7. Résultats obtenus avec le model DT.	39
Tableau. III. 8. La matrice de confusion des patients avec le model DT.....	40
Tableau. III. 9. Résultats obtenus avec le model RF.....	40
Tableau. III. 10. La matrice de confusion des patients avec le model RF.	41
Tableau. III. 11. Résultats obtenus avec le model GB.....	41
Tableau. III. 12. La matrice de confusion des patients avec le model GB.....	42
Tableau. III. 13. Comparaison avec l'état de l'art.....	45

INTRODUCTION GENERALE

Les maladies cardiaques et d'autres formes des maladies vasculaires ont joué un rôle dévastateur dans la vie humaine pendant de nombreuses années. Bien que les taux de mortalité aient diminué au cours des 30 dernières années, il reste l'une des principales causes de décès et d'invalidité dans le monde [1], représentant 17,7 millions de décès chaque année, soit 31 % de tous les décès dans le monde, comme l'a rapporté l'Organisation mondiale de la santé en 2017 [2]. Les maladies cardiovasculaires se développent tranquillement, mais elles peuvent être prévenues ou détectées tôt en identifiant les personnes « à risque ». La prédiction des maladies cardiaques est un facteur délicat, risqué et très important [3]. S'il est fait correctement, il peut être utilisé par le personnel médical pour sauver des vies [4]. Le diagnostic et le traitement des maladies cardiaques sont très complexes, en particulier dans les pays en développement, en raison du manque d'équipements de diagnostic, de médecins et d'autres ressources qui affectent la prédiction et le traitement appropriés des patients cardiaques. Avec cette préoccupation récemment, les technologies informatiques et les techniques d'apprentissage statistique sont utilisées pour développer des programmes visant à aider les cliniciens à prendre une décision concernant les maladies cardiaques au stade initial. La détection précoce de la maladie et la prédiction de la probabilité qu'une personne développe une maladie cardiaque peuvent réduire la mortalité [5].

En fait, plusieurs types de recherche se sont concentrés sur le développement de systèmes de prédiction des maladies cardiaques, Nous mentionnons de cette recherche :

Dans [6], les auteurs ont utilisé différents algorithmes de machine learning tels que LR, SVM, DT, KNN et GNB basé sur l'ensemble de données fourni par UCI heart disease. Les auteurs ont rapporté que la précision la plus élevée a été obtenue en utilisant l'algorithme DT. R. Bharti et al [7] ont utilisé le concept machine learning et deep learning pour la prédiction des maladies cardiaques, leur modèle était basé sur un ensemble de données sur les maladies cardiaques de l'UCI avec 14 attributs, les auteurs ont indiqué avoir atteint une précision de 94,2 % en utilisant l'approche de deep learning. Dans [8], les auteurs ont mis en œuvre l'algorithme SVM ,LR, Decision Tree, KNN et Random

Forest basé sur l'ensemble de données fourni par UCI heart disease, les auteurs ont rapporté que la précision la plus élevée a été obtenue en utilisant l'algorithme DT. Dans [9], les auteurs ont mis en œuvre plusieurs algorithmes de machine supervisés tels que Decision Tree, KNN, SVM et Random Forest, ils ont rapporté une précision de 79 %, 87 %, 83 % et 84 % respectivement. Dans [10], les auteurs ont découvert que différents algorithmes de classification donnaient une grande précision pour la prédiction des maladies cardiaques. Les auteurs ont également mentionné que pour la prédiction des maladies cardiaques, nous pouvons utiliser un nombre limité d'attributs au lieu de considérer tous les attributs de l'ensemble de données.

Dans cette mémoire, nous pouvons émettre l'hypothèse qu'un modèle de machine learning formé avec un ensemble de données de patients cardiaques peut être utilisé pour prédire le risque cardiaque potentiel chez les patients. Quatre modèles de machine learning (Support Vector Machine, Decision Tree, Random Forest et Gradient Boosting) ont été construits à l'aide de l'ensemble de données utilisée sur les maladies cardiaques.

Le travail réalisé est axé autour de trois chapitres qui sont présentés comme suit :

Le premier chapitre consiste à une introduction au domaine des maladies cardiaques a permis de présenter les types, symptômes, causes, facteur de risques et Comment les maladies cardiaques sont-elles diagnostiquées en médecine.

Le deuxième chapitre, nous allons décrire un état de l'art sur les méthodes appliquées à la prédiction. Quatre modèles (DT, RF, SVM et GB) fondés sur ce type d'apprentissage ont été exposés.

Enfin, le troisième et dernier chapitre est consacré à la conception et à la mise en œuvre d'un système basé sur l'ensemble des données utilisé. L'objectif est de valider et d'évaluer les performances de chacune des algorithmes présentés. Afin de réaliser une étude comparative permettant un choix définitif de la méthode de classification la plus appropriée dans la décision de présence ou d'absence de maladie cardiaque, nous évaluerons les modèles pour chaque algorithme. La discussion des résultats conclura cette étude de simulation pour sélectionner la technique la plus appropriée.

Une conclusion générale en fin de cette mémoire est prévue, elle retrace les différentes étapes réalisées et souligne les perspectives envisagées.

CHAPITRE I :

LES MALADIES CARDIAQUES

1 INTRODUCTION

Les maladies cardiaques sont l'une des principales causes de l'invalidité et la mort prématurée des personnes monde, on estime que 17,9 millions de personnes sont décédées des suites de maladies cardiovasculaires en 2019, ce qui représente 32 % de tous les décès dans le monde. Parmi ces décès, 85 % étaient dus à une crise cardiaque et à un accident vasculaire cérébral. La plupart des maladies cardiovasculaires peuvent être prévenues en s'attaquant aux facteurs de risque comportementaux tels que le tabagisme, une alimentation malsaine et l'obésité, la sédentarité et l'usage nocif de l'alcool, il est important de détecter les maladies cardiovasculaires le plus tôt possible afin que la prise en charge avec des conseils et des médicaments puisse commencer [2].

Ce premier chapitre se veut introductif, il a pour objectif de donner une idée générale du domaine des maladies cardiaques. La première partie est consacrée à la définition des types des maladies ainsi que de ses symptômes. Les causes et facteurs des risques sont particulièrement présentées. Sont cités dans ce cadre, les différentes méthodes de diagnostics.

2 TYPES DES MALADIES CARDIAQUES

Les maladies cardiaques sont une description d'un groupe de maladies qui affectent le cœur. Les maladies cardiaques comprennent[14,15]:

2.1 Arythmie

Une arythmie est un problème lié à la fréquence ou au rythme de votre rythme cardiaque, Cela signifie que votre cœur bat trop vite, trop lentement ou de manière irrégulière. Lorsque le cœur bat plus vite que la normale, on parle de tachycardie. Lorsque le cœur bat trop lentement, on parle de bradycardie. Le type d'arythmie le plus courant est la fibrillation auriculaire, qui provoque un rythme cardiaque irrégulier et rapide, de nombreux facteurs peuvent affecter le rythme de votre cœur, comme une crise cardiaque, le tabagisme, des malformations cardiaques congénitales et le stress. Certaines substances ou médicaments peuvent également provoquer des arythmies[13].

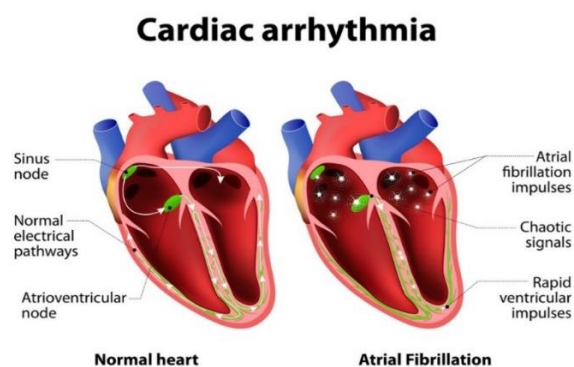


Fig. I. 1. Arythmie cardiaque.

2.2 Maladie cardiaque congénitale

La cardiopathie congénitale est un terme général désignant une gamme de malformations congénitales qui affectent le fonctionnement normal du cœur, le terme « congénital » signifie que la condition est présente dès la naissance, les cardiopathies congénitales sont l'un des types d'anomalies congénitales les plus courants, affectant près de 1 bébé sur 100 nés au Royaume-Uni [14].

2.3 Maladie vasculaire

La maladie vasculaire comprend toute condition qui affecte le système circulatoire, telle que la maladie artérielle périphérique. Cela va des maladies des artères, des veines et des vaisseaux lymphatiques aux troubles sanguins qui affectent la circulation.

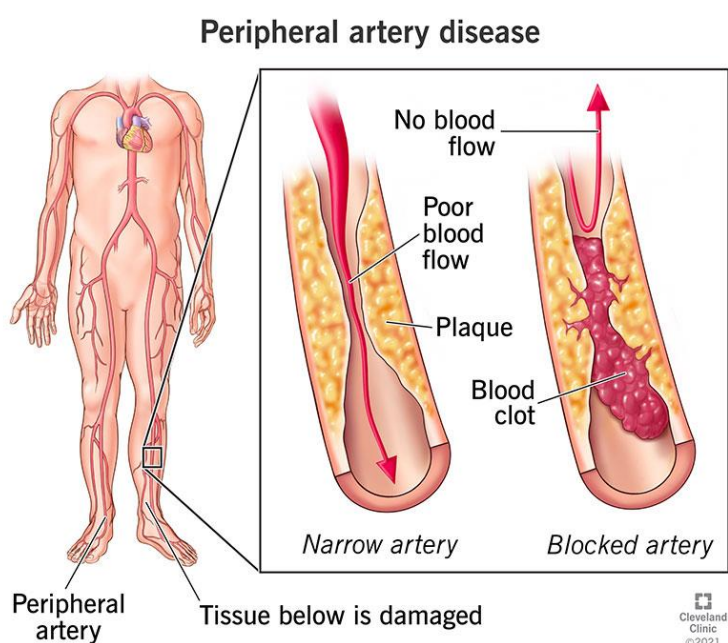


Fig. I. 2. Maladie vasculaire.

2.4 Valvulopathie cardiaque

Il y a quatre valves à l'intérieur de votre cœur qui permettent au sang de circuler dans la bonne direction en n'ouvrant qu'un seul chemin, quand c'est nécessaire. Pour s'assurer que les valves fonctionnent correctement, elles doivent être exemptes de malformations congénitales et capables de s'ouvrir complètement et de se fermer hermétiquement pour ne pas fuir. Les quatre soupapes sont [12]:

- Valve tricuspide ;
- la valve mitrale ;
- valve pulmonaire ;
- La valve aortique.

Dans certains cas, une ou plusieurs vannes ne s'ouvrent pas ou ne se ferment pas correctement. Cela peut perturber le flux sanguin de votre cœur vers votre corps.

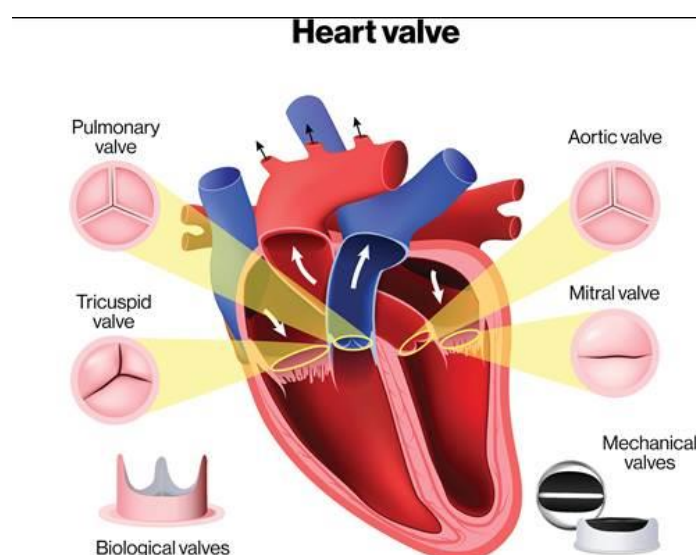


Fig. I. 3. Valvulopathie cardiaque.

2.5 Maladie du muscle cardiaque (cardiomyopathie)

La cardiomyopathie est une maladie du muscle cardiaque qui empêche votre cœur de pomper le sang vers le reste de votre corps. La cardiomyopathie peut entraîner une insuffisance cardiaque.

CARDIOMYOPATHY

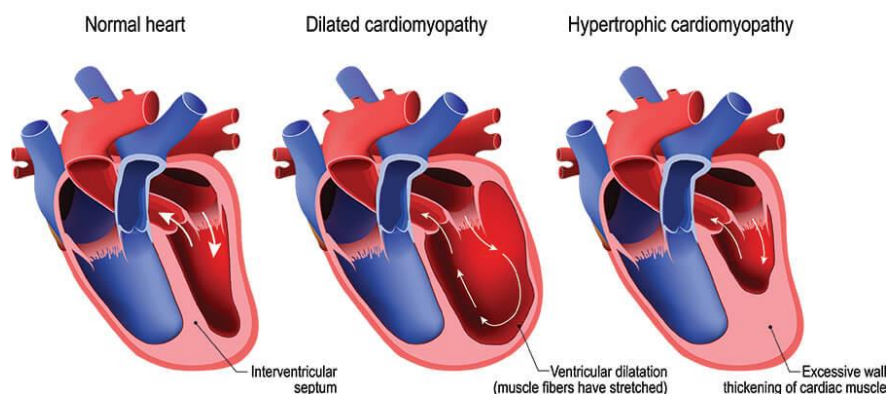


Fig. I. 4. Maladie du muscle cardiaque.

2.6 Insuffisance cardiaque

L'insuffisance cardiaque est un syndrome pouvant revêtir différentes formes et correspondant à l'évolution de la plupart des pathologies cardiaques. La définition retenue par la Société Européenne de Cardiologie est la présence de symptômes d'insuffisance cardiaque (au repos ou à l'effort) associés à une preuve (de préférence échocardiographie) de dysfonction cardiaque systolique et/ou diastolique (au repos) et en cas de doute diagnostique, associés à une réponse favorable au traitement habituel de l'insuffisance cardiaque. Les deux premiers critères doivent être obligatoirement présents pour en faire le diagnostic[15].

HEART FAILURE

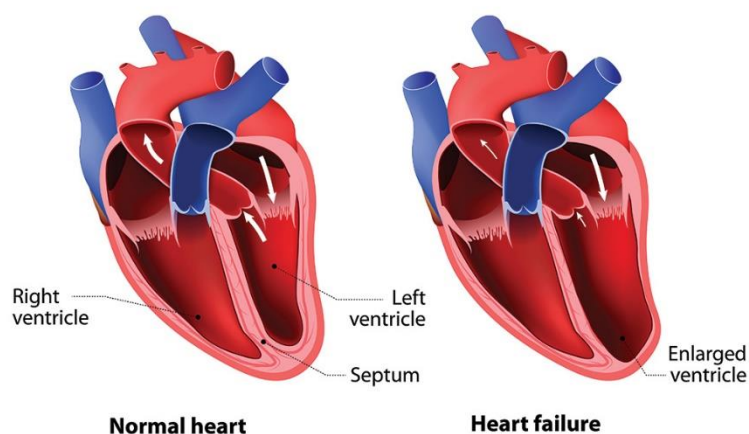


Fig. I. 5. Insuffisance cardiaque.

3 SYMPTOMES

Les symptômes de la maladie cardiaque dépendent du type de blessure que vous avez [16]:

3.1 Symptômes de MC dans les vaisseaux sanguins

L'accumulation de plaques graisseuses dans les artères ou l'athérosclérose peut endommager les vaisseaux sanguins et le cœur. Ce qui pourrait entraîner une crise cardiaque, des douleurs thoraciques (angine de poitrine) ou un accident vasculaire cérébral. Les signes et symptômes de la maladie peuvent inclure [17]:

- douleur et oppression thoracique (angine de poitrine) ;
- essoufflement ;
- Douleur, faiblesse ou froideur dans les jambes ou les bras si les vaisseaux sanguins de ces parties de votre corps sont étroits ;
- Douleur dans le cou, la mâchoire, la gorge, le haut de l'abdomen ou le dos.

3.2 Symptômes de MC dus à un rythme cardiaque irrégulier

Le rythme cardiaque peut être trop rapide, trop lent ou irrégulier. Les signes et symptômes des arythmies peuvent inclure [18]:

- palpitations dans la poitrine ;
- rythme cardiaque rapide ;
- rythme cardiaque lent ;
- Douleur dans la région de la poitrine ;
- Essoufflement ;
- Vertige ;
- Vertiges ;
- Perte de conscience ou approche de la perte de conscience.

3.3 Symptômes de MC causée par une cardiopathie congénitale

Les malformations cardiaques graves avec lesquelles vous êtes né sont généralement remarquées peu de temps après la naissance. Les signes et symptômes de malformations cardiaques chez les enfants peuvent inclure :

- Peau devenant gris pâle ou bleue ;
- Gonflement des jambes, de l'abdomen ou des zones autour des yeux ;
- Pour les nourrissons, essoufflement pendant la tétée.

3.4 Symptômes de MC causés par la cardiomyopathie

Aux premiers stades de la cardiomyopathie, vous pouvez ne présenter aucun symptôme.

À mesure que l'état s'aggrave, les symptômes peuvent inclure :

- Essoufflement pendant l'activité ou au repos ;
- Jambes, chevilles et pieds enflés ;
- Fatigue ;
- Arythmie ;
- Se sentir étourdi, étourdi et s'évanouir.

3.5 Symptômes de MC causés par des problèmes de valves cardiaques

Les symptômes de la cardiopathie valvulaire comprennent généralement :

- Fatigue ;
- Essoufflement ;
- Arythmie ;
- Pieds ou chevilles enflés ;
- Douleur thoracique ;
- Perte de connaissance.

4 CAUSES ET FACTEURS DE RISQUE

La maladie cardiaque se développe lorsqu'il y a :

- dommages à tout ou partie du cœur ;
- un problème avec les vaisseaux sanguins menant ou provenant du cœur ;
- un faible apport d'oxygène et de nutriments au cœur ;
- un problème de rythme cardiaque.

Dans certaines causes, il existe une cause génétique. Cependant, certains facteurs liés au mode de vie et certaines conditions médicales peuvent également augmenter le risque.

La plupart des gens meurent des suites d'une maladie cardiaque à 65 ans, Même si c'est risqué pour les deux sexes. Les femmes ont plus de chances d'avoir une maladie cardiaque [19], Environ 72 % des personnes qui sont des professionnels des affaires ont une possibilité d'avoir une maladie cardiaque [20], Les chômeurs ont environ 50% de chance mourir d'une maladie cardiaque, selon une étude, en 2017 [21]. Une plage de fréquence cardiaque normale est de 60 à 100 battements par minute, Le taux peut descendre en

dessous de 60 battements par minute qui font de l'exercice en continu ou qui prennent des médicaments. Si votre fréquence cardiaque est supérieure à 100 battements par minute, on parle de tachycardie, La pression artérielle systolique normale est de 120 et la pression artérielle diastolique normale est de 80 [22], Une lecture de plus de 200 mg/dl (11,1 mmol/L) après deux heures indique un diabète. Une lecture entre 140 et 199 mg/dl (7,8 mmol/L et 11,0 mmol/L) définit le prédiabète [23].

Un facteur de risque est toute caractéristique, symptôme d'un individu qui augmente la possibilité de développer une maladie. Plusieurs symptômes, habitudes alimentaires, mode de vie, âge, antécédents familiaux, diabète, tabagisme peuvent augmenter le risque de chances d'avoir une crise cardiaque. Ceux-ci sont appelés facteurs de risque. Certains facteurs de risque ne sont pas contrôlables et conduisent à une crise cardiaque ischémique. Ceux-ci inclus :

- **hypertension artérielle** : est la principale cause de maladie rénale chronique (MRC) qui augmente le risque de crise cardiaque, d'insuffisance cardiaque et d'accident vasculaire cérébral Une pression artérielle élevée non contrôlée peut entraîner un durcissement et un épaississement de vos artères, rétrécissant les vaisseaux par lesquels le sang circule[24].
- **taux de cholestérol élevé** : Des niveaux élevés de cholestérol dans le sang peuvent augmenter le risque de formation de plaque et d'athérosclérose[25].
- **fumeur** : La nicotine resserre les vaisseaux sanguins et le monoxyde de carbone peut endommager leur paroi interne, les rendant plus sensibles à l'athérosclérose. Les crises cardiaques sont plus fréquentes chez les fumeurs que chez les non-fumeurs, Toute durée est nocive pour le cœur. Plus une personne fume, plus elle risque d'avoir une maladie cardiaque. Il n'y a pas de telle limitation à la fumée qui est bonne pour le cœur [11].
- **une forte consommation d'alcool** : L'American Heart Association affirme qu'une forte consommation d'alcool entraîne une augmentation des triglycérides dans le sang, augmente la tension artérielle et augmente le risque d'insuffisance cardiaque et d'arythmie.
- **surpoids et obésité** : L'excès de poids aggrave généralement d'autres facteurs de risque de maladie cardiaque.

- **diabète** : Le diabète augmente votre risque de maladie cardiaque. Les deux conditions partagent des facteurs de risque similaires, tels que l'obésité et l'hypertension artérielle[26].
- **une histoire familiale de maladie cardiaque** : Des antécédents familiaux de maladie cardiaque augmentent votre risque de maladie coronarienne, surtout si un parent l'a développé à un âge précoce.
- **Une mauvaise alimentation** : Une alimentation riche en matières grasses, en sel, en sucre et en cholestérol peut contribuer au développement de maladies cardiaques.
- **âge** : Vieillir augmente le risque d'artères endommagées et rétrécies et d'un muscle cardiaque affaibli ou épaissi.
- **faibles niveaux d'activité** : Le manque d'activité physique est associé à l'exposition à de nombreuses maladies cardiaques et à certains des facteurs de risque qui y conduisent également.
- **niveaux élevés de stress et d'anxiété** : Un stress non soulagé peut endommager vos artères et aggraver d'autres facteurs de risque de maladie cardiaque.
- **Mauvaise santé dentaire** : Il est important de se brosser les dents et de passer du fil dentaire souvent et de passer des examens dentaires réguliers. Si vos dents et vos gencives ne sont pas saines, des germes peuvent pénétrer dans votre circulation sanguine et se rendre jusqu'à votre cœur, provoquant une endocardite.

5 DIAGNOSTIC

Certains des tests couramment utilisés pour diagnostiquer les maladies cardiovasculaires comprennent [27] :

5.1 Des analyses de sang

Les tests de laboratoire sont utilisés pour détecter les facteurs de risque des maladies cardiaques. Ceux-ci incluent la détection des graisses, du cholestérol et des composants lipidiques du sang, notamment les LDL, HDL et les triglycérides.

La glycémie et l'hémoglobine glycosylée sont mesurées pour la détection du diabète. La protéine C-réactive (CRP) et d'autres marqueurs protéiques tels que les apolipoprotéines A1 et B sont utilisés pour détecter l'inflammation pouvant entraîner des maladies cardiaques.

Lors d'une crise cardiaque, les cellules du muscle cardiaque meurent et libèrent des protéines dans la circulation sanguine. Des tests sanguins peuvent mesurer la quantité de ces protéines dans le sang. Des niveaux élevés de ces protéines sont le signe d'une crise cardiaque récente.

L'un des marqueurs de la crise cardiaque est la troponine-T cardiaque. D'autres biomarqueurs comprennent le fibrinogène et le PAI-1, des niveaux élevés d'homocystéine, une diméthylarginine asymétrique élevée et un peptide natriurétique cérébral élevé (également connu sous le nom de type B) (BNP) [28].

5.2 ECG (électrocardiogramme)

Il s'agit d'un test simple et indolore qui enregistre l'activité électrique du cœur. Le patient est attaché à l'instrument avec plusieurs patchs ou fils placés sur sa poitrine, ses poignets et ses chevilles. Une petite machine portable enregistre les activités du cœur sur une bande de papier quadrillé.x.

Le test montre à quelle vitesse le cœur bat et son rythme. La force et la synchronisation des signaux électriques lorsqu'ils traversent le cœur sont également observées. Un ECG peut aider à détecter une crise cardiaque, des crises d'angine de poitrine, des arythmies, etc.[29].

5.3 Tests de stress

Pour ce test, le patient est obligé de travailler dur, par ex. courir sur un tapis roulant ou faire de l'exercice pendant que les dérivations de l'ECG sont placées sur leur corps. Ceux qui ne peuvent pas faire d'exercice reçoivent des pilules pour augmenter leur fréquence cardiaque. Le test détecte les effets de l'exercice sur le cœur.

Chez les patients atteints d'athérosclérose et de maladies coronariennes, les artères rétrécies par des plaques ne peuvent pas fournir suffisamment de sang aux muscles cardiaques alors qu'ils battent plus vite. Cela peut entraîner un essoufflement et des douleurs thoraciques. Le schéma ECG, les arythmies, etc. montrent également la possibilité d'une maladie coronarienne[29].

5.4 Échocardiographie

Ce test utilise des ondes sonores pour créer une image animée du cœur. Il s'agit également d'un test indolore où une sonde est roulée sur la poitrine et la machine crée l'image du

cœur sur le moniteur. Cela fournit des informations sur la forme, la taille, le fonctionnement, les valves et les cavités du cœur.

L'échocardiographie peut également être combinée avec le Doppler pour montrer les zones de faible apport sanguin au cœur. Il montre les zones du muscle cardiaque qui ne se contractent pas normalement et les lésions antérieures du muscle cardiaque[28].

5.5 Angiographie coronarienne et cathétérisme cardiaque

Ce test est un test invasif. Un colorant est injecté dans les veines pour atteindre les artères coronaires. Cela se fait par cathétérisme coronaire. Ensuite, des images détaillées des vaisseaux sanguins du cœur sont prises à l'aide de méthodes d'imagerie spéciales. C'est ce qu'on appelle la coronarographie.

Le cathétérisme cardiaque consiste à enfiler un tube mince et flexible appelé cathéter via un vaisseau sanguin dans le bras, l'aîne (haut de la cuisse) ou le cou. Le tube est inséré sous guidage imaginaire jusqu'à ce qu'il atteigne le cœur. La coronarographie détecte les blocages dans les grosses artères coronaires[28].

5.6 Radiographie pulmonaire

Il s'agit d'un test qui montre la forme et la taille des poumons du cœur et des principaux vaisseaux sanguins. Il s'agit d'un test rarement utilisé dans le diagnostic des maladies cardiaques car il ne fournit pas d'informations supplémentaires par rapport à l'échocardiographie et à d'autres études d'imagerie.

5.7 Tomographie par faisceau d'électrons ou EBCT

L'EBCT aide à détecter les dépôts de calcium ou les calcifications dans les parois des artères coronaires. Ce sont des marqueurs précoces de l'athérosclérose et des maladies coronariennes. Ce n'est pas un test de routine dans les maladies coronariennes.

5.8 IRM cardiaque

IRM cardiaque (imagerie par résonance magnétique) qui utilise des ondes radio, des aimants et un ordinateur pour créer des images du cœur. Cela donne une image 3D du mouvement ainsi que des images fixes du cœur.

6 CONCLUSION

Ce chapitre a servi d'introduction au domaine de contrôle et de diagnostic des maladies cardiaques. Les différents types des maladies sont alors présentés. De même, les différentes méthodes de diagnostic actuellement utilisées sont scrupuleusement évoquées.

Il est d'ores et déjà apparu qu'un contrôle automatique et permanent basé sur ses paramètres descripteurs ne pouvait être qu'indispensable. Dans ce domaine la prédiction joue un rôle primordial, et l'utilisation de l'intelligence artificielle demeure incontournable. L'objet du chapitre suivant va être orienté vers le vaste domaine de l'apprentissage statistique, notamment ceux utilisés dans la prédiction des maladies cardiaques.

CHAPITRE II

METHODES D'APPRENTISSAGE

STATISTIQUES

1 INTRODUCTION

La résolution des problèmes par la construction de machines capables d'apprendre à partir des entrées et des sorties, caractérise l'approche fondamentale de la théorie d'apprentissage (*Machine Learning*). Depuis des décennies de nombreux sous problèmes ont été identifiés, l'extraction et la sélection des caractéristiques dans un but de réduction de dimensionnalité, la variété des espaces de représentations, la popularité, la complexité et toutes ces variantes du problème de l'apprentissage ont généré une multitude de méthodes de résolution. Pour traiter un problème dans ce sens, diverses méthodes ont été développées. Parmi celles-ci on trouve particulièrement la Decision Tree (DT), les machines à vecteurs de support (SVM), le Random Forest (RF) et le Gradient Boosting (GB).

Dans le cadre de notre travail, la solution adoptée rappelons-le une fois de plus, est une prédiction des maladies cardiaques par les techniques d'apprentissage statistiques. Dans ce chapitre, nous allons donc pouvoir passer en revue des méthodes d'apprentissage statistique appliquées à la classification.

La première partie de ce chapitre est donc dédiée à l'intelligence artificielle (IA) et décrit brièvement l'histoire de l'IA. L'apprentissage statistique ainsi que leurs type d'apprentissage, sont à cet effet décrits. Dans la seconde partie, un état de l'art sur les méthodes utilisées dans la prédiction des maladies est présenté. Enfin, les applications d'apprentissage statistique sont aussi rappelées et décrites.

2 L'INTELLIGENCE ARTIFICIELLE

2.1 Définition de l'Intelligence artificielle

Ce sont le comportement et les caractéristiques spécifiques des programmes informatiques qui les font imiter les capacités mentales et les schémas de travail humains. L'une des plus importantes de ces caractéristiques est la capacité d'apprendre, de déduire et de réagir à des situations qui n'ont pas été programmées dans la machine. Cependant, ce terme est controversé car il n'existe pas de définition définitive de l'intelligence[30].

2.2 Historique

Il est intéressant de revenir sur les origines et l'historique de l'intelligence artificielle afin de bien comprendre ses premières orientations et ses perspectives pour l'avenir [31]:

- 1950 : Alan M. Turing, mathématicien et théoricien précurseur de l'informatique, lance le concept d'intelligence artificielle.
- 1955 – 1956 : Lancement du premier programme d'intelligence artificielle par Allen newell, John C. Shaw et Herbert A. Simon, logic theorist.
- 1957 : Modélisation des jeux d'échec.
- 1958 : John McCarthy invente le Lisp (list processing), langage de programmation interactif (développement au MIT).
- 1958 : Construction du premier réseau neuronal, le Perceptron, de Frank Rosenbaltt, machine dite connexionniste. Chapitre I La biomédicale et l'apprentissage des machines 7
- 1959 : Élaboration du premier GPS (genaral problem solver) -fin de la première période de l'intelligence artificielle.
- 1970 : Néoconnectionnisme.
- 1989 : DeepThought, supercalculateur d'IBM, deux millions de coups par seconde.
- 1990 – 1997 : Développement de Deep Blue rebaptisé DeeperBlue: conception d'un système de 256 processeurs fonctionnant en parallèle, chaque processeur peut calculer environ trois millions de coups par seconde.
- 2009 : Le MIT a lancé un projet visant à repenser la recherche en intelligence artificielle.
- 2011 : Watson, le superordinateur d'IBM remporte deux des trois manches du jeu télévisé Jeopardy! La performance a consisté pour cette intelligence artificielle à répondre à des questions de culture générale.
- 2013 : Humain Brain Project. Google ouvre un laboratoire de recherches dans les locaux de la NASA.
- 2014 : DeepKnowlegdge Ventures : nomme à son conseil d'administration VITAL, un algorithme capable d'élaborer ses décisions en analysant les bilans comptables des entreprises potentiellement intéressantes, les tests cliniques, la propriété intellectuelle et les précédents investissements.
- 2015: Facebook Artificial Intelligence Research (FAIR). Google rend sa technologie d'intelligence artificielle TensorFlow accessible à tous.

Développement d'une crainte que l'intelligence artificielle dépasse à terme les performances de l'intelligence humaine.

- 2016 : Amelia d'IPSoft un agent virtuel. Aussi, AlphaGo bat trois fois consécutives le champion du monde de jeu de go, Lee Se-Dol en cinq manches.

3 L'APPRENTISSAGE STATISTIQUE

L'apprentissage statistique est l'un des champs d'étude de l'intelligence artificielle. C'est également une science qui concerne le développement et l'implémentation de procédures automatisables qui permettent à une machine d'apprendre et d'évoluer. L'être humain a tendance à améliorer sa situation en utilisant l'expérience obtenue durant des années. Nous pouvons également appliquer ce procédé sur des machines. C'est ce que nous appelons machine Learning ou apprentissage automatique qui vise à entraîner un algorithme en se basant sur des exemples. Cette technique permet à la machine d'apprendre sans avoir à modifier les algorithmes qui la constituent et donc, elle pourra exécuter des tâches pour lesquelles elle n'est pas programmée. Ci-dessous, un schéma qui résume le processus de L'apprentissage statistique [32]:

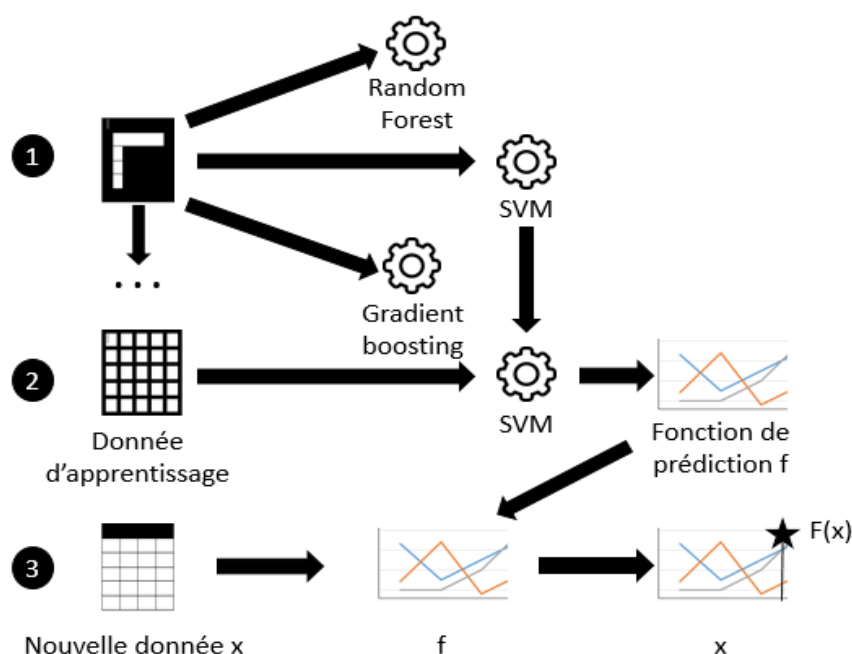


Fig. II. 1. Processus d'apprentissage automatique.

- 1) Sélection d'un algorithme de Machine Learning dans une librairie de ML.
- 2) L'entraînement de l'algorithme retenu à partir des données produit une fonction de prédiction f .

3) Prédiction pour une nouvelle observation.

D'après le schéma ci-dessus, nous distinguons 3 grandes étapes du processus d'apprentissage automatique. La première étape est la sélection d'un algorithme d'apprentissage statistique pour résoudre un problème donné. La seconde étape consiste à entraîner cet algorithme afin qu'il puisse minimiser l'erreur et construire la fonction de prédiction la plus optimale. La troisième et dernière étape consiste à donner de nouvelles entrées à notre algorithme afin de prédire une valeur quantitative ou bien une valeur qualitative. Il faut noter qu'avant la première étape, un prétraitement des données est nécessaire pour espérer l'obtention d'un bon modèle.

3.1 Types d'apprentissage

Les types d'apprentissage statistiques sont souvent classés en fonction de la manière dont les algorithmes apprennent à devenir plus précis pour prédire les événements, et le type d'algorithme que les programmeurs choisissent d'utiliser dépend du type de données qu'ils souhaitent prédire.

Quant aux types d'apprentissage statistique, ils sont :

3.1.1 L'apprentissage supervisé (Supervised Learning)

La majorité de l'apprentissage statistique pratique utilise l'apprentissage supervisé.

L'apprentissage supervisé est l'endroit où vous avez des variables d'entrée (x) et une variable de sortie (Y) et vous utilisez un algorithme pour apprendre la fonction de mappage de l'entrée à la sortie.

$$y = f(x)$$

2.1

L'objectif est d'approximer si bien la fonction de mappage que lorsque vous avez de nouvelles données d'entrée (x), vous pouvez prédire les variables de sortie (Y) pour ces données.

C'est ce qu'on appelle l'apprentissage supervisé parce que le processus d'apprentissage d'un algorithme à partir de l'ensemble de données d'apprentissage peut être considéré comme un enseignant supervisant le processus d'apprentissage. On connaît les bonnes réponses, l'algorithme fait itérativement des prédictions sur les données d'apprentissage et est corrigé par l'enseignant. L'apprentissage s'arrête lorsque l'algorithme atteint un niveau de performance acceptable[33].

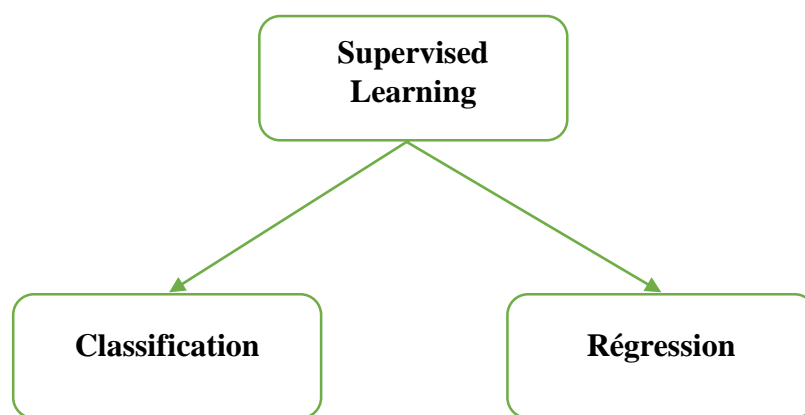


Fig. II. 2. Subdivisions Supervised Learning.

3.1.2 L'apprentissage non supervisé (Unsupervised Learning)

L'apprentissage non supervisé est l'endroit où vous n'avez que des données d'entrée (X) et aucune variable de sortie correspondante.

L'objectif de l'apprentissage non supervisé est de modéliser la structure sous-jacente ou la distribution dans les données afin d'en savoir plus sur les données.

Celles-ci sont appelées apprentissage non supervisé car, contrairement à l'apprentissage supervisé ci-dessus, il n'y a pas de bonnes réponses et il n'y a pas d'enseignant. Les algorithmes sont laissés à eux-mêmes pour découvrir et présenter la structure intéressante des données.

Les problèmes d'apprentissage non supervisé peuvent être regroupés en problèmes de regroupement et d'association[34].

3.1.3 Apprentissage semi-supervisé (Semi-Supervised Learning)

Les problèmes où vous avez une grande quantité de données d'entrée (X) et où seules certaines des données sont étiquetées (Y) sont appelés des problèmes d'apprentissage semi-supervisé.

Ces problèmes se situent entre l'apprentissage supervisé et non supervisé.

Un bon exemple est une archive de photos où seules certaines des images sont étiquetées (par exemple, un chien, un chat, une personne) et la majorité ne sont pas étiquetées.

De nombreux problèmes d'apprentissage automatique dans le monde réel relèvent de ce domaine. En effet, il peut être coûteux ou long d'étiqueter les données, car cela peut nécessiter l'accès à des experts du domaine. Alors que les données non étiquetées sont bon marché et faciles à collecter et à stocker.

Vous pouvez utiliser des techniques d'apprentissage non supervisé pour découvrir et apprendre la structure des variables d'entrée.

Vous pouvez également utiliser des techniques d'apprentissage supervisé pour faire des prédictions optimales pour les données non étiquetées, réintroduire ces données dans l'algorithme d'apprentissage supervisé en tant que données d'apprentissage et utiliser le modèle pour faire des prédictions sur de nouvelles données invisibles[34].

3.1.4 L'apprentissage par renforcement (Reinforcement Learning)

L'apprentissage par renforcement fonctionne en programmant des algorithmes avec un objectif distinct et un ensemble défini de règles pour atteindre cet objectif.

L'algorithme est programmé pour obtenir des récompenses positives lorsque quelque chose d'utile lié à l'objectif final est atteint. Et vous recevez des pénalités lorsque vous entreprenez une action qui est loin de l'objectif à atteindre. Cet apprentissage sert généralement à [33]:

- Robotique : Les robots peuvent apprendre à effectuer des tâches en utilisant ce type d'apprentissage.
- Jeux vidéo : L'apprentissage par renforcement est utilisé pour apprendre aux robots à jouer à des jeux vidéo.
- Gestion des ressources : ce type d'apprentissage peut aider à la planification des processus de gestion des ressources.

3.2 Algorithmes d'apprentissage statistique utilisées

Il n'existe pas d'algorithme d'apprentissage parfait ou d'algorithme meilleur que l'autre. Chacun peut être meilleur selon le domaine d'étude. C'est là que le facteur humain intervient pour faire le bon choix en tenant compte des données utilisées, de l'objet de la problématique et surtout des indicateurs de performances. Nous allons voir quelques algorithmes populaires d'apprentissage statistique.

3.2.1 Decision Tree

Lorsque la relation entre les caractéristiques et le résultat n'est pas linéaire ou lorsque caractéristiques interagissent les unes avec les autres, les modèles de régression linéaire et de régression logistique les performances se dégradent. Une solution simple à ces situations est les arbres de décision. Décision les arbres fonctionnent en divisant les données plusieurs fois en fonction d'une mesure telle que des informations gain qui

détermine la quantité d'informations pouvant être acquises par cette division. Le fractionnement La procédure aboutit à différents sous-ensembles de l'ensemble de données. Les sous-ensembles finaux sont appelés feuille nœuds. La prédiction se produit sur ces nœuds feuilles où le résultat prévu est le résultat moyen des données de formation dans ce sous-ensemble[35].

$$E(S) = -p_+ \log(p_+) - p_- \log(p_-) \quad 2.2$$

p_+ : Pourcentage des personnes souffrant de maladies cardiaques.

p_- : Pourcentage des personnes sans maladie cardiaque.

$$y = -\sum_{i=1}^k p_i \log_k(p_i) \quad 2.3$$

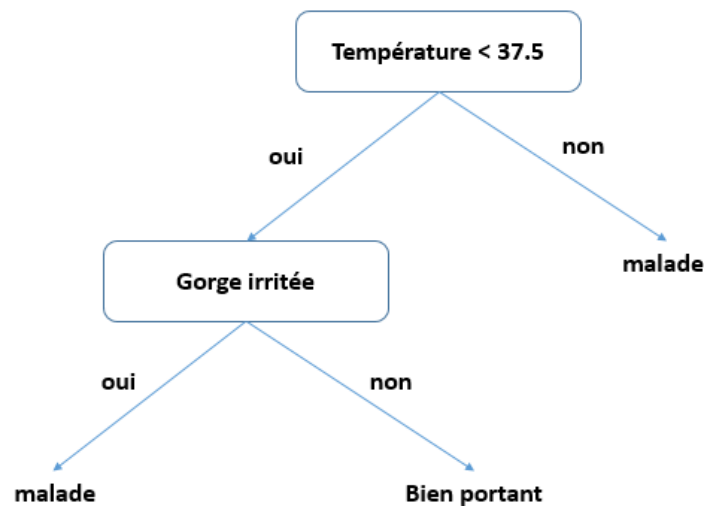


Fig. II. 3. Exemple de Decision Tree.

$$G = \text{sum} (pk * (1 - pk)) \quad 2.3$$

a. Type de Decision Tree

Les types de Decision Tree sont basés sur le type de variable cible dont nous disposons. Il peut être de deux types :

- Decision Tree à variable catégorielle : Decision Tree qui a une variable cible catégorielle, puis appelé arbre de décision à variable catégorielle.
- Decision Tree à variable continue : Decision Tree a une variable cible continue, il est alors appelé Decision Tree à variable continue.

b. Les avantages

Les avantages d'algorithme Decision Tree sont [36]:

- Simple à comprendre et à interpréter. Les arbres peuvent être visualisés.
- Nécessite peu de préparation des données. D'autres techniques nécessitent souvent une normalisation des données, des variables fictives doivent être créées et des valeurs vides doivent être supprimées.
- Le coût d'utilisation de l'arbre (c'est-à-dire la prédiction des données) est logarithmique dans le nombre de points de données utilisés pour former l'arbre.
- Capable de traiter des données numériques et catégorielles.
- Capable de gérer les problèmes multi-sortis.
- Possibilité de valider un modèle à l'aide de tests statistiques. Cela permet de rendre compte de la fiabilité du modèle.
- Fonctionne bien même si ses hypothèses sont quelque peu violées par le vrai modèle à partir duquel les données ont été générées.

c. Les inconvénients

Les inconvénients d'algorithme Decision Tree sont [36]:

- Les apprenants en Decision Tree peuvent créer des arbres trop complexes qui ne généralisent pas bien les données. C'est ce qu'on appelle le surajustement.
- Decision Tree peuvent être instables car de petites variations dans les données peuvent entraîner la génération d'un arbre complètement différent.
- Les prédictions de Decision Tree ne sont ni lisses ni continues, mais des approximations constantes par morceaux. Par conséquent, ils ne sont pas bons pour l'extrapolation.
- Il y a des concepts difficiles à apprendre car Decision Tree ne les expriment pas facilement,
- Les apprenants de Decision Tree créent des arbres biaisés si certaines classes dominent. Il est donc recommandé d'équilibrer l'ensemble de données avant de l'adapter à l'arbre de décision.

3.2.2 Support Vector Machine (SVM)

La méthode Support Vector Machine est basée sur les travaux théoriques de Vanpik et Cortes. Les SVM sont généralement utilisés pour les modèles de classification et appliqués à des problèmes linéairement et non linéairement séparables [37]. L'objectif est de trouver l'hyperplan séparateur optimal parmi une infinité d'hyperplans, qui sépare linéairement l'échantillon d'apprentissage en maximisant la marge (en termes de distance) entre l'hyperplan séparateur et les points les plus proches de la limite de séparation (voir figure). Ces points sont appelés vecteurs supports [38].

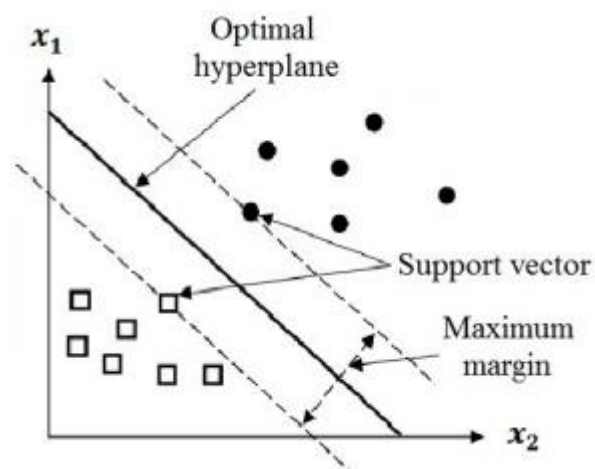


Fig. II. 4. Hyperplan optimal pour Support Vector Machine avec deux classes.

L'hyperplan séparateur est représenté par l'équation suivante :

$$H(x) = \omega^T x + b \quad 2.5$$

a. Les avantages

Les avantages d'algorithme SVM sont [36]:

- Efficace dans les espaces de grande dimension.
- Toujours efficace dans les cas où le nombre de dimensions est supérieur au nombre d'échantillons.
- Utilise un sous-ensemble de points d'apprentissage dans la fonction de décision (appelés vecteurs de support), il est donc également efficace en mémoire.

- Polyvalent : différentes fonctions du noyau peuvent être spécifiées pour la fonction de décision. Des noyaux communs sont fournis, mais il est également possible de spécifier des noyaux personnalisés.

b. Les inconvénients

Les inconvénients d'algorithme SVM sont [36]:

- Si le nombre de caractéristiques est bien supérieur au nombre d'échantillons, évitez le sur-ajustement dans le choix des fonctions du noyau et le terme de régularisation est crucial.
- Les SVM ne fournissent pas directement des estimations de probabilité, celles-ci sont calculées à l'aide d'une validation croisée quintuple coûteuse (voir Scores et probabilités, ci-dessous).

3.2.3 Random Forest (RF)

Random Forest est un algorithme d'apprentissage automatique supervisé largement utilisé dans les problèmes de classification et de régression. Il construit des arbres de décision sur différents échantillons et prend leur vote majoritaire pour le classement et la moyenne en cas de régression.

L'une des caractéristiques les plus importantes de l'algorithme Random Forest est qu'il peut gérer l'ensemble de données contenant des variables continues comme dans le cas de la régression et des variables catégorielles comme dans le cas de la classification. Il donne de meilleurs résultats pour les problèmes de classification[39].

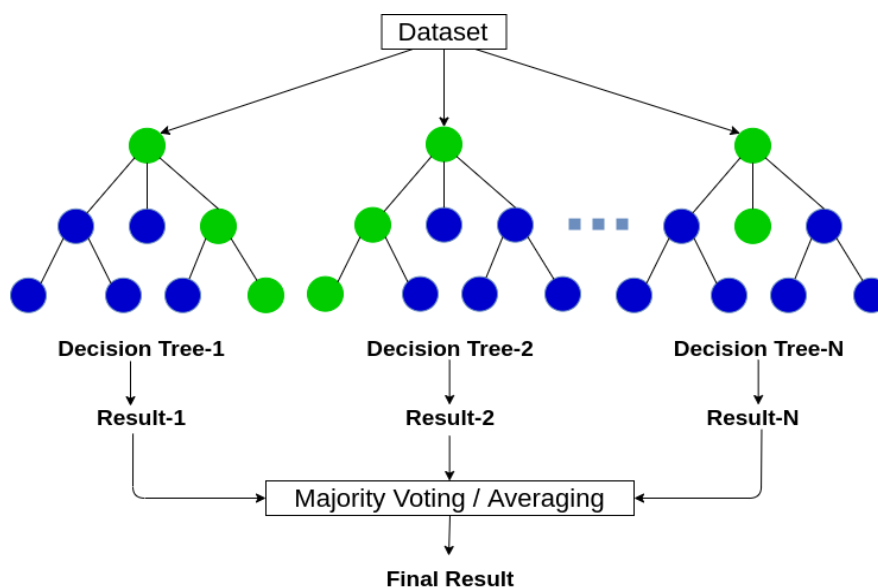


Fig. II. 5. Algorithme de Random Forest.

a. Caractéristiques

Les caractéristiques d'algorithme RF sont [39] :

- Il est plus précis que l'algorithme de Decision Tree.
- Il fournit un moyen efficace de gérer les données manquantes.
- Il peut produire une prédiction raisonnable sans réglage des hyperparamètres.
- Il résout le problème de surajustement dans Decision Tree.
- Dans chaque arbre forestier aléatoire, un sous-ensemble d'entités est sélectionné au hasard au point de division du nœud.

b. Les avantages

Les avantages d'algorithme RF sont [40] :

- Il peut effectuer à la fois des tâches de régression et de classification.
- Random Forest produit de bonnes prédictions qui peuvent être facilement comprises.
- Il peut gérer efficacement de grands ensembles de données.
- L'algorithme de Random Forest offre un niveau de précision plus élevé dans la prédiction des résultats par rapport à l'algorithme Decision Tree.

c. Les inconvénients

Les inconvénient d'algorithme RF sont [40] :

- Lorsque vous utilisez une forêt aléatoire, davantage de ressources sont nécessaires pour le calcul.
- Il consomme plus de temps qu'un algorithme Decision Tree.

3.2.4 Gradient Boosting

L'algorithme de Gradient Boosting est l'un des algorithmes les plus puissants dans le domaine de l'apprentissage automatique. Comme nous le savons, les erreurs dans les algorithmes d'apprentissage automatique sont généralement classées en deux catégories, à savoir l'erreur de biais et l'erreur de variance. Comme l'amplification de gradient est l'un des algorithmes d'amplification, il est utilisé pour minimiser l'erreur de biais du modèle.

L'algorithme de Gradient Boosting peut être utilisé pour prédire non seulement la variable cible continue (en tant que régresseur), mais également la variable cible catégorielle (en tant que classificateur). Lorsqu'il est utilisé comme régresseur, la fonction de coût est

l'erreur quadratique moyenne (MSE) et lorsqu'il est utilisé comme classificateur, la fonction de coût est la perte de journal [41].

$$F_0(x) = \mathbf{arg\ min} \sum_{i=1}^n L(y_i, \gamma) \quad 2.6$$

- **L** est notre fonction de perte ;
- **Gamma** est notre valeur prédite ;
- **argmin** signifie que nous devons trouver une valeur prédite/gamma pour laquelle la fonction de perte est minimale ;

Puisque la colonne cible est continue, notre fonction de perte sera :

$$L = \frac{1}{n} \sum_{i=0}^n (y_i - \gamma_i)^2 \quad 2.7$$

i est la valeur observée.

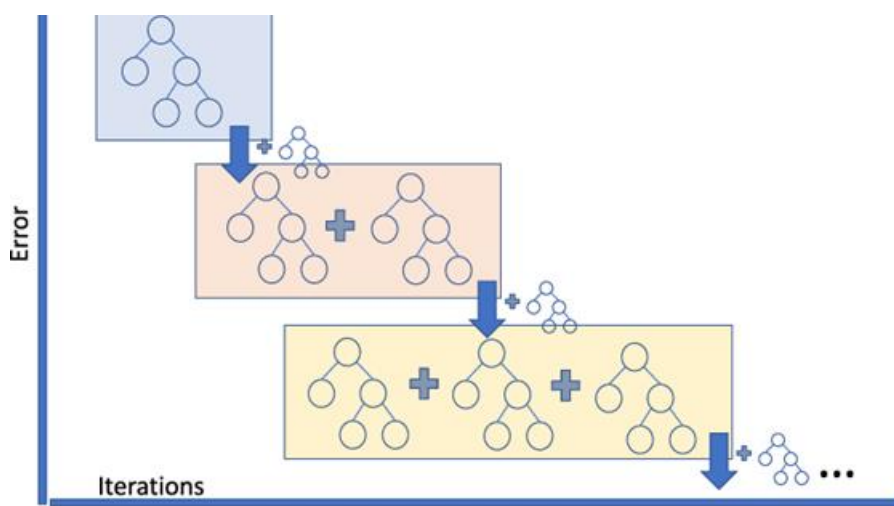


Fig. II. 6. Algorithme de Gradient Boosting.

4 APPLICATIONS D'APPRENTISSAGE STATISTIQUES

Machine learning est un mot à la mode pour la technologie d'aujourd'hui, et il se développe très rapidement de jour en jour. Nous utilisons l'apprentissage automatique dans notre vie quotidienne même sans le savoir, comme Google Maps, l'assistant Google, Alexa, etc. Vous trouverez ci-dessous quelques applications réelles les plus tendances de machine learning [42]:

4.1 Reconnaissance d'images

La reconnaissance d'images est l'une des applications les plus courantes de l'apprentissage automatique. Il est utilisé pour identifier des objets, des personnes, des lieux, des images numériques, etc.[43].

4.2 Reconnaissance de la parole

Lors de l'utilisation de Google, nous avons une option de "Recherche vocale", cela relève de la reconnaissance vocale et c'est une application populaire de l'apprentissage automatique[43].

4.3 Prévision de trafic

Si nous voulons visiter un nouveau lieu, nous prenons l'aide de Google Maps, qui nous montre le chemin correct avec l'itinéraire le plus court et prédit les conditions de circulation[43].

4.4 Recommandations de produits

L'apprentissage automatique est largement utilisé par diverses sociétés de commerce électronique et de divertissement telles qu'Amazon, Netflix, etc., pour la recommandation de produits à l'utilisateur. Chaque fois que nous recherchons un produit sur Amazon, nous avons commencé à recevoir une publicité pour le même produit tout en surfant sur Internet sur le même navigateur et cela est dû à l'apprentissage automatique[36].

4.5 Voitures autonomes

L'une des applications les plus intéressantes de l'apprentissage automatique est la voiture autonome. L'apprentissage automatique joue un rôle important dans les voitures autonomes. Il utilise une méthode d'apprentissage non supervisée pour entraîner les modèles de voiture à détecter les personnes et les objets pendant la conduite[43].

4.6 Filtrage des courriers indésirables et des logiciels malveillants

Chaque fois que nous recevons un nouvel e-mail, il est automatiquement filtré comme important, normal et spam. Nous recevons toujours un courrier important dans notre boîte de réception avec le symbole important et les courriers indésirables dans notre boîte de courrier indésirable, et la technologie derrière cela est l'apprentissage automatique[44].

4.7 Diagnostic médical

En science médicale, l'apprentissage automatique est utilisé pour diagnostiquer des maladies. Grâce à cela, la technologie médicale se développe très rapidement et est

capable de construire des modèles 3D capables de prédire la position exacte des lésions dans le cerveau[36].

5 CONCLUSION

Ce chapitre a fait l'objet d'une présentation des techniques de prédiction de maladies cardiaques vues comme moyen intéressant d'aider les médecins. Les techniques utilisées pour développer ce prédicteur sont des méthodes paramétriques non linéaires qui relient des variables d'entrée à des variables de sortie, appliquées à la prédiction et la classification. Des généralités sur les réseaux DT et RF ont été décrits à cet effet. Un exposé non exhaustif des bases théoriques de l'apprentissage statistique et de la méthode SVM en particulier, a été présenté. Une attention particulière à la technique GB, a été toutefois donnée en rappelant les concepts les plus importants pour sa mise en œuvre en tant qu'outil de prédiction et de classification.

Une étude et analyse des performances apportées par ces techniques appliquées au domaine de prédiction et classification des maladies cardiaques, constitue notre principal objectif dans ce travail. D'ailleurs, une simulation ayant pour but l'évaluation des performances de ces techniques, fera l'objet du chapitre suivant. L'évaluation des résultats obtenus, reflétera les performances attendues de chacune de ces méthodes et conduira au choix de la technique la mieux adaptée à l'application. L'architecture du système de prédiction à base des techniques d'apprentissage statistiques à proposer dans ce cadre, sera présentée.

CHAPITRE III

SIMULATION

1 INTRODUCTION

Ce dernier chapitre est consacré à une étude en simulation et vise principalement l'évaluation des performances des techniques proposées comme étant une solution dans la prédiction des maladies cardiaques. Une étude comparative entre les techniques DT, RF, SVM et GB en termes de précision et d'autres paramètres de classification est alors présentée. L'objectif est de mettre en place un système de prédiction à base de techniques d'apprentissage statistiques afin de prendre une décision adaptée au contrôle et au suivi les malades cardiaques.

2 PROBLÉMATIQUE

Aujourd'hui, la plupart des pays sont confrontés à des taux élevés et croissants de maladies cardiaques et elles sont devenues l'une des principales causes d'affaiblissement et de décès dans le monde chez les hommes et les femmes de plus de soixante-cinq ans et aujourd'hui, dans de nombreux pays, les maladies cardiaques sont considérées comme une « deuxième épidémie », remplaçant les maladies infectieuses comme première cause de décès. Un diagnostic précoce des maladies cardiaques peut aider à réduire le taux de mortalité.

Diagnostiquer une maladie cardiaque par des experts prend beaucoup de temps et coïncide avec le manque d'experts qui possèdent les connaissances. Ainsi, les méthodes automatisées peuvent résoudre les limites des méthodes de diagnostic traditionnelles et fournir des connaissances médicales à des fins de diagnostic. Pour résoudre ce problème et bien d'autres dans le secteur de la santé liés au diagnostic des maladies cardiaques, l'apprentissage automatique peut être une solution en créant des bases de données à partir d'énormes ensembles de données qui peuvent être utilisées pour diagnostiquer les maladies cardiaques.

Le but de cette partie est de développer un système de classification qui peut prédire et justifier avec précision la présence ou l'absence d'un problème cardiaque chez les patients. Il fournit des bases de données pour les patients cardiaques et peut être géré à l'aide d'algorithmes de classification d'intelligence artificielle. Les catégories binaires seront définies comme suit : « 1 » signifie « présence de maladie » et « 0 » signifie « absence de maladie. » Les algorithmes qui seront utilisés à des fins de classification sont SVM, DT, RF et GB. Ensuite, le meilleur modèle passera par une étape interprétable où diverses métriques seront examinées.

3 ARCHITECTURE DU SYSTEME

Le schéma présenté dans la figure III.1 Présente l'architecture du système de prédiction des maladies cardiaques, Il se compose d'une entrée des données obtenues en collectant quatre bases de données différentes , D'un système d'acquisition de données et logiciel permettant le traitement de données et la prise de décision concernant la prédiction de la présence ou de l'absence des maladies cardiaques.

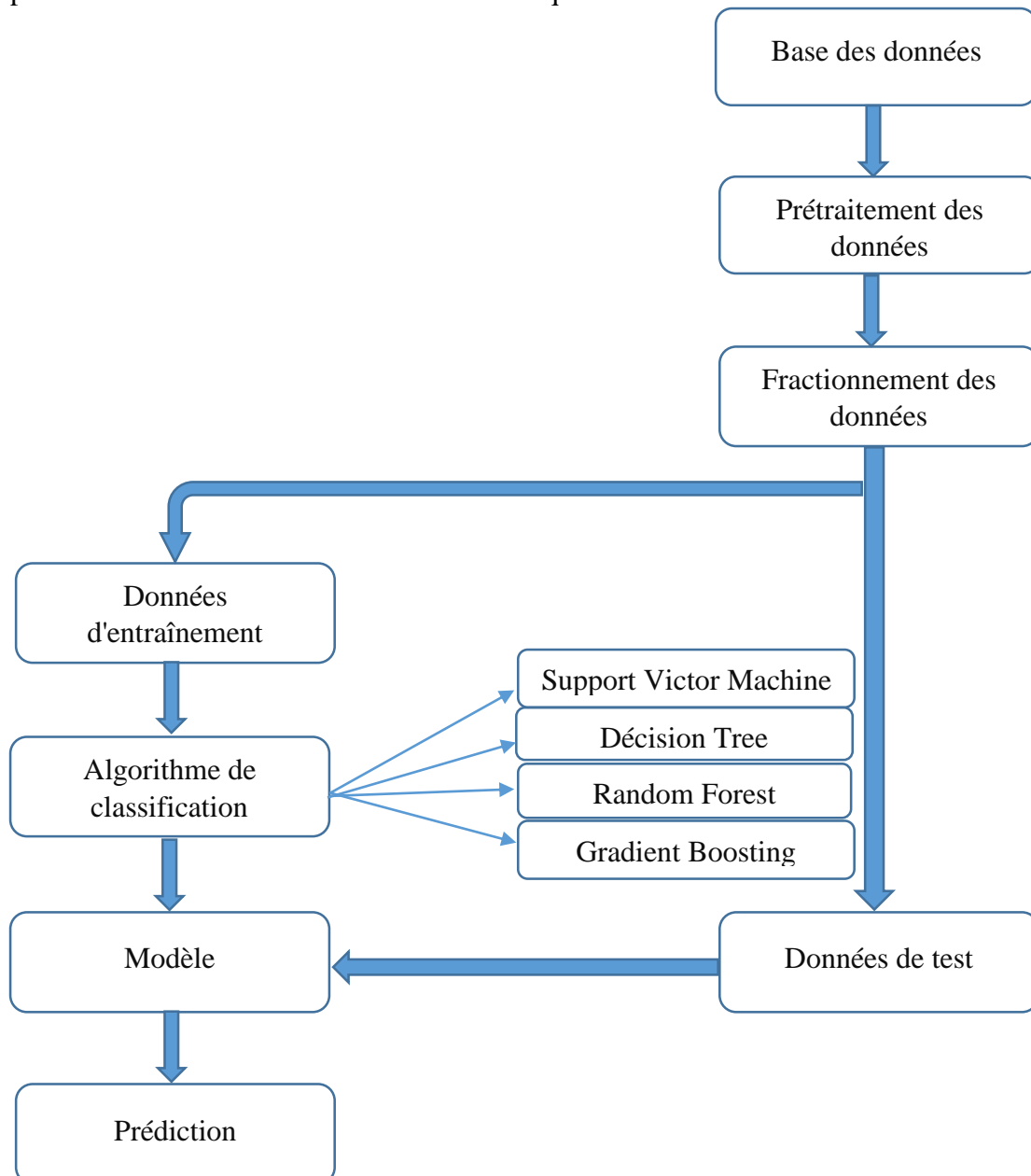


Fig. III. 1. Architecture du système de prédiction des maladies cardiaques.

4 BASE DES DONNÉES UTILISÉE

Cet ensemble de données date de 1988 et se compose de quatre bases de données :

- Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D.
- University Hospital, Zurich, Switzerland: William Steinbrunn, M.D.
- University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D.
- V.A. Medical Center, Long Beach and Cleveland Clinic Foundation: Robert Detrano, M.D., Ph.D.

Il contient 76 attributs, mais seuls 14 d'entre eux sont utilisés pour notre évaluation afin de rendre nos scores comparables à travaux antérieurs, y compris l'attribut prédit, mais toutes les expériences publiées se réfèrent à l'utilisation d'un sous-ensemble de 14 d'entre eux. Le champ "cible" fait référence à la présence d'une maladie cardiaque chez le patient. Il est de valeur entière : 0 : pas de maladie et 1 : maladie [45].

Dans le tableau suivant description des attributs de base des données utilisée :

N°	attributs	Signification	valeurs
1	Age	Age du patient en années	29 à 79
2	Sex	sexe du patient	0=F, 1=M
3	Cp	Type de douleur thoracique	1, 2, 3,4
4	Trestbps	tension artérielle au repos (mm/Hg)	94 à 200
5	Chol	sérum cholestoral en mg/dl	126 à 564
6	Fbs	(glycémie à jeun > 120 mg/dl)	0,1
7	Restecg	résultats électrocardiographiques au repos	0, 1,2
8	Thalach	fréquence cardiaque maximale atteinte	71 à 202
9	Exang	angine d'effort	0,1
10	Old peak	Dépression ST induite par l'exercice par rapport au repos	1 à 3
11	Slope	la pente du segment ST d'effort maximal	1, 2,3
12	Ca	Vaisseaux majeurs colorés par fluoroscopie	0 à 3
13	Thal	Représente la fréquence cardiaque du patient	3, 6,7
14	Target	Présence ou absence de maladie cardiaque	0,1

Tableau. III. 1. Description des attributs de base des données utilisée.

Dans le tableau suivant les statistiques descriptives des paramètres utilisés :

Entrées (Variables)	Minimum	Maximum	Moyenne	Ecart-types
Age	29	77	54.43	9.07
Sex	0	1	0.69	0.46
Cp	0	3	0.94	1.02
Trestbps	94	200	131.61	17.51
Chol	126	564	246	51.59
Fbs	0	1	0.14	0.35
Restecg	0	2	0.52	0.52
Thalach	71	202	149.11	23
Exang	0	1	0.33	0.47
Oldpeak	0	6.2	1.07	1.17
Slope	0	2	1.38	0.61
Ca	0	4	0.75	1.03
Thal	0	3	2.32	0.62
Target	0	1	0.51	0.50

Tableau. III. 2. Statistiques descriptives des paramètres utilisés.

5 PRETRAITEMENT DES DONNEES

Les données réelles sont constituées de valeurs redondantes et de beaucoup de bruit. Les données doivent être nettoyées, et les valeurs manquantes doivent être remplies avant que les données ne soient alimentées pour créer un modèle [46]. Dans le processus de prétraitement, ces problèmes sont pris en compte afin que la prédiction puisse être faite avec précision. Une fois que nous avons fini de nettoyer les données, c'est-à-dire de supprimer le bruit et de remplir les valeurs manquantes, nous devons les convertir. De nombreux algorithmes d'apprentissage supervisé fonctionnent sur des données nominales ou de base. Ainsi, la transformation de données est appliquée à l'ensemble de données dans le travail en cours. La réduction d'ensemble de données est appliquée pour

transformer un ensemble de données complexe en une forme plus simple, ce qui améliore la précision du modèle.

5.1 Normalisation

Les attributs (valeurs des variables du modèle) sont en général exprimés dans des unités différentes, et ont des ordres de grandeurs différentes. Il est donc nécessaire de prétraiter ces valeurs pour qu'elles aient la même influence sur la construction de modèles. Afin d'uniformiser l'importance de chaque entrée sur la détermination des paramètres doivent être normalisées.

La normalisation est très utile lorsque le système a plusieurs variables dont les amplitudes et variances varient dans des plages différentes. Elle consiste à mettre à l'échelle la matrice de données X en rendant les variables de la nouvelle matrice X_N comparables (la même plage de variation). Parmi les méthodes de normalisation de données [47]:

Normalisation par la méthode Min-Max

La normalisation **Min-Max** conserve la distribution de scores originale à un facteur d'échelle près et transforme tous les scores dans l'intervalle $[0,1]$. Le score normalisé Min-Max pour le score de test X_N est donné par l'équation suivante :

$$x_N = \frac{x - \min(X)}{\max(X) - \min(X)} \quad 3.1$$

6 PROTOCOLE EXPÉRIMENTAL

Après avoir décrit les caractéristiques de base de données. Maintenant, l'opération suivante consiste à implémenter un modèle et obtenir des résultats sous forme de probabilité ou de classe. L'étape suivante consiste à déterminer l'efficacité du modèle basé sur une métrique utilisant de données de test. Différentes mesures de performance sont utilisées pour évaluer différents algorithmes d'apprentissage statistique, métrique pour l'évaluation d'algorithmes d'apprentissage statistique est la précision (an Anglais : Accuracy), erreurs, intervenons pour voir ce que sont ces métriques. Voici quelques termes courants avec lesquels il faut être clair :

6.1 Matrice de confusion

Une matrice de confusion (ou matrice d'erreurs) est une disposition de tableau spécifique qui permet de visualiser les performances d'un algorithme d'apprentissage supervisé. Chaque ligne de la matrice représente les instances d'une classe réelle tandis que chaque colonne représente les instances d'une classe prédite. Le tableau ci-dessous est un exemple de matrice de confusion pour une classification binaire à partir de laquelle d'autres terminologies/métriques peuvent être dérivées.

		CLASSE PREDITE	
		Positive	Négative
CLASSE REELLE	Positive	TP	FN
	Négative	FP	TN

Tableau. III. 3. Matrice de confusion.

Avec :

Terme	Signification	Descriptions
TP	True positive	Cas positifs qui sont prédits comme positifs
FP	False positive	Cas négatifs qui sont prédits comme positifs
TN	True négative	Cas positifs qui sont prédits comme négatifs
FN	False négative	Cas négatifs qui sont prédits comme négatifs

Tableau. III. 4. Terme associés à la matrice de confusion.

Nous discuterons des différentes façons de vérifier les performances de notre modèle d'apprentissage statistique et pourquoi utiliser l'un à la place de l'autre. Nous aborderons des termes tels que :

6.2 Accuracy

La métrique la plus couramment utilisée pour juger un modèle et n'est en fait pas un indicateur clair de la performance. Le pire arrive lorsque les classes sont déséquilibrées.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad 3.2$$

6.3 Précision

Pourcentage d'instances positives par rapport au nombre total d'instances positives prévues. Ici, le dénominateur est la prédiction du modèle effectuée comme positive à partir de l'ensemble de données donné. Prenez-le pour savoir « dans quelle mesure le modèle est juste quand il dit qu'il est juste ».

$$\mathbf{Precision} = \frac{TP}{TP+FP} \quad \mathbf{3.3}$$

6.4 Sensitivity

Pourcentage d'instances positives par rapport au nombre total d'instances positives réelles. Par conséquent, le dénominateur (TP + FN) est ici le nombre réel d'instances positives présentes dans l'ensemble de données. Prenez-le pour savoir "combien de bons supplémentaires, le modèle a manqué quand il a montré les bons".

$$\mathbf{Sensitivity} = \frac{TP}{TP+FN} \quad \mathbf{3.4}$$

6.5 F1-score

C'est la moyenne harmonique de la précision et du rappel. Cela prend la contribution des deux, donc plus le score F1 est élevé, mieux c'est. Voyez qu'en raison du produit dans le numérateur si l'un devient bas, le score final de la F1 diminue considérablement. Ainsi, un modèle réussit bien dans le score F1 si les prédicts positifs sont en fait positifs (précision) et ne manquent pas de positifs et les prédisent négatifs (recall).

$$\mathbf{F1\ score} = \frac{2}{\frac{1}{precision} + \frac{1}{recall}} \quad \mathbf{3.5}$$

$$\mathbf{F1\ score} = \frac{2*precision*recall}{precision+recall} \quad \mathbf{3.6}$$

Donc :

$$\mathbf{F1\ score} = \frac{2*TP}{2*TP+FP+FN} \quad \mathbf{3.7}$$

7 ALGORITHMES DE PREDICTION

Les résultats ont été obtenus en appliquant différents algorithmes de classification à l'ensemble de données utilisé. Où nous avons utilisé 4 algorithmes d'apprentissage statistique qui est :

7.1 Support Vector machine

Quatre modèles différents obtenus en changeant les paramètres d'algorithme SVM ont été utilisés

Les modèles	Paramètre	Apprentissage				Test			
		Acc (%)	Sens (%)	Prec (%)	F1 score (%)	Acc (%)	Sens (%)	Prec (%)	F1 score (%)
Model 1	C=1	69	79	68	73	66	72	63	67
Model 2	C=10	75	80	75	77	72	73	70	71
Model 3	C=100	87	93	85	88	84	93	77	85
Model 4	C=500	90	93	88	90	85	93	80	86

Tableau. III. 5. Résultats obtenus avec le model SVM.

Le tableau précédent représente les résultats de quatre modèles de l'algorithme SVM, et dans chaque modèle le paramètre de l'algorithme est modifié pour obtenir une meilleure précision. Où l'on remarque qu'à chaque fois qu'on augmente le paramètre de l'algorithme (C) les résultats sont stables et précis.

Les meilleurs résultats étaient lorsque C = 500.

Note :

C : Paramètre de régularisation. La force de la régularisation est inversement proportionnelle à C. Doit être strictement positive.

Matrice de confusion de modèle SVM

		Prédite		
		No maladies	Maladies	
Réelle	No maladies	125	34	No maladies
	Maladies	10	139	Maladies

Tableau. III. 6. La matrice de confusion des patients avec le model SVM.

7.2 Décision Tree

Quatre modèles différents obtenus en changeant les paramètres d'algorithme DT ont été utilisés :

Les modèles	Paramètre	Apprentissage				Test			
		Acc (%)	Sens (%)	Prec (%)	F1 score (%)	Acc (%)	Sens (%)	Prec (%)	F1 score (%)
Model 1	Min samples split=20	95	94	97	95	90	85	94	89
Model 2	Min samples split=10	97	96	98	97	91	87	94	90
Model 3	Min samples split=8	98	98	98	98	94	92	96	94
Model 4	Min samples split=2	100	100	100	100	97	93	100	96

Tableau. III. 7. Résultats obtenus avec le model DT.

Le tableau précédent représente les résultats de quatre modèles de l'algorithme DT, Dans chaque modèle le paramètre de l'algorithme est modifié afin d'obtenir une meilleure précision. Où l'on remarque qu'à chaque fois qu'on réduit le paramètre de l'algorithme (**Min_sample_split**) les résultats sont stables et précis.

Les meilleurs résultats étaient lorsque $\text{Min_sample_split} = 2$.

Matrice de confusion de modèle DT

		Prédite		
		No maladies	Maladies	
réelle	No maladies	159	0	No maladies
	Maladies	9	140	Maladies
		No maladies	Maladies	

Tableau. III. 8. La matrice de confusion des patients avec le model DT.**7.3 Random Forest**

Quatre modèles différents obtenus en changeant les paramètres d'algorithme RF ont été utilisés :

Les modèles	Paramètre	Apprentissage				Test			
		Acc (%)	Sens (%)	Prec (%)	F1 score (%)	Acc (%)	Sens (%)	Prec (%)	F1 score (%)
Model 1	Min samples split=20	94	97	92	94	88	92	84	88
Model 2	Min samples split = 10	98	98	98	98	94	95	94	94
Model 3	Min samples split = 8	98	99	98	99	96	95	97	96
Model 4	Min samples split = 2	100	100	100	100	98	95	100	97

Tableau. III. 9. Résultats obtenus avec le model RF.

Le tableau précédent représente les résultats de quatre modèles de l'algorithme RF, Dans chaque modèle le paramètre de l'algorithme est modifié afin d'obtenir une meilleure précision. Où l'on remarque qu'à chaque fois qu'on réduit le paramètre de l'algorithme (**Min_sample_split**) les résultats sont stables et précis.

Les meilleurs résultats étaient lorsque $\text{Min_sample_split} = 2$.

Matrice de confusion de modèle RF

		Prédite		
		No maladies	Maladies	
réelle	No maladies	159	0	No maladies
	Maladies	6	143	Maladies
		No maladies	Maladies	

Tableau. III. 10. La matrice de confusion des patients avec le model RF.

7.4 Gradient Boosting

Quatre modèles différents obtenus en changeant les paramètres d’algorithme GB ont été utilisés :

Les modèles	Paramètre	Apprentissage				Test			
		Acc (%)	Sens (%)	Prec (%)	F1 score (%)	Acc (%)	Sens (%)	Prec (%)	F1 score (%)
Model 1	Number estimators=50	96	98	94	96	89	93	85	89
Model 2	Number estimators= 100	98	98	98	98	94	94	94	94
Model 3	Number estimators= 150	99	100	99	99	98	97	98	98
Model 4	Number estimators= 200	100	100	100	100	99	97	100	98

Tableau. III. 11. Résultats obtenus avec le model GB.

Le tableau précédent représente les résultats de quatre modèles de l’algorithme GB, et dans chaque modèle le paramètre de l’algorithme est modifié pour obtenir une meilleure précision. Où l’on remarque qu’à chaque fois qu’on augmente le paramètre de l’algorithme (n_estimators) les résultats sont stables et précis.

Les meilleurs résultats étaient lorsque n_estimators = 200.

Matrice de confusion de modèle GB

		Prédite		
		No maladies	Maladies	
réelle	No maladies	159	0	No maladies
	Maladies	3	146	Maladies
		No maladies	Maladies	

Tableau. III. 12. La matrice de confusion des patients avec le model GB.

8 EVALUATION DES PERFORMANCES

Les sections suivantes présentent les résultats obtenus en appliquant un modèle de classification différent à l'ensemble de données utilisé pour la cardiologie.

8.1 Accuracy

Cet histogramme représente les résultats obtenus au critère d'accuracy :

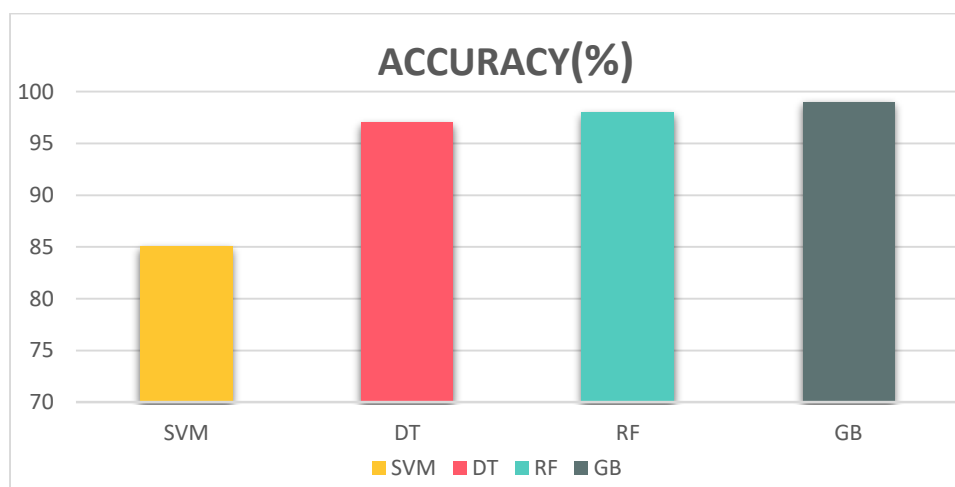


Fig. III. 2. Résultat obtenu selon l'accuracy.

La figure **III.2** montre la comparaison des différents modèles de prédiction utilisés pour prédire les maladies cardiaques selon l'accuracy. Les résultats étaient proches. Cependant, l'algorithme Gradient Boosting a donné la plus grande accuracy pour l'ensemble de données utilisé.

8.2 Précision

Cet histogramme représente les résultats obtenus au critère de précision :

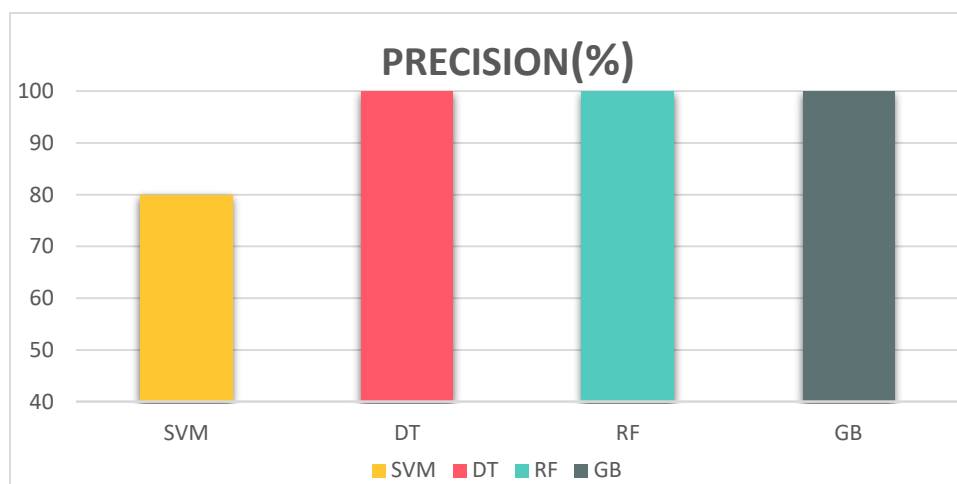


Fig. III. 3. Résultat obtenu selon la précision.

La figure **III.3** montre une comparaison des différents modèles de prédiction utilisés pour prédire les maladies cardiaques selon la précision. Les résultats étaient les mêmes pour les trois algorithmes (DT, RF, GB). L'algorithme SVM a donné la précision la plus faible pour l'ensemble de données utilisé.

8.3 Sensitivity

Cet histogramme représente les résultats obtenus au critère de Sensitivity :

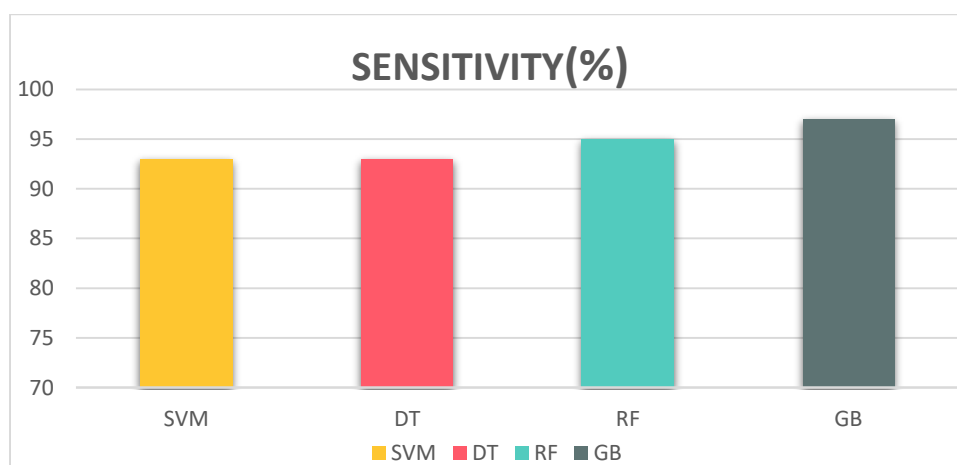


Fig. III. 4. Résultat obtenu selon la Sensitivity.

La figure **III.4** montre la comparaison des différents modèles de prédiction utilisés pour prédire les maladies cardiaques selon la Sensitivity. Les résultats étaient proches. Cependant, l'algorithme Gradient Boosting a donné la plus grande Sensitivity pour l'ensemble de données utilisé.

8.4 F1 Score

Cet histogramme représente les résultats obtenus au critère de f1 score :

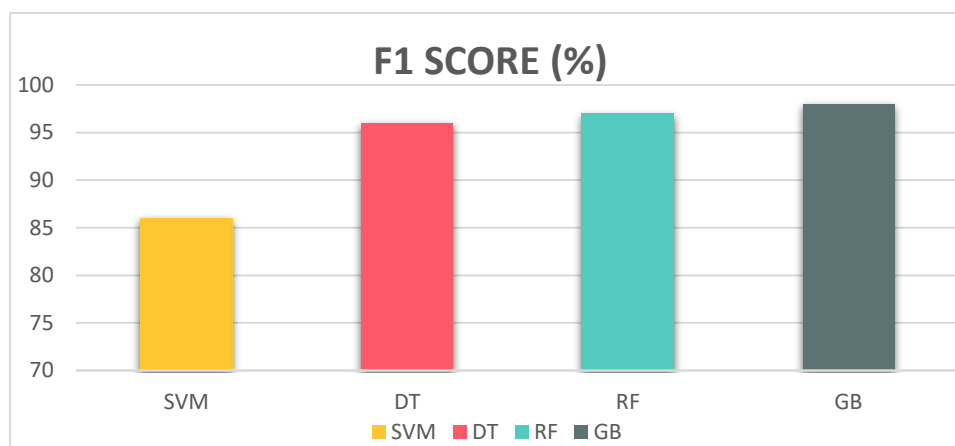


Fig. III. 5. Résultat obtenu selon F1 Score.

La figure **III.5** montre la comparaison des différents modèles de prédiction utilisés pour prédire les maladies cardiaques selon F1 Score. Les résultats étaient proches. Cependant, l'algorithme Gradient Boosting a donné la plus grande F1 Score pour l'ensemble de données utilisé.

9 COMPARAISON AVEC L'ETAT DE L'ART

Afin de donner une idée du classement de notre système de prédiction des maladies cardiaques en termes de performances, nous comparons des travaux qui ont utilisé le même protocole expérimental et les mêmes mesures de performances. Nous notons également que le temps d'exécution n'est pas pris en compte dans la comparaison en raison de l'absence de cette information dans les œuvres que nous comparons. La précision de la prédiction des maladies cardiaques calculée dans notre système proposé ainsi que dans d'autres travaux similaires est rapportée dans le tableau **III.13**. [4].

Recherche	Modèle utilisée	Accuracy (%)	Sensitivity (%)	Précision (%)	F1 score (%)
[9]	DT	93	98	90	94
[10]	DL	94	82	/	/
[11]	GNB-RF	98	98	98	98
[12]	KNN	87	73	83	77

[13]	KNN	99	/	/	/
Notre travail	GB	99	97	100	98

Tableau. III. 13. Comparaison avec l'état de l'art.

10 CONCLUSION

Ce dernier chapitre a fait l'objet d'une étude en simulation qui consistait à évaluer les performances de quatre techniques d'apprentissage statistique appliquées au domaine de prédiction des maladies cardiaques. Cette étude a permis la comparaison entre les modèles DT, RF, SVM et GB, en vue de choisir la technique la mieux adaptée pour la prédiction. L'objectif a été de mettre en place un système intelligent à base des techniques d'apprentissage statistiques permettant de prendre une décision adaptée au suivi des malades cardiaques.

Les paramètres liés aux Accuracy, Précision, Sensitivity et F1-score, ont été les facteurs pertinents qui ont permis d'évaluer les méthodes étudiées. Les résultats obtenus ont montré que ces quatre modèles étaient similaires en terme de précision ce qui leur confère l'avantage de s'intégrer dans un système de surveillance dynamique.

CONCLUSION GENERALE

Le travail présenté dans le cadre de cette mémoire a fait l'objet d'une contribution à la conception de systèmes intelligents de surveillance et de prédiction des maladies cardiaques à base des techniques d'apprentissage statistiques.

Le travail présenté a été structuré autour de trois chapitres essentiels. Le premier a été consacré à des généralités concernant les maladies cardiaques, les types des maladies cardiaques, les symptômes, les causes ainsi que les tests couramment utilisés pour diagnostiquer les maladies cardiovasculaires. Dans le deuxième chapitre, quatre modèles (DT, RF, SVM et GB) à base de la conception de ces prédicteurs ont été étudiés et évalués. Enfin, le troisième et dernier chapitre a été dédié à une application réelle justifiant la conception et la mise en œuvre d'un système intelligent pour la prédiction des maladies cardiaques. Plusieurs paramètres ont été évalués à cet effet, tels que l'accuracy, la précision etc.

Ce travail a donc permis la validation et l'évaluation des performances de chacune des techniques étudiées. L'étude comparative entreprise entre les modèles DT, RF, SVM et GB dans le but d'un choix décisif de la méthode la mieux adaptée à la prédiction et à la classification a donc été effectuée. Les résultats obtenus ont montré d'abord que ces quatre modèles étaient similaires en termes de précision, ce qui leur donne l'avantage de s'intégrer dans un système de prédiction en ligne.

Reference

- [1] V. F. Dr. Vladimir, *les maladies des coeur et les maladies vasculaire cérébrales*, vol. 1, no. 69. 1967.
- [2] “Cardiovascular diseases.” https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1 (accessed Apr. 02, 2022).
- [3] M. S. Amin, Y. K. Chiam, and K. D. Varathan, “Identification of significant features and data mining techniques in predicting heart disease,” *Telemat. Informatics*, vol. 36, pp. 82–93, Mar. 2019, doi: 10.1016/J.TELE.2018.11.007.
- [4] M. Djerioui, Y. Brik, M. Ladjal, and B. Attallah, “Neighborhood component analysis and support vector machines for heart disease prediction,” *Ing. des Syst. d’Information*, vol. 24, no. 6, pp. 591–595, 2019, doi: 10.18280/isi.240605.
- [5] J. Patel, A. A. Khaked, J. Patel, and J. Patel, “Heart Disease Prediction Using Machine Learning,” *Lect. Notes Networks Syst.*, vol. 203 LNNS, no. April, pp. 653–665, 2021, doi: 10.1007/978-981-16-0733-2_46.
- [6] B. Adhikari and S. Shakya, “Heart Disease Prediction Using Ensemble Model”, doi: 10.1007/978-981-16-7657-4_69.
- [7] R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, “Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning,” 2021, doi: 10.1155/2021/8387680.
- [8] J. Patel, A. A. Khaked, J. Patel, and J. Patel, “Heart Disease Prediction Using Machine Learning,” *Lect. Notes Networks Syst.*, vol. 203 LNNS, pp. 653–665, 2021, doi: 10.1007/978-981-16-0733-2_46.
- [9] K. Srivastava* and D. K. Choubey*, “Heart Disease Prediction using Machine Learning and Data Mining,” *Int. J. Recent Technol. Eng.*, vol. 9, no. 1, pp. 212–219, 2020, doi: 10.35940/ijrte.f9199.059120.
- [10] K. M. Almustafa, “Prediction of heart disease and classifiers’ sensitivity analysis,” *BMC Bioinformatics*, vol. 21, no. 1, pp. 1–18, Jul. 2020, doi: 10.1186/S12859-020-03626-Y/FIGURES/10.

- [11] “What is Cardiovascular Disease? | American Heart Association.” <https://www.heart.org/en/health-topics/consumer-healthcare/what-is-cardiovascular-disease> (accessed Apr. 02, 2022).
- [12] A. V. Ferry et al., “Presenting symptoms in men and women diagnosed with myocardial infarction using sex-specific criteria,” *J. Am. Heart Assoc.*, vol. 8, no. 17, Sep. 2019, doi: 10.1161/JAHA.119.012307.
- [13] M. J. Prust, W. G. Stevenson, G. R. Strichartz, and L. S. Lilly, “Mechanisms of cardiac arrhythmias,” *Pathophysiol. Hear. Dis. A Collab. Proj. Med. Students Fac.*, vol. 65, no. 2, pp. 268–286, 2015, doi: 10.1016/j.rec.2011.09.020.
- [14] B. A. Rn, “Blue is Not the New Pink : An Introduction to Congenital Heart Defects”.
- [15] L. E. Y. Bernard, J.L. Bonnet, “Item 250 : Insuffisance cardiaque de l ’ adulte,” *Collège Des Enseignants Cardiol. Mal. Vasc.*, p. 21, 2012.
- [16] “Heart disease - Symptoms and causes - Mayo Clinic.” <https://www.mayoclinic.org/diseases-conditions/heart-disease/symptoms-causes/syc-20353118> (accessed Apr. 02, 2022).
- [17] “maladie cardiaque chez les hommes : 12 symptômes dangereux, vérifiez immédiatement - flx2.com.” <https://www.flx2.com/post/6401> (accessed Apr. 18, 2022).
- [18] P. De Jonge, D. Mangano, and M. A. Whooley, “Differential Association of Cognitive and Somatic Depressive Symptoms With Heart Rate Variability in Patients With Stable Coronary Heart Disease: Findings From the Heart and Soul Study,” *Psychosom. Med.*, vol. 69, no. 8, p. 735, 2007, doi: 10.1097/PSY.0B013E31815743CA.
- [19] “Understand Your Risks to Prevent a Heart Attack | American Heart Association.” <https://www.heart.org/en/health-topics/heart-attack/understand-your-risks-to-prevent-a-heart-attack> (accessed Apr. 02, 2022).
- [20] “Some jobs seem riskier when it comes to heart health - CBS News.” <https://www.cbsnews.com/news/some-jobs-seem-riskier-when-it-comes-to-heart-health/> (accessed Apr. 02, 2022).

- [21] “Amiram Elwork | Lawyers With Depression.” <https://lawyerswithdepression.wordpress.com/tag/amiram-elwork/> (accessed Apr. 02, 2022).
- [22] A. Tverdal and K. Bjartveit, “Health consequences of reduced daily cigarette consumption,” *Tob. Control*, vol. 15, no. 6, pp. 472–480, Dec. 2006, doi: 10.1136/TC.2006.016246.
- [23] “Heart rate: What is a normal heart rate?” <https://www.medicalnewstoday.com/articles/235710> (accessed Apr. 02, 2022).
- [24] E. Strickland, “What Is High Blood Pressure,” *S.F. Wkly.*, no. 54, p. 1, 2007, [Online]. Available: https://www.lib.uwo.ca/cgi-bin/ezpauthn.cgi?url=http://search.proquest.com/docview/367778105?accountid=15115%5Cnhttp://vr2pk9sx9w.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQ%3Aaltpresswatch&rft_val_fmt=
- [25] NIH, “Lowering Your Cholesterol With TLC (Therapeutic Lifestyle Changes),” *Natl. Inst. Heal.*, pp. 1–85, 2005.
- [26] Z. Punthakee, M. Sc, R. Goldenberg, and P. Katz, “Et Du Syndrome Métabolique,” *Elsevier.can j diabetes*, vol. 42, 2018.
- [27] S. B. Dulak, “Your Guide To Living Well With Heart Disease,” *RN*, vol. 67, no. 11, 2004.
- [28] M. Yousef, A. Sulieman, C. Edward, and B. Ahmed, “A Review on Diagnostic Procedures for the Cardiovascular System,” *Int. J. Sci. Res.*, vol. 2, no. 2, pp. 2319–7064, 2013, [Online]. Available: www.ijsr.net
- [29] T. Jernberg, “Common investigations in cardiovascular disease,” *Hear. Educ. Rehabil. Toolkit*, vol. 1, no. 1, pp. 1–3, 2014, [Online]. Available: www.heartonline.org.au/resources
- [30] M. A. Boden, “1. Qu’est-ce au juste que l’intelligence artificielle?,” *L’intelligence Artif.*, pp. 13–32, 2021, doi: 10.1051/978-2-7598-2580-6.c003.
- [31] R. Algerienne and D. Et, “Remerciements,” 2019.

- [32] “Apprentissage artificiel - A.Cornuéjols, L.Miclet, V.Barra - 4e... - Éditions Eyrolles.”
- [33] A. Abdi, “Three types of Machine Learning Algorithms List of Common Machine Learning Algorithms,” no. November, 2016, doi: 10.13140/RG.2.2.26209.10088.
- [34] T. Oladipupo, “Types of Machine Learning Algorithms,” *New Adv. Mach. Learn.*, no. February 2010, 2010, doi: 10.5772/9385.
- [35] S. Umadevi and K. S. J. Marseline, “A survey on data mining classification algorithms,” *Proc. IEEE Int. Conf. Signal Process. Commun. ICSPC 2017*, vol. 2018-Janua, no. July, pp. 264–268, 2018, doi: 10.1109/CSPC.2017.8305851.
- [36] S. Angra and S. Ahuja, “Machine learning and its applications: A review,” *Proc. 2017 Int. Conf. Big Data Anal. Comput. Intell. ICBDACI 2017*, no. January, pp. 57–60, 2017, doi: 10.1109/ICBDACI.2017.8070809.
- [37] C. Cortes, V. Vapnik, and L. Saitta, “Support-Vector Networks Editor,” *Mach. Learning*, vol. 20, pp. 273–297, 1995.
- [38] V. N. Vapnik, “Statistical learning theory,” p. 736, 1998.
- [39] “Introduction to Random Forest in Machine Learning | Engineering Education (EngEd) Program | Section.” <https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/> (accessed Apr. 04, 2022).
- [40] M. Raihan, “Prediction on Ischemic Heart Disease using Machine Learning Approaches,” 2019.
- [41] N. M. Lemekhova and A. L. Petrushin, “A Gentle Introduction to Gradient Boosting,” *Nov. Khirurgii*, vol. 23, no. 5, pp. 566–569, 2015, doi: 10.18484/2305-0047.2015.5.566.
- [42] W. Apt, “Introduction to machine learning,” *Demogr. Res. Monogr.*, pp. 1–13, 2014, doi: 10.1007/978-94-007-6964-9_1.
- [43] J. Shapiro, “machine learning and its applications,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture*

- Notes in Bioinformatics), vol. 2049 LNAI. pp. 146–168, 2001. doi: 10.1007/3-540-44673-7_7.
- [44] “Top 10 Machine Learning Applications and Examples in 2022.” <https://www.simplilearn.com/tutorials/machine-learning-tutorial/machine-learning-applications> (accessed Apr. 23, 2022).
- [45] R. Algerienne, D. Et, and R. Scientifique, “Prédiction des maladies les plus fréquentes par les techniques de deep learning * Remerciements,” 2020.
- [46] V. Madaan and A. Goyal, “Predicting Ayurveda-Based Constituent Balancing in Human Body Using Machine Learning Methods,” *IEEE Access*, vol. 8, pp. 65060–65070, 2020, doi: 10.1109/ACCESS.2020.2985717.
- [47] M. Djerioui, “Contribution au Développement de Systèmes Multicapteurs Intelligents Dédiés à la Surveillance et au Contrôle de la Qualité des Eaux Propres.,” 2019.