

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique

Université de Relizane
Faculté des Sciences et Technologie
Département Informatique



جامعة غليزان
RELIZANE UNIVERSITY

MEMOIRE

En vue de l'obtention du diplôme de MASTER en :
Informatique
Spécialité : Réseaux Informatiques
Intitulé

***Contribution à l'amélioration de la
prévention des maladies cardiaques par les
techniques d'Optimisation***

Présenté par :

- Boucedra Ahlam
- Mekhdoum karima

Devant les membres de jury :

Président : Mr. Benyettou Abdelkader

Encadreur : Mr. Benyettou mohammed

Examineur : Mr. Belkadi Abdellah

Année universitaire : 2024/2025

Remerciements

Avant tout, nous rendons grâce à Dieu Tout-Puissant qui nous a donné la force, la patience et le courage nécessaires pour mener à bien ce travail.

Nous exprimons notre profonde reconnaissance à nos chers parents pour leur amour inconditionnel, leurs sacrifices, leurs prières et leur soutien moral tout au long de notre parcours universitaire. Sans eux, rien n'aurait été possible.

Nous tenons à remercier sincèrement monsieur [benyattou mohammed], notre encadrant, pour sa disponibilité, ses précieux conseils, son encadrement rigoureux et sa bienveillance tout au long de ce travail.

Nous remercions également les membres du jury qui ont accepté d'évaluer notre travail, pour l'intérêt qu'ils y portent ainsi que pour leurs remarques et suggestions constructives.

Enfin, nous souhaitons adresser nos remerciements à l'ensemble du personnel administratif et pédagogique du département d'informatique, pour les connaissances transmises, l'encadrement général, et l'environnement favorable à l'apprentissage tout au long de notre formation.

Résumé

Les maladies cardiaques, et en particulier la maladie coronarienne, figurent parmi les principales causes de décès dans le monde. Leur détection précoce reste un défi majeur en raison de la complexité des facteurs de risque et du coût des examens cliniques. Dans ce contexte, l'intelligence artificielle (IA) s'impose comme un outil prometteur pour améliorer la prévention et le diagnostic.

Ce mémoire propose une approche de prédiction du risque de maladie coronarienne en utilisant plusieurs modèles d'apprentissage automatique supervisé. Contrairement à d'autres études qui se limitent à un seul algorithme, nous avons opté pour une combinaison de trois modèles puissants : Decision Tree, Random Forest et Bagging Classifier. L'objectif est d'analyser leur comportement et leur complémentarité dans la détection de patients à risque.

Les expériences ont été menées sur une base de données médicale intitulée heartDiseaseTain, contenant des informations cliniques pertinentes sur divers patients. Un ensemble de techniques de prétraitement a été appliqué afin de nettoyer et préparer les données pour l'entraînement des modèles.

Les résultats ont permis de dégager les avantages de chaque modèle, sans privilégier l'un par rapport aux autres, tout en mettant en évidence l'importance des approches d'ensemble dans la stabilité et la précision de la prédiction. Cette étude confirme le potentiel de l'IA dans la lutte contre les maladies cardiovasculaires et ouvre la voie à des outils de prévention plus intelligents et personnalisés.

Abstract

Heart diseases, particularly coronary heart disease, are among the leading causes of death worldwide. Early detection remains a major challenge due to the complexity of risk factors and the high cost of clinical examinations. In this context, artificial intelligence (AI) emerges as a promising tool to improve prevention and diagnosis.

This work aims to propose a prediction approach for the risk of coronary heart disease using multiple supervised machine learning models. Unlike studies that rely on a single algorithm, we adopted three powerful models: Decision Tree, Random Forest, and Bagging Classifier, in order to analyze their behavior and combined effectiveness in identifying patients at risk.

Experiments were conducted using a medical dataset named heartDiseaseTain, which contains relevant clinical information about various patients. A set of preprocessing techniques was applied to clean and prepare the data for model training.

The results highlighted the strengths of each model individually, without favoring any specific one, while demonstrating the benefits of ensemble methods in improving prediction accuracy and stability. This study confirms the potential of AI as a decision support tool in the prevention of cardiovascular diseases and paves the way for more intelligent and personalized preventive system

Mots clés : Maladies cardiaque, prévention, intelligence artificielle (IA), Techniques d'optimisation, Apprentissage automatique, Modèles de classification, Prédiction du risque, Arbre de décision, Forêt aléatoire, Bagging.....

Table des matières

Remerciements

Résumé

Abstract

Table des matières

Table des figures

Liste des tableaux

Liste des abréviations

Introduction générale	1
1 Contexte global	1
2 Problématique	1
3 Objectifs de notre travail	1
4 Organisation du rapport	2
Chapitre 1	3
Généralités sur les maladies cardiaques	3
1 Introduction :	4
2 Anatomie du cœur	4
2.1 Le cœur :	4
2.2 Structure interne du cœur :	4
2.3 Les valves cardiaques :	4
2.4 Le myocarde :	5
2.5 Le péricarde :	5
2.6 Les vaisseaux sanguins associés au cœur :	5
2.7 Circulation sanguine :	5
3 Fonctionnement du cœur :	5
4 Les maladies cardiaques	6
5 Types de maladies du cœur	6
5.1 La coronaropathie (Chd) :	6
5.2 Les maladies vasculaires :	7
5.3 Les troubles du rythme cardiaque (arythmies)	8
5.4 La cardiopathie congénitale (malformations cardiaques)	8
5.5 L'insuffisance cardiaque :	9
5.6 Autres maladies du cœur :	10
6 Facteurs de risque des maladies cardiaque :	10

7	Diagnostic des maladies cardiaques :	11
8	Prévention des maladies cardiaques:	12
8.1	Prévention individuelle	12
8.2	Prévention collective:	12
	Conclusion	13
	Chapitre 2	14
	Apprentissage automatique	14
1.	Introduction :	15
2.	Apprentissage automatique :	15
3.	Les différentes méthodes d'apprentissage automatique :	16
3.1	Apprentissage supervisé :	16
3.2	Apprentissage non supervisé :	16
3.3	Apprentissage par renforcement :	17
4	Les algorithmes d'apprentissage supervisé utilisés :	17
4.1	Arbre de Décision :	17
4.1.1	Définition :	17
4.1.2	Classification des Arbres de Décision :	18
4.1.3	Fonctionnement des Arbres de Décision :	18
4.1.4	Avantage des Arbres de Décision	19
4.1.5	Inconvénients des Arbres de Décision	19
4.1.6	Applications des Arbres de Décision	19
4.2	Random Forest Classificateur :	19
4.2.1	Définition :	19
4.2.2	Fonctionnement de Random Forest :	19
4.2.3	Avantages de Random Forest :	20
4.2.4	Inconvénients de Random Forest :	20
4.2.5	Applications de Random Forest :	21
4.3	Bagging Classificateur :	21
4.3.1	Définition :	21
4.3.2	Fonctionnement du Bagging Classifie Bagging Classificateur :	21
4.4	AdaBoost (Adaptive Boosting)	22
4.5	Gradient Boosting	23
4.6	SVM (Support Vector Machine)	23
4.7	KNN (K-Nearest Neighbors)	23
4.8	ANN (Artificial Neural Networks)	23
5	Applications de l'apprentissage automatique dans le secteur de la santé	23
5.1	Le diagnostic des maladies cardiaques	23

5.2	La prédiction du diabète	23
5.3	La découverte de nouveaux médicaments.....	24
6	Des travaux étudiés	24
	Conclusion :	25
	Chapitre 3	26
	Implémentation et résultats expérimentaux	26
1	Introduction :	27
2	L'architecture du système :	27
2.1	L'ensemble de données :.....	28
2.2	Analyse et Exploration des données :.....	28
2.3	Le prétraitement de données :.....	29
2.3.1	Vérification des valeurs manquantes :	29
2.3.2	Nettoyage des données :.....	30
2.3.3	Transformation et encodage :.....	30
2.3.4	Standardisation des données :	31
2.3.5	Extraction des caractéristiques:	31
2.3.6	Le fractionnement des données	31
3	Test du modèle :	32
4	Optimisation et évaluation des modelés :	33
4.1	Les mesures d'évaluation de performance :	33
4.2	Test sur le sous-ensemble de teste:	33
4.3	Matrice de confusion de modèle :.....	34
4.4	Test sur un échantillon de teste :.....	35
	Conclusion.....	37
	Chapitre 4	37
	Conception et fonctionnalités de l'application	37
1	Introduction	38
2	Environnement de développement	38
2.1	Plateforme de développement	38
2.1.1	Google Colab.....	38
2.1.2	Jupyter Notebook	38
2.2	Langage de développement	38
2.2.1	Python	38
2.3	Bibliothèques utilisées	39
2.3.1	Pandas	39
2.3.2	NumPy.....	39
2.3.3	Matplotlib	39

2.3.4	Seaborn	39
2.3.5	SciPy.stats	39
3	Réalisation de l'interface utilisateur	39
3.1	Choix des technologies utilisées	39
3.1.1	ReactJS (JavaScript)	39
3.1.2	Next.js (JavaScript)	39
3.1.3	Prisma (JavaScript/TypeScript)	39
3.1.4	PostgreSQL (Base de données)	40
3.1.5	FastAPI (Python)	40
4	Aperçu de l'Interface Utilisateur	40
4.1	Écran de connexion utilisateur	40
4.2	Tableau de bord principal	41
4.3	Formulaire d'enregistrement de patient	42
4.4	Liste des patients	43
4.5	Formulaire d'évaluation du risque de maladie cardiaque	43
4.6	Liste des rapports de diagnostic récents	44
5	Rapport d'évaluation du risque de maladie cardiaque	45
	Conclusion :	47
	Conclusion générale	48
	Bibliographies	49

Table des figures

Figure 1: fonctionnement de cœur	6
Figure 2: La maladie coronaire	7
Figure 3 : Les maladies vasculaires	7
Figure 4: Les troubles du rythme cardiaque	8
Figure 5: La cardiopathie congénitale (malformations cardiaques)	9
Figure 6: La différence entre les deux cœurs	9
Figure 7: relation de IA et machine Learning	15
Figure 8: Les différentes méthodes d'apprentissage automatique	16
Figure 9: Apprentissage supervisé vs Apprentissage non supervisé.....	17
Figure 10: Structure d'un arbre de décision.	18
Figure 11: Arbre de décision.	18
Figure 12: Illustration d'un ensemble d'arbres dans une forêt aléatoire avec agrégation des prédictions.	20
Figure 13: Schéma du Bagging (Bootstrap Aggregating) pour les Classifieurs.....	21
Figure 14: Évaluation Out-of-Bag (OOB) dans la méthode Bagging montrant la prédiction sur des échantillons non vus.	22
Figure 15– Architecture du système	27
Figure 16– les types de données.	28
Figure 17– Graphe illustrant les deux classes	29
Figure 18– Les valeurs manquantes	30
Figure 19– Fractionnement de données	32
Figure 20– Mesures d'évaluation des performances de l'arbre de décision.....	34
Figure 21– Mesures d'évaluation des performances de forêt aléatoire.....	34
Figure 22– Mesures d'évaluation des performances de Bagging	34
Figure 23– Matrices de confusion pour les différents modèles.....	35
Figure 24: la fenêtre de connexion de l'application.....	41
Figure 25: fenêtre des statistiques des patients.....	42
Figure 26: fenêtre d'information personnelles	42
Figure 27: fenêtre de gestion des patients	43
Figure 28: fenêtre d'ajoute des données médicales.....	44
Figure 29: fenêtre de suivis des patients.	45
Figure 30: Résultat d'évaluation du risque cardiaque.....	46

Liste des tableaux

Tableau 1 –Comparaison des travaux étudiés	24
Tableau 2 :Comparaison préliminaire des performances des modèles	32
Tableau 3 – Résultats de l'échantillon.....	36

Liste des abréviations

IA	Intelligent artificielle
ML	Machine learning
CHD	Coronary Heart Disease
TP	True positive
TN	True négative
FP	False positive
FN	False négative
OMS	l'Organisation mondiale de la santé
OOB	Out-of-Bag
SVM	Support à vecteur machine
KNN	K plus proches voisins
ANN	Artificial Neural Networks
DT	Decision Tree
RF	random forest

Introduction générale

1 Contexte global

Le cœur est un organe vital assurant la circulation du sang et l'apport en oxygène ainsi qu'en nutriments essentiels à l'ensemble du corps humain. Malgré son importance, il est souvent affecté par diverses pathologies cardiovasculaires, qui représentent aujourd'hui la première cause de mortalité dans le monde. Parmi celles-ci, la maladie coronarienne (Coronary Heart Disease - CHD) constitue l'une des affections les plus fréquentes et les plus graves. Elle est causée par l'accumulation progressive de plaques athéromateuses dans les artères coronaires, réduisant ainsi l'apport sanguin au muscle cardiaque.

Cette maladie peut engendrer des complications sévères telles que l'angine de poitrine, l'infarctus du myocarde ou encore l'insuffisance cardiaque. Son dépistage précoce et sa prise en charge restent toutefois des défis majeurs dans le domaine médical.

2 Problématique

Le diagnostic des maladies coronariennes repose souvent sur des examens cliniques complexes et coûteux, nécessitant l'intervention de spécialistes. De plus, les signes cliniques peuvent varier d'un patient à un autre, rendant le diagnostic encore plus difficile. Dans ce contexte, comment peut-on améliorer la détection précoce de la maladie coronarienne à l'aide des technologies modernes ? Et comment tirer profit des données médicales disponibles pour anticiper les risques chez les patients ?

C'est ici qu'intervient l'intelligence artificielle, et plus précisément l'apprentissage automatique, qui permet d'extraire des connaissances à partir de grandes bases de données médicales, souvent hétérogènes et multidimensionnelles.

3 Objectifs de notre travail

Ce travail vise à contribuer à la prévention des maladies coronariennes à travers l'analyse de données médicales en utilisant des techniques d'apprentissage automatique. Les objectifs principaux sont les suivants :

- Étudier les caractéristiques de la maladie coronarienne à partir de données cliniques.
- Comparer les performances des modèles afin d'identifier ceux offrant le meilleur compromis entre précision et interprétabilité.
- Appliquer trois algorithmes d'apprentissage supervisé (Decision Tree, Random Forest et Bagging) pour prédire la présence ou non de la maladie.
- Concevoir une application permettant de faciliter l'utilisation des résultats obtenus par les professionnels de la santé.

4 Organisation du rapport

Ce mémoire est structuré en plusieurs parties, permettant de présenter le contexte, la problématique, la méthodologie adoptée ainsi que les résultats obtenus dans le cadre de notre contribution à la prévention des maladies cardiaques à l'aide des techniques d'intelligence artificielle.

- **introduction générale**

Présente le contexte global, la problématique de recherche, les objectifs du travail, ainsi que la méthodologie générale suivie.

- **Chapitre 1** : Généralités sur les maladies cardiaques

Ce chapitre présente les bases anatomiques du cœur, les types de maladies cardiovasculaires, en particulier la maladie coronarienne, ainsi que les facteurs de risque et les approches de prévention.

- **Chapitre 2** : Apprentissage automatique et méthodes utilisées

Ce chapitre introduit les concepts fondamentaux de l'intelligence artificielle, notamment l'apprentissage automatique, et décrit en détail les algorithmes supervisés utilisés dans ce travail : Decision Tree, Random Forest, et Bagging. Il présente également les critères de performance adoptés pour évaluer les modèles.

- **Chapitre 3** : Implémentation et résultats expérimentaux

Décrit les étapes de préparation des données, l'entraînement des modèles sur la base de données choisie, l'évaluation des performances et l'interprétation des résultats obtenus.

- **Chapitre 4** : Conception et fonctionnalités de l'application

Présente le développement d'une application de bureau interactive permettant d'utiliser les modèles entraînés pour prédire les risques cardiovasculaires à partir de données cliniques.

- **Conclusion générale**

Résume les apports du travail, les limites rencontrées, et propose des perspectives d'amélioration dans le cadre de l'exploitation des outils d'IA en milieu médical



Chapitre 1

Généralités sur les maladies cardiaques



1 Introduction :

Les maladies cardiaques constituent l'une des principales préoccupations sanitaires mondiales, représentant une cause majeure de mortalité. Selon l'Organisation mondiale de la santé (OMS), les maladies cardiovasculaires sont responsables de près de 31 % de l'ensemble des décès dans le monde. Malgré les avancées en matière de traitements médicaux, la prévention demeure l'approche la plus efficace pour réduire le fardeau de ces pathologies.

Dans ce contexte, il devient crucial d'améliorer les stratégies de prévention des maladies cardiaques en utilisant des technologies modernes telles que les techniques d'optimisation et l'intelligence artificielle, qui peuvent contribuer à un diagnostic précoce plus précis et à des solutions personnalisées pour les individus à risque. Cette étude a pour objectif de proposer un cadre visant à améliorer la prévention des maladies cardiaques à travers l'utilisation des techniques d'optimisation basées sur des modèles mathématiques et l'analyse des données.

À travers cette recherche, nous examinerons en détail la structure et le fonctionnement du cœur, ainsi que les maladies cardiaques courantes et les méthodes actuelles de prévention. De plus, l'accent sera mis sur la manière dont les techniques d'optimisation peuvent être appliquées à la prévention des maladies cardiaques pour améliorer les stratégies de détection précoce et de réduction des risques. [1]

2 Anatomie du cœur

2.1 Le cœur :

Le cœur est un organe musculaire en forme de cône, de la taille approximative d'un poing, situé dans la cavité thoracique, légèrement à gauche du sternum. Il est responsable de la circulation sanguine, assurant l'apport en oxygène et en nutriments aux cellules du corps.

2.2 Structure interne du cœur :

Le cœur est divisé en quatre cavités :

1. Oreillette droite : Elle reçoit le sang désoxygéné provenant du corps via les veines caves.
2. Oreillette gauche : Elle reçoit le sang oxygéné des poumons via les veines pulmonaires.
3. Ventricule droit : Il pompe le sang désoxygéné vers les poumons via l'artère pulmonaire.
4. Ventricule gauche : Il pompe le sang oxygéné vers le reste du corps via l'aorte.

2.3 Les valves cardiaques :

Le cœur contient quatre valves qui régulent le flux sanguin et empêchent le reflux :

- Valve tricuspide : Située entre l'oreillette droite et le ventricule droit.
- Valve pulmonaire : Située entre le ventricule droit et l'artère pulmonaire.
- Valve mitrale (ou bicuspid) : Située entre l'oreillette gauche et le ventricule gauche.
- Valve aortique : Située entre le ventricule gauche et l'aorte.



2.4 Le myocarde :

Est le tissu musculaire qui compose la majeure partie du cœur. C'est ce muscle qui se contracte pour pomper le sang. Il est composé de cellules musculaires appelées cardiomyocytes, qui ont la capacité de se contracter de manière autonome, ce qui permet au cœur de battre sans intervention du système nerveux central.

2.5 Le péricarde :

Le cœur est entouré par une membrane appelée le péricarde, qui est une double couche de tissu. Le péricarde contient du liquide pour permettre au cœur de se déplacer en douceur pendant ses contractions.

2.6 Les vaisseaux sanguins associés au cœur :

- L'aorte : L'artère principale qui transporte le sang oxygéné du ventricule gauche vers tout le corps.
- Les veines caves : Elles transportent le sang désoxygéné du corps vers l'oreillette droite.
- Les artères pulmonaires : Elles transportent le sang désoxygéné du ventricule droit vers les poumons.
- Les veines pulmonaires : Elles rapportent le sang oxygéné des poumons à l'oreillette gauche.

2.7 Circulation sanguine :

Le cœur assure deux circulations distinctes :

- La circulation systémique : Le sang est pompé par le ventricule gauche vers l'ensemble du corps pour fournir de l'oxygène et des nutriments.
- La circulation pulmonaire : Le sang est pompé par le ventricule droit vers les poumons pour y recevoir de l'oxygène et éliminer le dioxyde de carbone.

En résumé, le cœur est un organe complexe composé de différentes structures et de mécanismes de régulation pour assurer une circulation sanguine efficace et continue. [2]

3 Fonctionnement du cœur :

1. Le sang pauvre en oxygène arrive dans l'oreillette droite par les veines caves,
2. L'oreillette droite se contracte et le sang est expulsé dans le ventricule droit par la valve tricuspide,
3. Le ventricule droit se contracte et expulse le sang dans les artères pulmonaires par la valve pulmonaire.
4. Ce sang va aller directement vers les poumons pour se charger en oxygène.
5. Le sang oxygéné arrive dans l'oreillette gauche par les veines pulmonaires,
6. La contraction de l'oreillette gauche permet le passage du sang vers le ventricule gauche par la valve mitrale,
7. Le ventricule gauche se contracte à son tour pour expulser le sang dans l'aorte par la valve aortique.
8. Le sang riche en oxygène peut alors être distribué à tous les organes, muscles et tissus de l'organisme.

Ce cycle recommence environ 60 fois par minute au repos. [3]

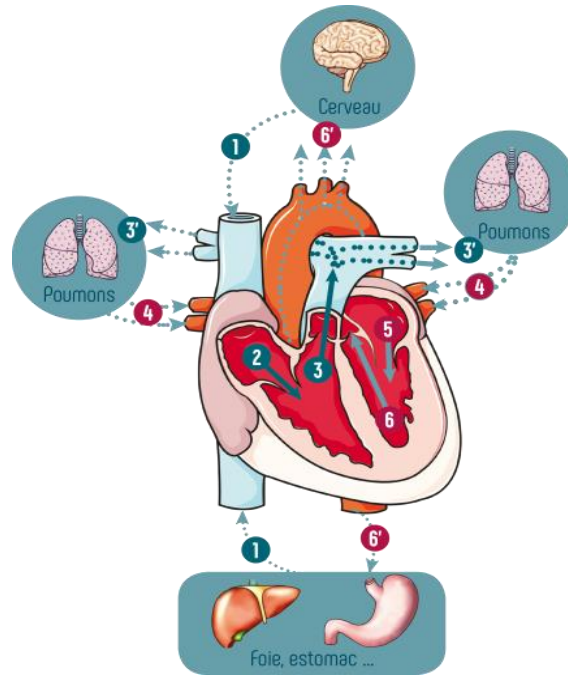


Figure 1: fonctionnement de cœur

4 Les maladies cardiaques

Le terme générique « maladie du cœur » est utilisé pour désigner tout trouble qui touche la structure ou le fonctionnement du cœur. Les maladies du cœur peuvent faire partie de n'importe quel groupe de problèmes de santé dont les causes profondes sont nombreuses. [4]

5 Types de maladies du cœur

Il existe de nombreux types de maladies du cœur. Certains de ces troubles cardiaques peuvent être regroupés en fonction de la manière dont ils touchent la structure ou le fonctionnement du cœur.

5.1 La coronaropathie (Chd) :

La coronaropathie se caractérise par le rétrécissement ou l'obstruction des artères coronaires, qui sont responsables de l'apport en sang, en oxygène et en nutriments au muscle cardiaque. Cette obstruction est généralement causée par une accumulation de plaque (dépôts de matières grasses, de cholestérol et d'autres substances), un phénomène appelé athérosclérose.

C'est la forme la plus fréquente de maladie cardiaque, et elle est la principale cause de crises cardiaques ainsi que de la douleur thoracique connue sous le nom d'angine de poitrine.

Symptômes courants :

- Douleur ou oppression thoracique (angine de poitrine), souvent déclenchée par l'effort ou le stress.
- Essoufflement.
- Fatigue inhabituelle.
- Palpitations.

- Dans les cas graves, douleur intense pouvant irradier vers le bras gauche, le cou ou la mâchoire (symptômes d'infarctus).

Cette maladie demeure la principale cause de mortalité cardiovasculaire dans le monde

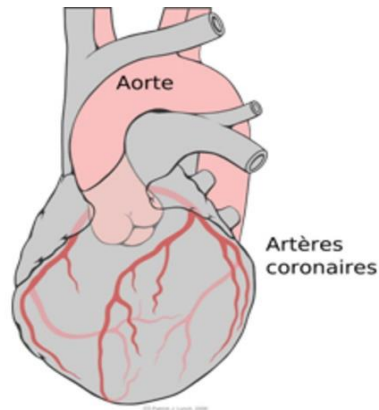


Figure 2: La maladie coronaire

5.2 Les maladies vasculaires :

Touchent les vaisseaux sanguins (artères ou veines) et peuvent se manifester n'importe où dans le corps. Elles peuvent réduire l'apport de sang vers certaines parties de l'organisme, ce qui entraîne de la douleur ou d'autres symptômes selon l'emplacement. Par exemple, les personnes atteintes de maladie artérielle périphérique peuvent ressentir de la douleur dans les bras ou les jambes. L'une des causes fréquentes des maladies vasculaires est le rétrécissement ou l'obstruction des vaisseaux sanguins en raison d'une accumulation de plaque (du cholestérol, des dépôts de matières grasses et d'autres substances), qu'on appelle aussi « athérosclérose ».

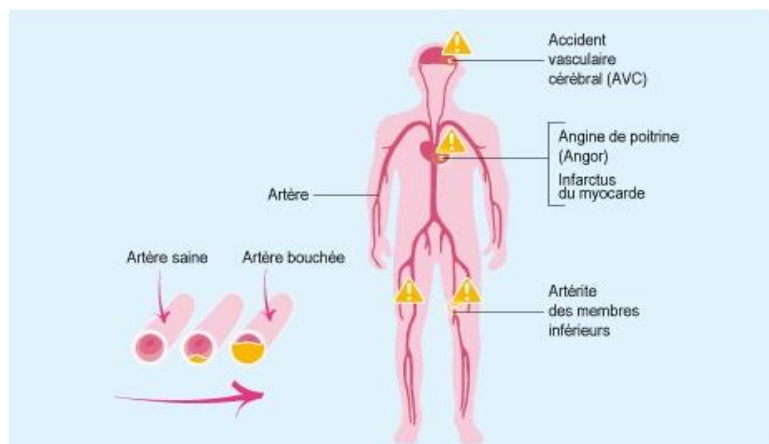


Figure 3 : Les maladies vasculaires



5.3 Les troubles du rythme cardiaque (arythmies)

Les arythmies sont des anomalies du rythme cardiaque. Le cœur peut battre :

- Trop lentement (bradycardie),
- Trop rapidement (tachycardie),
- Ou de façon irrégulière.

Certaines arythmies sont bénignes, tandis que d'autres peuvent provoquer :

- Palpitations,
 - Douleurs thoraciques,
 - Essoufflement,
- et dans les cas graves, représenter un risque vital

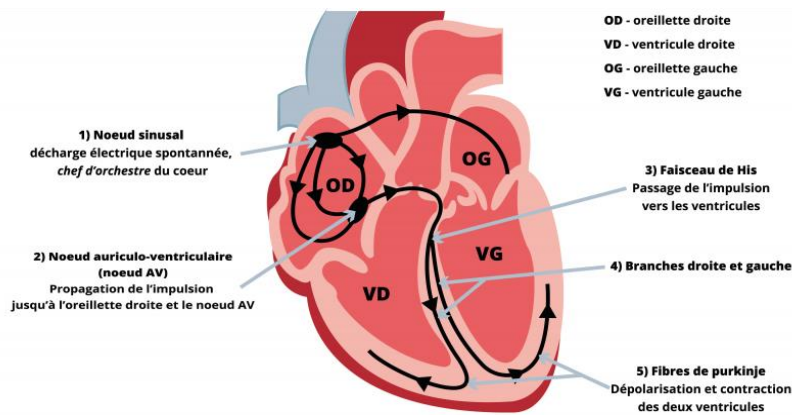


Figure 4: Les troubles du rythme cardiaque

5.4 La cardiopathie congénitale (malformations cardiaques)

Se caractérise par divers problèmes structurels qui affectent le cœur du fœtus pendant son développement. Il en existe de nombreux types, y compris des anomalies des valvules, des grands vaisseaux, des parois et des cavités du cœur. Elles peuvent être causées par des facteurs génétiques ou autres. Les personnes atteintes d'une cardiopathie congénitale et leur famille ont besoin de soutien tout au long de leur vie. Elles peuvent continuer à recevoir des soins médicaux et à subir des interventions chirurgicales une fois adultes.

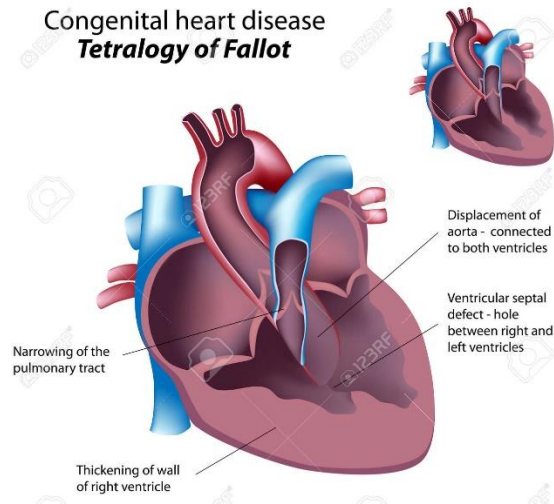


Figure 5: La cardiopathie congénitale (malformations cardiaques)

5.5 L'insuffisance cardiaque :

Est une maladie chronique due au fait que le cœur ne fonctionne pas comme Il le devrait ou présente un problème de structure. Elle peut se produire si le cœur est trop faible ou trop rigide, ou les deux. Elle peut entraîner de la fatigue, une enflure des jambes et de l'abdomen et de l'essoufflement, qui peut être causé par la présence de liquide dans les poumons. Les deux causes les plus courantes de l'insuffisance cardiaque sont la crise cardiaque et l'hypertension artérielle. Il n'existe aucun moyen d'en guérir, mais un diagnostic précoce, des modifications du mode de vie et des médicaments peuvent aider les personnes qui en sont atteintes à mener une vie active, à éviter les hospitalisations et à vivre plus longtemps.

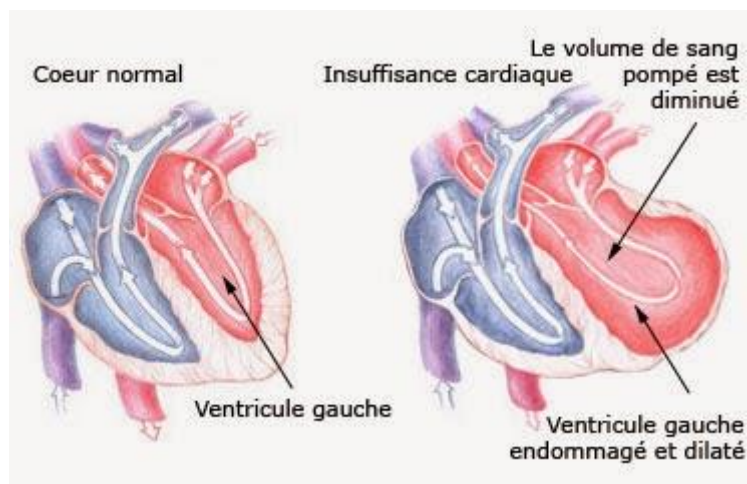


Figure 6: La différence entre les deux cœurs



5.6 Autres maladies du cœur :

Parmi les autres affections cardiaques, on peut citer :

- Les infections cardiaques,
- L'hypertrophie cardiaque (augmentation anormale de la taille du cœur),
- Certains troubles héréditaires. [5]

6 Facteurs de risque des maladies cardiaque :

Un facteur de risque est un élément clinique ou biologique associé à une augmentation du risque de développer une maladie avec une relation de causalité entre le facteur et la maladie .Ces facteurs sont multiples, Ils sont classés en facteurs non modifiables et facteurs modifiables.

• Facteurs de risque non modifiables :

- Âge et sexe :

Le nombre absolu de décès cardiovasculaires est plus important chez les femmes (54 %) que chez les hommes mais avant 65 ans, la mortalité cardiovasculaire des hommes est 3 à 4 fois supérieure à celle des femmes. En pratique, les accidents cardiovasculaires surviennent en moyenne 10 ans plus tôt chez l'homme que chez la femme.

- Héritéité :

Son évaluation repose sur la notion d'événements précoces (< 55 ans chez les hommes, < 65 ans chez les femmes) chez les parents ou dans la fratrie. Ceux-ci peuvent être liés à la transmission génétique de facteurs de risque modifiables (hypercholestérolémie familiale, hypertension artérielle, diabète, etc.). Mais c'est souvent la seule présence de facteurs environnementaux familiaux défavorables (tabagisme, alimentation déséquilibrée et sédentarité, etc.) qui explique les accidents sur plusieurs générations. [6]

• Facteurs de risque modifiables :

— Tabagisme:

Il augmente les lésions athérosclérotiques en modifiant la fonction endothéliale, en perturbant la vasomotion, en activant l'agrégation plaquettaire et en réduisant le cholestérol des lipoprotéines de haute densité. Elle est athérogène et prothrombotique. Le risque relatif de crise cardiaque était de 5, et le risque relatif de maladie artérielle des membres inférieurs était de 2. Ce risque relatif existe également lors du tabagisme passif. Le risque est proportionnel à l'exposition au tabac et s'évalue en paquets-années. Les bénéfices de l'arrêt du tabac sont rapides : l'augmentation du risque relatif disparaît en 3 ans, et le risque de récurrence est réduit de 50% chez les patients coronariens.

— Hypertension artérielle :

Elle est définie par une valeur de pression de > 140 mmHg pour la pression artérielle systolique (PAS) ou de > 90 mmHg pour la pression artérielle diastolique (PAD). Tous les types d'hypertension sont des facteurs de risque : hypertension permanente, hypertension paroxystique, traitée ou non. Le risque relatif était de 7 points pour les accidents vasculaires cérébraux, 3 points pour les maladies coronariennes et 2 points pour les maladies artérielles des membres inférieurs. Avant 55 ans, ce risque était associé aux valeurs de tension artérielle systolique et diastolique.



Après 60 ans, la corrélation est plus forte avec la pression différentielle (PAS- PAD), et donc la pression artérielle systolique surtout chez les personnes âgées. Le traitement de l'hypertension artérielle réduit le risque d'AVC de 40 % et le risque d'infarctus de 15 %.

— Dyslipidémies :

Parmi les dyslipidémies circulantes, le principal facteur de risque de maladie cardiovasculaire est l'élévation du cholestérol LDL, avec un cholestérol lié au LDL 1,60 g/L (4,1 mmol/L). Le cholestérol LDL était positivement associé au risque de maladie cardiovasculaire, tandis que le cholestérol HDL était associé négativement s'il était > 0,40 g/L (1 mmol/L). Des taux élevés de triglycérides (> 2,0 g/L) seuls ne sont pas un facteur de risque (indépendant), mais peuvent être un facteur de risque en combinaison avec d'autres facteurs (voir Syndrome métabolique). Le cholestérol à lipoprotéines de basse densité a un rôle direct dans la croissance des plaques d'athérosclérose et leur rupture due à l'instabilité. Le risque relatif d'hypercholestérolémie pour les maladies coronariennes était 3 fois plus élevé que pour les maladies artérielles et les accidents vasculaires cérébraux. L'efficacité du traitement de l'hypercholestérolémie est un facteur majeur de réduction de la mortalité cardiovasculaire (30% à 20 ans).

— Diabète :

Le diabète a été défini comme deux mesures à jeun > 1,26 g/L (7 mmol/L) ou une seule mesure de glycémie > 2 g/L (11 mmol/L). Le diabète de type I ou de type II est associé à un risque cardiovasculaire accru. Les complications cardiovasculaires du diabète I commencent plus tôt dans la trentaine, mais l'apparition rapide du diabète II en fait un facteur de risque très préoccupant. Il a un risque relatif de >2, causant principalement plus de maladies artérielles que de maladies coronariennes et d'accidents vasculaires cérébraux. Mais le diabète se complique plus souvent de microangiopathie (rétinopathie et maladie rénale). Ce risque relatif augmente avec les anomalies rénales. Le traitement du diabète avec un objectif de 6,5 % d'hémoglobine glyquée (HbA1c) réduit les complications cardiovasculaires.

-Insuffisance rénale :

L'insuffisance rénale chronique est associée à un taux élevé de complications cardiovasculaires, comparable à la sévérité du diabète sur le système cardiovasculaire. [7]

7 Diagnostique des maladies cardiaques :

Le diagnostic des maladies cardiaques repose sur une approche multidimensionnelle qui inclut des interrogatoires détaillés, ainsi que des procédures diagnostiques non invasives et certaines, peu invasives :

- **interrogatoire et examen clinique :**

La première étape vers un diagnostic précis de maladie cardiaque commence par un interrogatoire approfondi du patient. Le cardiologue recueille des informations essentielles sur les symptômes ressentis, l'historique médical, les antécédents familiaux de maladies cardiaques, et les habitudes de vie. Cet interrogatoire est suivi d'un examen clinique minutieux qui comprend la vérification des paramètres vitaux, la palpation du pouls, l'auscultation du cœur et des poumons, ainsi que l'examen des veines, du thorax, des poumons, de l'abdomen et des extrémités.



- **l'électrocardiogramme :**

Il s'agit d'un examen indispensable permettant de poser de nombreux diagnostics de pathologie cardiaque. Cet examen est indolore et réalisé au cabinet du cardiologue, lors de la consultation. Le spécialiste applique des électrodes sur le thorax et les membres du patient. Cet examen mesure la fréquence cardiaque et permet de déceler de potentiels troubles du rythme cardiaque, ainsi que d'autres anomalies ;

- **les analyses de sang :**

La glycémie et le taux de cholestérol sont des paramètres importants à explorer. D'autres marqueurs peuvent compléter l'examen biologique, comme le taux de protéine de l'inflammation, ou encore le dosage de la Troponine. S'il le juge nécessaire, le cardiologue pourra compléter le bilan par des examens complémentaires comme :

- la scintigraphie
- l'échocardiogramme
- l'angiographie
- la coronarographie
- l'holter
- l'échographie doppler
- le test d'effort [8]

8 Prévention des maladies cardiaques:

Les maladies cardiovasculaires constituent une cause majeure de morbidité et de mortalité. La prévention, qu'elle soit individuelle ou collective, vise à réduire les facteurs de risque et à améliorer la santé cardiovasculaire à travers différentes stratégies.

8.1 Prévention individuelle

Elle repose sur des actions personnelles visant à limiter le risque de maladies cardiaques par des changements de mode de vie et une surveillance médicale régulière.

- Adopter une alimentation saine : Privilégier les aliments riches en fibres, en oméga-3 et limiter les sucres et graisses saturées.
- Pratiquer une activité physique régulière : Au moins 30 minutes par jour d'exercice modéré.
- Arrêter le tabac : Réduction significative du risque cardiovasculaire.
- Limiter l'alcool : Une consommation excessive peut entraîner de l'hypertension et des troubles du rythme cardiaque.
- Surveiller les facteurs de risque : Suivi de la tension artérielle, du cholestérol et du diabète.
- Gérer le stress et le sommeil : Techniques de relaxation, sommeil réparateur.
- Observance des traitements : Respecter les prescriptions médicales pour les personnes à risque.

8.2 Prévention collective:

Elle implique des politiques de santé publique et des actions collectives visant à réduire l'incidence des maladies cardiovasculaires à l'échelle de la population.



- Campagnes de sensibilisation : Information sur les risques et les comportements protecteurs (OMS, ministères de la santé).
- Accès facilité aux soins : Dépistage précoce, suivi médical et remboursement des traitements.
- Législation et régulation : Taxes sur le tabac et l'alcool, régulation des publicités alimentaires, normes nutritionnelles.
- Aménagements urbains : Espaces favorisant l'activité physique, limitation de la pollution.
- Programmes d'éducation thérapeutique : Rôle des professionnels de santé, dont les pharmaciens, dans l'accompagnement des patients. [9]

Conclusion

Dans ce chapitre, nous avons compris que les maladies cardiaques sont très fréquentes et peuvent être très dangereuses. Elles avancent souvent sans symptômes, ce qui les rend encore plus graves. C'est pour cela que la prévention est très importante : elle peut aider à éviter beaucoup de problèmes si elle est bien suivie.

Ces éléments montrent pourquoi il est utile d'utiliser des solutions intelligentes, comme l'intelligence artificielle, pour mieux détecter les risques et protéger la santé des personnes.



Chapitre 2

Apprentissage automatique



1. Introduction :

L'apprentissage automatique (ou machine Learning) est une discipline de l'intelligence artificielle qui permet aux systèmes d'apprendre à partir de données, d'identifier des patterns complexes et de prendre des décisions prédictives avec une intervention humaine minimale. Dans le domaine de la santé, et plus particulièrement dans la prévention des maladies cardiaques, l'apprentissage automatique offre des perspectives innovantes pour améliorer le diagnostic précoce et la gestion des risques. Les maladies cardiaques, telles que les cardiopathies ischémiques et l'insuffisance cardiaque, restent l'une des principales causes de mortalité à l'échelle mondiale, et leur prévention nécessite une analyse approfondie de multiples facteurs de risque (âge, génétique, mode de vie, etc.). Grâce à sa capacité à traiter de vastes ensembles de données et à détecter des corrélations non évidentes, l'apprentissage automatique se positionne comme un outil essentiel pour anticiper les risques cardiovasculaires et personnaliser les stratégies de prévention.

Ce chapitre explore les fondements de l'apprentissage automatique, ses applications dans le domaine cardiovasculaire, et les défis liés à son utilisation pour la prévention des maladies cardiaques. Nous examinerons également comment ces techniques peuvent être optimisées pour améliorer leur efficacité et leur précision.

2. Apprentissage automatique :

L'apprentissage automatique, aussi appelé apprentissage machine ou encore auto-apprentissage, désigne une branche de l'intelligence artificielle (IA) qui utilise des algorithmes afin d'imiter la capacité de raisonnement de l'être humain. Ce champ d'études est aujourd'hui appliqué dans de nombreux domaines et permet d'assister les humains dans la prise de décision et la résolution de problèmes complexes.

L'apprentissage automatique fonctionne à l'aide d'algorithmes capables de traiter une quantité considérable de données, impossibles à analyser à l'échelle humaine. Avec le temps, les capacités de stockage et de traitement des données ont connu une croissance exponentielle, offrant une puissance de calcul toujours plus importante et des prédictions toujours plus précises. [10]



Figure 7: relation de IA et machine Learning



3. Les différentes méthodes d'apprentissage automatique :

Les algorithmes de machine Learning possèdent différents niveaux de sophistication selon le mode d'apprentissage utilisé. Celui-ci dépend notamment de la complexité des tâches à effectuer, ainsi que de la quantité de données disponibles pour entraîner l'ordinateur. Il existe 4 types d'apprentissage automatique : l'apprentissage supervisé, non supervisé, semi-supervisé, et l'apprentissage par renforcement. [10]

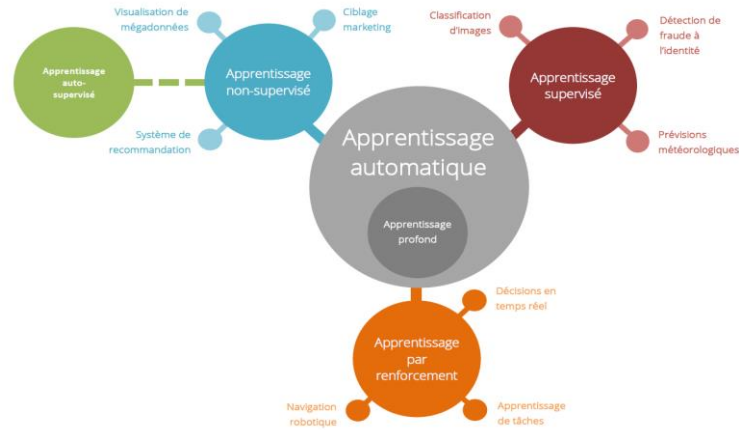


Figure 8: Les différentes méthodes d'apprentissage automatique

3.1 Apprentissage supervisé :

Apprentissage supervisé est une méthode d'apprentissage automatique, caractérisée par la création d'un algorithme qui apprend une fonction prédictive. Ceci est possible grâce à un entraînement à partir d'exemples annotés, qui incluent un groupe de variables d'entrée, accompagnées de leurs variables de sortie respectives. Ce processus d'entraînement est répété jusqu'à l'obtention d'une performance satisfaisante. Lors de chaque itération, la machine crée un certain nombre de règles, reliant les variables d'entrée aux variables de sortie. Ce processus permet au modèle d'apprendre à partir des données et d'appliquer les règles afin de prédire, de façon précise, la valeur de sortie lorsqu'une valeur d'entrée est donnée. [11]

En apprentissage supervisé, on distingue entre deux types de tâches :

-La classification : Quand la variable cible (à prédire) est discrète. Ce qui revient à attribuer une classe (ou étiquette) à chaque entrée. C'est le cas si on cherche à prédire la tendance d'un mouvement futur d'un actif (haut, neutre, bas).

-La régression : Quand la variable cible à prédire est continue.

Voici quelques algorithmes d'apprentissage supervisé les plus importants :

- K plus proches voisins.
- Régression linéaire.
- Régression logistique.
- Machines à vecteurs de support (SVM).

3.2 Apprentissage non supervisé :

Pour ce type d'apprentissage la base de données d'apprentissage ne contient pas de variable cible (comme on l'a vu en apprentissage supervisé). Il y a seulement un ensemble de données collectées en entrée.

L'algorithme doit découvrir par lui-même la structure en fonction des données. On utilise cette technique



pour partitionner les données en groupes d'éléments homogènes. La distance est souvent la plus utilisée comme mesure de similarité entre les groupes. [12]

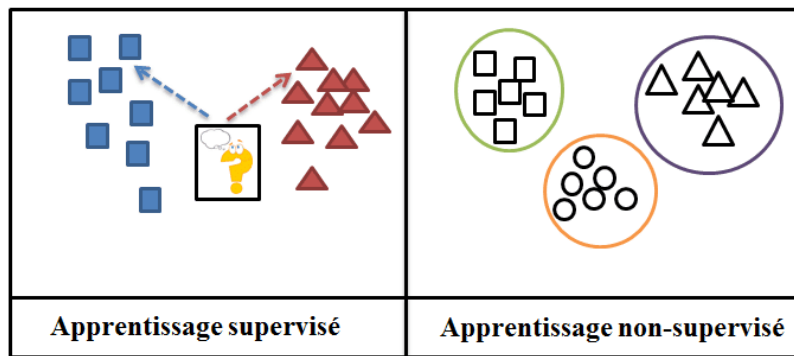


Figure 9: Apprentissage supervisé vs Apprentissage non supervisé

3.3 Apprentissage par renforcement :

L'apprentissage par renforcement est utile lorsque l'agent d'apprentissage automatique, tel qu'un robot, tente d'accomplir une tâche dans un environnement inexploré ou difficile à prévoir. L'apprentissage par renforcement fait appel au modèle mathématique appelé processus de décision de Markov, qui permet de comprendre comment les décisions sont prises lorsqu'un élément aléatoire affecte le résultat.

Un système d'apprentissage par renforcement doit comporter quatre éléments :

- L'agent ou l'algorithme d'apprentissage automatique
- L'environnement dans lequel l'agent travaille
- La politique ou les instructions de base que l'agent suit
- Un signal de récompense que l'agent observe en fonction de ses actions. [13]

4 Les algorithmes d'apprentissage supervisé utilisés :

4.1 Arbre de Décision :

4.1.1 Définition :

Un arbre de décision est une représentation graphique des différentes options permettant de résoudre un problème, montrant comment les différents facteurs sont liés. Il a une structure hiérarchique qui commence par une question principale en haut appelée nœud, qui se ramifie ensuite en différents résultats possibles :

-Nœud Racine : le point de départ qui représente l'ensemble du jeu de données.

Branches : les lignes qui relient les nœuds. Elles montrent le flux d'une décision à une autre.

-Nœuds Internes : les points où des décisions sont prises en fonction des caractéristiques d'entrée.

-Feuilles (Nœuds Terminaux) : les nœuds à la fin des branches représentant les résultats ou prédictions finaux.

Les arbres de décision facilitent la prise de décision en visualisant les différents résultats possibles.

Vous pouvez rapidement évaluer et comparer les « branches » pour déterminer la meilleure voie à suivre.

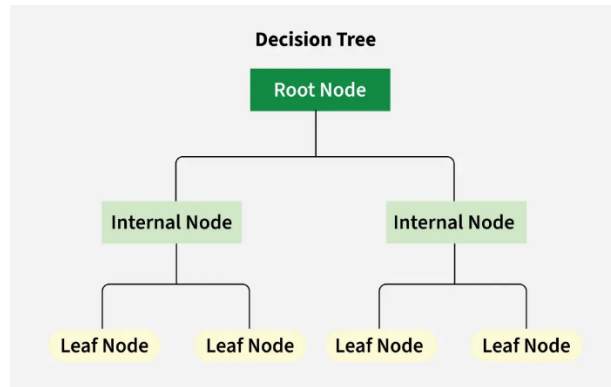


Figure 10: Structure d'un arbre de décision.

Exemple :

Pour décider si vous devez boire du café selon l'heure de la journée et votre niveau de fatigue. Si c'est le matin, l'arbre vous demande si vous êtes fatigué.

Si oui : boire du café.

Sinon : pas besoin.

Si c'est l'après-midi, la question est répétée avec les mêmes conclusions

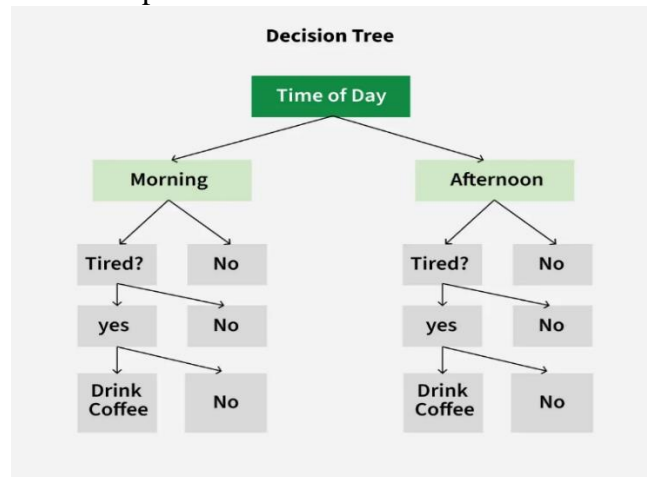


Figure 11: Arbre de décision.

4.1.2 Classification des Arbres de Décision :

On distingue principalement deux types d'arbres de décision selon la nature de la variable cible :

-Arbres de Classification : conçus pour prédire des résultats catégoriels.

Par exemple, déterminer si un e-mail est "spam" ou "non spam".

-Arbres de Régression : utilisés lorsque la variable cible est continue. Ils prédisent des valeurs numériques, comme estimer le prix d'une maison en fonction de sa taille, de son emplacement, etc.

4.1.3 Fonctionnement des Arbres de Décision :

Le fonctionnement d'un arbre de décision commence par une question principale (le nœud racine) issue des caractéristiques du jeu de données.

À partir de ce nœud, l'arbre pose une série de questions oui/non, chacune servant à diviser les données en sous-groupes selon certaines caractéristiques.



Exemple : “Pleut-il ?”

Si la réponse est oui, on suit une branche.

Sinon, une autre.

Ce processus se poursuit jusqu’à atteindre un nœud terminal où il n’y a plus de question à poser.

Le nœud final contient la prédiction ou la classification.

4.1.4 Avantage des Arbres de Décision

- Simplicité et Interopérabilité : faciles à comprendre, on peut les visualiser comme un organigramme.
- Polyvalence : adaptés aux tâches de classification et de régression.
- Pas besoin de normalisation des données ou de mise à l’échelle.
- Gestion des relations non linéaires entre les variables.

4.1.5 Inconvénients des Arbres de Décision

- Sur apprentissage (Overfitting) : l’arbre peut s’adapter trop aux données d’entraînement, perdant sa capacité à généraliser.
- Instabilité : de légers changements dans les données peuvent produire un arbre très différent.
- Biais vers les variables avec plusieurs niveaux : l’arbre peut favoriser ces variables et négliger d’autres facteurs importants.

4.1.6 Applications des Arbres de Décision

-Approbation de prêts bancaires : en utilisant des données comme le revenu, le score de crédit, l’emploi et l’historique des prêts, l’arbre prédit si un prêt sera approuvé ou non.

-Diagnostic médical : pour détecter si un patient est diabétique à partir de caractéristiques comme la glycémie, l’IMC, la tension artérielle, etc.

-Prévision des résultats scolaires : l’arbre peut prédire si un élève va réussir ou échouer en fonction de son assiduité, son temps d’étude et ses notes passées. [14]

4.2 Random Forest Classificateur :

4.2.1 Définition :

L’algorithme Random Forest est une extension de la méthode de Bagging, car il utilise à la fois le Bagging et l’aléa dans la sélection des variables pour créer une « forêt » non corrélée d’arbres de décision. L’aléa dans les variables, aussi appelé feature Bagging ou « méthode des sous-espaces aléatoires », génère un sous-ensemble aléatoire de caractéristiques, ce qui assure une faible corrélation entre les arbres. C’est la différence clé entre les arbres de décision simples et les forêts aléatoires. Tandis que les arbres de décision considèrent toutes les divisions possibles des caractéristiques, les forêts aléatoires n’en sélectionnent qu’un sous-ensemble.

Si l’on revient à l’exemple de « dois-je faire du surf ? », les questions que je pourrais poser pour faire une prédiction ne seraient pas aussi complètes que celles d’une autre personne. En tenant compte de toute la variabilité potentielle des données, on réduit les risques de sur apprentissage (overfitting), de biais et de variance globale, ce qui permet d’obtenir des prédictions plus précises.

4.2.2 Fonctionnement de Random Forest :

Les algorithmes Random Forest ont trois principaux hyper paramètres à définir avant l’entraînement

- la taille minimale des nœuds,
- le nombre d'arbres,
- et le nombre de caractéristiques échantillonnées.

L'algorithme est constitué d'un ensemble d'arbres de décision. Chaque arbre est entraîné sur un échantillon de données tiré avec remise du jeu de données initial, appelé échantillon bootstrap. Un tiers de cet échantillon est mis de côté comme données de test, connues sous le nom d'échantillon out-of-bag (oob).

Un second niveau d'aléa est introduit par le feature Bagging, qui augmente la diversité entre les arbres et réduit leur corrélation.

- Pour un problème de régression, la prédiction finale est obtenue en moyennant les résultats des arbres.
- Pour un problème de classification, la prédiction est obtenue par vote majoritaire (la classe la plus fréquente).

Enfin, l'échantillon oob est utilisé pour faire une validation croisée et affiner la prédiction.

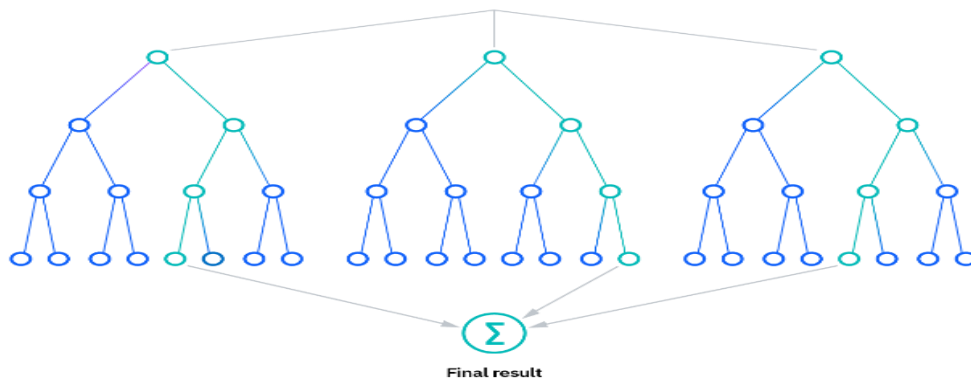


Figure 12: Illustration d'un ensemble d'arbres dans une forêt aléatoire avec agrégation des prédictions.

4.2.3 Avantages de Random Forest :

- Réduction du sur apprentissage : Contrairement aux arbres de décision qui peuvent s'adapter trop étroitement aux données d'entraînement, les forêts aléatoires réduisent ce risque grâce à la moyenne de plusieurs arbres non corrélés, ce qui réduit la variance et l'erreur de prédiction.
- Grande flexibilité : Peut gérer aussi bien des tâches de régression que de classification avec une grande précision. Le feature Bagging rend également la forêt aléatoire efficace pour estimer les valeurs manquantes.
- Évaluation facile de l'importance des variables : L'algorithme permet d'évaluer facilement l'importance de chaque variable.

4.2.4 Inconvénients de Random Forest :

- Temps de traitement long : Bien que très précis, l'algorithme peut être lent car il traite chaque arbre séparément.
- Consommation de ressources : Le traitement de grands ensembles de données nécessite plus de mémoire et de puissance de calcul.

- Moins interprétable : Contrairement à un seul arbre de décision facile à expliquer, une forêt d'arbres est plus difficile à interpréter.

4.2.5 Applications de Random Forest :

Finance : Préférée à d'autres algorithmes car elle réduit le temps consacré à la gestion et au prétraitement des données. Elle peut être utilisée pour :

- évaluer le risque de crédit.
- détecter la fraude.
- résoudre des problèmes de tarification d'options financières.

Santé : Utilisée en biologie computationnelle, elle aide les médecins à :

- classer les expressions géniques,
- identifier des bios marqueurs,
- annoter des séquences.
- Cela permet de mieux prédire la réponse d'un patient à un médicament.

E-commerce : Sert dans les moteurs de recommandation pour proposer des produits à vendre en complément. [15]

4.3 Bagging Classificateur :

4.3.1 Définition :

Le Bagging (ou Bootstrap Aggregating) est une méthode d'apprentissage ensembliste dans laquelle plusieurs modèles de base sont entraînés indépendamment et en parallèle sur différents sous-ensembles des données d'entraînement. Chaque sous-ensemble est généré par un échantillonnage bootstrap, c'est-à-dire en sélectionnant aléatoirement des points de données avec remise. Dans un classificateur Bagging, la prédiction finale est obtenue par vote majoritaire entre les prédictions de tous les modèles de base.

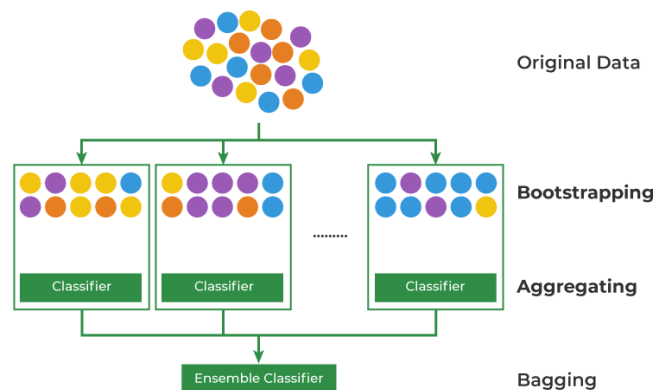


Figure 13: Schéma du Bagging (Bootstrap Aggregating) pour les Classifieurs.

4.3.2 Fonctionnement du Bagging Classifieur Bagging Classificateur :

1. Échantillonnage Bootstrap :

Les données sont échantillonnées de manière aléatoire avec remise pour générer plusieurs sous-ensembles de taille identique. Cela signifie que certains échantillons peuvent apparaître plusieurs fois dans un sous-ensemble, alors que d'autres peuvent ne pas

Apparaître du tout. Cela permet d'obtenir des modèles diversifiés, réduisant ainsi le sur-apprentissage et améliorant la précision.

Exemple :

- Ensemble d'entraînement original : [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
- Échantillon bootstrap 1 : [2, 3, 3, 5, 6, 1, 8, 10, 9, 1]
- Échantillon bootstrap 2 : [1, 1, 5, 6, 3, 8, 9, 10, 2, 7]
- Échantillon bootstrap 3 : [1, 5, 8, 9, 2, 10, 9, 7, 5, 4]

2. Entraînement des modèles de base :

Chaque sous-ensemble bootstrap est utilisé pour entraîner un modèle de base indépendant (comme des arbres de décision, des machines à vecteurs de support ou des Réseaux de neurones). Ces modèles, appelés apprenants faibles, ne sont pas très performants individuellement. Entraînement étant parallèle, il est rapide et efficace.

3. Agrégation :

Une fois les modèles entraînés, ils prédisent l'étiquette d'une instance sur de nouvelles Données non vues. Le Bagging classifieur détermine alors la prédiction finale par vote Majoritaire : la classe la plus prédite est choisie.

4. Evaluation Out-of-Bag (OOB):

Certains échantillons ne sont pas utilisés dans l'entraînement de certains modèles (ceux exclus lors du bootstrap). Ces échantillons hors sac servent à évaluer les performances du Modèle sans avoir besoin de validation croisée. [16]

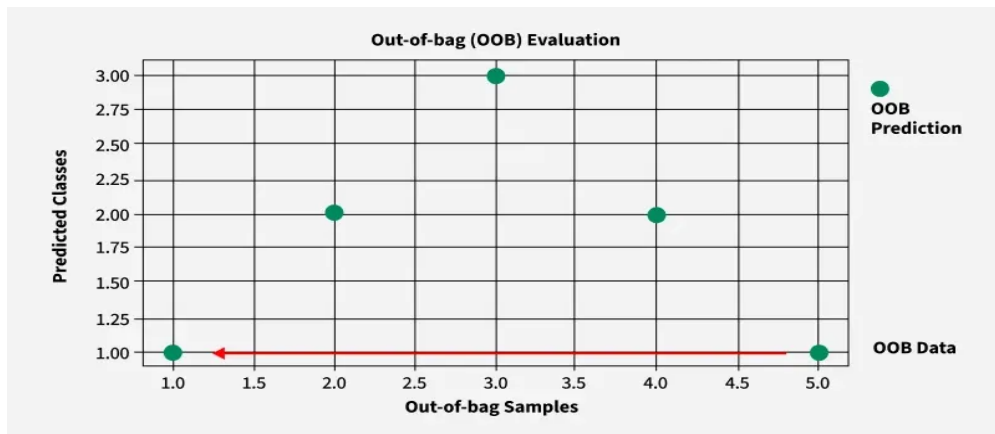


Figure 14: Évaluation Out-of-Bag (OOB) dans la méthode Bagging montrant la prédiction sur des échantillons non vus.

4.4 AdaBoost (Adaptive Boosting)

AdaBoost est un algorithme d'ensemble qui combine plusieurs "apprenants faibles" (comme des petits arbres de décision) pour former un "apprenant fort". À chaque itération, il accorde plus d'importance aux exemples que les modèles précédents ont mal classés, afin d'améliorer les performances globales. [17]



4.5 Gradient Boosting

Gradient Boosting est une technique d'ensemble qui construit des modèles de manière séquentielle. Chaque nouveau modèle corrige les erreurs du précédent en utilisant la descente de gradient pour minimiser une fonction de perte. Il est particulièrement puissant pour les problèmes de régression et de classification. [18]

4.6 SVM (Support Vector Machine)

Le SVM est un algorithme de classification qui cherche à trouver l'hyperplan optimal qui sépare les classes de données. Il est efficace dans les espaces de grande dimension et utilise des "vecteurs de support" pour maximiser la marge entre les classes. [19]

4.7 KNN (K-Nearest Neighbors)

KNN est un algorithme simple qui classe un nouvel exemple en fonction de la majorité des "k" exemples les plus proches dans l'espace des caractéristiques. Il ne nécessite pas d'apprentissage préalable, car il stocke simplement les données d'entraînement. [20]

4.8 ANN (Artificial Neural Networks)

Les réseaux de neurones artificiels sont inspirés du fonctionnement du cerveau humain. Ils sont composés de couches de "neurones" interconnectés qui traitent les données d'entrée pour apprendre des relations complexes. Les ANN sont très utilisées pour les tâches comme la reconnaissance d'images ou la prédiction. [21]

5 Applications de l'apprentissage automatique dans le secteur de la santé

L'objectif de l'apprentissage automatique en informatique est de rendre la machine plus efficace et plus fiable. Dans le domaine des soins de santé, la machine est une extension et un multiplicateur de force pour le cerveau du médecin. Après tout, un patient aura toujours besoin du contact et des soins d'un être humain, ce qu'une machine ne peut fournir. Le travail d'une machine ne consiste donc pas à remplacer le médecin, mais plutôt à l'aider à fournir un meilleur service et de meilleurs soins.

5.1 Le diagnostic des maladies cardiaques

Le cœur est l'un des organes principaux de notre corps. Il existe une variété de maladies cardiaques dont nous souffrons, telles que la cardiopathie coronarienne, la maladie des artères coronaires, etc. Des chercheurs sont en train de développer des algorithmes d'apprentissage automatique pour faciliter le diagnostic des maladies cardiaques. Il s'agit d'un sujet très étudié à l'échelle mondiale, et la mise en place d'un système automatisé pour le diagnostic des maladies cardiaques serait l'un des plus grands exploits de l'humanité au 21ème siècle.

5.2 La prédiction du diabète

Le diabète est non seulement une maladie dangereuse, mais aussi l'une des maladies les plus répandues dans le monde. Il s'agit également d'une maladie d'entrée, qui est elle-même l'une des principales causes d'autres maladies et qui conduit inexorablement ses victimes vers la mort.

De nombreux algorithmes peuvent être utilisés pour prédire le diabète, notamment Naïve Bayes, Decision Trees, Random Forests et KNNs. L'algorithme de Naïve Bayes est plus performant que les autres en termes de précision, en raison de ses bonnes performances et du peu de temps qu'il faut pour effectuer les calculs.



5.3 La découverte de nouveaux médicaments

L'apprentissage automatique présente plusieurs avantages lorsqu'il est appliqué à la science des soins de santé. Il rendra le processus de découverte de nouveaux médicaments plus rapide et moins sujet aux erreurs en réduisant considérablement le taux d'échec. Il permettra également de réduire le coût de la découverte de médicaments en optimisant le processus de fabrication des médicaments.

L'apprentissage automatique a révolutionné le domaine médical grâce à sa capacité à analyser de grandes quantités de données complexes. Son champ d'application est très vaste dans le secteur de la santé. Outre les maladies cardiovasculaires et le diabète, il est également utilisé dans la détection et la classification des maladies hépatiques, du cancer, ainsi que dans de nombreuses autres pathologies. Cette diversité d'usages témoigne de l'importance croissante de l'intelligence artificielle pour améliorer les diagnostics, personnaliser les traitements et optimiser les soins de santé. [22]

6 Des travaux étudiés

Références	Date	Algorithmes	Accuracy
P. Sujatha and K. Maha lakshmi [39]	2020	RF LR DT KNN	83,51% 80,21% 79,90% 72,52%
Jessical.Gupta, Michael [40]	2020	NaiveBayes KNN SVM RF DT	84% 81,31% 81,42% 77,14% 82,28%
BEHLOULI L.A., HAMDI H [41]	2022	DT	83.9%

Tableau 1–Comparaison des travaux étudiés

Conclusion :

Ce chapitre a permis de poser les bases théoriques et pratiques de l'apprentissage automatique, en mettant en lumière ses différents types — supervisé, non supervisé, semi-supervisé et par renforcement — ainsi que les algorithmes majeurs qui en découlent. Grâce à sa capacité à analyser de grandes quantités de données, à détecter des motifs complexes et à fournir des prédictions fiables, l'apprentissage automatique s'affirme comme un pilier central de l'intelligence artificielle moderne. En particulier, son application dans des domaines sensibles comme la santé illustre son potentiel à transformer la prise de décision, en offrant des outils prédictifs puissants, capables d'anticiper des événements critiques comme les maladies cardiovasculaires. Toutefois, ces techniques ne sont pas exemptes de défis : sur apprentissage, interopérabilité des modèles, ou encore la nécessité d'un prétraitement adapté des données.

Ainsi, l'apprentissage automatique constitue une discipline en constante évolution, dont la maîtrise repose sur une compréhension fine des algorithmes, de leurs avantages, de leurs limites, et des contextes dans lesquels ils peuvent être appliqués efficacement.



Chapitre 3
Implémentation et résultats
expérimentaux

1 Introduction :

Dans Ce chapitre présent l'implémentation pratique du système de prédiction des maladies cardiaques nous décrivons l'environnement de travail, les bibliothèques utilisées, les étapes de prétraitement des données, la construction et l'entraînement des modèles, ainsi que l'évaluation des performances obtenues.

2 L'architecture du système :

Afin de construire un modèle d'apprentissage automatique efficace pour la classification, nous proposons une démarche structurée illustrée dans le schéma ci-dessous. Cette approche a été appliquée à trois algorithmes : l'arbre de décision (Decision Tree), la forêt aléatoire (Random Forest) et le Bagging, afin d'améliorer la précision des prédictions.

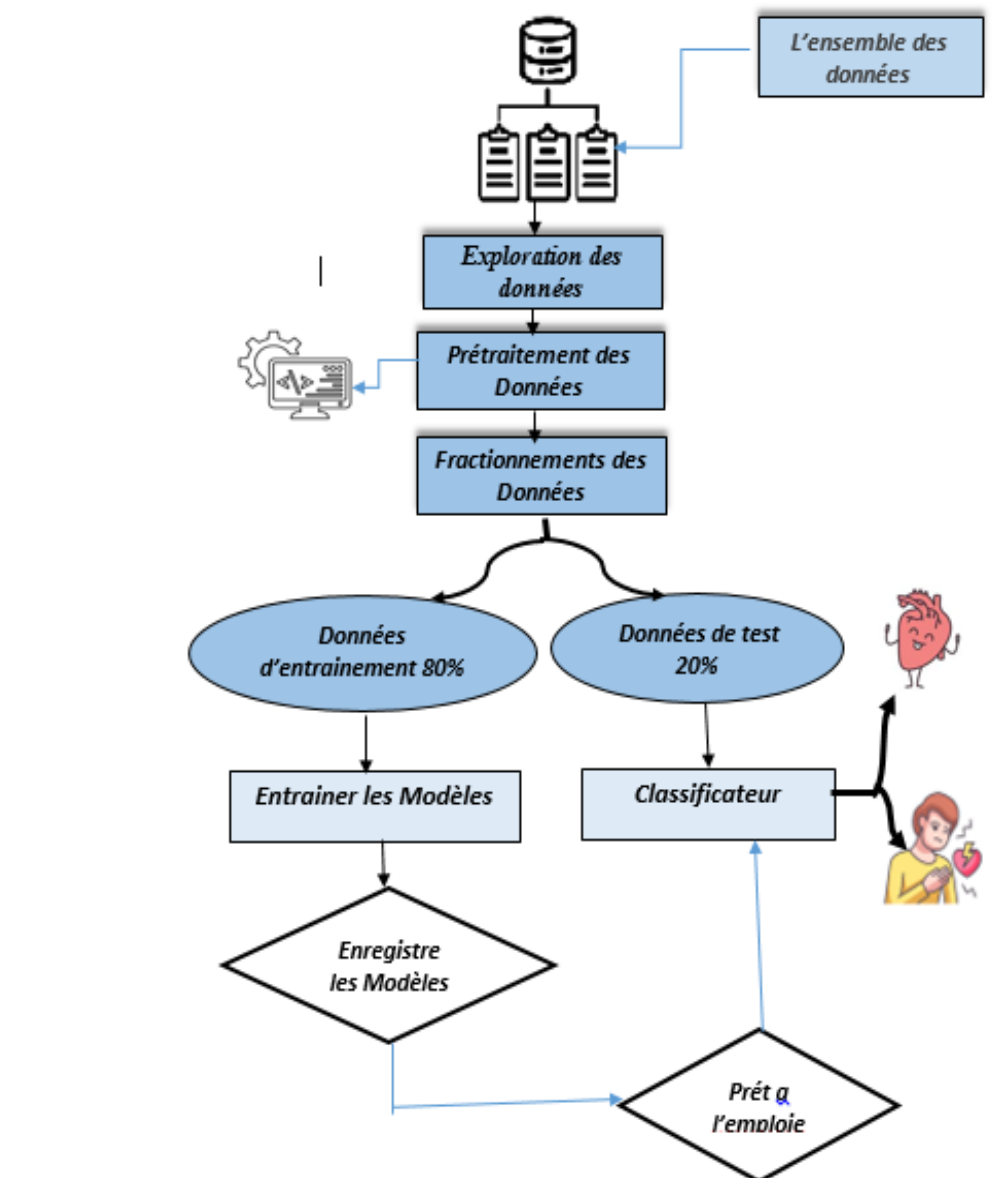


Figure 15– Architecture du système



2.1 L'ensemble de données :

Nous avons utilisé une base de données publique connue sous le nom de 'Heart Disease Dataset'. Cette base contient 1025 enregistrements médicaux, chacun représentant un patient, avec 14 attributs médicaux (tels que l'âge, le sexe, la tension artérielle, le cholestérol, etc.).

L'attribut cible est nommé target, et il indique :

- 0 : absence de maladie coronarienne (CHD)
- 1 : présence de maladie coronarienne (CHD)

Les principales caractéristiques incluses sont :

- Âge
- Sexe
- Pression artérielle au repos
- Cholestérol sérique
- Glycémie à jeun
- Résultats de l'électrocardiogramme
- Fréquence cardiaque maximale atteinte
- Angine de poitrine induite par l'effort

2.2 Analyse et Exploration des données :

Nous avons effectué une analyse exploratoire des données à l'aide de plusieurs commandes Python, notamment :

df.info () : pour afficher les types de données et vérifier les valeurs manquantes

```
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1025 entries, 0 to 1024
Data columns (total 14 columns):
#   Column                                     Non-Null Count  Dtype
---  ---
0   age                                         1025 non-null   int64
1   sex                                         1025 non-null   object
2   chest_pain_type                             1025 non-null   object
3   resting_blood_pressure                       1025 non-null   int64
4   cholestoral                                 1025 non-null   int64
5   fasting_blood_sugar                         1025 non-null   object
6   rest_ecg                                    1025 non-null   object
7   Max_heart_rate                             1025 non-null   int64
8   exercise_induced_angina                    1025 non-null   object
9   oldpeak                                     1025 non-null   float64
10  slope                                       1025 non-null   object
11  vessels_colored_by_flourosopy              1025 non-null   object
12  thalassemia                                1025 non-null   object
13  target                                      1025 non-null   int64
dtypes: float64(1), int64(5), object(8)
memory usage: 112.2+ KB
```

Figure 16– les types de données.

df.shape : pour vérifier les dimensions du dataset (1025, 14)

```
df.shape
(1025, 14)
```

- **Visualisation de la classe cible :**

La visualisation de la classe cible est importante pour voir l'équilibrage des données qui est essentiel pour avoir un résultat précis. La Figure 17- représente les graphiques de visualisation de la classes cible. Elle nous montre que le nombre de personnes sans maladie cardiaque est de 499 avec une proportion de 48.7% et le nombre de personnes atteintes de maladie cardiaque est de 526 avec une proportion de 51.3%. Nous notons que le pourcentage de personnes atteintes de maladies cardiaques et de personnes sans maladie est presque proche c'est-à-dire un dataset presque équilibré.

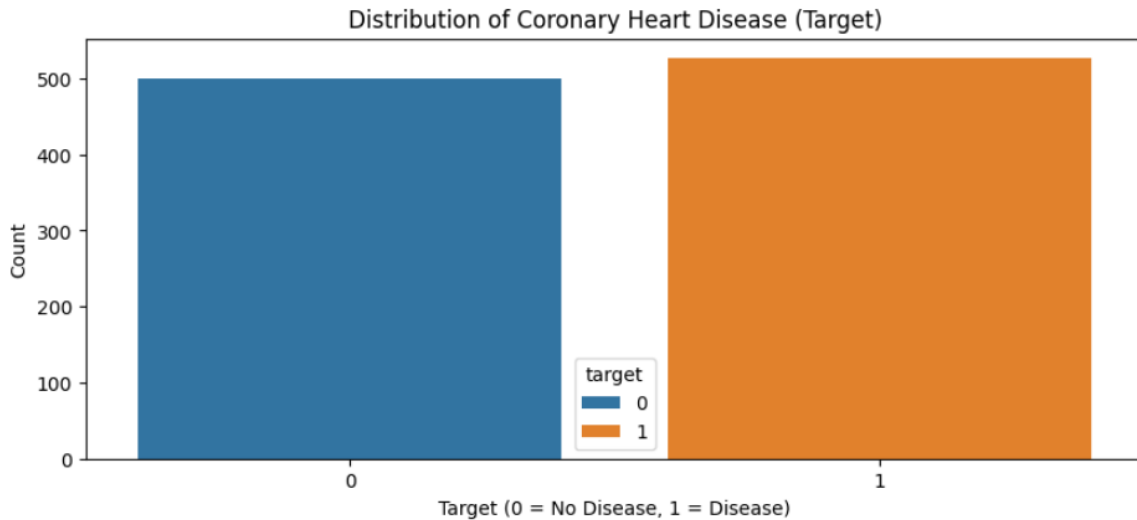


Figure 17– Graphe illustrant les deux classes

2.3 Le prétraitement de données :

Le prétraitement des données constitue une étape primordiale pour garantir la qualité du jeu de données et maximiser la performance des modèles d'apprentissage automatique. Plusieurs opérations ont été réalisées afin de rendre les données exploitables et fiables pour la modélisation.

2.3.1 Vérification des valeurs manquantes :

La première étape a consisté à vérifier l'existence éventuelle de valeurs manquantes dans l'ensemble de données. Cette vérification a été effectuée à l'aide de la fonction `isna()` de la bibliothèque Pandas.

Les résultats de cette analyse ont révélé l'absence totale de valeurs manquantes dans le jeu de données. Cela signifie que l'intégralité des 1025 enregistrements a pu être conservée, sans nécessiter ni suppression ni opération d'imputation. Cette conservation complète est un avantage considérable pour la robustesse de l'entraînement des modèles.

age	0
sex	0
chest_pain_type	0
resting_blood_pressure	0
cholesterol	0
fasting_blood_sugar	0
rest_ecg	0
Max_heart_rate	0
exercise_induced_angina	0
oldpeak	0
slope	0
vessels_colored_by_flourosopy	0
thalassemia	0
target	0

Figure 18– Les valeurs manquantes

2.3.2 Nettoyage des données :

Le nettoyage des données constitue une étape essentielle pour garantir la précision et la fiabilité d'un modèle prédictif. Avant toute phase d'entraînement, il est indispensable d'effectuer un prétraitement rigoureux des données. Cela implique notamment la détection et le traitement des valeurs aberrantes, qui peuvent être soit supprimées, soit remplacées par des valeurs plus représentatives. Une fois ces opérations de nettoyage terminées, les données sont alors prêtes à être exploitées pour l'entraînement du modèle.

2.3.3 Transformation et encodage :

Avant de poursuivre l'entraînement du modèle, il est essentiel de traiter les variables catégoriques, car les algorithmes d'apprentissage automatique ne sont généralement pas capables de gérer directement ce type de variables en combinaison avec des variables numériques.

Pour surmonter cette limitation, plusieurs techniques de transformation des variables catégoriques en représentations numériques peuvent être utilisées. Parmi les méthodes les plus courantes, on peut citer le Label Encoding et l'One-Hot Encoding, chacune adaptée à un type spécifique de données et de modèle

-Le Label Encoding: est une méthode de transformation des variables catégoriques consistant à attribuer un entier unique à chaque catégorie. Cette technique est simple et utile lorsque les catégories ont un ordre implicite ou explicite.

Exemple :

La variable "taille" avec les catégories ["petite", "moyenne", "grande"] peut être encodée comme [0, 1, 2]



- **Le One-Hot Encoding** : est une méthode de transformation des variables catégoriques consistant à créer une nouvelle colonne pour chaque catégorie unique. Chaque observation est représentée par un vecteur binaire où seule la colonne correspondant à sa catégorie vaut 1

Exemple :

La variable "animal" avec les catégories ["chat", "chien", "oiseau"] devient : [38]

- "chat" → [1, 0, 0]
- "chien" → [0, 1, 0]
- "oiseau" → [0, 0, 1]

2.3.4 Standardisation des données :

La standardisation des données consiste à transformer les données en un format et une structure cohérents pour permettre une analyse et une intégration transparentes entre différents systèmes. Ce processus consiste à convertir des données provenant de différentes sources dans un format standardisé avec une représentation uniforme des mêmes points de données, quelle que soit leur origine. L'objectif est d'éliminer les écarts et les incohérences, ce qui facilite l'agrégation, la comparaison et l'analyse des données.

2.3.5 Extraction des caractéristiques:

L'extraction des caractéristiques représente une étape cruciale dans le processus de traitement des données. Elle consiste à identifier, sélectionner et préparer les variables les plus pertinentes à partir de l'ensemble de données initial. Cela inclut la collecte des données brutes, la sélection des caractéristiques significatives, l'extraction de nouvelles caractéristiques dérivées, ainsi que leur évaluation afin de garantir leur pertinence et leur impact positif sur les performances du modèle prédictif.

Dans le cas d'analyses complexes, où les ressources informatiques (mémoire, temps de calcul) sont sollicitées de manière intensive, cette étape permet d'optimiser l'utilisation des algorithmes de classification, en réduisant la dimensionnalité des données. Elle contribue ainsi à alléger le coût computationnel tout en conservant l'essentiel de l'information utile. L'issue de cette phase est l'obtention d'un sous-ensemble de caractéristiques sélectionnées, appelées caractéristiques extraites, qui serviront à alimenter les modèles d'apprentissage automatique pour une meilleure prédiction du risque de CHD.

2.3.6 Le fractionnement des données

Le fractionnement des données est une étape essentielle dans le processus de modélisation en apprentissage automatique. Il consiste à diviser l'ensemble des données disponibles en deux sous-ensembles distincts :

- **un ensemble d'apprentissage**, représentant généralement 80 % des données, utilisé pour entraîner le modèle ;

- **un ensemble de test**, représentant les 20 % restants, destiné à évaluer les performances du modèle sur des données inédites

Cette séparation permet de mesurer la capacité du modèle à généraliser ses prédictions sur de nouvelles données, en évitant les biais liés à un apprentissage uniquement sur les données connues.



Figure 19– Fractionnement de données

3 Test du modèle :

Dans une première étape, nous avons réalisé une comparaison préliminaire entre plusieurs algorithmes d'apprentissage automatique, notamment Decision Tree, Random Forest, AdaBoost, Gradient Boosting, Bagging, SVM, K-Nearest Neighbors (KNN) et Artificial Neural Network (ANN)

Cette comparaison nous a permis d'identifier les modèles les plus performants en termes de précision et de robustesse.

Modèle	Accuracy	Précision (macro avg)	Rappel (macro avg)	F1-score (macro avg)
Decision Tree	0.99	0.99	0.99	0.99
Random Forest	0.99	0.99	0.99	0.99
Ada Boost	0.80	0.80	0.79	0.79
Gradient Boost	0.93	0.93	0.93	0.93
Bagging	0.99	0.99	0.99	0.99
SVM Normal	0.68	0.68	0.68	0.68
SVM + Standard Scaler	0.85	0.86	0.85	0.85
SVM + Robust Scaler	0.84	0.85	0.84	0.84
KNN Normal	0.68	0.68	0.68	0.68
KNN + Standard Scaler	0.85	0.85	0.85	0.85
KNN + Robust Scaler	0.83	0.84	0.83	0.83
ANN Normal	0.79	0.79	0.79	0.78
ANN + Standard Scaler	0.88	0.88	0.88	0.88
ANN + Robust Scaler	0.84	0.86	0.84	0.84

Tableau 2: Comparaison préliminaire des performances des modèles

Les résultats ont montré que les algorithmes Decision Tree, Random Forest et Bagging se distinguaient par une performance supérieure sur nos données.

Ces trois modèles ont donc été retenus pour la suite du travail, car ils offraient le meilleur compromis entre simplicité, précision et capacité de généralisation

4 Optimisation et évaluation des modèles :

Afin d'entraîner nos trois modèles sélectionnés de manière optimale, nous avons utilisé la méthode de validation croisée (Cross Validation) Cette technique nous a permis d'évaluer la robustesse des modèles sur différentes partitions des données et de réduire le risque d'overfitting, en s'assurant que le modèle généralise bien à des données non vues

Pour chaque modèle, nous avons appliqué GridSearchCV afin de rechercher les meilleurs hyperparamètres possibles Cette étape d'optimisation a contribué à améliorer les performances de chaque algorithme tout en maintenant un bon équilibre entre biais et variance.

Finalement, au lieu de sélectionner un seul modèle, nous avons choisi d'intégrer les trois modèles optimisés dans l'application, et d'afficher leurs prédictions simultanément pour permettre une interprétation comparative par l'utilisateur

4.1 Les mesures d'évaluation de performance :

L'évaluation des performances d'un modèle d'apprentissage automatique repose sur un ensemble de mesures permettant de juger sa qualité prédictive. Ces indicateurs comparent les prédictions du modèle aux données réelles afin de quantifier son efficacité.

***Précision (Accuracy) :** Elle indique la proportion d'échantillons correctement classés parmi l'ensemble des observations. Cette métrique fournit une vue d'ensemble de la justesse globale du modèle, particulièrement utile lorsque les classes sont équilibrées.

***Rappel (Recall) :** Aussi connu sous le nom de sensibilité ou taux de vrais positifs, le rappel mesure la capacité du modèle à détecter correctement les cas positifs. Il est essentiel dans les contextes sensibles où il est crucial de minimiser les faux négatifs, comme dans le domaine médical.

***F-mesure (F1-score) :** Il s'agit de la moyenne harmonique entre la précision et le rappel. La F-mesure est particulièrement adaptée aux situations où l'on cherche un compromis équilibré entre ces deux métriques, notamment en présence de classes déséquilibrées.

Ces mesures sont souvent complétées par des outils comme la matrice de confusion, qui permet une analyse plus fine des erreurs de classification. Le choix des indicateurs d'évaluation dépend des priorités spécifiques au problème étudié. [23]

4.2 Test sur le sous-ensemble de teste:

Après l'entraînement, les modèles optimisés ont été testés sur un sous-ensemble de données jamais vues.

L'objectif est d'évaluer leur capacité à bien généraliser, Nous avons utilisé plusieurs métriques : Accuracy, Recall, précision, et AUC pour mesurer la performance globale.

```

Decision Tree:
Accuracy: 0.8293
AUC: 0.9293
Classification Report:
              precision    recall  f1-score   support

     0           0.88       0.76       0.82         102
     1           0.79       0.89       0.84         103

 accuracy          0.83          0.83          0.83         205
 macro avg         0.83          0.83          0.83         205
 weighted avg      0.83          0.83          0.83         205

```

Figure 20– Mesures d'évaluation des performances de l'arbre de décision

```

Random Forest:
Accuracy: 0.8829
AUC: 0.9788
Classification Report:
              precision    recall  f1-score   support

     0           0.91       0.85       0.88         102
     1           0.86       0.91       0.89         103

 accuracy          0.88          0.88          0.88         205
 macro avg         0.88          0.88          0.88         205
 weighted avg      0.88          0.88          0.88         205

```

Figure 21– Mesures d'évaluation des performances de forêt aléatoire

```

Bagging:
Accuracy: 0.8732
AUC: 0.9676
Classification Report:
              precision    recall  f1-score   support

     0           0.91       0.82       0.87         102
     1           0.84       0.92       0.88         103

 accuracy          0.87          0.87          0.87         205
 macro avg         0.88          0.87          0.87         205
 weighted avg      0.88          0.87          0.87         205

```

Figure 22– Mesures d'évaluation des performances de Bagging

4.3 Matrice de confusion de modèle :

Le Machine Learning consiste à nourrir un algorithme à l'aide de données afin qu'il apprenne par lui-même à effectuer une certaine tâche. Dans les problèmes de classification, il prédit des résultats



que l'on doit comparer à la réalité pour mesurer son degré de performance. On utilise généralement la matrice de confusion, appelée aussi tableau de contingence. Elle mettra non seulement en valeur les prédictions correctes et incorrectes mais nous donnera surtout un indice sur le type d'erreurs commises. Pour calculer une matrice de confusion, on a besoin d'un ensemble de données de test et un autre de validation qui contient les valeurs des résultats obtenus.

Chaque colonne du tableau contient une classe prédite par l'algorithme et les lignes des classes réelles.

On classe les résultats en 4 catégories :

***True Positive (TP)** : la prédiction et la valeur réelle sont positives.

Exemple : Une personne malade et prévu malade.

***True Negative (TN)** : la prédiction et la valeur réelle sont négatives.

Exemple : Une personne saine et prévu saine.

***False Positive (FP)** : la prédiction est positive alors que la valeur réelle est négative.

Exemple : Une personne saine et prévu malade.

***False Negative (FN)** : la prédiction est négative alors que la valeur réelle est négative.

Exemple : Une personne malade et prévu saine. [24]

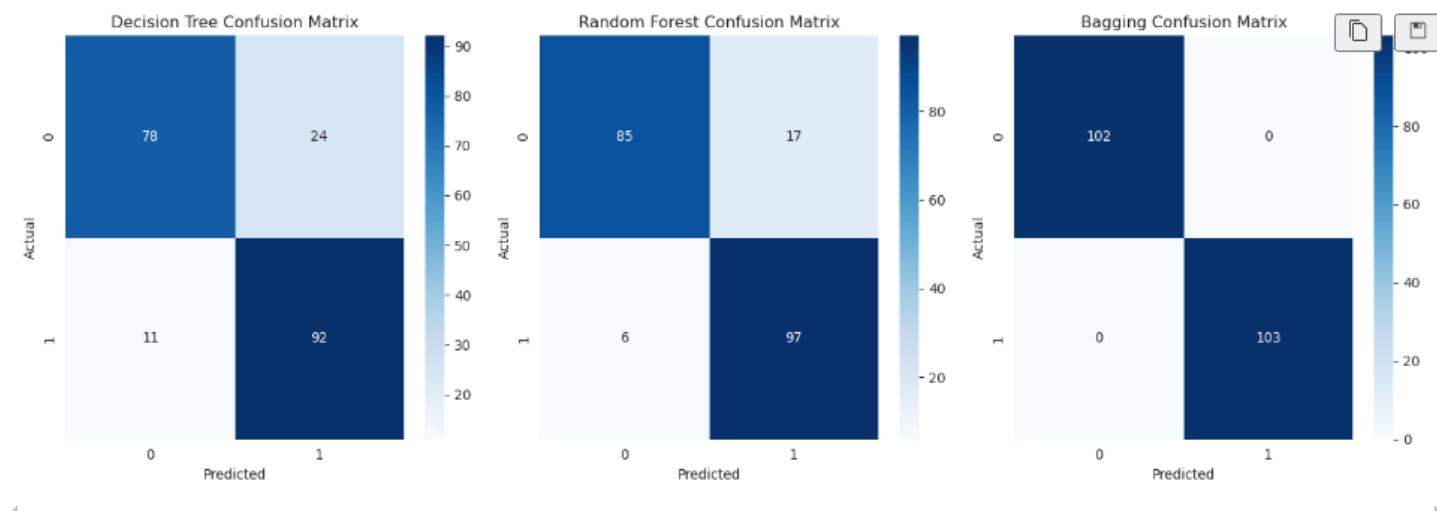


Figure 23– Matrices de confusion pour les différents modèles

4.4 Test sur un échantillon de teste :

Afin de tester l'efficacité de nos modèles dans la prédiction de la maladie coronarienne (CHD), nous avons sélectionné trois patients (P1, P2, P3).

Ces échantillons ont été évalués à l'aide des trois modèles retenus : Arbre de Décision, Forêt Aléatoire et Bagging.

Pour l'interprétation des résultats, un seuil de décision de 50% a été utilisé. Ainsi, une probabilité supérieure ou égale à 50% indique que le patient est considéré comme malade, tandis qu'une probabilité inférieure à ce seuil signifie qu'il est sain. Les résultats prédictifs, présentés dans le



tableau [2], montrent la capacité des modèles à détecter les profils à risque, ce qui peut favoriser un dépistage précoce et une intervention préventive adapté.

modelé patients	l'arbre de décision	foret aléatoire	Bagging	résultats
P1	100.0%	81.9%	68.5%	Présence de CHD (83.5%)
P2	0.0%	20.1%	31.3%	Absence de CHD
P3	0.0%	24.1%	36.7%	Absence de CHD

Tableau 3– Résultats de l'échantillon



Conclusion

Ce chapitre a résumé le processus de construction des modèles de prédiction appliqués à notre problématique. Après une préparation appropriée des données, plusieurs algorithmes ont été testés. Trois modèles se sont révélés les plus efficaces : l'arbre de décision, la forêt aléatoire et le Bagging. Ils ont été optimisés et évalués afin de garantir des résultats fiables. Ces modèles sont désormais intégrés dans l'application pour appuyer le diagnostic médical.



Chapitre 4
Conception et fonctionnalités de
l'application



1 Introduction

Ce chapitre est consacré à la présentation des outils et technologies utilisés dans le cadre de notre travail. Nous y décrirons brièvement les environnements de développement, le langage utilisé, ainsi que les bibliothèques principales. Une partie sera également réservée à la définition et la conception de l'interface développée dans notre système.

2 Environnement de développement

Pour implémenter notre système, nous avons utilisé des plateformes de développement, un langage de programmation et des bibliothèques

2.1 Plateforme de développement

2.1.1 Google Colab

Google Colaboratory, ou Google Colab en abrégé, est un environnement de développement interactif basé sur le cloud qui permet d'écrire et d'exécuter du code Python directement dans un navigateur web. C'est un outil gratuit qui offre un accès à des ressources de calcul puissantes, notamment des GPU et des TPU, sans nécessiter de configuration locale. [25]

2.1.2 Jupyter Notebook

Jupyter Notebook est une application web open source qui permettent de créer et de partager des documents contenant du code exécutable, des équations, du texte narratif et des visualisations. Il est largement utilisé en data science, apprentissage automatique, et analyse de données. [26]

2.2 Langage de développement

Nous avons utilisé pour le développement de notre modèle le langage de programmation python.

2.2.1 Python

Python est un langage de programmation (au même titre que le C, C++, fortran, java . . .), développé en 1989. Ses principales caractéristiques sont les suivantes :

- «open-source» : son utilisation est gratuite et les fichiers sources sont disponibles et modifiables.
- simple et très lisible.
- doté d'une bibliothèque de base très fournie.
- importante quantité de bibliothèques disponibles : pour le calcul scientifique, les statistiques, les bases de données, la visualisation
- grande portabilité : indépendant vis à vis du système d'exploitation (linux, Windows, MacOS).
- orienté objet.
- typage dynamique : le typage (association à une variable de son type et allocation zone mémoire en conséquence) est fait automatiquement lors de l'exécution du programme, ce qui permet une grande flexibilité et rapidité de programmation, mais qui se paye par une surconsommation de mémoire et une perte de performance.
- présente un support pour l'intégration d'autres langages. [27]



2.3 Bibliothèques utilisées

2.3.1 Pandas

Pandas est une bibliothèque du langage de programmation Python, entièrement dédiée à la Data Science. Découvrez à quoi sert cet outil, et pourquoi il est incontournable pour les Data Scientists. [28]

2.3.2 NumPy

NumPy (Numerical Python) est une bibliothèque fondamentale pour le calcul scientifique en Python. Elle fournit un objet tableau multidimensionnel puissant, ainsi qu'un large éventail de fonctions mathématiques pour effectuer des opérations rapides sur ces tableaux. [29]

2.3.3 Matplotlib

Matplotlib est une bibliothèque Python open source permettant de créer des visualisations de données. Découvrez tout ce que vous devez savoir : définition, fonctionnement, difficultés, formations... [30]

2.3.4 Seaborn

Seaborn est une bibliothèque de visualisation de données statistiques en Python, conçue pour créer des graphiques informatifs et esthétiquement attrayants. [31]

2.3.5 SciPy.stats

Le SciPy est une extension de Numpy (Numerical Python), et permet donc un traitement de données extrêmement rapide et efficace. Cette bibliothèque est écrite en C, C++, Fortran et Python.

Les commandes et classes de haut niveau permettent de manipuler et visualiser les données de façon simple. [32]

3 Réalisation de l'interface utilisateur

3.1 Choix des technologies utilisées

3.1.1 ReactJS (JavaScript)

React est une bibliothèque JavaScript open-source qui est utilisée pour construire des interfaces utilisateur spécifiquement pour des applications d'une seule page. Elle est utilisée pour gérer la couche d'affichage des applications web et mobiles. React nous permet également de créer des composants d'interface utilisateur réutilisables. [33]

3.1.2 Next.js (JavaScript)

Next.js est un Framework JavaScript open-source pour créer des applications web avec React. Il fournit des fonctionnalités pour faciliter le développement d'applications web statiques et dynamiques, comme la prise en charge de l'exécution côté serveur pour les composants React, la gestion automatique des routes, la pré-rendu, la prise en charge des médias statiques. [34]

3.1.3 Prisma (JavaScript/TypeScript)

Prisma est un ORM, une bibliothèque permettant de gérer une BDD de manière grandement simplifiée en représentant les tables sous forme d'entités. Comme nous allons le voir, il est loin d'être le premier à proposer cette approche, mais il apporte un certain nombre d'innovations qui le démarquent des autres bibliothèques du même genre. [35]



3.1.4 PostgreSQL (Base de données)

PostgreSQL est un système de base de données open source, très stable, qui prend en charge différentes fonctions de SQL, comme les clés étrangères, les sous-requêtes, les déclencheurs, ainsi que différents types et fonctions définis par l'utilisateur. Il augmente encore le langage SQL en proposant plusieurs fonctionnalités qui permettent de mettre à l'échelle et de réserver méticuleusement les charges de travail de données. Il est principalement utilisé pour stocker les données de nombreuses applications mobiles, web, géospatiales et analytiques. [36]

3.1.5 FastAPI (Python)

FastAPI est un Framework web haute performance, open source, permettant de créer des APIs avec Python à partir de la version 3.6. Lancé en 2018, il fut créé par Sébastian Ramirez qui n'était pas satisfait par les Framework existants comme Flaks et DRF.

C'est la raison pour laquelle il a créé son propre Framework à l'aide d'outils comme Starlette et Pydantic. Aujourd'hui, de nombreuses entreprises de renommée mondiale comme Uber, Netflix et Microsoft utilisent FastAPI pour créer leurs applications.

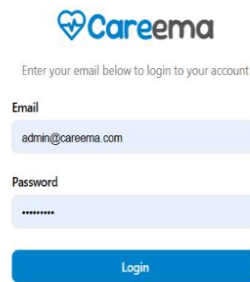
Son premier point fort est un haut niveau de performances, comparables avec NodeJS et Go grâce à Starlette et pydantic. Ce Framework est également très rapide à coder, offrant une accélération du développement. [37]

4 Aperçu de l'Interface Utilisateur

L'interface utilisateur de l'application a été conçue dans un souci de simplicité, d'accessibilité et d'efficacité. Elle permet aux utilisateurs, principalement les professionnels de santé, d'interagir facilement avec le système : enregistrer de nouveaux patients, saisir des données cliniques, lancer une prédiction à l'aide du modèle de machine learning, et consulter les résultats. L'objectif est d'offrir une expérience fluide et intuitive, tout en mettant en avant les informations médicales essentielles.

4.1 Écran de connexion utilisateur

Cette page présente un formulaire de connexion permettant aux utilisateurs d'accéder à leur compte. Le formulaire contient deux champs de saisie : un champ "Email" où l'adresse "admin@careema.com" est affichée par défaut, et un champ "Mot de passe" pour la saisie du mot de passe, avec une indication visuelle que des caractères sont entrés. Un bouton "Login" en bleu est positionné sous ces champs, servant de point d'action principal pour soumettre les informations de connexion. La page inclut également le logo de l'application, "Careema", et un court message invitant à entrer l'email pour se connecter. Visuellement, une illustration d'une doctoresse tenant une tablette affichant un cœur est présente sur le côté, ajoutant un contexte médical à l'application.



The login form for the Careema application is displayed on a white background. At the top left is the Careema logo, which consists of a blue heart icon with a white pulse line and the word "Careema" in a blue sans-serif font. Below the logo is the instruction "Enter your email below to login to your account". There are two input fields: the first is labeled "Email" and contains the text "admin@careema.com"; the second is labeled "Password" and contains a series of dots. A blue "Login" button is positioned at the bottom of the form.



Figure 24: la fenêtre de connexion de l'application.

4.2 Tableau de bord principal

Cette page représente le tableau de bord principal de l'application. En haut, on trouve des indicateurs clés affichant le nombre total de patients, le nombre de diagnostics en attente, et le nombre de diagnostics complétés aujourd'hui. La section centrale présente deux graphiques circulaires. À gauche, un graphique "Sex Statistics" indique la répartition des patients par sexe (bleu pour les hommes, rose pour les femmes). À droite, un graphique "Prevalence of Disease X in Total Patients" illustre la proportion de patients avec (rouge) et sans (vert) la maladie X. En bas, un histogramme intitulé "Patient Age Distribution" affiche la distribution des patients par tranche d'âge, indiquant le nombre de patients pour chaque groupe d'âge. Sur la gauche de la page, une barre de navigation verticale en bleu foncé permet d'accéder aux différentes sections de l'application : "Home", "Patients", "Reports", et un bouton de "Logout" en bas. Le logo "Careema" est également présent en haut de cette barre de navigation.

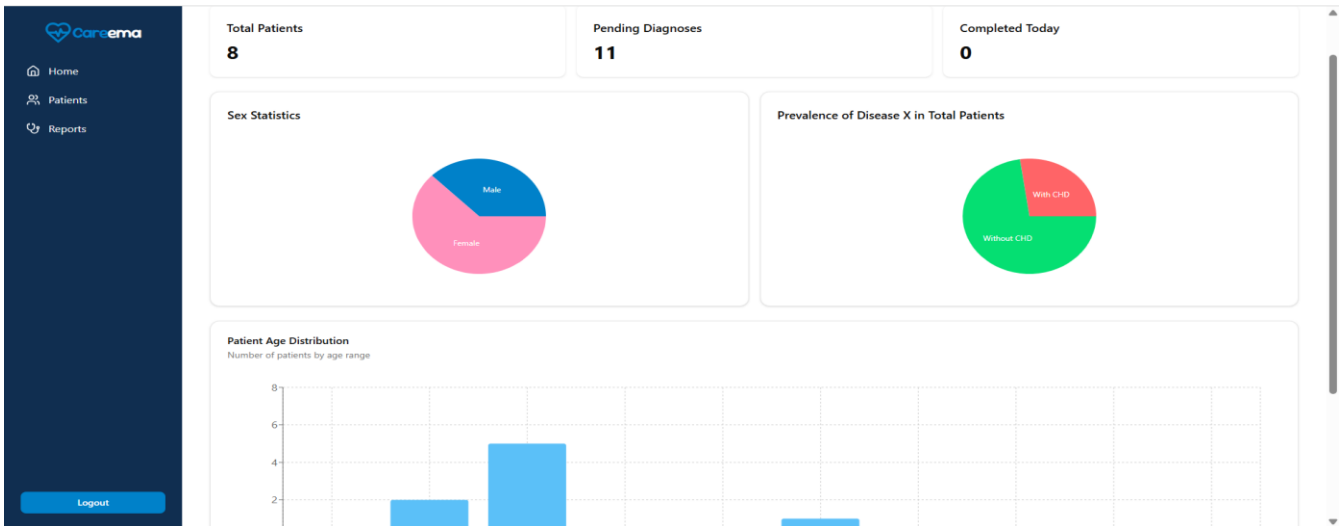


Figure 25: fenêtre des statistiques des patients.

4.3 Formulaire d'enregistrement de patient

Cette page affiche un formulaire destiné à l'enregistrement de nouveaux patients. Le formulaire contient les champs suivants : "Name" (Nom) pour saisir le nom du patient, un menu déroulant "Sex" (Sexe) permettant de sélectionner le sexe du patient, un champ "Age" (Âge) pour indiquer l'âge, un champ "Phone" (Téléphone) pour entrer le numéro de téléphone, et un champ "Address" (Adresse) pour renseigner l'adresse du patient. Un bouton "Register" (Enregistrer) en bleu est situé en bas du formulaire pour soumettre les informations du nouveau patient. Sur la gauche, la barre de navigation bleue habituelle permet d'accéder à "Home", "Patients", "Reports", et offre la possibilité de se déconnecter via le bouton "Logout". Le logo "Careema" est présent en haut de cette barre. Le titre principal de la page est "Patient Registration" (Enregistrement de patient).

Profile Information

Name

Sex

Age

Phone

Must be a 10-digit phone number

Figure 26: fenêtre d'information personnelles

4.4 Liste des patients

Cette page affiche une table listant les patients enregistrés. En haut, un champ de filtre permet de rechercher des patients. La table présente des colonnes pour le "Name" (Nom), le "Sex" (Sexe) affiché avec une puce colorée (bleue pour masculin, rose pour féminin), l'"Age" (Âge), le "Phone" (Téléphone), et une colonne "actions". Pour chaque patient, dans la colonne "actions", un menu déroulant (activé par un icône "...") offre des options telles que "diagnose" (diagnostiquer), "Edit" (Modifier), et "Delete" (Supprimer). En haut à droite, un bouton "add patient" (ajouter un patient) est disponible. La navigation en bas de la table permet de passer aux pages "Previous" (Précédente) et "Next" (Suivante). La barre de navigation latérale bleue à gauche donne accès aux sections "Home", "Patients", "Reports", et au bouton "Logout", avec le logo "Careema" en haut. Le titre principal de la page est "Patients".

Name ↕	Sex ↕	Age ↕	Phone ↕	actions
gggg	male	52	1234567890	...
somia	female	22	1234567890	<ul style="list-style-type: none"> diagnose Edit Delete
ahlem	female	22	1234565894	...
test	female	10	1234567890	...
karima	female	23	4654546220	...
cill	male	21	1234567800	...
sss	female	20	1234567811	...
nadjib	male	11	1234567855	...

Figure 27: fenêtre de gestion des patients

4.5 Formulaire d'évaluation du risque de maladie cardiaque

Cette page présente un formulaire permettant d'évaluer le risque de maladie cardiaque d'un patient. Le formulaire est divisé en plusieurs sections et comprend les champs suivants :

- Chest Pain Type (Type de douleur thoracique) : un menu déroulant pour sélectionner le type de douleur.
- Resting Blood Pressure (mmHg) (Pression artérielle au repos) : un champ pour entrer la valeur de la pression artérielle au repos.
- Cholesterol (mg/dl) (Cholestérol) : un champ pour entrer le niveau de cholestérol.
- Fasting Blood Sugar (Glycémie à jeun) : un menu déroulant pour sélectionner le niveau de glycémie à jeun.
- Resting ECG Results (Résultats de l'ECG au repos) : un menu déroulant pour sélectionner les résultats de l'ECG au repos.

- Maximum Heart Rate (Fréquence cardiaque maximale) : un champ pour entrer la fréquence cardiaque maximale atteinte.
- Exercise Induced Angina (Angine de poitrine induite par l'exercice) : un menu déroulant pour indiquer si l'angine est induite par l'exercice.
- ST Depression (Oldpeak) (Sous-décalage du segment ST) : un champ pour entrer la valeur du sous-décalage du segment ST.
- Slope of Peak Exercise ST Segment (Pente du segment ST à l'exercice maximal) : un menu déroulant pour sélectionner la pente.
- Major Vessels Colored by Fluoroscopy (Nombre de vaisseaux principaux colorés par fluoroscopie) : un menu déroulant pour sélectionner le nombre de vaisseaux.
- Thalassemia (Thalassémie) : un menu déroulant pour sélectionner le type de thalassémie.

Un bouton bleu "Submit Assessment" (Soumettre l'évaluation) est situé en bas du formulaire pour lancer l'évaluation du risque. La barre de navigation latérale bleue à gauche permet d'accéder aux sections "Home", "Patients", "Reports", et au bouton "Logout", avec le logo "Careema" en haut. Le titre principal de la page est "Heart Disease Risk Assessment Form" (Formulaire d'évaluation du risque de maladie cardiaque).

The screenshot shows the 'Heart Disease Risk Assessment Form' interface. On the left is a dark blue sidebar with the Careema logo at the top, followed by navigation icons and text for 'Home', 'Patients', and 'Reports'. At the bottom of the sidebar is a 'Logout' button. The main content area is white and contains the following fields:

- Chest Pain Type**: A dropdown menu with the text 'Select chest pain type'.
- Cholesterol (mg/dl)**: A text input field with the placeholder 'Enter cholesterol level'.
- Resting ECG Results**: A dropdown menu with the text 'Select resting ECG result'.
- Exercise Induced Angina**: A dropdown menu with the text 'Select option'.
- Slope of Peak Exercise ST Segment**: A dropdown menu with the text 'Select slope type'.
- Thalassemia**: A dropdown menu with the text 'Select thalassemia type'.
- Resting Blood Pressure (mmHg)**: A text input field with the placeholder 'Enter resting blood pressure'.
- Fasting Blood Sugar**: A dropdown menu with the text 'Select fasting blood sugar'.
- Maximum Heart Rate**: A text input field with the placeholder 'Enter maximum heart rate'.
- ST Depression (Oldpeak)**: A text input field with the placeholder 'Enter ST depression value'.
- Major Vessels Colored by Fluoroscopy**: A dropdown menu with the text 'Select number of vessels'.

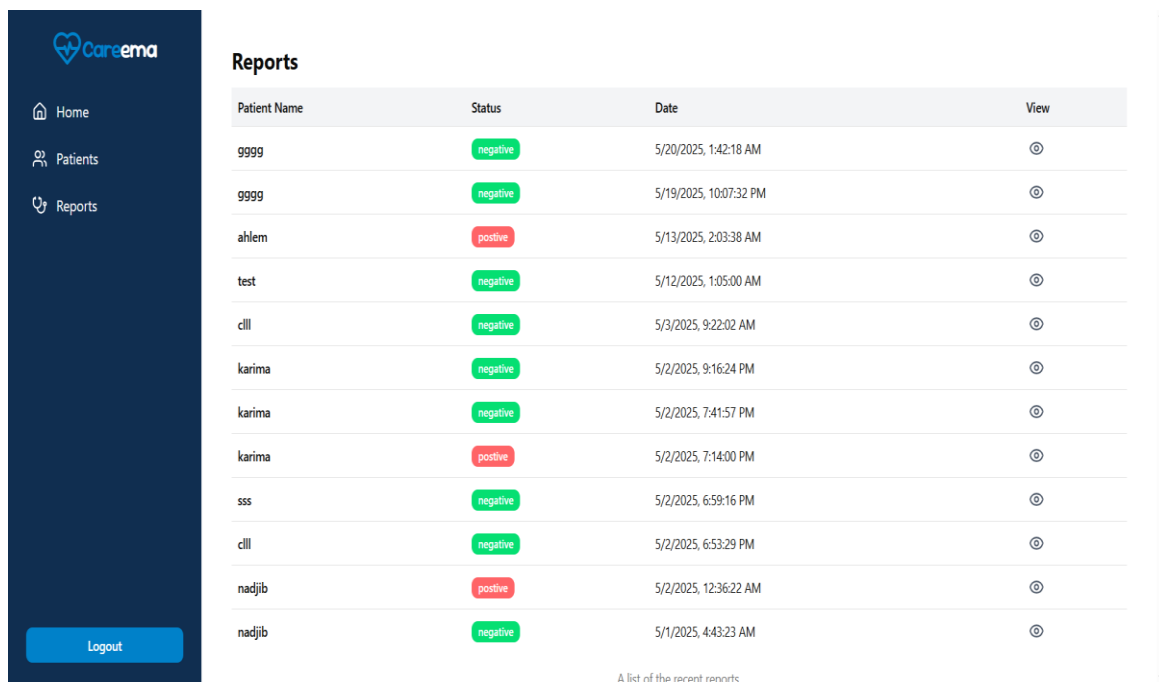
At the bottom left of the form area is a blue button labeled 'Submit Assessment'.

Figure 28: fenêtre d'ajout des données médicales.

4.6 Liste des rapports de diagnostic récents

Cette page affiche une table listant les rapports de diagnostic récents. La table comprend les colonnes suivantes : "Patient Name" (Nom du patient), "Status" (Statut) indiquant le résultat du diagnostic (affiché avec une puce verte pour "negative" et rouge pour "positive"), "Date" (Date) et heure du rapport, et "View" (Voir) avec une icône permettant de consulter le rapport détaillé. Les rapports sont listés pour différents patients avec leurs statuts et dates correspondants. En bas de la table, une mention indique "A list of the recent reports." (Une liste des rapports récents). La barre

de navigation latérale bleue à gauche donne accès aux sections "Home", "Patients", "Reports", et au bouton "Logout", avec le logo "Careema" en haut. Le titre principal de la page est "Reports" (Rapports).



Patient Name	Status	Date	View
9999	negative	5/20/2025, 1:42:18 AM	⊙
9999	negative	5/19/2025, 10:07:32 PM	⊙
ahlem	positive	5/13/2025, 2:03:38 AM	⊙
test	negative	5/12/2025, 1:05:00 AM	⊙
cill	negative	5/3/2025, 9:22:02 AM	⊙
karima	negative	5/2/2025, 9:16:24 PM	⊙
karima	negative	5/2/2025, 7:41:57 PM	⊙
karima	positive	5/2/2025, 7:14:00 PM	⊙
sss	negative	5/2/2025, 6:59:16 PM	⊙
cill	negative	5/2/2025, 6:53:29 PM	⊙
nadjib	positive	5/2/2025, 12:36:22 AM	⊙
nadjib	negative	5/1/2025, 4:43:23 AM	⊙

A list of the recent reports.

Figure 29: fenêtre de suivis des patients.

5 Rapport d'évaluation du risque de maladie cardiaque

Cette page présente le rapport d'évaluation du risque de maladie cardiaque. Elle commence par les "Patient Information" (Informations du patient), affichant le nom, l'âge et le sexe.

La section "Heart Disease Risk Assessment" (Évaluation du risque de maladie cardiaque) fournit le verdict global, ici "No Heart Disease Détecté" (Aucune maladie cardiaque détectée), avec une probabilité globale de la maladie (ex: 19.6% de risque faible) et le consensus des modèles (ex: 0 prédiction positive sur 3 modèles utilisés).

La section "Individuel Model Probabilities" (Probabilités des modèles individuels) détaille les prédictions de chaque modèle (Bagging, Decision Tree, Random Forest) via un histogramme et des probabilités individuelles avec leur niveau de risque associé. Une mention finale rappelle que les prédictions sont basées sur des modèles ML et nécessitent une confirmation professionnelle.

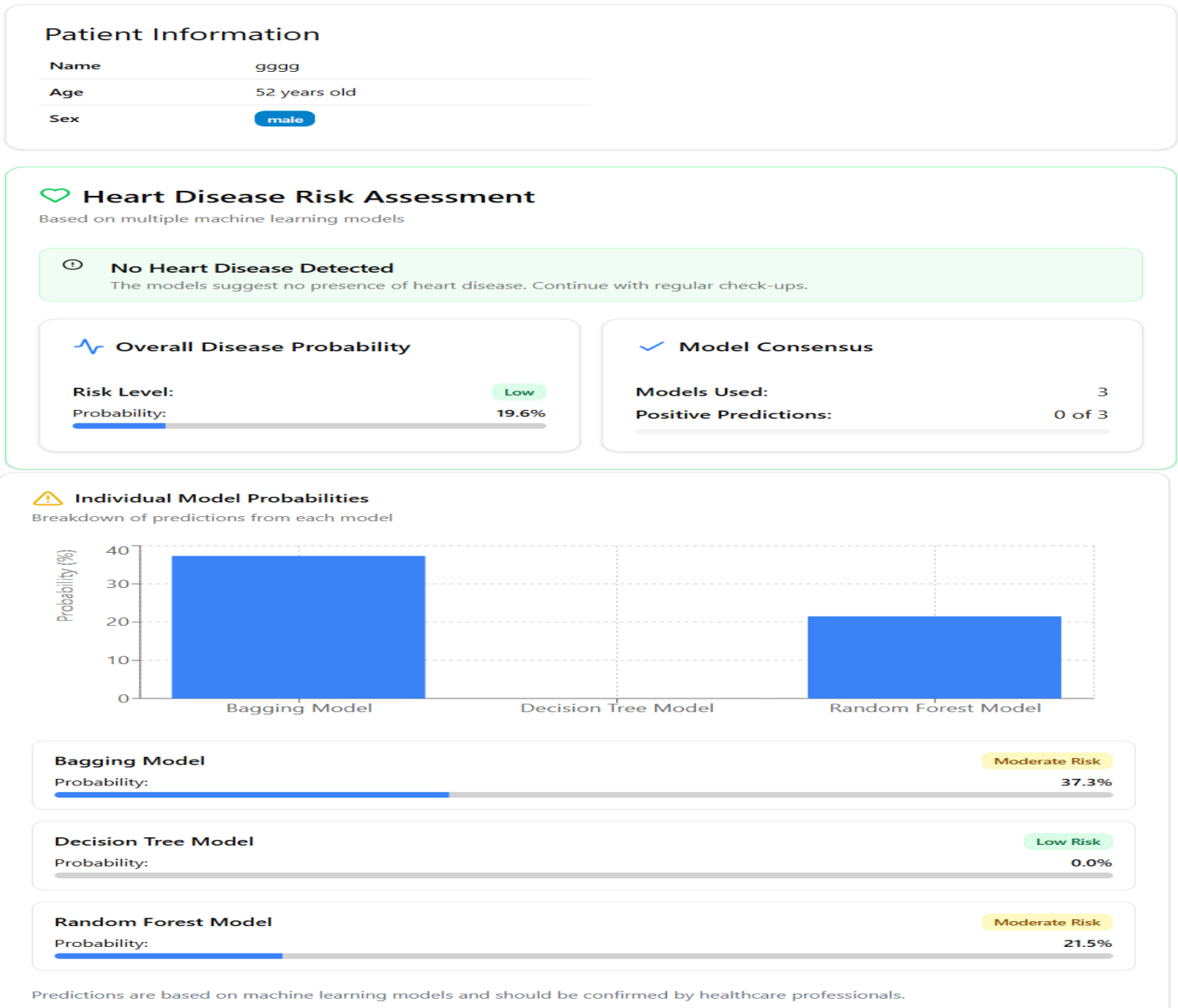


Figure 30: Résultat d'évaluation du risque cardiaque

**Conclusion :**

Ce chapitre nous a permis de poser les bases concrètes de notre application en explorant l'environnement technique dans lequel elle a été développée. Le choix réfléchi des langages, des bibliothèques et des interfaces n'est pas anodin : Il reflète notre volonté de construire un système à la fois performant, accessible et adapté aux besoins réels du domaine médical. Chaque composant technique s'inscrit dans une logique d'ensemble, orientée vers une exploitation optimale des données. Cette étape représente une transition clé entre la conception théorique et l'application pratique, ouvrant la voie à l'analyse des résultats et à l'évaluation de l'efficacité du système dans les chapitres suivants.

Conclusion générale

En conclusion, cette étude met en lumière l'importance cruciale de l'utilisation des techniques d'apprentissage automatique dans le domaine de la cardiologie, et plus précisément dans l'analyse et la prédiction de la maladie coronarienne. Face à la complexité et à la diversité des données médicales, ces méthodes offrent une solution innovante et prometteuse pour améliorer la précision diagnostique et la prise de décision clinique.

Les modèles développés, basés sur les algorithmes de l'arbre de décision, de la forêt aléatoire et du Bagging, démontrent leur efficacité dans la classification et la prédiction du risque chez les patients. L'intégration de ces outils dans le système de santé peut contribuer à une meilleure détection précoce, une prévention adaptée et une gestion personnalisée des patients, réduisant ainsi les complications et la mortalité associées à cette pathologie.

Toutefois, malgré ces avancées, il reste essentiel de poursuivre les recherches afin d'optimiser ces modèles, d'en améliorer la robustesse et de garantir leur applicabilité dans différents contextes cliniques. De plus, la collaboration étroite entre informaticiens, chercheurs et professionnels de la santé demeure indispensable pour assurer une intégration efficace et éthique des technologies d'intelligence artificielle.

En définitive, ce mémoire contribue modestement à l'évolution vers une médecine plus prédictive, proactive et personnalisée, où la puissance des données et des algorithmes soutient les efforts humains pour combattre les maladies cardiovasculaires, enjeu majeur de santé publique mondial

Bibliographies

- [1] World Health Organization. (2021). Cardiovascular diseases (CVDs). Récupéré de : [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))
- [2] Guyton, A. C., & Hall, J. E. (2016). Physiologie médicale. *Elsevier Masson*
- [3] CHU de Rouen. (n.d.). *Comment le cœur fonctionne-t-il ?* Récupéré le 28 mai 2025, de <https://www.chu-rouen.fr/services/cardiologie/valves-cardiaques/fonctionnement-du-coeur>
- [4] Fondation des maladies du cœur et de l'AVC. (n.d.). *Les types de maladies du cœur*. Récupéré le 28 mai 2025, de <https://www.coeuretavc.ca/maladies-du-coeur/qu-est-ce-que-les-maladies-du-coeur/les-types-de-maladies-du-coeur>
- [5] Fondation des maladies du cœur et de l'AVC
- [6] Société Française de Cardiologie. (2024). *Chapitre 02 – Item 222 : Facteurs de risque cardiovasculaire et prévention*. Récupéré le 28 mai 2025, de <https://www.sfcadio.fr/publication/chapitre-02-item-222-facteurs-de-risque-cardiovasculaire-et-prevention/#facteurs>
- [7] Université de Bouira. (n.d.). *Mémoire de Master (Mémoire de master)*. Université de Bouira. Récupéré de <http://dspace.univbouira.dz:8080/jspui/bitstream/123456789/13458/1/M%C3%A9moire%20de%20Master.pdf>
- [8] Médecin Direct. (2025, 3 février). *Maladies cardiaques : comprendre, prévenir et traiter*. Récupéré le 28 mai 2025, de <https://www.medecindirect.fr/maladies/maladies-cardiaques-comprendre-prevenir-et-traiter#part-4>
- [9] Anin, É. (2015). *Participation du pharmacien d'officine à l'éducation thérapeutique des patients après un programme de réadaptation cardiaque : Pérennisation de l'observance médicamenteuse, du respect des règles hygiéno-diététiques et du maintien de l'activité physique* (Thèse de doctorat). Université de Lorraine
- [10] Lacour, D. (2023, février 1). *Apprentissage automatique : définition, méthodes et applications*. NordVPN <https://nordvpn.com/fr/blog/apprentissage-automatique/>
- [11] Linedata. (s.d.). *Qu'est-ce que l'apprentissage supervisé ?* Récupéré de <https://fr.linedata.com/quest-ce-que-lapprentissage-supervise>
- [12] Mifdal, R. (2019). *Application des techniques d'apprentissage automatique pour la prédiction de la tendance des titres financiers* [Mémoire de maîtrise, École de technologie supérieure]. Espace ÉTS. https://espace.etsmtl.ca/id/eprint/2422/1/MIFDAL_Rachid_.pdf
- [13] Coursera Staff. (2024, 25 octobre). *Qu'est-ce que l'apprentissage par renforcement ?* Coursera. <https://www.coursera.org/fr-FR/articles/reinforcement-learning>

- [14] GeeksforGeeks. (2025, 16 janvier). *Decision Tree*. <https://www.geeksforgeeks.org/decision-tree/>
- [15] IBM. (2024, 25 octobre). *What is random forest?* <https://www.ibm.com/think/topics/random-forest>
- [16] GeeksforGeeks. (2025, 15 mai). *What is bagging classifier?* <https://www.geeksforgeeks.org/ml-bagging-classifier/>
- [17] GeeksforGeeks. (s.d.). *AdaBoost Algorithm*. <https://www.geeksforgeeks.org/adaboost-algorithm>
- [18] Singh, H. (2018, 3 novembre). *Understanding Gradient Boosting Machines*. Towards Data Science. <https://towardsdatascience.com/understanding-gradient-boosting-machines-9be756fe76ab>
- [19] GeeksforGeeks. (2025, 28 mai). *Support Vector Machine (SVM) Algorithm*. <https://www.geeksforgeeks.org/support-vector-machine-algorithm/>
- [20] GeeksforGeeks. (2025, 14 mai). *K-Nearest Neighbor (KNN) Algorithm*. <https://www.geeksforgeeks.org/k-nearest-neighbour-algorithm/>
- [21] GeeksforGeeks. (s.d.). *Artificial Neural Networks (ANN)*. <https://www.geeksforgeeks.org/artificial-neural-networks-ann>
- [22] Walker, N. (2019, 18 novembre). *Top applications de l'apprentissage automatique dans le domaine de la santé*. Kolabtree. <https://www.kolabtree.com/blog/fr/top-applications-of-machine-learning-in-healthcare/>
- [23] Sammut, C., & Webb, G. I. (2011). *Encyclopedia of Machine Learning*. Springer. Disponible en ligne : <https://link.springer.com/referencework/10.1007/978-0-387-30164-8>
- [24] Kassel, R. (2021, 16 février). *Matrice de confusion : qu'est-ce que c'est et comment l'utiliser ?* DataScientest. <https://datascientest.com/matrice-de-confusion>
- [25] Wild Code School. (s.d.). *Définition Google Colaboratory – Qu'est-ce que Google Colab ?* <https://www.wildcodeschool.com/lexique/google-colaboratory>
- [26] Project Jupyter. (s.d.). *Project Jupyter*. <https://jupyter.org/>
- [27] Hivert, F. (2015). *Introduction à la programmation en langage Python* [Support de cours, Université Paris-Sud]. <https://www.lri.fr/~hivert/COURS/Methodo/python.pdf>
- [28] Robert, J. (2022, 10 janvier). *Pandas : la bibliothèque Python dédiée à la Data Science*. DataScientest. <https://datascientest.com/pandas-python-data-science>
- [29] Wild Code School. (s.d.). *Définition NumPy – Qu'est-ce que NumPy ?* <https://www.wildcodeschool.com/lexique/definition-numpy>
- [30] Kassel, R. (2021, 22 février). *Matplotlib : Tout savoir sur la bibliothèque Python de Dataviz*. DataScientest. <https://datascientest.com/matplotlib-tout-savoir>
- [31] Wild Code School. (s.d.). *Définition Seaborn – Qu'est-ce que Seaborn ?* <https://www.wildcodeschool.com/lexique/definition-seaborn>
- [32] Kassel, R. (2022, 8 juin). *SciPy : tout savoir sur la bibliothèque Python de Machine Learning*. DataScientest. <https://datascientest.com/scipy>

- [33] Ibraci Links. (s.d.). *Qu'est-ce que ReactJS et pourquoi devrions-nous utiliser ReactJS ?* <https://ibracilinks.com/blog/quest-ce-que-reactjs-et-pourquoi-devrions-nous-utiliser-reactjs>
- [34] Bility. (s.d.). *Définition NextJS*. <https://bility.fr/definition-nextjs/>
- [35] Lonestone. (2023, 21 décembre). *Introduction à Prisma, un ORM novateur*. <https://www.lonestone.io/blog/introduction-a-prisma>
- [36] Kinsta. (2024, 3 décembre). *Qu'est-ce que PostgreSQL ?* <https://kinsta.com/fr/base-de-connaissances/qu-est-postgresql/>
- [37] Kassel, R. (2022, 4 mai). *FastAPI : tout savoir sur le framework web Python le plus utilisé pour le Machine Learning*. DataScientest. <https://datascientest.com/fastapi>
- [38] Éron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd Ed.). O'Reilly Media. Chapitre 2 – Data Preprocessing
- [39] Sujatha, p. et mahalakshmi, k. Évaluation des performances d'algorithmes d'Apprentissage automatique supervisé dans la prédiction des maladies cardiaques. Dans : Conférence internationale ieeee 2020 pour l'innovation technologique (inocon). Ieee, 2020. p. 1-7
- [40] El Hamlaoui, Halima, boujraf, saïd, chaoui, nour el houda, et al. L'invention concerne un système d'assistance clinique pour la prédiction des maladies cardiaques à l'aide de techniques d'apprentissage automatique. Dans: 2020 5th international conference on advanced technologies for signal and image processing (atsip). Ieee, 2020. p. 1-5.
- [41] BEHLOULI L.A., HAMDY H., *Une approche hybride pour la prédiction des maladies cardiaques*, Mémoire de Master, Univ. AMO de Bouira, 2025