



MÉMOIRE DE FIN DE CYCLE

En vue de l'obtention du diplôme de MASTER recherche en Informatique  
**Option : Intelligence artificielle**

## Thème

---

**IoT et Machine Learning pour la réalisation d'un  
cadre méthodologique dans le domaine de la  
santé : suivi des personnes cardiaques**

---

*Présenté par :*

**M. AISSAOUI Sid Ahmed**  
**Mlle. AIT MOUHOU B Sonia**

*Présentée et soutenue le 30 juin 2024*

*Devant les jurés composés de :*

<b>Président :</b>	Pr. YAZID Mohand	Professeur
<b>Examineur :</b>	Dr. HOCINI Kenza	Maître de Conférences B
<b>Encadrante :</b>	Dr. AIT ABDELOUHAB Karima	Maître de Conférences B
<b>Co-Encadrante :</b>	Dr. AZOUI Aicha	Maître de Assistante A

## *✧ Remerciements ✧*

### ***Tout d'abord,***

*nous remercies Allah Tout Puissant, qui nous a donné la force et la patience pour mettre en œuvre ce projet.*

### ***Nous exprimons notre profonde gratitude***

*envers notre encadrante, Mme AIT ABDELOUHAB Karima, MCB à l'université de BEJAIA, ainsi qu'à notre co-encadrante, Mme AZOUI Aicha, MAA à l'université de BEJAIA, pour leur accompagnement bienveillant et leur soutien tout au long de la réalisation de ce mémoire.*

*Leur présence constante, leur écoute attentive, leur patience, leur enthousiasme, leur confiance et leurs précieuses orientations ont enrichi chaque étape de ce travail. Sans leur contribution, ce projet n'aurait jamais vu le jour.*

### ***En second lieu,***

*nous remercions les membres du jury, Mr. YAZID et Mme K. HOCINI, pour avoir consacré du temps à l'examen attentif de notre mémoire et pour l'honneur qu'ils nous font en participant à l'évaluation de ce travail.*

### ***Nous tenons également à remercier***

*chaleureusement toutes les personnes qui ont contribué, de près ou de loin, à l'aboutissement de ce mémoire.*

*Leur engagement et leurs conseils ont été une source d'inspiration et ont enrichi notre parcours de recherche de manière significative.*

***Sid Ahmed & Sonia***

## *\* Dédicaces \**

*Je remercie Allah de m'avoir donné le courage pour accomplir ce modeste travail*

*que je dédie : À mon père et ma mère, Je souhaite exprimer ma profonde gratitude envers mes parents pour leur amour, leur soutien inconditionnel et leurs sacrifices constants.*

*Votre encouragement a été une source d'inspiration pour moi. A mon frère Sid Ali à qui Je souhaite du succès dans sa vie personnelle et académique.*

*A mon grand père et ma grand-mère.*

*A mon oncle Hamza, sa femme et ses filles Yasmine, Lydia et Dina.*

*A mon oncle abdelmalek et sa fille Khadidja.*

*A mes oncles mounir et Walid.*

*A mes tantes Souad, Halima, Miled, Djamila, Nadia et Sonia .*

*A toute la famille.*

*A une personne très chère pour moi qui m'a encouragé toujours et m'a donné la volonté et le courage pour arriver à ce jour-là je la remercie infiniment pour son soutien.*

*Je tiens également à remercier mes chers amis : Rahim, Chamsou, Abderezzak, Haider, Aghiles, Issam, Rabah, Ilyes-Zoo, Akram, Abderrahmane, Youcef, Zaki, Yazid, Adel et toutes mes amis de promos SIA et IA . Votre présence, vos encouragements et vos moments de détente ont rendu cette aventure encore plus mémorable*

***Mr. Sid Ahmed***

## *\* Dédicaces \**

*Derrière chaque réussite, il y a des personnes qui nous entourent, nous soutiennent et nous aident à atteindre nos objectifs, de près ou de loin.*

*Je tiens à exprimer ma profonde gratitude envers chacune d'elles.*

*À mes chers parents, guides bienveillants dans ce voyage qu'est la vie, dont l'amour infini a été ma boussole et ma force dans les moments de doute. Vos sacrifices ont été le fondement de ma réussite.*

*Je souhaite exprimer ma profonde gratitude envers mon frère Bachir, qui a été ma force. Ma sœur Yasmine, ta détermination m'a été une source d'inspiration pour avancer et ma petite ange sœur Asma, dont la joie de vivre et la bienveillance ont illuminé mes journées, apportant une touche de bonheur à chaque instant. Je vous remercie pour votre présence tout au long de mon parcours.*

*Je leur suis infiniment reconnaissant pour tout l'amour, le soutien et la positivité qu'ils m'ont apportés.*

*Ensuite, un immense merci à tous mes amis et mes copines, en particulier à Zahia, qui est bien plus qu'une amie, qui a été ma sœur de cœur à travers les hauts et les bas.*

*Leurs mots d'encouragement et leur présence dans les moments difficiles ont illuminé mon chemin.*

*À ma famille, je dédie ce mémoire avec un immense sentiment de gratitude et d'amour éternel. Votre soutien inconditionnel a été le moteur de mes réalisations et la raison de ma persévérance. Que ces mots transmettent toute l'admiration et la reconnaissance que je porte à chacun de vous.*

**Mlle Sonia**

# Table des matières

Table des matières	i
Liste des figures	v
Liste des tableaux	vii
Liste des abréviations	viii
Introduction générale	1
<b>1 Introduction à la révolution technologique : IA, IoT et Cloud</b>	<b>3</b>
1.1 Introduction . . . . .	3
1.2 Internet des objets . . . . .	4
1.2.1 Définition . . . . .	4
1.2.2 Technologie de l'IOT . . . . .	5
1.2.2.1 Capteurs ou Dispositifs Connectés . . . . .	5
1.2.2.2 Réseaux de communication IoT . . . . .	5
1.2.2.3 Plateformes de traitement de données . . . . .	6
1.2.3 Défi de L'IOT . . . . .	6
1.3 Cloud Computing . . . . .	8
1.3.1 Définition . . . . .	8
1.3.2 Modèles de service du cloud computing . . . . .	9
1.3.2.1 IaaS (Infrastructure as a Service) . . . . .	9
1.3.2.2 PaaS (Platform as a Service) . . . . .	9
1.3.2.3 SaaS (Software as a Service) . . . . .	10
1.3.3 Modèles de déploiement du cloud computing . . . . .	10
1.3.3.1 Cloud privé . . . . .	10
1.3.3.2 Cloud public . . . . .	11
1.3.3.3 Cloud communautaire . . . . .	11
1.3.3.4 Cloud hybride . . . . .	11
1.3.4 défis du cloud . . . . .	12
1.3.5 Aventure du Cloud . . . . .	12

1.4	Intelligence artificielle . . . . .	13
1.4.1	Définition . . . . .	13
1.4.2	Machine learning . . . . .	14
1.4.2.1	Définition . . . . .	14
1.4.2.2	Types de machine learning . . . . .	15
1.4.3	Différence entre le Machine Learning et le Deep Learning . . . . .	18
1.4.4	Domaines d'applications de l'intelligence artificielle . . . . .	20
1.4.4.1	Soins de santé . . . . .	20
1.4.4.2	Banques . . . . .	20
1.4.4.3	E-Commerce . . . . .	20
1.4.4.4	Éducation . . . . .	20
1.4.4.5	Agriculture . . . . .	21
1.4.5	Avantages et inconvénients de l'IA . . . . .	21
1.4.5.1	Avantages . . . . .	21
1.4.5.2	Inconvénients . . . . .	22
1.5	Combinaison IA, Cloud et IoT . . . . .	22
1.6	Conclusion . . . . .	23

<b>2</b>	<b>IoT et IA dans les soins de santé pour le suivi des patients atteints d'une maladie cardiovasculaire : État de l'art</b>	<b>24</b>
2.1	Introduction . . . . .	24
2.2	Compréhension des maladies cardiovasculaires . . . . .	24
2.2.1	Définition . . . . .	24
2.2.2	Types et symptômes des maladies cardiovasculaires . . . . .	25
2.2.2.1	Maladies des artères coronaires . . . . .	25
2.2.2.2	Maladies cérébrovasculaires . . . . .	26
2.2.2.3	Maladies cardiaques Congénitales . . . . .	26
2.2.2.4	Maladie artérielle périphérique . . . . .	27
2.2.3	causes des maladies cardiovasculaires . . . . .	28
2.2.3.1	Facteur de risques modifiables des maladies cardiaques . . . . .	28
2.2.3.2	Facteurs de risque non modifiables des maladies cardiaques . . . . .	30
2.2.4	Statistiques sur les personnes atteintes de maladies cardiovasculaires dans le monde . . . . .	31
2.3	Travaux sur l'utilisation des techniques de machine learning supervisé pour le suivi des personnes atteintes des maladies cardiovasculaires . . . . .	32
2.3.1	Amélioration de la prédiction des maladies cardiovasculaires basée sur les concepts d'IA et d'IoT . . . . .	32
2.3.2	Soins de santé numériques basés sur l'IA intégrée . . . . .	33

2.3.3	Un cadre basé sur l’IoT pour détecter les pathologies cardiaques à l’aide de l’apprentissage automatique . . . . .	35
2.3.4	Système de prédiction des maladies cardiovasculaires basé sur l’Internet des Objets (IoT) et l’apprentissage automatique . . . . .	36
2.4	Tableau comparatif des travaux existant : . . . . .	37
2.5	Synthèse . . . . .	39
2.6	Conclusion . . . . .	39

<b>3</b>	<b>Architecture proposée et implémentation des modèles de Machine Learning</b>	<b>40</b>
3.1	Introduction . . . . .	40
3.2	Architecture proposée . . . . .	40
3.2.1	Phase 1 : Collecte des données . . . . .	41
3.2.2	Phase 2 : Stockage et traitement des données (Cloud) . . . . .	42
3.2.3	Phase 3 : Visualisation des données . . . . .	42
3.3	Outils de développement . . . . .	43
3.3.1	Environnement de développement . . . . .	43
3.3.1.1	Google Golab . . . . .	43
3.3.1.2	Langage python . . . . .	44
3.3.1.3	Le python et l’IA . . . . .	44
3.3.2	Les bibliothèques utilisées . . . . .	44
3.3.2.1	Numpy . . . . .	44
3.3.2.2	Pandas . . . . .	44
3.3.2.3	Matplotlib . . . . .	45
3.3.2.4	Seaborn . . . . .	45
3.3.2.5	Scikit-learn . . . . .	45
3.4	Dataset . . . . .	45
3.4.1	Description du dataset . . . . .	45
3.4.2	Description des attributs du dataset : . . . . .	46
3.5	Prétraitement des données . . . . .	47
3.5.1	Vérifier les valeurs manquantes . . . . .	47
3.5.2	Normalisation de données . . . . .	48
3.5.2.1	Définition de StandardScaler . . . . .	48
3.5.3	Visualisation des données . . . . .	49
3.5.3.1	Fréquence de la cardiologie sexuelle : . . . . .	49
3.5.3.2	Fréquence des maladies cardiaques par type de douleur thoracique . . . . .	50
3.5.3.3	Visualisation des corrélations entre les variables . . . . .	51
3.6	Étapes de création des modèles de machine learning . . . . .	51

3.6.1	Séparation des ensembles de données . . . . .	51
3.6.2	Algorithme de ML utilisés . . . . .	52
3.6.2.1	Algorithme des K plus proches voisins(KNN) . . . . .	52
3.6.2.2	Algorithme de Régression logistique . . . . .	53
3.6.2.3	Les machines à vecteurs de support (SVM) . . . . .	53
3.6.2.4	Arbres de décisions (Decision Tree) . . . . .	54
3.6.2.5	Random Forest . . . . .	55
3.6.2.6	GridSearchCv . . . . .	56
3.6.3	Évaluation des résultats . . . . .	57
3.6.3.1	Sans GridSearchCv . . . . .	57
3.6.3.2	Avec GridSearchCv . . . . .	62
3.7	Interprétation des résultats et discussion . . . . .	66
3.7.1	Sans l'utilisation de GridSearchCV . . . . .	66
3.7.1.1	Comparaison des Performances des Algorithmes . . . . .	66
3.7.1.2	Taux d'Erreur . . . . .	66
3.7.1.3	Temps d'Exécution . . . . .	67
3.7.1.4	Évaluation Globale . . . . .	67
3.7.2	Avec l'utilisation de GridSearchCv . . . . .	67
3.7.2.1	Comparaison des Performances des Algorithmes . . . . .	67
3.7.2.2	Taux d'Erreur . . . . .	67
3.7.2.3	Temps d'Exécution . . . . .	68
3.7.2.4	Évaluation Globale . . . . .	68
3.7.3	Prédictions Des résultats . . . . .	68
3.7.3.1	Prédiction sans utiliser GridSearchCv . . . . .	68
3.7.3.2	Prédiction avec l'utilisation de GridSearchCv . . . . .	70
3.8	Conclusion . . . . .	71
	<b>Conclusion et perspectives</b>	<b>72</b>
	<b>Bibliographie</b>	<b>74</b>

# Table des figures

1.1	les 10 top des technologies de l'IOT[31]. . . . .	5
1.2	Concepts du cloud computing [32]. . . . .	9
1.3	Modèles de services cloud [33]. . . . .	10
1.4	Modèles de déploiement du cloud computing [34]. . . . .	11
1.5	Représentation complète des sous-ensembles de l'IA. . . . .	14
1.6	L'apprentissage supervisé. . . . .	15
1.7	L'apprentissage non supervisé. . . . .	16
1.8	L'apprentissage par renforcement. . . . .	17
1.9	Les Types de machine learning. . . . .	17
1.10	Domaines d'applications de l'intelligence artificielle [35]. . . . .	21
2.1	Maladies des artères coronaires (CAD) [38]. . . . .	25
2.2	Maladies cérébrovasculaires [39]. . . . .	26
2.3	Maladies cardiaques Congénitales [40]. . . . .	27
2.4	Maladie artérielle périphérique [41]. . . . .	28
2.5	Facteur de risques modifiables des maladies cardiaques [42]. . . . .	30
2.6	Facteurs de risque non modifiables des maladies cardiaques [43]. . . . .	31
2.7	Statistique des Maladies Cardiovasculaires Mondiale en 2021 [44]. . . . .	32
2.8	L'approche utilisée pour améliorer la prédiction des maladies cardiovasculaires [46]. . . . .	33
2.9	Architecture du système proposé [47]. . . . .	34
2.10	Méthodologie de développement des modules de prédiction de l'état de santé [47]. . . . .	34
2.11	Organigramme du modèle de détection des crises cardiaques [48]. . . . .	35
2.12	Organigramme du modèle de détection des crises cardiaques proposé (HADM)[48]. . . . .	36
2.13	diagramme de la méthodologie proposée [49]. . . . .	37
3.1	Architecture globale du système. . . . .	41
3.2	logo de Google Colab. . . . .	43
3.3	Extrait du dataset Cleveland. . . . .	46
3.4	Vérification des valeurs manquantes. . . . .	47

---

3.5	avant la normalisation. . . . .	48
3.6	après la normalisation. . . . .	49
3.7	Visualisation de Fréquence de la cardiologie sexuelle. . . . .	50
3.8	Visualisation de Fréquence des maladies cardiaques par type de douleur thoracique. . . . .	50
3.9	Matrice de corrélation. . . . .	51
3.10	Division des Ensembles de Données. . . . .	52
3.11	Classification avec l’algorithme KNN [60]. . . . .	53
3.12	Fonctionnement de la Régression Logistique [62]. . . . .	53
3.13	Illustration du principe de SVM [64]. . . . .	54
3.14	Exemple d’Arbre de décision pour évaluer les risques de crise cardiaque [66].	55
3.15	Schéma de fonctionnement de Random Forest [67]. . . . .	56
3.16	Évaluation des performances des modèles. . . . .	59
3.17	Évaluation des métriques des modèles. . . . .	61
3.18	Comparaison du temps d’exécution des modèles utilisés. . . . .	62
3.19	Implémentation de GridSearchCv. . . . .	63
3.20	Comparaison des métriques de performance avec GridSearchCv. . . . .	64
3.21	Analyse des erreurs des modèles suite à l’utilisation de GridSearchCv. . . . .	65
3.22	Analyse de la durée d’exécution des modèles employés. . . . .	66
3.23	Prédiction avec le KNN sans utiliser GridSearchCV. . . . .	69
3.24	Prédiction avec le RandomForest avec l’utilisation de GridSearchCV. . . . .	70

# Liste des tableaux

1.1	table comparative entre machine learning et deep learning [27]. . . . .	19
2.1	table comparative de quelques approches des travaux précédents. . . . .	38
3.1	Comparaison des Performances de Divers modèles de Classification . . . . .	58
3.2	Évaluation des performances des modèles basée sur les métriques d'erreur. . .	60
3.3	Évaluation des performances des modèles Selon le temps d'exécution. . . . .	61
3.4	Comparaison des Performances de Divers modèles de Classification. . . . .	63
3.5	Analyse comparative des modèles selon les métriques d'erreurs. . . . .	64
3.6	Évaluation des résultats des modèles En fonction de la durée d'exécution. . .	65

# Liste des abréviations

<b>Adam</b>	Adaptive Moment Estimation Algorithm
<b>AIT</b>	Attaque Ischémique Transitoire
<b>AVC</b>	Accident Vasculaire Cérébral
<b>AWS</b>	Amazon Web Services
<b>CAD</b>	Maladie Cardiaque Coronaryenne (Coronary Artery Disease)
<b>CERP-IoT</b>	Cluster of European Research Projects on IoT
<b>CNN</b>	Réseau de Neurones Convolutifs (Convolutional Neural Network)
<b>DL</b>	Deep Learning (Apprentissage profond)
<b>DNN</b>	Deep Neural Network (Réseau de Neurones Profonds)
<b>DT</b>	Decision Tree (Arbre de Décision)
<b>ECG</b>	Électrocardiogramme
<b>GridSearchCV</b>	Grid Search Cross-Validation
<b>GUI</b>	Graphical User Interface
<b>HADM</b>	Modèle de détection des crises cardiaques
<b>HDL</b>	High-Density Lipoprotein (Bon Cholestérol)
<b>IA</b>	Intelligence Artificielle
<b>IaaS</b>	Infrastructure as a Service
<b>IdO</b>	Internet des Objets
<b>IEEE</b>	Institute of Electrical and Electronics Engineers
<b>IMC</b>	Indice de Masse Corporelle
<b>IoMT</b>	Internet of Medical Things (Internet des Objets Médicaux)
<b>IoT</b>	Internet des Objets (Internet of Things)

---

<b>KNN</b>	K-Nearest Neighbors (Algorithme des K plus proches voisins)
<b>LDL</b>	Low-Density Lipoprotein (Mauvais Cholestérol)
<b>LPWAN</b>	Low-Power Wide Area Network
<b>LSTM</b>	Long Short-Term Memory
<b>MAE</b>	Mean Absolute Error (Erreur Moyenne Absolue)
<b>MAP</b>	Maladie Artérielle Périphérique
<b>MCV</b>	Maladies Cardiovasculaires
<b>ML</b>	Machine Learning (Apprentissage automatique)
<b>MSE</b>	Mean Squared Error (Erreur Quadratique Moyenne)
<b>PaaS</b>	Platform as a Service
<b>R2 Score</b>	Coefficient de Détermination (R Square)
<b>RF</b>	Random Forest (Forêt Aléatoire)
<b>RL</b>	Apprentissage par Renforcement (Reinforcement Learning)
<b>RMS</b>	Root Mean Square
<b>RNN</b>	Réseau de Neurones Récurents (Recurrent Neural Network)
<b>SaaS</b>	Software as a Service
<b>SSL</b>	Apprentissage Semi-Supervisé (Semi-Supervised Learning)
<b>SVC</b>	Support Vector Classifier
<b>SVM</b>	Support Vector Machine (Machine à Vecteurs de Support)
<b>UIT</b>	Union Internationale des Télécommunications
<b>XGBoost</b>	eXtreme Gradient Boosting

# Introduction générale

Le monde est témoin d'une révolution technologique majeure qui touche divers domaines. L'intelligence artificielle, et le Machine Learning et L'internet des objets transforment notre quotidien et les secteurs industriels. Leur interconnexion ouvre des opportunités inédites pour optimiser les systèmes existants et améliorer la qualité de vie. À l'ère de l'IA et grâce aux avancées des technologies informatiques, il est désormais possible d'exploiter de vastes quantités de données pour créer des modèles d'apprentissage automatique et d'apprentissage profond, facilitant ainsi la prédiction, la classification et la prise de décision. L'intégration de ces technologies à l'Internet des objets a conduit à des développements importants dans divers domaines de la vie. Surtout dans le domaine de la santé. Cependant, la santé reste un secteur particulièrement impacté par ces technologies, offrant de nouvelles perspectives pour le suivi, la prévention et la gestion des maladies chroniques.

Les problèmes de santé cardiaque constituent un enjeu majeur pour la santé publique à l'échelle internationale. Avec un nombre croissant et un impact important sur la morbidité et la mortalité, il est essentiel de mettre au point des systèmes novateurs pour leur gestion. C'est à cet égard que l'IoT et l'IA se révèlent pertinents, en offrant un suivi continu et une analyse approfondie des données de santé des patients.

Notre travail se concentre sur l'application des technologies de l'IA et de l'IoT pour la surveillance des personnes atteintes de maladies cardiaques, répondant ainsi à un besoin urgent de suivi continu de ces patients, dont le nombre de décès ne cesse d'augmenter quotidiennement. L'intégration de l'IA avec l'IoT est envisagée comme une solution efficace pour améliorer la surveillance des patients cardiaques en analysant leurs données de manière continue et précise.

Le travail que nous avons réalisé est structuré en trois chapitres principaux, chacun traitant de sujets particuliers de recherche et d'application des technologies étudiées.

1. **Révolution Technologique :IA, Machine Learning et Internet des Objets :** Ce chapitre introduit les concepts fondamentaux de l'IA, du Machine Learning et de l'IoT, expliquant comment ces technologies interconnectées ont transformé divers aspects de la vie moderne.

2. **IoT et IA dans les Soins de Santé** :Ce chapitre examine l'état de l'art concernant l'application de l'IoT et de l'IA dans le domaine médical, en particulier pour le suivi des maladies cardiovasculaires. Il explore les défis et les opportunités offerts par ces technologies pour améliorer la détection précoce et la gestion des maladies cardiaques.
3. **Architecture Proposée et Implémentation des Modèles de Machine Learning** :Ce dernier chapitre détaille l'architecture proposée pour la surveillance des patients cardiaques, incluant la collecte, le stockage et le traitement des données ainsi que la visualisation des résultats. Il présente également les outils de développement utilisés et les algorithmes de machine learning implémentés.

Enfin, Dans notre mémoire, nous proposons une approche innovante pour améliorer la gestion des maladies cardiovasculaires en combinant l'IA, le ML et l'IoT. Les résultats obtenus montrent que ces technologies ont le potentiel de proposer un suivi sur mesure et précis, ce qui ouvre la voie à des soins de santé plus efficaces et proactifs.

# Introduction à la révolution technologique : IA, IoT et Cloud

## 1.1 Introduction

Le monde connaît une révolution technologique massive qui entraîne des changements radicaux dans divers aspects de la vie. Cette révolution est menée par trois techniques principales : Intelligence artificielle, Machine Learning et l'internet des objets.

Ces techniques forment un système interconnecté et se renforçant mutuellement. Les algorithmes d'IA ont permis aux systèmes d'internet des objets de fonctionner efficacement et l'apprentissage automatique a permis à l'IA d'extraire des connaissances et de prendre des décisions en analysant les méga données collectées par les technologies IoT.

Dans ce chapitre, nous explorerons en profondeur l'Internet des Objets et le Cloud Computing, avant d'approfondir notre compréhension de l'Intelligence Artificielle. Nous débuterons par une analyse détaillée de l'IoT, décrivant son concept et ses technologies associées, suivie d'une exploration du Cloud Computing, avec une définition claire de ses services et modèles de déploiement. Ensuite, nous nous pencherons sur l'IA, en définissant ses fondements et ses applications, puis nous aborderons le Machine Learning et le Deep Learning, en mettant en évidence leurs distinctions essentielles.

Enfin, nous démontrons la combinaison entre les trois technologies. Pour conclure ce chapitre, nous dresserons un bref résumé des points abordés et soulignerons leur importance dans le contexte de notre étude.

## 1.2 Internet des objets

### 1.2.1 Définition

Le terme ‘Internet des Objets’(IdO ou IoT en anglais) ne fait pas encore consensus sur sa définition. Certaines insistent sur ses aspects techniques. Tandis que d’autres se concentrent plutôt sur les usages et les fonctionnalités :

- Le groupe de travail Internet of Things Global Standards Initiative (IoT-GSI), piloté par l’Union Internationale des Télécommunications (UIT), considère l’IoT comme [1] : « Une infrastructure mondiale au service de la société de l’information permettant d’offrir des services évolués en interconnectant des objets (physiques et virtuels) grâce à l’interopérabilité de technologies de l’information et de la communication existantes ou en évolution ». L’institut d’ingénieurs en électricité et électronique (Institute of Electrical and Electronics Engineers (IEEE)) définit l’IoT comme [2] : « Un réseau d’éléments chacun muni de capteurs qui sont connectés à Internet ».

- Le cluster de projets de recherche européens sur l’IoT (Cluster of European Research Projects on IoT (CERP-IoT)) définit l’Internet des objets comme [3] : « Une infrastructure dynamique d’un réseau global. Ce réseau global a des capacités d’auto-configuration basées sur des standards et des protocoles de communication interopérables. Dans ce réseau, les objets physiques et virtuels ont des identités, des attributs physiques, des personnalités virtuelles et des interfaces intelligentes, et ils sont intégrés au réseau d’une façon transparente ». L’IoT peut être divisé en 3 catégories, en fonction de l’utilisation et de la clientèle :

- L’IoT grand public comprend les appareils connectés tels que les voitures intelligentes, les téléphones, les montres, les ordinateurs portables, les appareils connectés et les systèmes de divertissement.

- L’IoT commercial comprend des éléments tels que le contrôle des stocks, le suivi des appareils et les appareils médicaux connectés.

- L’IoT industriel couvre des éléments tels que les compteurs électriques connectés, les systèmes de traitement des eaux usées, les débitmètres, les moniteurs de pipelines, les robots de fabrication et d’autres types d’appareils et de systèmes industriels connectés [3].

En résumé, l’IoT est un ensemble d’objets connectés permettant de rendre le monde réel plus intelligent.

## 1.2.2 Technologie de l'IOT

Les technologies de l'Internet des Objets comprennent un large éventail de composants et de systèmes qui permettent la connectivité, la collecte de données et le traitement des informations dans un environnement IoT (Figure 1.1). Parmi ces technologies, on trouve. [4].

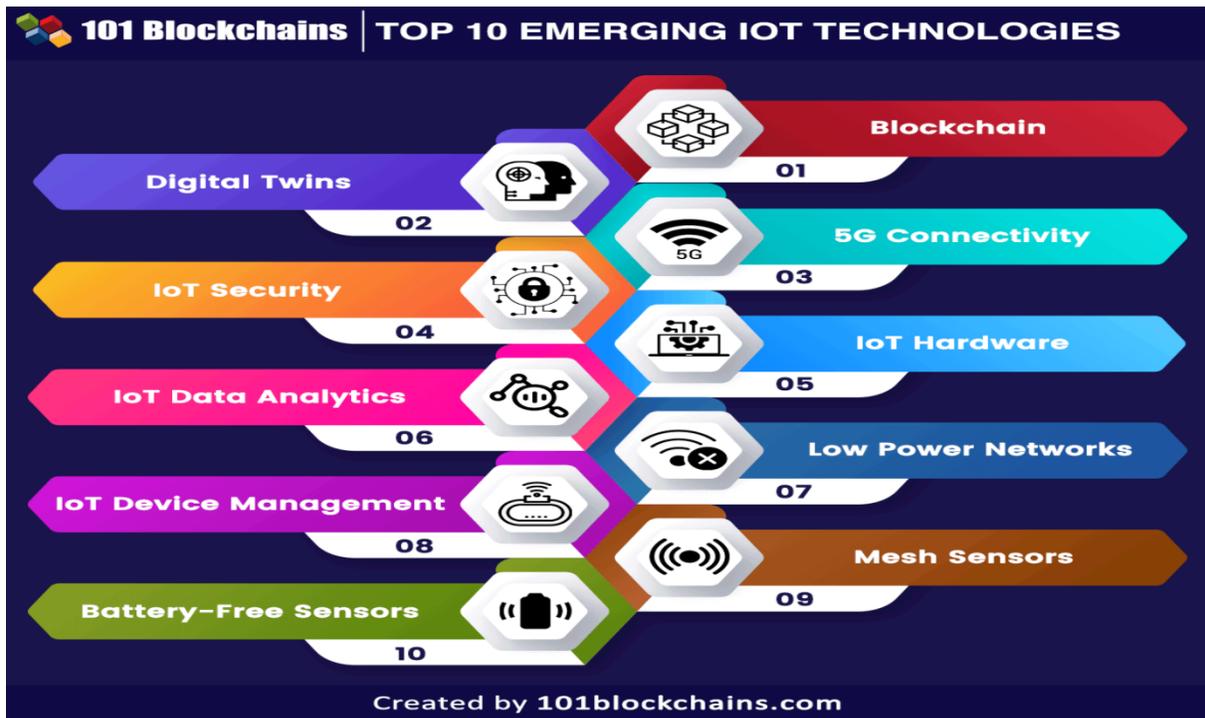


FIGURE 1.1 – les 10 top des technologies de l'IOT[31].

### 1.2.2.1 Capteurs ou Dispositifs Connectés

Les capteurs ou dispositifs connectés constituent l'élément le plus visible et crucial de l'IoT. Dotés de capteurs et d'appareils spécialisés, ces objets sont chargés de collecter des données environnementales variées. Ces données sont ensuite transmises à un serveur central, soit localement, soit via le cloud, pour être traitées et analysées. Ces capteurs peuvent mesurer une gamme étendue de variables, allant des paramètres physiques tels que la température, l'humidité et la pression, aux données plus complexes comme les images, les signatures acoustiques ou vibratoires.

### 1.2.2.2 Réseaux de communication IoT

- Les réseaux sans fil à courte distance, comme le Wi-Fi, le Bluetooth ou le Zigbee, sont faciles à installer, abordables, capables de gérer un débit de données élevé, mais leur portée est restreinte et leur consommation d'énergie est limitée.

- Les réseaux de téléphonie cellulaire, comme la 4G ou la 5G, offrent une portée accrue, sont capables de transporter un volume considérable de données, mais sont plus énergivores et fonctionnent grâce à des abonnements payants auprès de fournisseurs de connectivité.

- Les réseaux LPWAN (Low-Power Wide Area Network), comme Sigfox, le LoRaWAN ou le NB-IoT, ont été spécialement développés pour les objets connectés qui nécessitent la transmission de données sur de longues distances, tout en réduisant la consommation de la batterie. Par conséquent, ces réseaux ne supportent que des messages (ou payloads) de faible volume qui peuvent être transmis régulièrement ou suite à un événement spécifique. Il est aussi nécessaire de souscrire à un abonnement et cela reste moins coûteux que les abonnements cellulaires.

- Les réseaux cellulaires ou LPWAN privés, qui reposent sur des infrastructures supplémentaires, connues sous le nom de gateways, offrent une connectivité locale en cas de manque de couverture sur site des réseaux publics.

- Les réseaux filaires, comme l’Ethernet ou la fibre optique, est souvent excessive pour les utilisations IoT, que ce soit en termes de débit de données ou de coûts associés.

### 1.2.2.3 Plateformes de traitement de données

Les instruments de traitement et de visualisation des données sont le dernier élément de la chaîne de l’IoT ; ce sont ces instruments qui permettent aux utilisateurs de tirer pleinement parti de la valeur ajoutée des données recueillies. Il existe une multitude de choix disponibles sur le marché, avec une forte tendance à la simplicité de mise en place et de configuration, notamment grâce aux interfaces low-code ou no-code.

## 1.2.3 Défi de L’IOT

Les défis de l’IoT (Internet des Objets) sont nombreux et variés, reflétant la complexité de ce domaine en constante évolution. Voici quelques-uns des principaux défis auxquels font face les acteurs de l’IoT [5].

- **Évolutivité Spontanément** : divers nouveaux objets ou appareils intelligents se connectent au réseau. L’IoT devrait donc être capable de résoudre des problèmes tels que l’adressage, la gestion des informations et la gestion des services, et devrait également prendre en charge les environnements à petite et à grande échelle.

- Autoconfiguration : Les objets IoT doivent être programmés pour une auto-configuration adaptée à un environnement particulier sans configuration manuelle par l'utilisateur.
- Interopérabilité (hétérogénéité des appareils) Dans l'IoT : de nombreux objets intelligents sont connectés et chaque objet intelligent possède sa propre capacité de collecte d'informations, de traitement et de communication. Pour la communication et la coopération entre les objets intelligents de différents types ; ils devraient avoir une norme de communication commune.
- Complexité logicielle :Étant donné que les systèmes logiciels des objets intelligents fonctionnent avec un minimum de ressources, il est nécessaire de disposer d'une infrastructure logicielle pour prendre en charge le réseau et nécessite un serveur en arrière-plan pour gérer et prendre en charge les objets intelligents du réseau.
- Volume de stockage : En fonction du scénario et du contexte, les objets intelligents collectent soit une petite quantité de données, soit un énorme volume de données. Ainsi, en fonction de la quantité de données, le stockage doit être alloué.
- Interprétation des données : Il est très important d'interpréter le contexte, que le capteur doit détecter. Le contexte joue un rôle important pour générer des informations utiles et tirer une conclusion à partir des données envoyées par les capteurs.
- Sécurité et confidentialité personnelle Dans l'IoT : réseau constitué d'objets intelligents via Internet, assurer la sécurité et la confidentialité constitue donc un défi de taille. Dans l'IoT, l'utilisateur empêche parfois un autre utilisateur d'accéder à certaines informations particulières à un moment donné ou empêche une communication ou une transaction pour protéger les informations secrètes des concurrents. Gérer toute cette situation est donc un grand défi.
- Tolérance aux pannes dans l'IoT : les objets ou dispositifs intelligents sont dynamiques et le contexte peut changer rapidement. Mais le réseau doit encore fonctionner correctement et automatiquement pour s'adapter aux nouvelles conditions. L'IoT doit donc être structuré pour être tolérant aux pannes et robuste.
- Échange de données omniprésent grâce aux technologies sans fil : problèmes tels que la disponibilité, les retards du réseau et la congestion, etc. des technologies sans fil ; qui est utilisé pour la communication des appareils intelligents constitue un grand défi.

- Solutions optimisées sur le plan énergétique : Le réseau est constitué de nombreux appareils interconnectés ; ce qui nécessite beaucoup d'énergie pour maintenir le réseau actif. L'optimisation énergétique est donc l'aspect majeur de l'IoT.

## 1.3 Cloud Computing

### 1.3.1 Définition

Dans la littérature, il existe plusieurs définitions concernant le concept du Cloud Computing,

Les auteurs de [6] définissent le cloud computing comme un modèle qui offre aux utilisateurs du réseau un accès à la demande, à un ensemble de ressources informatiques partagées et configurables (par exemple, réseaux serveurs, stockage, applications et services) qui peuvent être rapidement mises à la disposition du client et publiées avec un effort de gestion minimal et sans interaction directe avec le prestataire de service.

Selon [11], le cloud est présenté comme une approche permettant de disposer d'applications, de puissance de calcul, de moyens de stockage, etc., en tant que services. Ceux-ci sont mutualisés, dématérialisés (indépendants de toutes contingences matérielles, logicielles et de communication), contractualisés en termes de performances, de niveau de sécurité et de coûts, et évolutifs en volume, en fonction et en caractéristiques, tout en étant accessibles en libre-service.

Le cloud computing est vu par [13] comme un réseau informatique distribué composé de serveurs virtualisés dont les ressources peuvent être approvisionnées de manière dynamique. Ce modèle repose sur un contrat de service (SLA, Service Level Agreement) entre un fournisseur et un client.

Tandis que dans [29] les auteurs considèrent le cloud computing comme une technologie qui fournit des services informatiques, stockage des données, calcul et mise en réseau à plusieurs utilisateurs à la fois, à l'emplacement et selon la quantité de travail qu'ils souhaitent consommer, avec des coûts basés uniquement sur les ressources utilisées. On peut déduire que le cloud computing n'est pas en soi une nouvelle technologie, il provient de l'aboutissement de plusieurs technologies existantes antérieurement : internet et la virtualisation, le tout appuyé sur un réseau fiable et à haut débit. La (Figure 1.2) illustre le concept du cloud computing

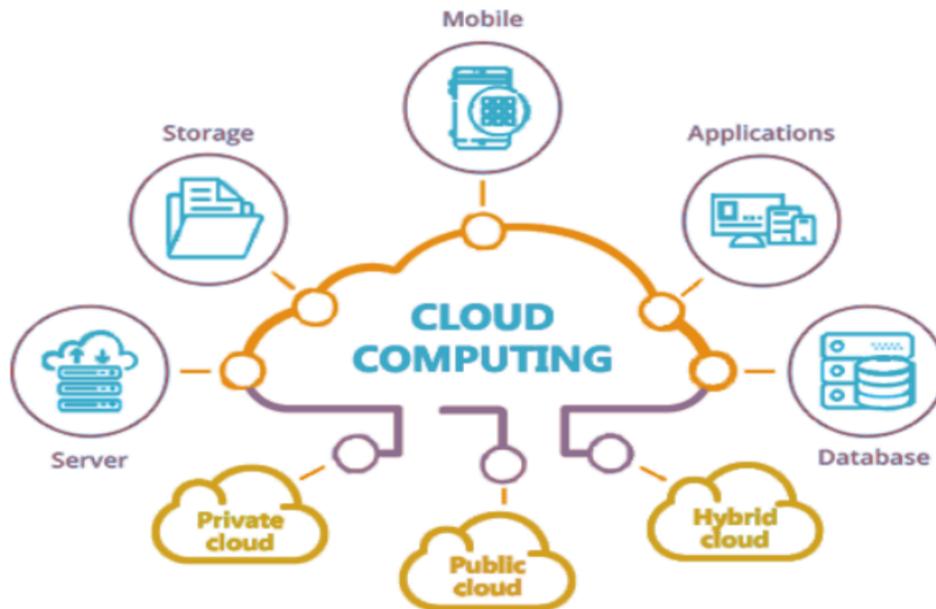


FIGURE 1.2 – Concepts du cloud computing [32].

### 1.3.2 Modèles de service du cloud computing

Différents types de services (Figure 1.3) couvrent la majorité des besoins des consommateurs, de ce fait plusieurs modèles peuvent être utilisés sur un cloud selon les ressources allouées aux clients. D’après la définition donnée dans [6] du Cloud, on peut distinguer trois modèles de services tels que :

#### 1.3.2.1 IaaS (Infrastructure as a Service)

Les services d’infrastructures (Infrastructure as a Service : IaaS) offrent des ressources informatiques générales à la demande telles que des serveurs virtualisés ou diverses formes de stockage en tant que ressources mesurées. Les clients de services achètent des ressources au lieu de configurer un ensemble de serveurs, logiciels, ou d’espaces dans les data center eux-mêmes. De plus, les utilisateurs de ce type de service sont facturés en fonction des ressources consommées[26].

#### 1.3.2.2 PaaS (Platform as a Service)

Les fournisseurs de service PaaS (platform as a service : PaaS) proposent une infrastructure logicielle gérée de niveau supérieur, où les clients peuvent construire et déployer des classes d’applications particulières et de services en utilisant des outils, environnements et langages de programmation qui sont pris en charge par le fournisseur. Le consommateur ne

gère et ne contrôle pas l'infrastructure cloud sous-jacente, y compris le réseau, les serveurs, les systèmes d'exploitation ou stockage, mais contrôle les applications déployées [6][26].

### 1.3.2.3 SaaS (Software as a Service)

Les services Applications (Software as a Service : SaaS) offrent des applications déjà créées qui s'exécutent dans une infrastructure Cloud. La plupart des services Cloud Computing en tant que logiciels sont des applications web qui peuvent accéder à travers différents périphériques clients afin d'envoyer des données et de recevoir des résultats. De plus, les clients qui utilisent ces services n'ont pas le droit de gérer ou de contrôler l'infrastructure sous-jacente du Cloud ou la plate-forme d'application[6].

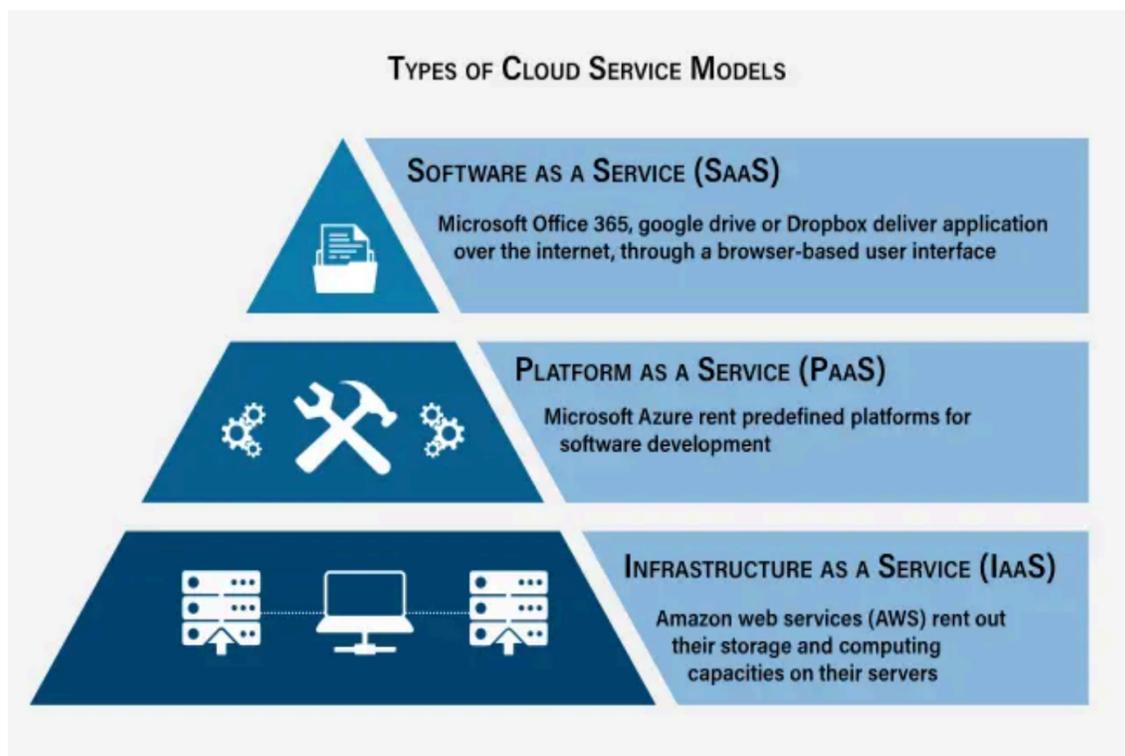


FIGURE 1.3 – Modèles de services cloud [33].

## 1.3.3 Modèles de déploiement du cloud computing

Le cloud computing offre quatre modèles de déploiement (Figure 1.4) en fonction de l'utilisation de ses services soit par des entreprises ou autre type d'utilisateurs[6].

### 1.3.3.1 Cloud privé

L'infrastructure cloud est provisionnée pour une utilisation exclusive par une seule organisation, entièrement dédiée et accessible via des réseaux sécurisés. Elle peut être hébergée

chez un tiers, mutualisée entre les différentes entités d'une même entreprise, et peut être située soit sur site, soit en dehors des locaux.

### 1.3.3.2 Cloud public

L'infrastructure cloud est externe à l'organisation, accessible via Internet, et mise à disposition pour une utilisation ouverte pour le grand public. Il peut être détenu, géré et exploité par une entreprise, un établissement universitaire ou un organisme gouvernemental, ou une combinaison de ceux-ci.

### 1.3.3.3 Cloud communautaire

L'infrastructure cloud est provisionnée pour une utilisation exclusive par une communauté spécifique de consommateurs d'organisations partageant des préoccupations communes (mission, exigences de sécurité, politique et considérations de conformité). Il peut être détenu, géré et exploité par une ou plusieurs organisations de la communauté, par un tiers ou une combinaison de ceux-ci, et il peut exister dans ou hors des locaux.

### 1.3.3.4 Cloud hybride

L'infrastructure cloud est une composition de deux ou plusieurs infrastructures cloud distinctes (privées, communautaires ou publiques) qui restent des entités uniques, mais sont liées par une technologie standardisée ou propriétaire qui permet la portabilité des données et des applications.

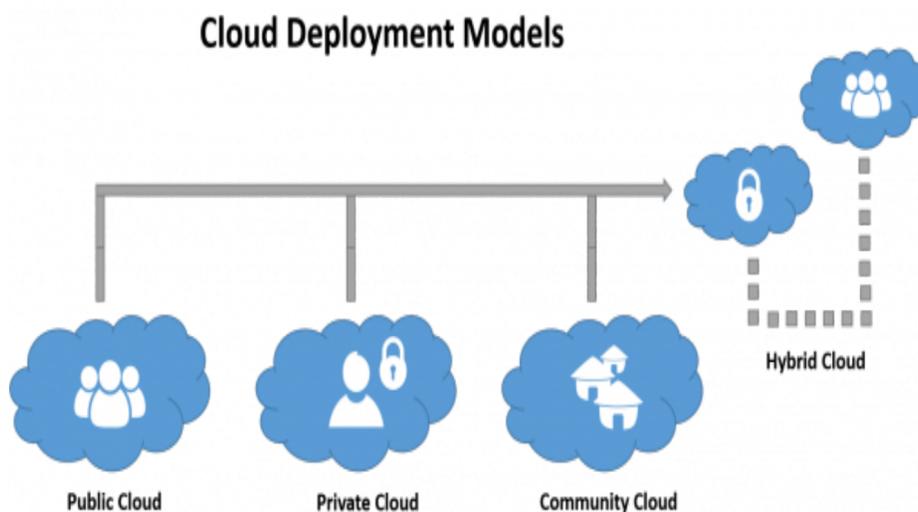


FIGURE 1.4 – Modèles de déploiement du cloud computing [34].

### 1.3.4 défis du cloud

Le Cloud présente plusieurs défis [30] :

- **Sécurité** : L'utilisation des réseaux publics, dans le cas du Cloud public, entraîne des risques liés à la sécurité du Cloud. En effet, la connexion entre les postes et les serveurs applicatifs passent par le réseau Internet, et exposent à des risques supplémentaires de cyberattaques, et de violation de confidentialité.

- **Disponibilité** : Le client d'un service de Cloud computing devient très dépendant de la qualité du réseau pour accéder à ce service. Aucun fournisseur de service Cloud ne peut garantir une disponibilité de 100

- **Piratage** : Tout comme les logiciels installés localement, les services de Cloud computing sont utilisables pour lancer des attaques (craquage de mots de passe, déni de service...). En 2009, par exemple, un cheval de Troie a utilisé illégalement un service du Cloud public d'amazone pour infecter des ordinateurs.

- **Cadre légal** : Des questions juridiques peuvent se poser, notamment par l'absence de localisation précise des données du Cloud computing.

- **Réversibilité** : En cas de rupture de contrat ou de changement de fournisseur, le client doit s'assurer de la récupération et de la destruction de ses données sur l'infrastructure du fournisseur après sa migration.

### 1.3.5 Aventure du Cloud

Cloud computing offre de nombreux avantages, notamment :

- **Évolutivité** : Le cloud permet de s'adapter rapidement aux besoins changeants en augmentant ou en diminuant les ressources selon la demande.

- **Coûts réduits** : Le cloud élimine le besoin d'investir dans du matériel coûteux et permet de payer uniquement pour les ressources utilisées, ce qui peut réduire considérablement les coûts d'infrastructure.

- **Accessibilité** : Les services cloud sont accessibles de n'importe où, à tout moment, tant que vous avez une connexion Internet, ce qui facilite le travail à distance et la collaboration.

- **Sécurité** : Les fournisseurs de cloud investissent massivement dans la sécurité pour protéger les données de leurs clients. Cela inclut le chiffrement, la surveillance continue et des protocoles de sécurité avancés.

**Fiabilité** : Les services cloud offrent généralement une haute disponibilité et des solutions de récupération après sinistre, garantissant ainsi une continuité des activités.

**Maintenance simplifiée** : Les fournisseurs de cloud gèrent la maintenance des infrastructures, ce qui permet aux entreprises de se concentrer sur leur cœur de métier sans se soucier des mises à jour et des correctifs.

**Innovation** : Les fournisseurs de cloud proposent souvent des services avancés comme l'intelligence artificielle, l'apprentissage automatique, l'analyse de données, et l'Internet des objets (IoT), permettant aux entreprises de tirer parti des dernières technologies sans avoir à les développer en interne.

**Collaboration améliorée** : Les outils et services cloud permettent une collaboration en temps réel, facilitant le travail en équipe, le partage de fichiers et la communication.

Ces avantages font du cloud computing une solution attrayante pour de nombreuses entreprises cherchant à améliorer leur efficacité, à réduire leurs coûts et à rester compétitives.

## 1.4 Intelligence artificielle

### 1.4.1 Définition

L'intelligence artificielle, ou IA, résulte de la fusion de deux termes : "artificielle" et "intelligence". "Artificielle" implique quelque chose de non naturel ou de non réel, tandis que "intelligence" fait référence à la capacité de raisonner, de générer de nouvelles idées, de percevoir et d'apprendre. Dans le domaine de l'informatique, l'IA se concentre sur la création de machines capables de fonctionner et de réagir de manière similaire à des êtres humains. Cela englobe diverses activités telles que la reconnaissance vocale, l'apprentissage, la planification et la résolution de problèmes. Un système est considéré comme intelligent lorsqu'il peut s'adapter à son environnement en fonction des situations auxquelles il est confronté [7].

Selon l'Encyclopædia Britannica, L'IA est la capacité des ordinateurs numériques ou des robots contrôlés par ordinateur à résoudre des problèmes qui sont normalement associés aux capacités de traitement intellectuel supérieures des humains[8].

Ainsi, l'IA englobe un large éventail de techniques visant à doter les machines de compétences intellectuelles semblables à celles des humains, leur permettant de prendre des décisions, de résoudre des problèmes et d'interagir avec leur environnement de manière autonome.

L'intelligence artificielle englobe différentes approches et techniques pour permettre aux machines de simuler des comportements humains ou de résoudre des problèmes de manière autonome. Voici quelques-unes des principales approches de l'IA : Apprentissage automatique

(Machine Learning), Apprentissage profond (Deep Learning). Tel que présenté dans la Figure 1.5 .

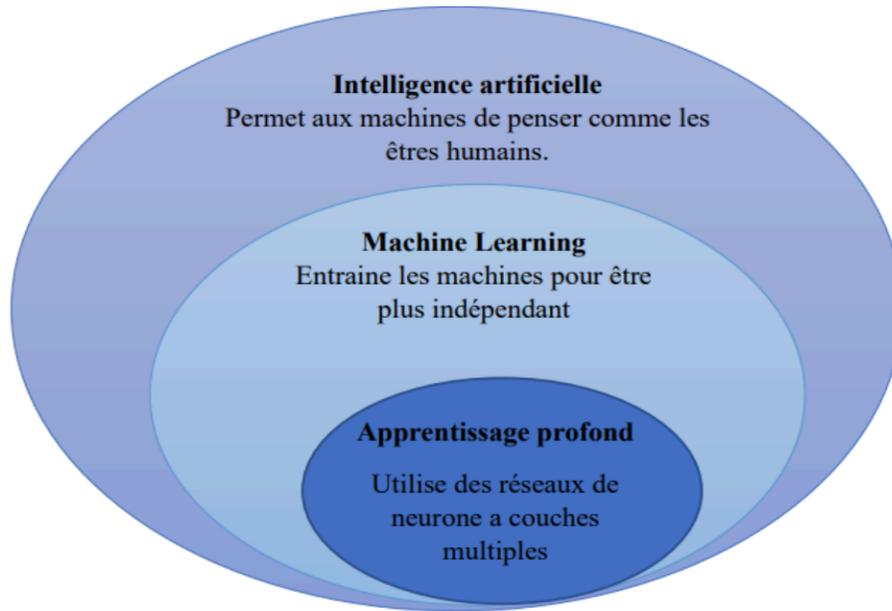


FIGURE 1.5 – Représentation complète des sous-ensembles de l'IA.

## 1.4.2 Machine learning

### 1.4.2.1 Définition

Les techniques d'apprentissage automatique (ML) permettent aux systèmes d'acquérir et d'apprendre des connaissances à partir de l'expérience et les intégrer à grande échelle. Contrairement à une programmation traditionnelle avec des instructions spécifiques, les systèmes ML apprennent de nouvelles informations par eux-mêmes, s'adaptant ainsi à leur environnement[9].

Les algorithmes d'apprentissage automatique sont largement utilisés dans divers domaines, tels que la publicité, les recommandations de produits, l'analyse d'images, le traitement du langage naturel et l'analyse du comportement des utilisateurs. Leur popularité découle de leur flexibilité et de leur efficacité dans le traitement de données complexes [10].

Ces algorithmes sont divers et peuvent être adaptés à différents types de problèmes ou ensembles de données. Ils sont utilisés pour comprendre les modèles dans les données, prédire des tendances futures et détecter des comportements anormaux, contribuant ainsi à la résolution de divers problèmes analytiques [14].

### 1.4.2.2 Types de machine learning

#### 1. Apprentissage supervisé

L'apprentissage supervisé, comme illustré dans la Figure 1.6, est une méthode d'apprentissage automatique où un modèle apprend à partir d'exemples étiquetés, c'est-à-dire des paires d'entrées-sorties. Il infère une fonction à partir des données d'entraînement étiquetées, qui consistent en un ensemble d'exemples où chaque exemple est associé à une sortie connue.

Les algorithmes d'apprentissage supervisé nécessitent une aide extérieure sous forme d'exemples étiquetés. L'ensemble de données d'entrée est généralement divisé en ensembles de données d'entraînement et de test. L'ensemble d'entraînement contient des exemples avec des variables de sortie étiquetées, que le modèle utilise pour apprendre. Une fois entraîné, le modèle est évalué sur l'ensemble de données de test pour évaluer ses performances dans la prédiction ou la classification de nouvelles données.

Parmi les algorithmes couramment utilisés en apprentissage supervisé, on trouve les arbres de décision, les machines à vecteurs de support (SVM), le classificateur Naïve Bayes, et d'autres encore. Ces algorithmes sont sélectionnés en fonction de la nature des données et de la tâche à accomplir[15].

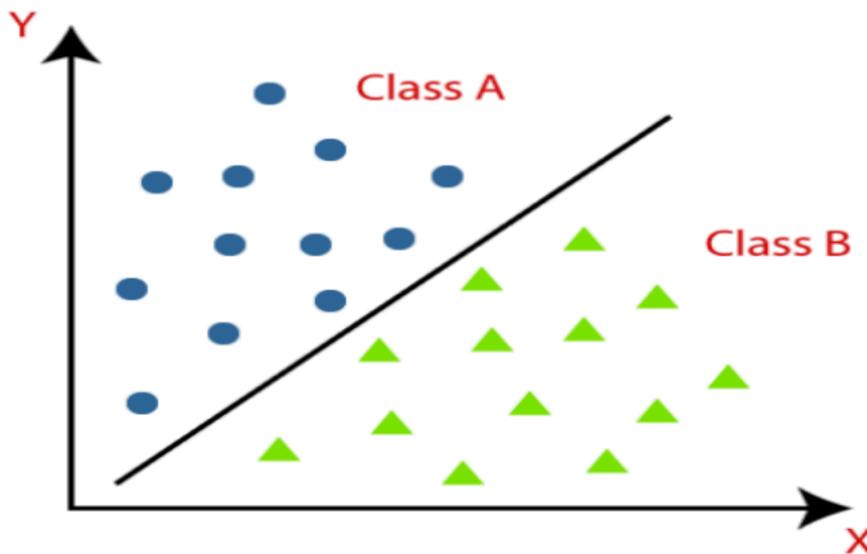


FIGURE 1.6 – L'apprentissage supervisé.

#### 2. Apprentissage non supervisé

L'apprentissage non supervisé (aussi appelé découverte des connaissances) utilise des données d'entraînement non étiquetées, non classifiées et non catégorisées (Figure 1.7). L'objectif principal de l'apprentissage non supervisé est de découvrir des patterns cachés et intéressants dans les données non étiquetées. Contrairement à l'apprentissage

supervisé, les méthodes d'apprentissage non supervisées ne peuvent pas être directement appliquées à une régression ou à un problème de classification, car on n'a aucune idée de ce que pourraient être les valeurs de la sortie. Le clustering est l'algorithme d'apprentissage non supervisé le plus couramment utilisé pour explorer l'analyse des données afin de trouver des modèles ou des regroupements cachés dans les données [16].

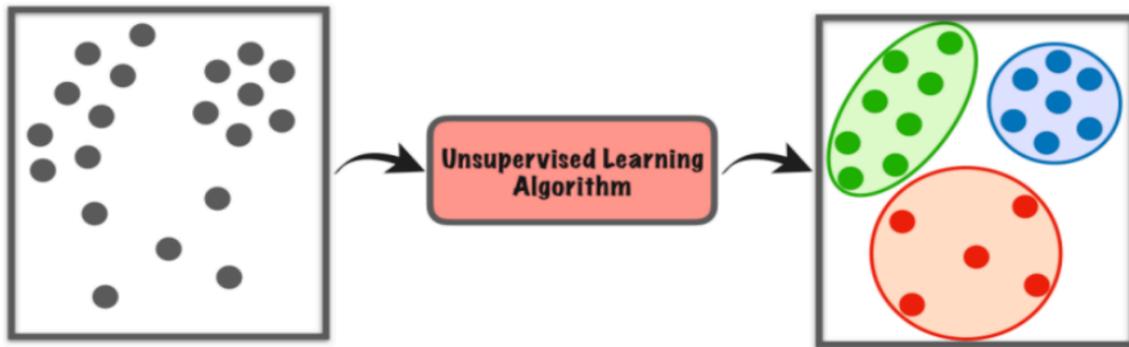


FIGURE 1.7 – L'apprentissage non supervisé.

### 3. Apprentissage semi-supervisé

L'apprentissage semi-supervisé (SSL) est une technique d'apprentissage automatique (ML) qui se situe à mi-chemin entre l'apprentissage supervisé et non supervisé.

L'objectif principal du SSL est de surmonter les inconvénients des deux méthodes. L'apprentissage supervisé nécessite une énorme quantité de données d'entraînement étiquetées pour classer les données de test, ce qui est un processus coûteux et long. En revanche, l'apprentissage non supervisé ne nécessite aucune donnée étiquetée et regroupe les données en fonction de la similitude des points de données, en utilisant des approches comme le clustering ou le maximum de vraisemblance. Le SSL utilise à la fois une petite quantité de données étiquetées et une grande quantité de données non étiquetées pour entraîner les modèles. Cela permet de réduire le coût et le temps d'étiquetage des données tout en améliorant la précision du modèle[17]

### 4. Apprentissage par renforcement

L'apprentissage par renforcement (RL) est une branche de l'apprentissage automatique où un agent apprend à interagir avec un environnement pour maximiser une récompense cumulative. L'agent prend des actions dans l'environnement, reçoit des récompenses ou des pénalités en fonction de ses actions, et ajuste son comportement en conséquence pour maximiser les récompenses au fil du temps (Figure 1.8). Cette approche repose sur l'apprentissage par essais et erreurs, où l'agent apprend une politique d'action optimale

en observant les conséquences de ses actions dans l’environnement. L’apprentissage par renforcement trouve des applications dans divers domaines, tels que les jeux, la robotique et la gestion des ressources, offrant un cadre puissant pour l’adaptation à des environnements complexes et changeants[18]

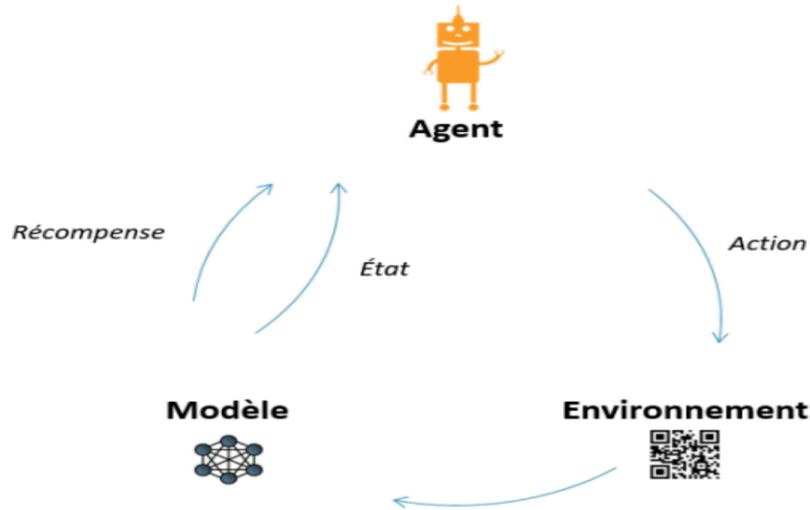


FIGURE 1.8 – L’apprentissage par renforcement.

La Figure 1.9 montre un schéma illustrant les différentes catégories d’apprentissage machine, avec quelques exemples d’algorithmes pour chaque type.

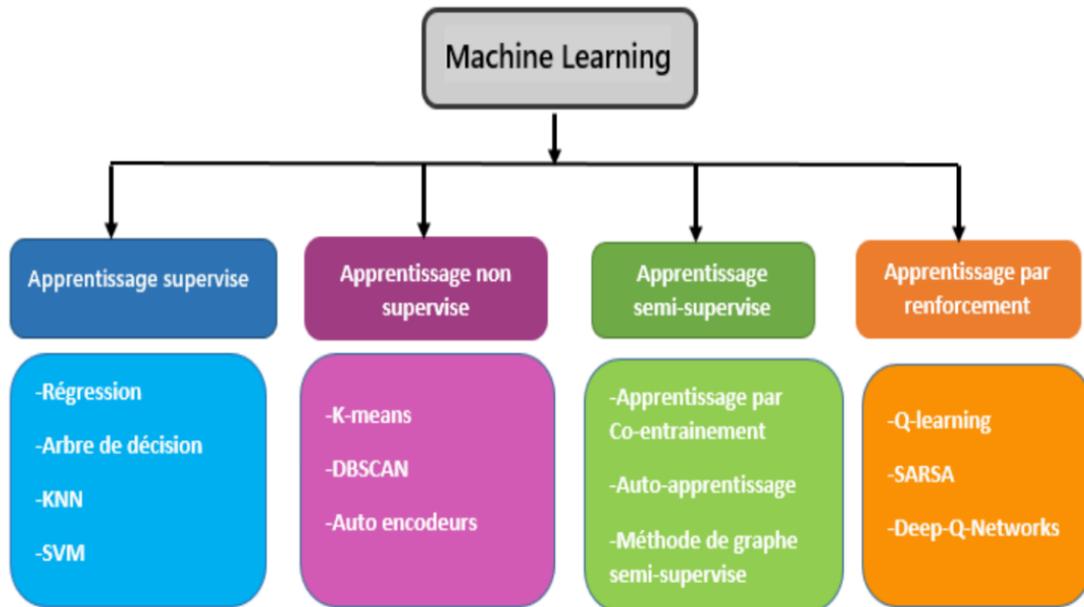


FIGURE 1.9 – Les Types de machine learning.

### 1.4.3 Différence entre le Machine Learning et le Deep Learning

L'apprentissage automatique (ML) et l'apprentissage profond (DL) sont deux sous-domaines de l'intelligence artificielle (IA) qui partagent des similitudes, mais présentent également des différences significatives.

#### Machine Learning (Apprentissage automatique)

- Le ML est une branche de l'IA qui se concentre sur le développement de modèles et d'algorithmes permettant aux ordinateurs d'apprendre à partir de données sans être explicitement programmés.
- Les méthodes de ML reposent souvent sur des statistiques et des modèles mathématiques pour détecter des schémas dans les données et prendre des décisions en fonction de ces schémas.
- Les techniques de ML incluent la régression linéaire, les arbres de décision, les machines à vecteurs de support (SVM), ainsi que des modèles de réseaux de neurones artificiels peu profonds.
- Le ML est utilisé dans une variété de domaines pour des tâches telles que la classification d'images, la prédiction de séries temporelles, et la recommandation de produits.

#### Deep Learning (Apprentissage profond)

- Le DL est une sous-discipline du ML qui se concentre sur l'utilisation de réseaux de neurones artificiels profonds pour apprendre des représentations de données hiérarchiques.
- Contrairement aux méthodes traditionnelles de ML, le DL utilise généralement des réseaux de neurones profonds composés de plusieurs couches de neurones interconnectés.
- Les réseaux de neurones profonds sont capables d'apprendre des caractéristiques complexes et abstraites des données, ce qui leur permet d'atteindre des performances exceptionnelles dans des tâches telles que la reconnaissance d'images, la traduction automatique et la synthèse de la parole.
- Les architectures de réseaux de neurones profonds populaires incluent les réseaux de neurones convolutionnels (CNN), les réseaux de neurones récurrents (RNN), ainsi que les

réseaux de neurones récurrents à mémoire à court et long terme (LSTM).

En résumé, le Machine Learning est une approche plus générale et polyvalente pour l'apprentissage à partir de données, tandis que le Deep Learning se distingue par sa capacité à apprendre des représentations complexes à partir de données non structurées à l'aide de réseaux de neurones profonds. Ces différences doivent être prises en compte lors du choix de la méthode appropriée pour une tâche spécifique en intelligence artificielle [25].

Voici un tableau comparatif entre machine learning et deep learning :

Caractéristique	DEEP LEARNING	MACHINE LEARNING
Sous domaine de L'IA	OUI	OUI
Exigences en matière de données	Nécessite généralement des données étiquettes	Peut-être étiqueté ou non
Types des algorithmes	Apprentissage principalement supervisé	Apprentissage supervisé, non supervisé et par renforcement
Structure du réseau de neurone	Réseaux profonds avec plusieurs couches	Des réseaux plus simples avec moins de couches
Applications	Reconnaissance d'images, traitement du langage naturel, reconnaissance vocale, diagnostic médical	Classification, régression, Clustering, Détection d'anomalie, traitement du langage naturel, vision par ordinateur, reconnaissance vocale
Forces	Peut apprendre de modèles complexes avec les données	Plus Flexible et polyvalent
Faiblesses	Plus complexe et nécessite plus de données pour s'entraîner	Moins puissant et moins capable d'apprendre des modèles complexes

TABLEAU 1.1 – table comparative entre machine learning et deep learning [27].

## 1.4.4 Domaines d'applications de l'intelligence artificielle

L'intelligence artificielle s'applique sur plusieurs domaines, Comme on peut le voir dans la figure 13, Parmi les domaines que l'IA s'applique [24] :

### 1.4.4.1 Soins de santé

L'intelligence artificielle (IA) trouve de nombreuses applications dans les établissements de soins de santé, où elle peut être utilisée pour diverses tâches courantes. Par exemple, les systèmes d'IA peuvent analyser les données médicales personnalisées et les images radiographiques des patients pour détecter différentes maladies (comme le cancer du sein sur les mammographies). De plus, l'IA peut agir comme un assistant personnel de soins de santé, en servant de coach de vie pour rappeler aux patients de prendre leurs médicaments. En déployant des solutions innovantes et précises, l'IA contribue à améliorer la prise en charge des patients atteints de diverses maladies.

### 1.4.4.2 Banques

L'intelligence artificielle améliore la vitesse, la précision et l'efficacité des opérations humaines dans les institutions financières. Les techniques d'IA peuvent être utilisées pour détecter les transactions bancaires potentiellement frauduleuses.

### 1.4.4.3 E-Commerce

L'intelligence artificielle (IA) est largement déployée dans le commerce électronique et les secteurs financiers pour améliorer l'expérience client, optimiser la gestion de la chaîne d'approvisionnement et renforcer l'efficacité opérationnelle. L'objectif principal est de développer des méthodes de contrôle de la qualité des produits fiables tout en recherchant de nouvelles façons de servir les clients de manière efficace à moindre coût.

### 1.4.4.4 Éducation

Selon l'UNESCO, l'intelligence artificielle (IA) offre un potentiel considérable pour relever certains des plus grands défis de l'éducation contemporaine, en innovant dans les pratiques d'enseignement et d'apprentissage et en accélérant les progrès vers l'Objectif de Développement Durable 4 (ODD 4) visant à assurer une éducation de qualité pour tous. Les applications de l'IA dans l'éducation ouvrent de nouvelles perspectives pour la conception d'activités d'apprentissage productives et le développement de technologies éducatives améliorées.

### 1.4.4.5 Agriculture

L'intelligence artificielle est utilisée pour identifier les carences en nutriments et les défauts du sol, en utilisant notamment la vision par ordinateur et la robotique. Les systèmes d'IA peuvent également détecter les mauvaises herbes et aider à la récolte des cultures de manière plus efficace que les travailleurs humains. Cette utilisation de l'IA dans l'agriculture vise à améliorer la productivité, l'efficacité et la durabilité de ce secteur, tout en minimisant l'impact environnemental.



FIGURE 1.10 – Domaines d'applications de l'intelligence artificielle [35].

## 1.4.5 Avantages et inconvénients de l'IA

### 1.4.5.1 Avantages

L'intelligence artificielle offre de nombreux avantages qui peuvent améliorer notre vie quotidienne et transformer nos industries. Voici quelques-uns des principaux avantages :

- **Facilite le travail** : L'IA peut faire des tâches ennuyeuses et répétitives à notre place, ce qui nous laisse plus de temps pour des choses plus intéressantes.
- **Précision** : Les systèmes d'IA exécutent des tâches avec une grande précision, minimisant ainsi les erreurs humaines et garantissant des résultats fiables.
- **Analyse des données** : L'IA peut analyser rapidement de grandes quantités de données pour nous donner des informations utiles.

- **Personnalisation** : Personnalisation des expériences : Grâce à l'IA, nous recevons des recommandations et des suggestions personnalisées, adaptées à nos préférences individuelles, que ce soit pour choisir un film à regarder ou un produit à acheter.
- **Innovation** : L'IA nous aide à créer de nouvelles technologies passionnantes, comme les voitures autonomes et les assistants virtuels.

#### 1.4.5.2 Inconvénients

- **Risque de dépendance** : Si nous comptons trop sur l'IA, nous pourrions avoir du mal lorsque les systèmes tombent en panne ou sont piratés.
- **Impact sur l'emploi** : L'automatisation par l'IA peut entraîner la disparition de certains emplois, ce qui peut créer des difficultés économiques pour les personnes concernées.
- **Risque de partialité** : Les systèmes d'IA peuvent produire des résultats injustes s'ils sont influencés par des données préjugées, ce qui peut avoir des conséquences néfastes dans des domaines comme le recrutement ou la justice.
- **Protection des données** : L'utilisation de l'IA peut poser des problèmes de confidentialité et de sécurité des données personnelles.
- **Questions éthiques** : L'IA soulève des questions complexes sur la responsabilité et l'éthique, notamment en ce qui concerne l'utilisation des technologies autonomes telles que les véhicules sans conducteur et les drones, ce qui nécessite une réflexion approfondie et des réglementations appropriées pour assurer une utilisation responsable et éthique de ces technologies.

En résumé, l'IA offre de nombreuses opportunités passionnantes, mais il est important de comprendre ses limites et ses implications pour pouvoir l'utiliser de manière responsable et bénéfique pour tous.

## 1.5 Combinaison IA, Cloud et IoT

La combinaison de l'Internet des Objets (IoT), du Cloud et de l'Intelligence Artificielle (IA) forme un écosystème technologique puissant, capable de transformer de nombreux aspects de notre quotidien et des activités commerciales. Ensemble, ces trois domaines peuvent créer des solutions innovantes :

L'IoT basé sur le cloud connecte une variété d'appareils tels que les véhicules, les téléphones cellulaires, les capteurs et les machines de fabrication, permettant ainsi la construction de systèmes intelligents dans divers secteurs comme les réseaux, les villes, les maisons et l'industrie.

En combinant l'IA, le Cloud et l'IoT, les possibilités et les avantages potentiels peuvent être pleinement exploités. Par exemple, l'IA associée à l'apprentissage automatique peut analyser les données IoT pour anticiper les activités futures. De plus, l'IA peut être intégrée à l'apprentissage automatique pour créer des systèmes de maison intelligente avancés, tandis que l'IoT peut être utilisé avec l'IA pour étudier le comportement humain à travers divers signaux.

Dans le monde des affaires, cette combinaison révolutionne de nombreux domaines, notamment l'analyse émotionnelle, la sécurité, les véhicules intelligents, l'automatisation de la maison intelligente, la surveillance environnementale, les villes intelligentes et l'industrie des soins de santé.

Malgré les défis, cette fusion de technologies représente un investissement unique garantissant des avantages à long terme [28].

## 1.6 Conclusion

Nous avons exploré les fondements de l'Intelligence Artificielle, du Machine Learning, du Deep Learning, de l'Internet des Objets et du Cloud Computing. Ces technologies révolutionnaires ouvrent de vastes possibilités, allant de l'automatisation des tâches à l'innovation dans divers domaines. Cependant, leur utilisation soulève des défis tels que la protection des données, la dépendance technologique et les implications éthiques. Il est crucial d'aborder ces défis de manière équilibrée pour maximiser les avantages de ces avancées tout en atténuant leurs inconvénients potentiels.

En cultivant une approche équilibrée entre l'innovation et la responsabilité, nous pouvons façonner un avenir avec lequel l'IA et les technologies associées contribuent de manière positive à notre monde. Cela nous amène à considérer comment ces avancées peuvent être appliquées dans le domaine spécifique de la détection et de la prévention des maladies cardiovasculaires. Dans le chapitre suivant, nous approfondirons notre compréhension de ces maladies, en examinant leurs caractéristiques et leur impact à l'échelle mondiale et locale. Ensuite, nous explorerons comment l'Internet des Objets et l'Intelligence Artificielle peuvent être intégrés pour améliorer la détection précoce et la prévention des maladies cardiovasculaires.

# IoT et IA dans les soins de santé pour le suivi des patients atteints d'une maladie cardiovasculaire : État de l'art

## 2.1 Introduction

Les maladies cardiovasculaires sont un défi majeur pour la santé publique à l'échelle mondiale. Dans ce contexte, l'Internet des Objets (IoT) et l'Intelligence Artificielle (IA) offrent de nouvelles perspectives dans la détection et la prévention de ces pathologies. Ce chapitre se penche sur l'utilisation de l'IoT et de l'IA dans ce domaine, en mettant en lumière leurs applications et en identifiant les défis associés.

Il examine l'intégration de l'IoT et de l'IA dans le domaine médical, en se concentrant sur la santé cardiovasculaire et en explorant les opportunités et les obstacles rencontrés.

L'objectif principal est d'analyser comment l'IoT et l'IA peuvent améliorer la détection précoce, le suivi et la gestion des maladies cardiovasculaires. En se basant sur les travaux existants, ce chapitre offre un aperçu des avancées récentes et des pistes pour la recherche future.

## 2.2 Compréhension des maladies cardiovasculaires

### 2.2.1 Définition

Les maladies cardiovasculaires, également désignées sous le terme de maladies du cœur et des vaisseaux sanguins, englobent un large éventail de problèmes de santé liés au système circulatoire. Parmi ces affections, on retrouve les cardiopathies coronariennes, les maladies vasculaires cérébrales, les maladies vasculaires périphériques, les cardiomyopathies, les malformations cardiaques congénitales, et d'autres encore. Ces maladies se manifestent souvent

suite à l'accumulation de plaque dans les artères, entraînant un rétrécissement ou un blocage qui perturbe la circulation sanguine vers le cœur, le cerveau ou d'autres parties du corps. En tant que première cause de décès dans le monde, elles représentent un fardeau significatif pour la santé publique, avec près de 17,9 millions de décès chaque année, soit environ 31% de l'ensemble des décès mondiaux. Parmi ces décès, environ 85% sont attribués à des crises cardiaques et à des accidents vasculaires cérébraux[12].

## 2.2.2 Types et symptômes des maladies cardiovasculaires

Les maladies cardiovasculaires se présentent sous plusieurs formes, chacune associée à des symptômes spécifiques. Nous nous concentrerons sur quelques-uns des types les plus courants de maladies cardiovasculaires, en mettant en évidence leurs symptômes distinctifs.

### 2.2.2.1 Maladies des artères coronaires

La maladie coronarienne (CAD), ou maladie cardiaque coronarienne, est la forme la plus répandue des maladies cardiovasculaires (CVD). Elle se caractérise par le rétrécissement des vaisseaux sanguins qui alimentent le cœur en sang oxygéné, en raison du dépôt de plaque à l'intérieur des artères coronaires, phénomène connu sous le nom d'athérosclérose. Lorsque ces plaques se rompent, des caillots sanguins peuvent obstruer partiellement ou complètement le flux sanguin vers le muscle cardiaque, entraînant des symptômes tels que l'essoufflement, l'infarctus du myocarde et l'angine de poitrine. Les facteurs de risque des CAD comprennent le cholestérol élevé, le tabagisme, l'obésité, la pression artérielle élevée et le diabète. Les études montrent que l'hyperglycémie liée au diabète accélère le processus d'athérosclérose. Parmi ces facteurs, le cholestérol LDL et l'hypertension artérielle de stade 1 sont particulièrement préoccupants en termes de risque de CAD[36].

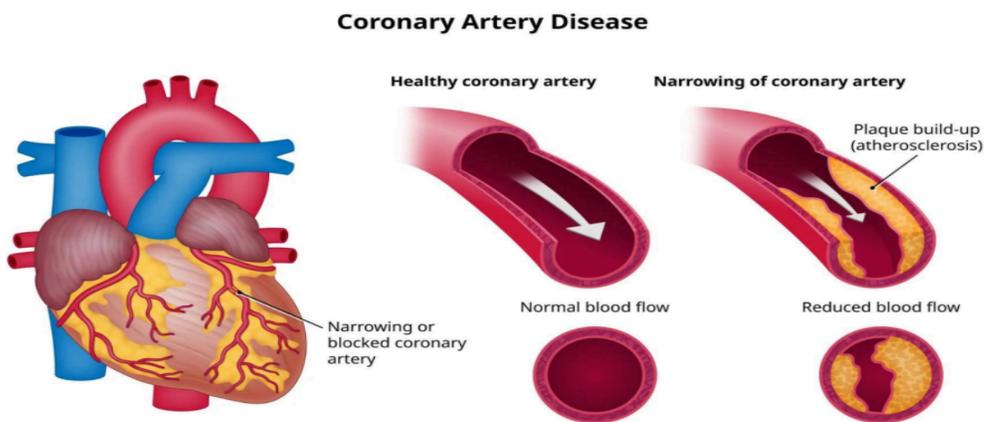


FIGURE 2.1 – Maladies des artères coronaires (CAD) [38].

### 2.2.2.2 Maladies cérébrovasculaires

Les maladies cérébrovasculaires, liées aux vaisseaux sanguins alimentant le cerveau, peuvent entraîner des accidents vasculaires cérébraux (AVC) figure(15). L'hypertension est la cause la plus fréquente de ces maladies, endommageant la paroi interne des artères et favorisant la formation de caillots sanguins. Les quatre principaux types de maladies cérébrovasculaires sont l'AVC, l'attaque ischémique transitoire (AIT), l'hémorragie sous-arachnoïdienne et la démence vasculaire. Les AVC surviennent lorsque le flux sanguin vers le cerveau est bloqué, entraînant des dommages cérébraux et des symptômes tels que la faiblesse soudaine d'un côté du corps, des difficultés d'élocution et des troubles de la vision. Les causes des AVC comprennent les arythmies, les troubles valvulaires cardiaques et les anomalies des chambres et des parois cardiaques. La fibrillation atriale, une forme d'arythmie, est une cause majeure d'AVC, favorisant la formation de caillots sanguins qui peuvent obstruer les artères cérébrales. Les AIT sont des AVC temporaires présentant des symptômes similaires. L'hémorragie sous-arachnoïdienne résulte d'une fuite de sang sur la surface du cerveau, causant des dommages tissulaires. Outre l'hypertension, l'obésité, le diabète et le tabagisme sont des facteurs de risque importants des maladies cérébrovasculaires[36].



FIGURE 2.2 – Maladies cérébrovasculaires [39].

### 2.2.2.3 Maladies cardiaques Congénitales

Les maladies cardiaques congénitales sont liées à la structure du cœur. Cette condition est le plus souvent identifiée comme des défauts de naissance chez les nouveaux nés. Les défauts peuvent varier, tels que des défauts structurels, des parois du cœur, des valves

cardiaques ou même des veines et des artères autour du cœur qui pourraient entraîner un blocage du flux sanguin, forcer le sang à circuler dans la mauvaise direction et ralentir le flux sanguin. Les symptômes de la MCV sont le plus souvent identifiés à la naissance, mais dans certains cas, les patients peuvent rester non diagnostiqués pendant longtemps, voire toute leur vie. Les symptômes courants de la MCV incluent un souffle au cœur, des membres sous-développés et un essoufflement, de la fatigue et une cyanose. Les causes des maladies cardiaques congénitales peuvent ne pas être directement identifiables. Elles pourraient être causées par différents facteurs tels que des infections durant la grossesse (rubéole), l'utilisation de certains médicaments, l'alcool et le tabac, une prédisposition génétique ou même une mauvaise nutrition. Le traitement des MCV peut dépendre de la gravité du défaut. Alors que dans certains cas, aucun traitement n'est nécessaire, certains pourraient nécessiter une chirurgie cardiaque pour réparer les défauts ou même des transplantations cardiaques[36].

### Cardiopathie congénitale *Communication interauriculaire (CIA)*

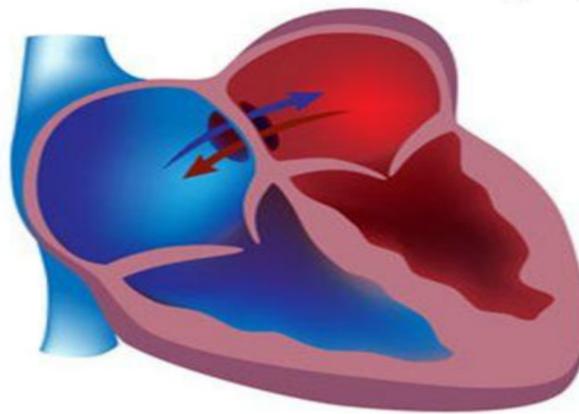


FIGURE 2.3 – Maladies cardiaques Congénitales [40].

#### 2.2.2.4 Maladie artérielle périphérique

Une condition causée par une diminution de l'apport sanguin aux membres due à l'athérosclérose (dépôts graisseux) dans les artères est appelée maladie artérielle périphérique (MAP). Cela est couramment associé aux jambes. Les symptômes communs de la MAP incluent la décoloration des jambes, des crampes dans les muscles de la hanche et du mollet, et une perte de cheveux sur les membres.

Cependant, dans de nombreux cas, ces symptômes peuvent passer inaperçus. Les facteurs de risque les plus courants de la MAP incluent l'hypertension artérielle, le tabagisme, le diabète, les lipides sanguins élevés, et des niveaux élevés d'homo cystéine. Parmi ceux-ci, le tabagisme et le diabète contribuent le plus à la MAP, car ils réduisent le flux sanguin vers les membres. La maladie artérielle périphérique pourrait entraîner des complications

supplémentaires telles que l'ischémie critique des membres où des plaies ouvertes apparaissent sur les membres et sont irrécupérables. Ces plaies peuvent provoquer la mort des tissus du membre, ce qui pourrait ultimement mener à l'amputation du membre [36].

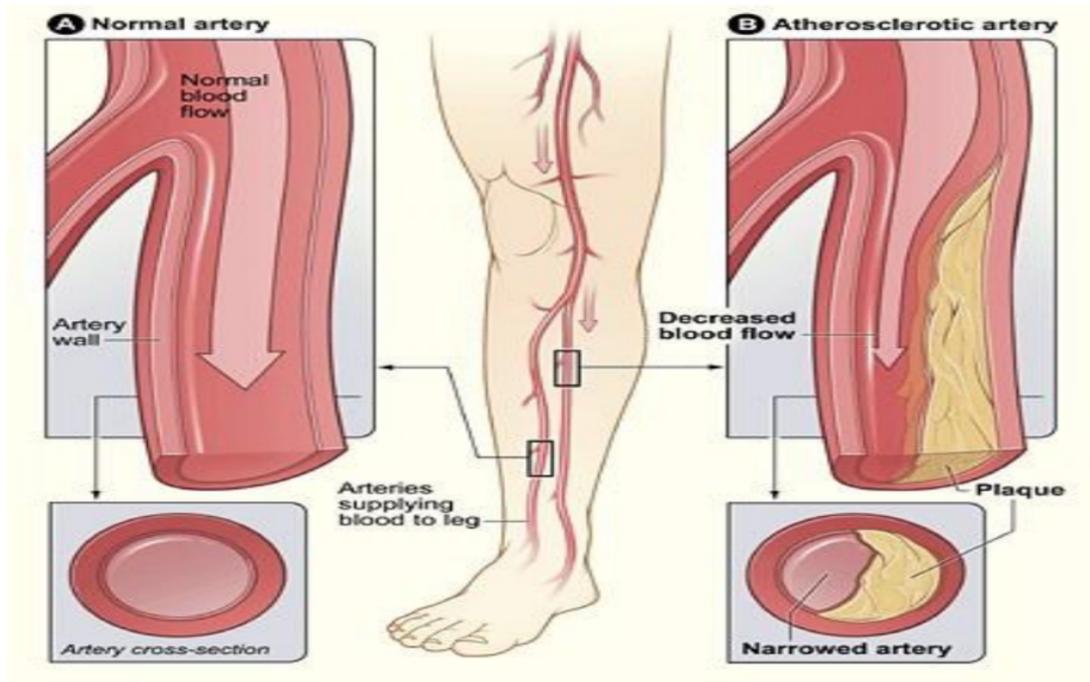


FIGURE 2.4 – Maladie artérielle périphérique [41].

### 2.2.3 causes des maladies cardiovasculaires

Voici quelques principaux éléments qui augmentent le risque de maladie cardiaque.

#### 2.2.3.1 Facteur de risques modifiables des maladies cardiaques

##### Haute Pression Artérielle

L'hypertension ou haute pression artérielle est un facteur de risque majeur pour les maladies cardiaques. Une pression sanguine élevée oblige le cœur à travailler plus pour pomper le sang et peut aussi endommager les artères.

Chez un adulte en bonne santé, une lecture normale de la pression artérielle devrait être de 120/80 mmHg ou moins. Les professionnels de la santé diagnostiquent l'hypertension si la pression artérielle est constamment supérieure à 130 mmHg systolique ou 80 mmHg diastolique.

L'hypertension est souvent silencieuse, ce qui signifie qu'elle peut ne présenter aucun symptôme ou signe d'avertissement [37].

## **Cholestérol Sanguin Élevé**

Des niveaux élevés de cholestérol sanguin sont un facteur de risque majeur pour les maladies cardiaques. Un régime alimentaire malsain riche en graisses saturées peut conduire à l'accumulation de plaques sur les parois des artères, amorçant le processus d'athérosclérose dans un vaisseau sanguin critique.

Un niveau total de cholestérol inférieur à 200 mg/dL est considéré comme souhaitable chez les adultes. Votre médecin en médecine cardiovasculaire examinera également d'autres types de cholestérol comme les triglycérides, le LDL (mauvais cholestérol) et le HDL (bon cholestérol) pour évaluer votre santé cardiaque [37].

## **Diabète**

Les problèmes cardiaques sont courants chez les personnes diabétiques. Selon l'American Heart Association, environ 65% des personnes diabétiques meurent d'une forme de maladie cardiovasculaire. Si vous avez le diabète, vous devriez être sous la surveillance d'un médecin. Contrôler vos niveaux de sucre dans le sang peut réduire votre risque de maladie cardiovasculaire [37].

## **Surpoids et Obésité**

Être en surpoids (IMC de 25 à 29,9) ou obèse (IMC de 30 ou plus) peut entraîner une haute pression artérielle, des niveaux élevés de cholestérol et du diabète, qui sont tous des facteurs de risque majeurs pour les maladies cardiaques. Vous pouvez calculer votre IMC (indice de masse corporelle) sur le site web de l'Institut national du Cœur, des Poumons et du Sang. Adopter une alimentation saine et mener une vie active peut vous aider à perdre du poids et réduire votre risque de maladies cardiaques [37].

## **Tabagisme**

Fumer des cigarettes et du tabac augmente votre risque de nombreuses maladies graves, y compris le cancer du poumon, la maladie vasculaire périphérique et les maladies cardiaques. Fumer augmente la pression artérielle et provoque le rétrécissement des artères en favorisant l'accumulation de plaque. Cela augmente également la fréquence cardiaque et peut causer un rythme cardiaque irrégulier [37].

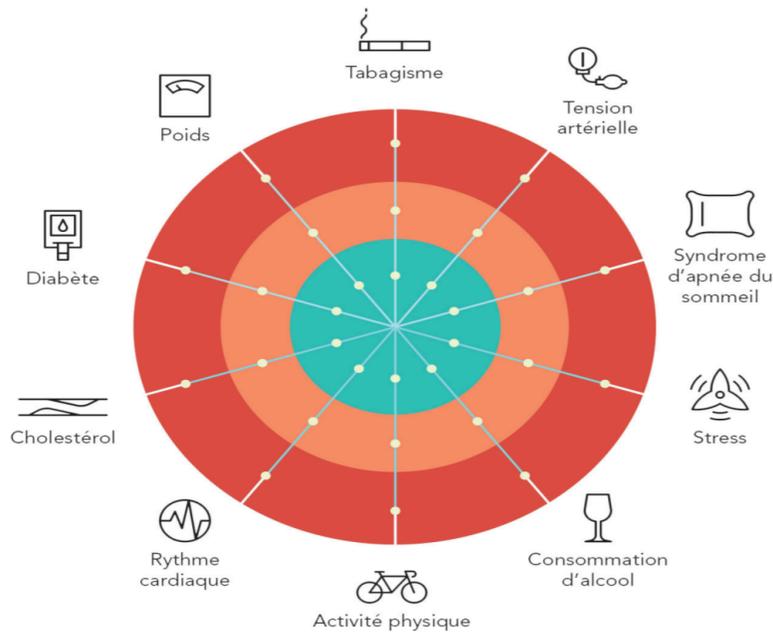


FIGURE 2.5 – Facteur de risques modifiables des maladies cardiaques [42].

### 2.2.3.2 Facteurs de risque non modifiables des maladies cardiaques

**Âge :** Le risque de maladie cardiaque augmente avec l'âge, avec environ 80% des décès liés aux maladies cardiaques survenant chez les personnes de plus de 65 ans [37].

**Genre :** Les hommes ont un risque accru de crise cardiaque par rapport aux femmes, bien que cette disparité diminue après l'âge de 65 ans. Il est cependant essentiel de souligner que les crises cardiaques sont souvent plus sévères chez les femmes que chez les hommes [37].

**Hérédité :** Les maladies cardiovasculaires ont généralement une composante héréditaire. Si vos parents ou frères et sœurs ont connu des problèmes cardiaques ou circulatoires à un jeune âge (avant 55 ans), votre risque est plus élevé que celui d'une personne sans antécédents familiaux. De nombreux facteurs de risque des maladies cardiaques, comme l'obésité, l'hypertension artérielle et le diabète, peuvent être transmis de génération en génération [37].

**Ethnicité :** Certains types de maladies cardiaques sont plus courants chez certains groupes ethniques et raciaux. Les Afro-Américains ont un risque plus élevé de maladie cardiaque que les Caucasiens [37].

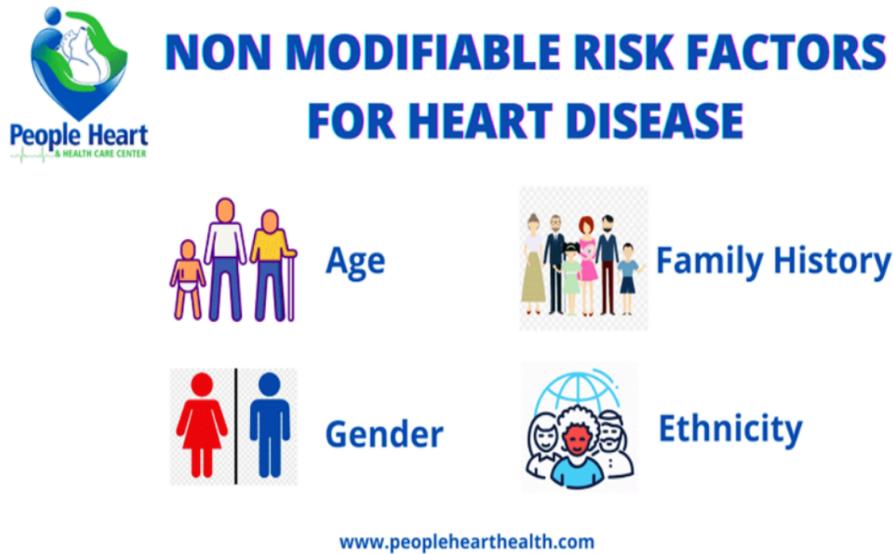


FIGURE 2.6 – Facteurs de risque non modifiables des maladies cardiaques [43].

#### 2.2.4 Statistiques sur les personnes atteintes de maladies cardiovasculaires dans le monde

- Ces maladies représentent environ un tiers des décès à l'échelle mondiale, soit environ 20,5 millions de décès estimés en 2021, soit en moyenne 56 000 personnes chaque jour, soit un décès toutes les 1,5 seconde. Elles sont les principales causes de décès dans le monde.
- En 2019, on estime qu'elles ont causé le décès d'approximativement 9,8 millions d'hommes et de 9,2 millions de femmes à l'échelle mondiale.
- La part des décès attribués à ces maladies a augmenté ; en 1990, elles représentaient un peu plus d'un quart (27 %) de tous les décès mondiaux.
- Le nombre de décès dus à ces affections devrait continuer à augmenter à l'échelle mondiale.
- Les taux de mortalité standardisés par âge ont diminué dans le monde entier en raison des améliorations de l'espérance de vie, mais cela a également entraîné une augmentation du nombre de personnes atteignant un âge où les maladies cardiaques et circulatoires sont plus courantes.

### Global Heart & Circulatory Disease Prevalence in 2021



FIGURE 2.7 – Statistique des Maladies Cardiovasculaires Mondiale en 2021 [44].

## 2.3 Travaux sur l'utilisation des techniques de machine learning supervisé pour le suivi des personnes atteintes des maladies cardiovasculaires

Plusieurs chercheurs ont fait des articles sur l'application des techniques de L'IA et L'IOT dans le domaine de la santé des maladies cardiovasculaires.

### 2.3.1 Amélioration de la prédiction des maladies cardiovasculaires basée sur les concepts d'IA et d'IoT

Selon Debasish et al. [46] Leurs études étaient sur la combinaison des technologies de l'intelligence artificielle avec l'internet des objets pour améliorer la prédiction des maladies cardiovasculaires. Cette étude a utilisé un modèle de réseaux de neurones profonds intégrant à la fois les technologies IA et IoT pour analyser les données de santé des patients en temps réel.

En passant par des étapes, de la Collection des données recueillies via des dispositifs IoT qui surveillaient divers indicateurs de santé des patients. Ensuite, le prétraitement des données collectées inclut la normalisation, la sélection des caractéristiques et l'encodage en temps réel pour garantir que les données étaient non redondantes et pertinentes.

Le modèle proposé comporte deux phases différentes. Dans la première phase, c'est que les informations en direct ont été recueillies en utilisant Arduino.

Ensuite, Le modèle DNN était entraîné sur les données prétraitées qui utilise la méthode de rétro propagation. L'algorithme DNN est utilisé pour tester les données.

Trois fonctions d'activation, Relu, Sigmoid et Tanh, sont utilisées pour tester la précision de l'algorithme. L'optimiseur utilisé pour le modèle est Adam (Adaptive Moment Estimation Algorithm) ; cet algorithme combine la descente de gradient avec la technique du momentum et la méthode RMS (Root Mean Square). L'optimiseur Adam s'adapte le mieux au modèle et produit une grande précision dans les cas de l'entraînement et de test.

Enfin, Le modèle formé était ensuite déployé pour faire des prédictions en temps réel sur la santé des patients, en utilisant de nouvelles données provenant des dispositifs IoT. Le modèle DNN s'est révélé très précis dans la prédiction des maladies cardiovasculaires, avec une précision de 95,34 %.

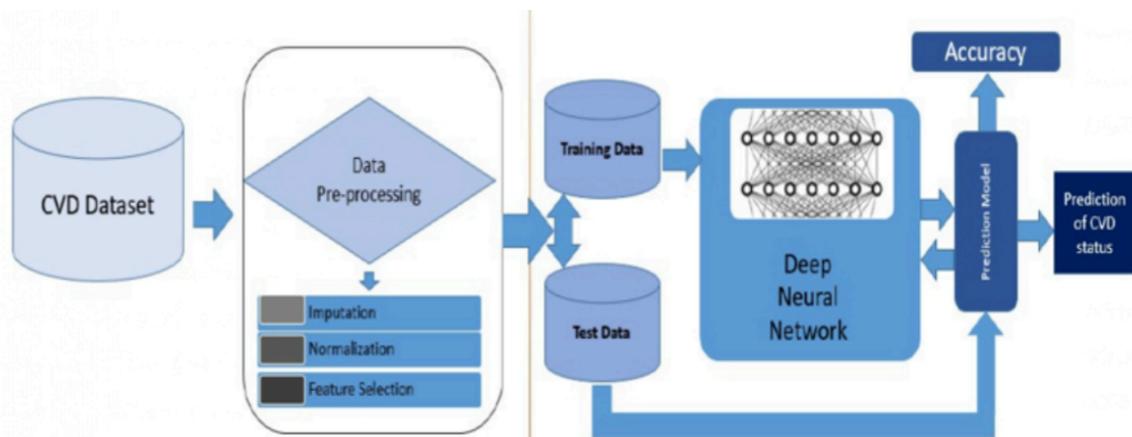


FIGURE 2.8 – L'approche utilisée pour améliorer la prédiction des maladies cardiovasculaires [46].

### 2.3.2 Soins de santé numériques basés sur l'IA intégrée

Selon Ashfaq et al. [47] qui ont proposé un système de surveillance à distance des patients basé sur l'Internet des objets médicaux (IoMT), qui s'appuie sur l'intelligence artificielle. Il permet de développer un prototype embarqué qui peut être utilisé pour la surveillance à distance des personnes cardiaques. Les principaux paramètres surveillés incluent la température corporelle, la fréquence cardiaque et la saturation en oxygène du sang. Pour la surveillance continue de la santé des patients, Les capteurs sont utilisés pour surveiller les signes vitaux du corps et ils envoient les données à la périphérie.

Le réseau de capteurs utilise la technologie WIFI pour mesurer les signes vitaux du corps et transmettre les données. En cas d'anomalie détectée par le capteur, il est nécessaire de l'informer au patient et aux soignants... La conception du matériel est basée sur l'Internet

des objets médicaux (IoMT) qui administre les facteurs physiologiques affectant la santé d'un patient.

La Figure 2.9 montre le système proposé. Le développement logiciel commence par acquérir des données vitales à partir de la base de données clinique MIMIC-III. Puis le prétraitement et le nettoyage des données.

Et enfin, La création du modèle qui utilise neuf algorithmes d'apprentissage automatique qui ont été testés pour trouver le modèle qui fonctionne mieux (KNN, SVM, DECISION TREES, NEURAL NETWORKS). Le modèle final était le KNN optimisé qui a obtenu les meilleurs résultats avec une précision de 96,26%. Et enfin, il a été déployé à la périphérie pour la prédiction et le suivi des patients. À l'aide des services cloud d'Amazon Web Services (AWS), le modèle peut être téléchargé sur les Raspberry Pi. Une application web a été créée et élaborée dans le but de faciliter l'accès aux données du patient.

La Figure 2.10 montre les étapes de création du modèle.

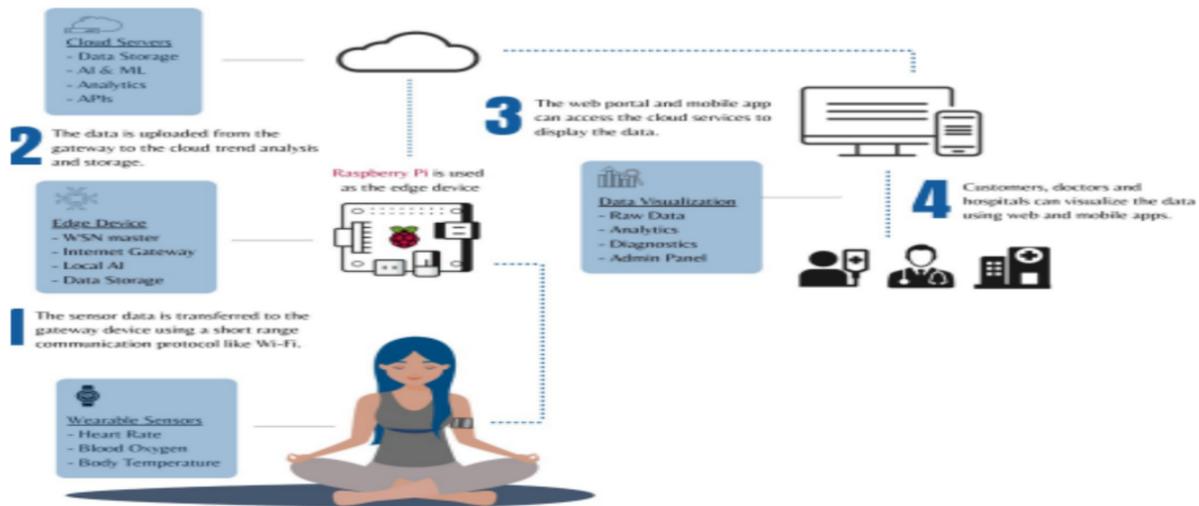


FIGURE 2.9 – Architecture du système proposé [47].



FIGURE 2.10 – Méthodologie de développement des modules de prédiction de l'état de santé [47].

### 2.3.3 Un cadre basé sur l'IoT pour détecter les pathologies cardiaques à l'aide de l'apprentissage automatique

Selon El Naggar et al.[48] Qui ont proposé un système utilisant l'IoT pour repérer les problèmes cardiaques. Le modèle de détection des crises cardiaques (HADM) et le modèle de classification multi-classe des battements cardiaques de l'ECG (ECG-HMCM) sont exposés en utilisant des méthodes de machine learning. Ces modèles utilisent des signaux de battements cardiaques de l'ECG et des méthodes de ML telles que le SVM, RANDOM FOREST et K-Nearest Neighbor (KNN) optimisé par Grid Search afin d'affiner la classification et la prédiction.

La Figure 2.11 montre l'approche utilisée.

Le premier modèle utilise un dataset qui comprend 1190 patients présentant 14 caractéristiques distinctes. Ensuite, ils ont effectué le prétraitement et le nettoyage du Dataset, puis ils ont appliqué les algorithmes suivants : KNN avec Grid Search pour optimiser les hyperparamètres, SVM et Random Forests.

Dans le deuxième modèle, ECG-HMCM, l'intégration de ResNet-18 avec l'algorithme K Nearest Neighbor (KNN) se déroule en deux étapes principales. Dans une première étape, ResNet-18 est employé pour extraire des caractéristiques des signaux ECG, tandis que KNN est utilisé pour classer ces caractéristiques en différentes catégories de battements cardiaques.

En utilisant des ensembles de données d'entraînement, le modèle complet est élaboré et ajusté, où les hyperparamètres de ResNet-18 et KNN sont optimisés avec l'optimiseur Grid Search afin d'optimiser la précision de classification.

Le modèle a obtenu une précision globale de classification très élevée, proche de 97,5%.

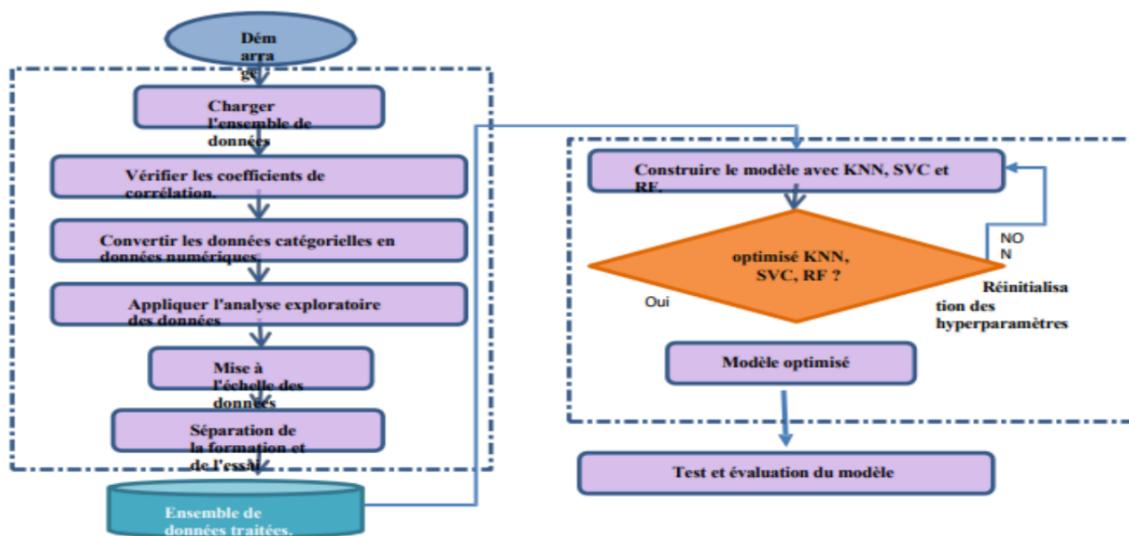


FIGURE 2.11 – Organigramme du modèle de détection des crises cardiaques [48].

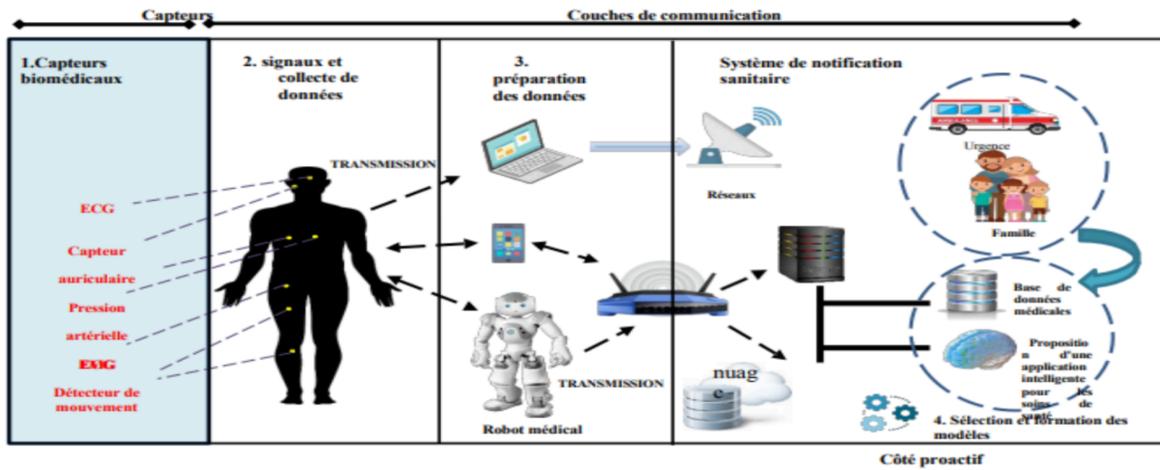


FIGURE 2.12 – Organigramme du modèle de détection des crises cardiaques proposé (HADM)[48].

### 2.3.4 Système de prédiction des maladies cardiovasculaires basé sur l’Internet des Objets (IoT) et l’apprentissage automatique

Selon Jameel.A et al [49]. L’article propose une méthode novatrice pour prédire les maladies cardiovasculaires en utilisant les données cliniques collectées via des appareils IoT. Ces données sont ensuite analysées dans le cloud avec des techniques d’apprentissage automatique pour offrir une prédiction en temps réel. Les étapes de l’approche utilisée pour le développement du modèle de prédiction des maladies cardiovasculaires sont structurées en quatre niveaux principaux :

- Collecte des Données : Les données cliniques sont collectées à Jammu et au Cachemire (Inde) grâce à des capteurs médicaux IoT portables qui enregistrent des paramètres vitaux. Ces données sont essentielles pour entraîner le modèle prédictif.
- Stockage des Données : Les données collectées sont stockées dans le cloud à l’aide de HEROKU PostgreSQL, ce qui permet un accès et une gestion centralisés.
- Analyse des Données : Application de divers algorithmes d’apprentissage automatique pour analyser les données stockées et prédire les maladies cardiovasculaires.
- Interface Utilisateur Graphique (GUI) : Développement d’une interface utilisateur permettant aux médecins et aux patients d’accéder aux résultats des analyses et aux prédictions des maladies.

Les chercheurs ont utilisé plusieurs techniques d'apprentissage automatique, notamment Random Forest (Forêt Aléatoire), Decision Tree (l'Arbre de Décision), les k plus proches voisins (KNN), la Machine à Vecteurs de Support (SVM) et le Bayésien Naïf (Naïve Bayes). Après ajustement minutieux des hyperparamètres, Random Forest (Forêt Aléatoire) s'est avéré être le modèle le plus performant, avec des scores élevés en précision(87.7%). Les autres modèles ont également affiché de bonnes performances, mais étaient légèrement inférieurs en comparaison.

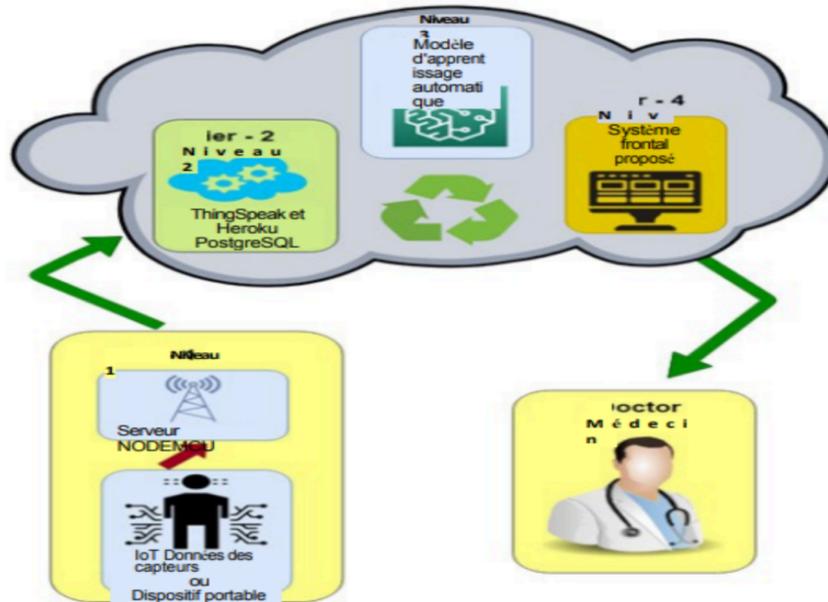


FIGURE 2.13 – diagramme de la méthodologie proposée [49].

## 2.4 Tableau comparatif des travaux existant :

Pour une meilleure compréhension des travaux étudiés précédemment, nous allons les comparer sous forme d'un tableau comparatif comme présenté ci-dessous :

Critères & travail	Datas set utilisé	Algorithmes utilisés	Type de l'apprentissage	Résultats Obtenues (Précision)
Debasish Swapnes h et al [46], 2023	Integration de 4 datasets : Cleveland, Suisse, LongBeach et Starlog	L'algorithme DNN (DEEP NEURAL NETWORK)	Apprentissage Supervisé	L'algorithme DNN a obtenu une précision de 95,34 %.

Zarlish Ashfaq et al [47], 2022	Dataset MIMIC-III clinical	9 algorithmes utilisés : Réseaux de neurones, Bagging, Random Forest, Decision Trees, KNN, Naive Bayes, Logistic Regression, Linear Regression	Apprentissage supervisé	L'algorithme KNN a obtenu la meilleure précision de 96,26%.
Mona Alnaggar et al [48], 2023	Le modèle initial a employé un ensemble de données comprenant 1190 patients et 14 caractéristiques. Le deuxième modèle a employé deux datasets : MIT-BIH Ahythmia Dataset Et PTB Diagnostic ECG Dataset	Le modèle de détection de l'attaque cardiaque (HADM) a employé trois algorithmes de machine learning : KNN, SVM et Random Forest *Le deuxième modèle de classification des signaux ECG (ECG-HMCM) a intégré Resnet 18 pour extraire les caractéristiques des signaux ECG et le KNN pour classifier ces signaux.	Apprentissage supervisé pour les 2 modèles.	*Le premier modèle (HADM) : La précision de l'algorithme KNN a été supérieure de 92,31 % par rapport aux algorithmes SVM et RANDOM FORESTS. *Le KNN a permis au deuxième modèle (ECG-HMCM) d'atteindre une précision de 97,5%.
Jameel Ahamed, et al [50] 2021	Jammu et cache-mire (Inde)	Random Forest, Decision Trees, KNN, SVM Naive Bayes,	Apprentissage supervisé	Random Forest montrant la meilleure précision de 87,72 %

TABLEAU 2.1 – table comparative de quelques approches des travaux précédents.

## 2.5 Synthèse

Les articles examinés explorent une variété de modèles d'intelligence artificielle (IA) et de techniques de l'Internet des objets (IoT) pour améliorer la prédiction des maladies cardiovasculaires (MCV). Une approche courante consiste à utiliser des réseaux de neurones profonds pour analyser en temps réel les données de santé collectées par des capteurs IoT, permettant ainsi de détecter les signes précoces de MCV. Certains articles se concentrent sur l'utilisation de méthodes d'apprentissage automatique classiques telles que les  $k$  plus proches voisins (KNN), les machines à vecteurs de support (SVM) et les forêts aléatoires pour classer et prédire les risques de MCV avec précision. En parallèle, l'utilisation de l'IoT permet une surveillance continue des signes vitaux des patients, facilitant une intervention médicale rapide en cas de détection d'anomalies. Cependant, la mise en œuvre de ces systèmes nécessite une gestion complexe des données et des algorithmes, ainsi qu'une attention particulière à la confidentialité et à la sécurité des données médicales. En somme, les progrès réalisés dans l'intégration de l'IA et de l'IoT ouvrent de nouvelles perspectives dans la prédiction et la gestion des MCV, mais soulèvent également des défis techniques et éthiques qui doivent être pris en compte lors de leur déploiement dans les environnements cliniques.

## 2.6 Conclusion

Nous avons exploré l'application de l'IoT et de l'IA dans la détection et la prévention des maladies cardiovasculaires. Cette section débute par une analyse approfondie des maladies cardiovasculaires, depuis leur définition jusqu'à leur prévalence mondiale et locale. Ensuite, une revue détaillée des travaux existants est présentée, examinant chaque étude pour mettre en évidence les tendances émergentes et les résultats significatifs. Un tableau comparatif des études examinées synthétise les principales caractéristiques des approches. Il est clair que l'avenir de la santé cardiovasculaire réside dans une approche équilibrée entre l'innovation et la responsabilité, où l'IA et l'IoT peuvent contribuer significativement à améliorer la santé et le bien-être mondial.

# Architecture proposée et implémentation des modèles de Machine Learning

## 3.1 Introduction

Au cours du chapitre précédent, nous avons examiné en détail les maladies cardiovasculaires. De plus, nous avons analysé des travaux existants dans l'utilisation de l'IA et de l'IoT pour le suivi des personnes atteintes de ces maladies. Dans ce dernier chapitre, nous passons de la théorie à la pratique en détaillant l'architecture proposée pour notre projet, présentant le schéma architectural, les outils de développement utilisés, et la description du dataset. Nous analysons les attributs, effectuons le prétraitement des données, et mettons en œuvre plusieurs algorithmes de classification, sans et avec l'optimiseur GridSearchCV. Les résultats obtenus sont évalués et interprétés, avec une discussion sur les performances, les limitations de l'étude, et les pistes pour des travaux futurs, offrant ainsi une vue complète de notre démarche et de nos découvertes sur les maladies cardiovasculaires.

## 3.2 Architecture proposée

Durant cette section, nous allons présenter notre architecture proposée afin de permettre un suivi fiable et précis des patients atteints de maladies cardiovasculaires, à partir des données collectées régulièrement. Cette architecture permet également aux médecins de surveiller à distance l'état de leurs patients. Notre architecture est composée de trois phases principales pour la gestion et l'analyse des données : (1) Collecte des données, (2) Stockage et Traitement des données, et (3) Visualisation des données, comme illustré dans la Figure 3.1.

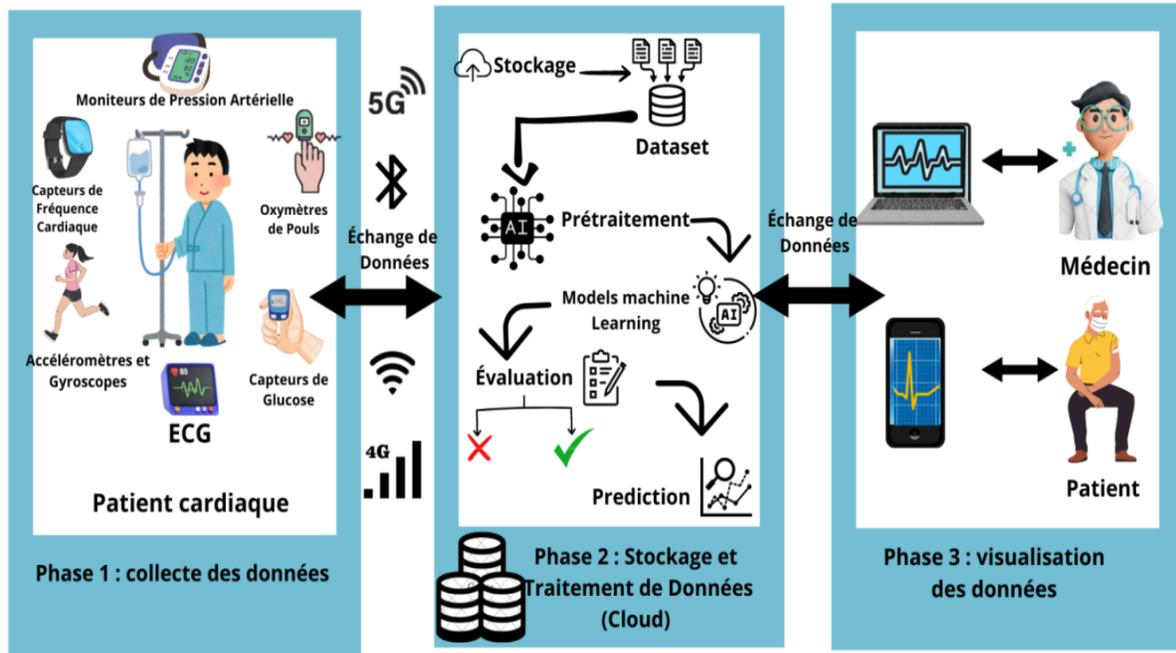


FIGURE 3.1 – Architecture globale du système.

Dans ce qui suit, nous présentons une description des trois phases de l'architecture proposée.

### 3.2.1 Phase 1 : Collecte des données

À cette étape, nous recueillons différentes informations sur la santé des patients atteints de maladies cardiaques en utilisant divers capteurs et équipements médicaux. Les capteurs principaux employés sont :

- **Moniteurs de pression artérielle** : Ils mesurent la pression sanguine.
- **Capteurs de fréquence cardiaque** : Ils surveillent la fréquence cardiaque en temps réel.
- **Oxymètres de pouls** : Ils mesurent la saturation en oxygène dans le sang.
- **Accéléromètres et gyroscopes** : Ils capturent les mouvements et les activités physiques.
- **Capteurs de glucose** : Ils surveillent le taux de glucose dans le sang.

- **ECG (électrocardiogramme)** : Ils enregistrent les signaux électriques produits par le cœur.

Ces dispositifs envoient les données collectées via des connexions sans fil (comme 5G, Bluetooth, WiFi) vers la phase suivante.

### 3.2.2 Phase 2 : Stockage et traitement des données (Cloud)

Dans cette phase, nous stockons les données collectées dans le cloud. Les étapes suivantes sont réalisées :

- **Stockage** : Les données sont centralisées et stockées sur des serveurs sécurisés dans le cloud, ce qui permet une gestion efficace et un accès facile pour le traitement ultérieur.

- **Prétraitement** : Les données brutes sont nettoyées, normalisées et transformées pour éliminer les anomalies et préparer les données pour l'analyse. Cela englobe la sélection des caractéristiques pertinentes et la suppression des informations redondantes ou inutiles.

- **Modélisation avec des algorithmes de machine learning** : Les données prétraitées sont utilisées pour entraîner des modèles de machine learning capables de faire des prédictions sur l'état de santé des patients. Divers modèles peuvent être appliqués, et leur performance est évaluée pour sélectionner le meilleur.

- **Évaluation** : On évalue l'efficacité des modèles en utilisant des indicateurs de performance tels que la précision, le rappel, le F1-score, etc. Les modèles inefficaces sont écartés, tandis que les modèles les plus performants sont choisis pour la prédiction.

- **Prédiction** : Les modèles approuvés sont employés afin de prédire l'état de santé à venir du patient, ce qui permet de détecter précocement les anomalies ou les problèmes cardiaques.

### 3.2.3 Phase 3 : Visualisation des données

Dans cette phase, nous partageons les résultats des prédictions et les données analysées avec les parties concernées :

- **Médecins** : Les données et les analyses sont transmises aux médecins via des interfaces utilisateurs, comme des ordinateurs ou des tablettes, leur permettant de consulter facilement les informations pertinentes sur l'état de santé de leurs patients.

- **Patients** : Les patients peuvent également accéder à certaines visualisations de leurs données de santé via des applications mobiles ou des dispositifs connectés, leur permettant de suivre leur propre état de santé et d'adopter des mesures préventives si nécessaire.

Nous donnerons une importance primordiale au prétraitement des données dans cette architecture.

Au cours de la procédure, les informations sont constamment transférées entre les différentes étapes grâce à des connexions sans fil sécurisées (4G, 5G, WiFi, Bluetooth). Cela facilite l'accès rapide aux informations pour le traitement et l'analyse, ainsi que la transmission efficace des résultats avec les professionnels de santé et les patients. La surveillance continue et proactive des patients cardiaques est assurée grâce à cette architecture intégrée, ce qui facilite des interventions médicales rapides et informées, ce qui contribue à améliorer les soins et les résultats de santé des patients.

## 3.3 Outils de développement

Durant cette partie, nous présenterons l'ensemble des outils utilisés pour la mise en œuvre de notre système de création des modèles de machine learning.

### 3.3.1 Environnement de développement

#### 3.3.1.1 Google Colab

Nous avons utilisé Google Colab (Figure 3.2) comme environnement de développement. Google Colab est une plateforme gratuite fournie par Google qui vous permet d'écrire et d'exécuter du code Python dans votre navigateur. Sa particularité réside dans sa capacité à exécuter des notebooks Jupyter sans avoir à vous préoccuper de votre matériel ou des logiciels installés sur votre ordinateur. Google Colab offre également une facilité d'accès à des ressources de calcul et aux bibliothèques d'apprentissage automatique couramment utilisées [53].



FIGURE 3.2 – logo de Google Colab.

### 3.3.1.2 Langage python

Python est un langage de programmation puissant et facile à apprendre. Il dispose de structures de données de haut niveau et permet une approche simple, mais efficace de la programmation orientée objet. Il dispose d'une syntaxe élégante, un typage dynamique, ce qui le rend simple à interpréter et à utiliser pour une variété de tâches de programmation [54].

### 3.3.1.3 Le python et l'IA

Python est aussi extrêmement prisé dans le domaine de l'intelligence artificielle en raison de ses bibliothèques de grande qualité et de sa capacité de calcul. De nombreux scientifiques et programmeurs font appel à Python afin de mettre en place et d'expérimenter de nouveaux algorithmes d'apprentissage automatique et de deep learning, notamment grâce aux frameworks comme TensorFlow et PyTorch. Grâce à Python, il est envisageable de concevoir rapidement des modèles Complexes et puissants, ce qui en fait une option parfaite pour les projets d'intelligence artificielle de toutes dimensions [55].

## 3.3.2 Les bibliothèques utilisées

Les bibliothèques utilisées sont [56] :

### 3.3.2.1 Numpy

Les modules de gestion de données et de calcul sont inclus dans Numpy. Elle facilite la gestion de bases de données : on les désigne sous le nom de numpy arrays. Il s'agit soit de listes, soit des listes de listes. Numpy présente l'avantage de permettre la création rapide d'une base de données, en utilisant des instructions simples que Python comprend.

### 3.3.2.2 Pandas

Pandas représente la bibliothèque la plus exhaustive en matière de gestion de données. Pandas peuvent être comparés à un "Excel sous stéroïdes" : ils offrent la possibilité de travailler avec. Les DataFrames sont des tableaux de données en deux dimensions (lignes et colonnes) et les Panels sont des ensembles de données en trois ou quatre dimensions. Grâce à la bibliothèque Pandas, il est possible d'importer des données d'un fichier.csv pour les nettoyer (par exemple pour éliminer toutes les lignes vides), les modifier ou les compléter.

### 3.3.2.3 Matplotlib

Matplotlib est vraiment satisfaisant comme outil de visualisation de données. Avec seulement une ligne de code, tu peux créer des graphiques qui mettent en valeur tes données. Tu peux l'utiliser avec Numpy ou Pandas pour une meilleure manipulation des données. Matplotlib propose une variété de graphiques adaptés à différents besoins : des histogrammes aux boîtes à moustache en passant par les courbes, les scatter plots et même les camemberts.

### 3.3.2.4 Seaborn

Seaborn est une alternative à Matplotlib, offrant des fonctionnalités avancées pour créer des graphiques relationnels entre deux vecteurs ainsi que des graphiques de distribution. Il est réputé pour sa praticité, car ses thèmes par défaut sont élégants et peuvent être directement utilisés pour créer des tableaux de bord, une fonctionnalité qui n'est pas toujours présente dans Matplotlib. De plus, avec Seaborn, il est possible de simplifier la création de graphiques grâce à l'automatisation, ce qui facilite le travail et permet de gagner du temps lors de la présentation visuelle des résultats.

### 3.3.2.5 Scikit-learn

Le module de machine learning le plus apprécié est Scikit learn, également connu sous le nom de sklearn. La bibliothèque permet de réaliser des opérations supervisées ou non supervisées. Elle a aussi des compétences en modélisation statistique, notamment en ce qui concerne la classification et la régression. Parmi les différentes classes de modèles proposées par Sklearn, il est possible de sélectionner un modèle, de le configurer, de l'entraîner et enfin de le tester sur tes données. Cette bibliothèque convient parfaitement aux usages professionnels, en particulier pour la création de modèles prédictifs.

## 3.4 Dataset

### 3.4.1 Description du dataset

L'ensemble de données utilisées dans notre projet est l'ensemble de données sur les maladies cardiaques de Cleveland, extraites du dépôt de l'UCI. Le fichier heart.csv contient le détail des attributs et des variables. Nous avons utilisé 14 attributs comme illustré dans la Figure 3.3. (y compris la caractéristique "target"). L'attribut "target" indique la présence ou l'absence de maladies cardiaques chez les patients. Il est représenté par une valeur entière, où 0 ne signifie aucune présence de maladies cardiaques et 1 signifie présence de maladies cardiaques [57].

La figure 29 montre une description de quelques lignes du dataset.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	0	145	233	1	2	150	0	2.3	2	0	2	0
1	67	1	3	160	286	0	2	108	1	1.5	1	3	1	1
2	67	1	3	120	229	0	2	129	1	2.6	1	2	3	1
3	37	1	2	130	250	0	0	187	0	3.5	2	0	1	0
4	41	0	1	130	204	0	2	172	0	1.4	0	0	1	0

FIGURE 3.3 – Extrait du dataset Cleveland.

### 3.4.2 Description des attributs du dataset :

- **Age** : affiche l'âge de l'individu.
- **Sexe** : affiche le sexe de l'individu selon le format ; 1 = homme, 0 = féminin.
- **Type de douleur à la poitrine** : affiche le type de douleur ressentie par l'individu à la poitrine en utilisant le format ; 0 = angine de poitrine typique, 1 = angine de poitrine atypique, 2 = non-douleur angineuse, 3 = asymptotique.
- **Tension artérielle au repos** : affiche la valeur de la tension artérielle au repos de l'individu en mmHg (unité).
- **Cholestérol sérique** : indique le cholestérol sérique en mg/dl (unité).
- **Glycémie à jeun** : compare la valeur de la glycémie à jeun d'un individu à 120 mg/dl. Si la glycémie à jeun est supérieure à 120 mg/dl, alors afficher 1 (vrai), sinon afficher 0 (faux).
- **ECG au repos** : affichage des résultats de l'électrocardiogramme au repos. 0 = normal, 1 = présence d'une anomalie de l'onde ST-T, 2 = hypertrophie ventriculaire gauche.
- **Fréquence cardiaque maximale atteinte** : affiche la fréquence cardiaque maximale atteinte par un individu.

- **Angine de poitrine induite par l'exercice** : 1 = oui, 0 = non.
- **ST dépression induite par l'exercice par rapport au repos** : affiche la valeur qui peut être un entier ou un flottant
- **Segment ST de l'exercice de pointe** : 1 = montée, 2 = plat, 3 = en pente descendante.
- **Nombre de vaisseaux principaux (0-3) colorés par fluoroscopie** : affiche la valeur sous forme d'un nombre entier ou d'un flottant.
- **Thalassémie** : affiche la thalassémie ; 3 = normal, 6 = défaut fixe, 7 = défaut réversible.
- **Diagnostic d'une maladie cardiaque** : indique si la personne souffre ou non d'une maladie cardiaque ; 0 = absence, 1 = présence.

## 3.5 Prétraitement des données

les étapes du prétraitement des données comprennent :

### 3.5.1 Vérifier les valeurs manquantes

Le Dataset ne présente aucune valeur manquante.

```
] df.isnull().sum()
age          0
sex          0
cp          0
trestbps    0
chol        0
fbs         0
restecg     0
thalach     0
exang       0
oldpeak     0
slope       0
ca          0
thal        0
target      0
dtype: int64
```

FIGURE 3.4 – Vérification des valeurs manquantes.

## 3.5.2 Normalisation de données

Nous avons appliqué `StandardScaler` pour mettre à l'échelle les données et les rendre comparables, facilitant ainsi l'analyse.

### 3.5.2.1 Définition de `StandardScaler`

La classe `StandardScaler` de Scikit Learn standardise les variables d'entrée (features), assurant qu'elles ont une moyenne proche de 0 et un écart type d'environ 1. Cette transformation ajuste les données pour qu'elles aient une distribution standardisée, les rendant ainsi adaptées à la modélisation. Cela permet également d'éviter qu'une caractéristique n'influence de manière disproportionnée l'algorithme en raison de différences d'échelle [58].

Dans cette figure, les données sont représentées avant la normalisation.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	0	145	233	1	2	150	0	2.3	2	0	2	0
1	67	1	3	160	286	0	2	108	1	1.5	1	3	1	1
2	67	1	3	120	229	0	2	129	1	2.6	1	2	3	1
3	37	1	2	130	250	0	0	187	0	3.5	2	0	1	0
4	41	0	1	130	204	0	2	172	0	1.4	0	0	1	0

FIGURE 3.5 – avant la normalisation.

Cette figure présente les données après la normalisation.

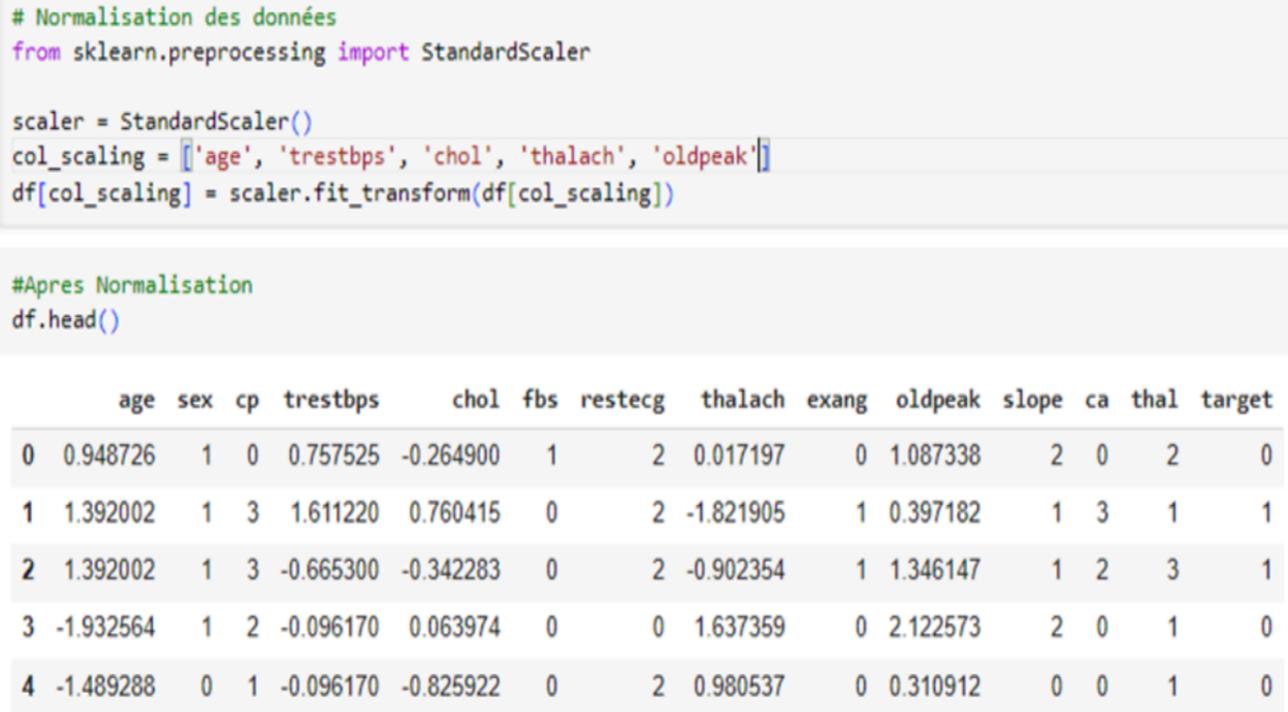


FIGURE 3.6 – après la normalisation.

### 3.5.3 Visualisation des données

Dans cette section, nous aborderons seulement 3 des 14 attributs du dataset, ainsi que la matrice de corrélation.

#### 3.5.3.1 Fréquence de la cardiologie sexuelle :

Le diagramme de barres ci-dessus représente la fréquence des cas de maladies cardiaques selon le sexe. Il montre le nombre d'hommes et de femmes malades et non malades comme le montre la Figure 3.7.

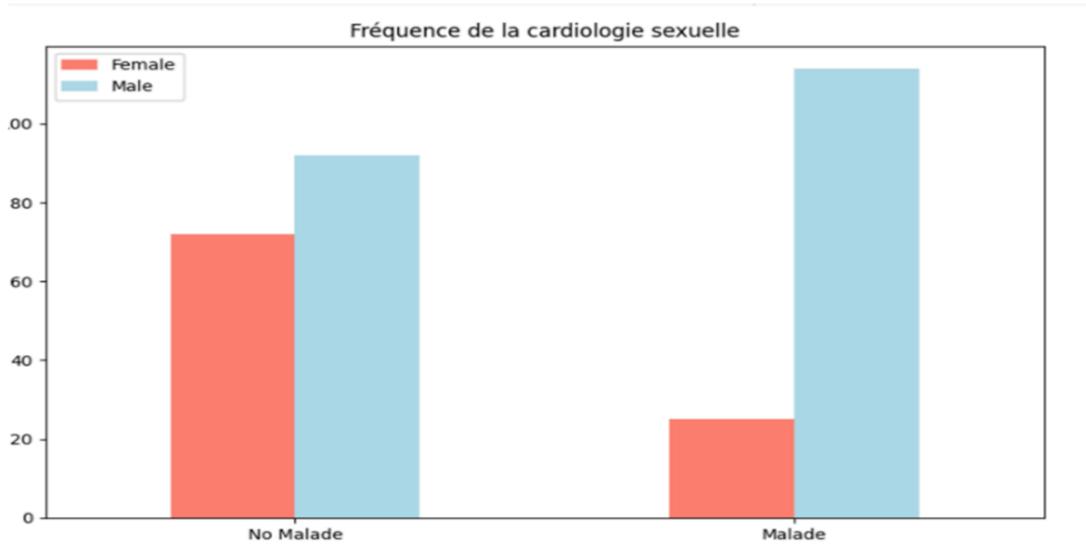


FIGURE 3.7 – Visualisation de Fréquence de la cardiologie sexuelle.

### 3.5.3.2 Fréquence des maladies cardiaques par type de douleur thoracique

Visualisation des types de douleurs à la poitrine chez les personnes malades et non malades en utilisant des catégories suivantes pour le type de douleur ressentie par l'individu à la poitrine.

0 = angine de poitrine typique, 1 = angine de poitrine atypique, 2 = douleur non angineuse, 3 = asymptomatique. La figure 34 montre cette visualisation.

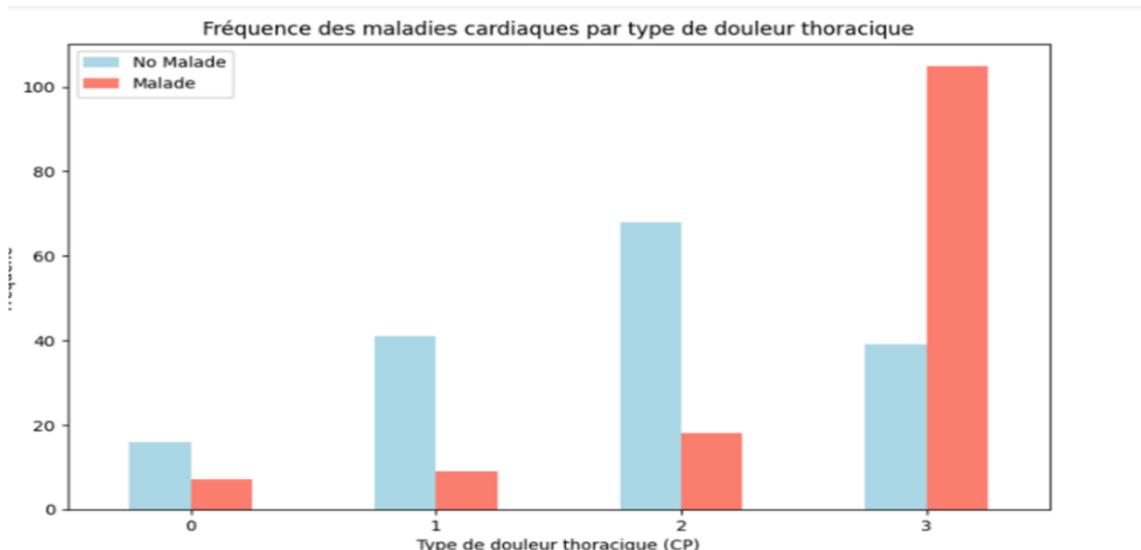


FIGURE 3.8 – Visualisation de Fréquence des maladies cardiaques par type de douleur thoracique.

### 3.5.3.3 Visualisation des corrélations entre les variables

La matrice de corrélation est un outil statistique qui permet de mesurer la corrélation entre les différentes variables d'un ensemble de données. La Figure 35 représente la matrice de corrélation de notre Dataset.

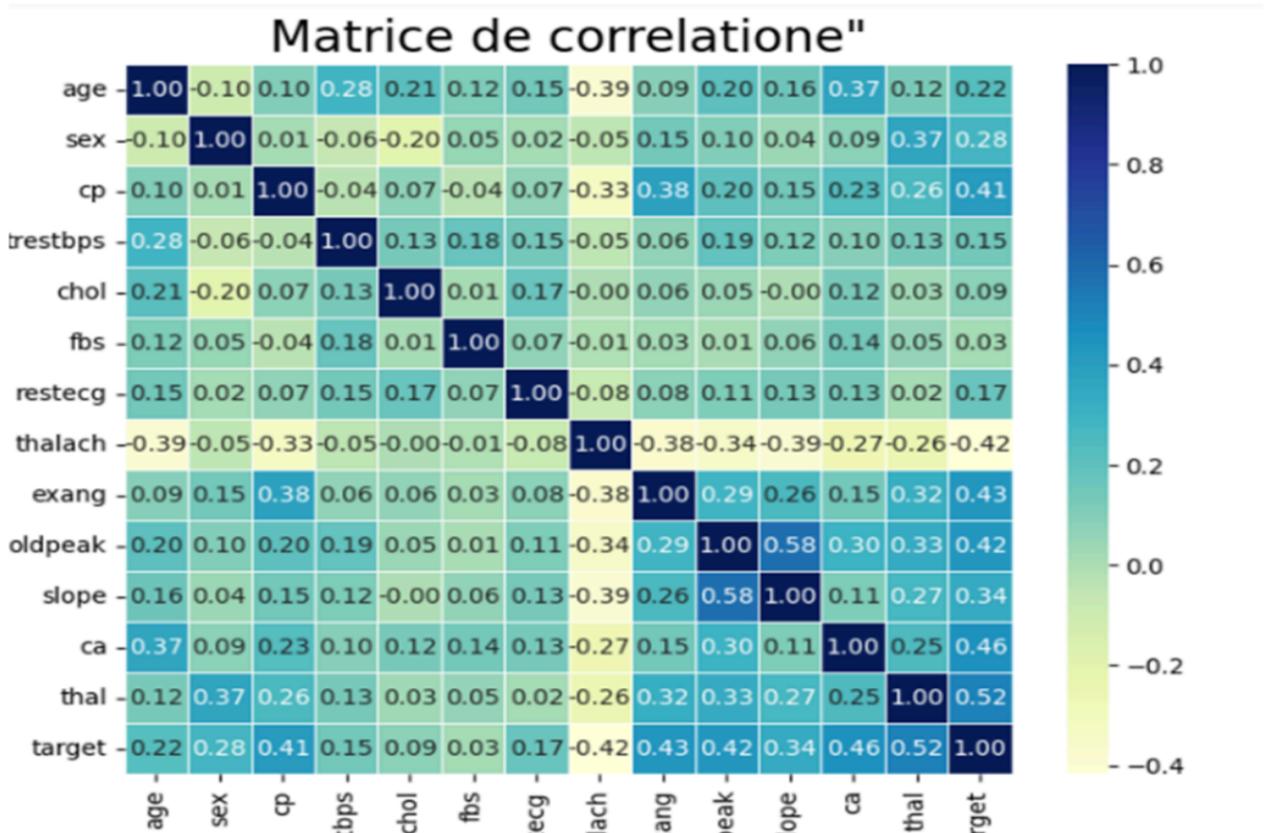


FIGURE 3.9 – Matrice de corrélation.

## 3.6 Étapes de création des modèles de machine learning

### 3.6.1 Séparation des ensembles de données

Les données sont divisées en ensembles d'entraînement et de test. L'ensemble d'entraînement est utilisé pour entraîner le modèle, l'ensemble de test est utilisé pour évaluer les performances finales du modèle. Nous avons opté pour le protocole 80 :20 qui consiste à réserver 80% des données pour l'ensemble d'entraînement et 20% pour l'ensemble de tests. La figure suivante montre un morceau de code de cette étape.

```
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.neighbors import KNeighborsClassifier
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split

np.random.seed(42)

# Diviser les données en ensembles d'entraînement et de test avec une graine fixe
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

FIGURE 3.10 – Division des Ensembles de Données.

### 3.6.2 Algorithme de ML utilisés

Afin de mettre en place notre modèle de prédiction, nous avons employé 5 algorithmes d'apprentissage automatique supervisé, ainsi que GridSearchCv pour optimiser les hyper paramètres des modèles d'apprentissage automatique. L'entraînement des modèles de machine learning a été effectué avec et sans GridSearchCv. Le test a été fait sur les hyper-paramètres les plus performants identifiés grâce à la méthode de GridSearchCv.

Voici une brève définition de chaque algorithme utilisé :

#### 3.6.2.1 Algorithme des K plus proches voisins(KNN)

K-Nearest Neighbours est une méthode et un algorithme d'apprentissage automatique qui peut être employé dans le cadre de tâches de régression et de classification. Les voisins les plus proches analysent les étiquettes d'un certain nombre de points de données entourant un point de données cible pour prédire la classe à laquelle il appartient. Un algorithme KNN passe par trois phases principales lors de son exécution. D'abord, Fixer K au nombre de voisins choisi et Calculer la distance entre un exemple fourni de test et les exemples de Dataset. Ensuite, Trier les distances calculées. Et enfin, en obtient les étiquettes des entrées K supérieures et envoyer une prédiction sur l'exemple de test [59].

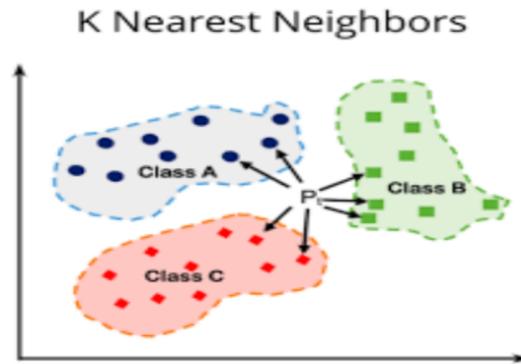


FIGURE 3.11 – Classification avec l’algorithme KNN [60].

### 3.6.2.2 Algorithme de Régression logistique

La régression logistique est un algorithme d’apprentissage automatique supervisé qui réalise des tâches de classification binaire en prédisant la probabilité d’un résultat, événement ou observation. Le modèle fournit un résultat binaire ou dichotomique limité à deux issues possibles : oui/non, 0/1, ou vrai/faux. Il est Plus facile à mettre en œuvre les méthodes d’apprentissage automatique. Il est utilisé pour Identifier les variables dépendantes pour garantir la cohérence du modèle et pour estimer le modèle et évaluer la qualité de l’ajustement. La figure 38 montre le fonctionnement de la Régression Logistique [61].

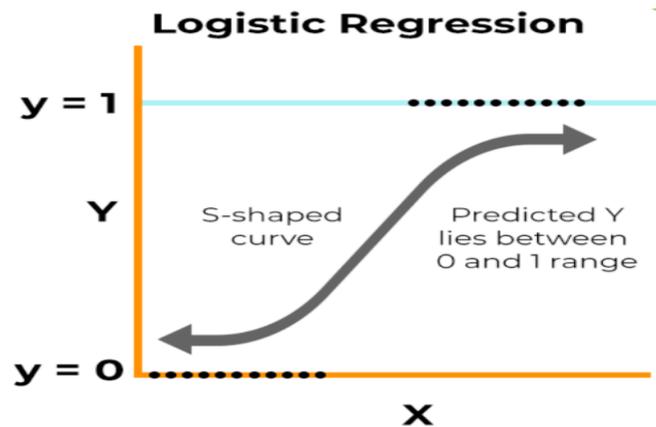


FIGURE 3.12 – Fonctionnement de la Régression Logistique [62].

### 3.6.2.3 Les machines à vecteurs de support (SVM)

Les SVM, également appelées machines à vecteurs de support, sont des modèles de machine learning supervisés qui se concentrent sur la résolution de problèmes de discrimination et de régression mathématiques. En raison de sa capacité à gérer des données de grande

taille, de ses garanties théoriques et des résultats satisfaisants obtenus en pratique, ce modèle a rapidement été adopté. Les SVM, qui nécessitent peu de paramètres, sont appréciées pour leur facilité d'utilisation. Le principe des SVM est de réduire un problème de classification ou de discrimination à un hyperplan où les données sont divisées en plusieurs classes dont la frontière est la plus éloignée possible des points de données (ou "marge maximale"). C'est pourquoi on donne un autre nom aux SVM : les séparateurs à large marge. Le concept de frontière nécessite que les données soient séparées de manière linéaire. Afin d'atteindre cet objectif, les machines support vector utilisent des noyaux, c'est-à-dire des fonctions mathématiques qui permettent de projeter et de séparer les données dans l'espace vectoriel, les « vecteurs de support » étant les données les plus proches de la limite. Il s'agit de la limite la plus éloignée de tous les points d'entraînement qui est la plus élevée, et donc qui offre la meilleure performance [63]. La Figure 3.13 montre une illustration sur le SVM.

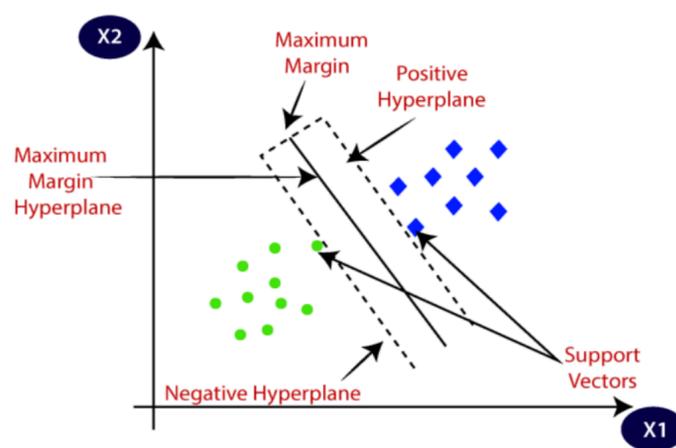


FIGURE 3.13 – Illustration du principe de SVM [64].

#### 3.6.2.4 Arbres de décisions (Decision Tree)

Un arbre de décision est une structure similaire à un organigramme utilisé pour prendre des décisions ou des prédictions. Il se compose de nœuds représentant des décisions ou des tests sur des attributs, de branches représentant le résultat de ces décisions, et de nœuds feuilles représentant des résultats finaux ou des prédictions. Chaque nœud interne correspond à un test sur un attribut, chaque branche correspond au résultat du test, et chaque nœud feuille correspond à une étiquette de classe ou à une valeur continue. Le processus de création d'un arbre de décision implique :

- Sélectionner le meilleur attribut : En utilisant une métrique comme l'impureté de Gini, l'entropie ou le gain d'information, le meilleur attribut pour diviser les données est sélectionné.

- Diviser l'ensemble de données : L'ensemble de données est divisé en sous-ensembles en fonction de l'attribut sélectionné.
- Répéter le processus : Le processus est répété de manière récursive pour chaque sous-ensemble, créant un nouveau nœud interne ou un nœud feuille jusqu'à ce qu'un critère d'arrêt soit atteint [65].

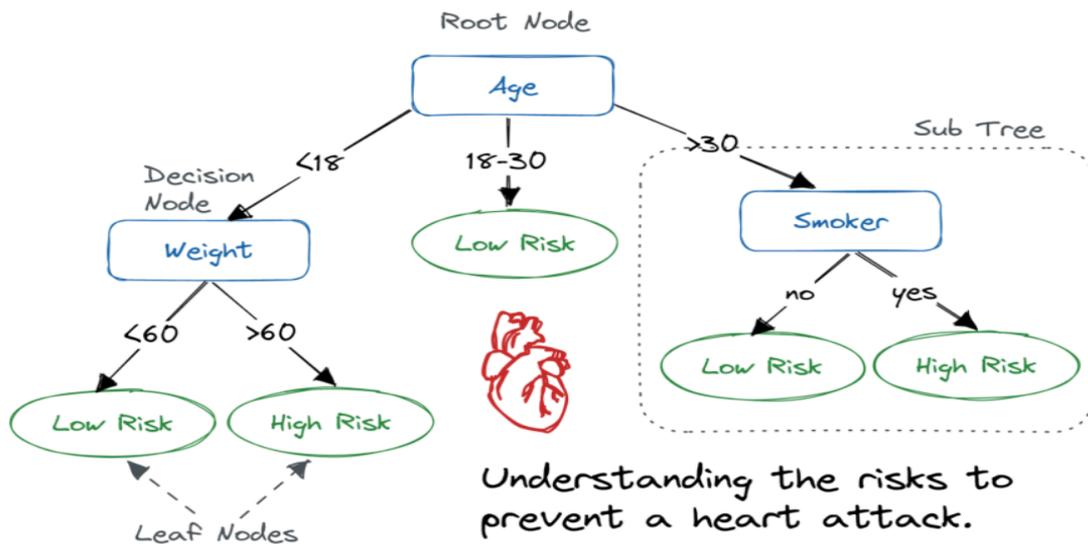


FIGURE 3.14 – Exemple d'Arbre de décision pour évaluer les risques de crise cardiaque [66].

### 3.6.2.5 Random Forest

L'algorithme Random Forest est une puissante technique d'apprentissage par arbre dans le domaine de l'apprentissage automatique. Il fonctionne en créant un certain nombre d'arbres de décision pendant la phase d'apprentissage. Chaque arbre est construit en utilisant un sous-ensemble aléatoire de l'ensemble de données pour mesurer un sous-ensemble aléatoire de caractéristiques dans chaque partition. Ce caractère aléatoire introduit une variabilité entre les arbres individuels, ce qui réduit le risque de surajustement et améliore les performances globales en matière de prédiction. Lors de la prédiction, l'algorithme agrège les résultats de tous les arbres, soit par vote (pour les tâches de classification), soit par calcul de la moyenne (pour les tâches de régression). Ce processus décisionnel collaboratif, soutenu par de multiples arbres avec leurs idées, fournit un exemple de résultats stables et précis. Les forêts aléatoires sont largement utilisées pour les fonctions de classification et de régression.

Elles sont connues pour leur capacité à traiter des données complexes, à réduire le surajustement et à fournir des prévisions fiables dans différents environnements [67].

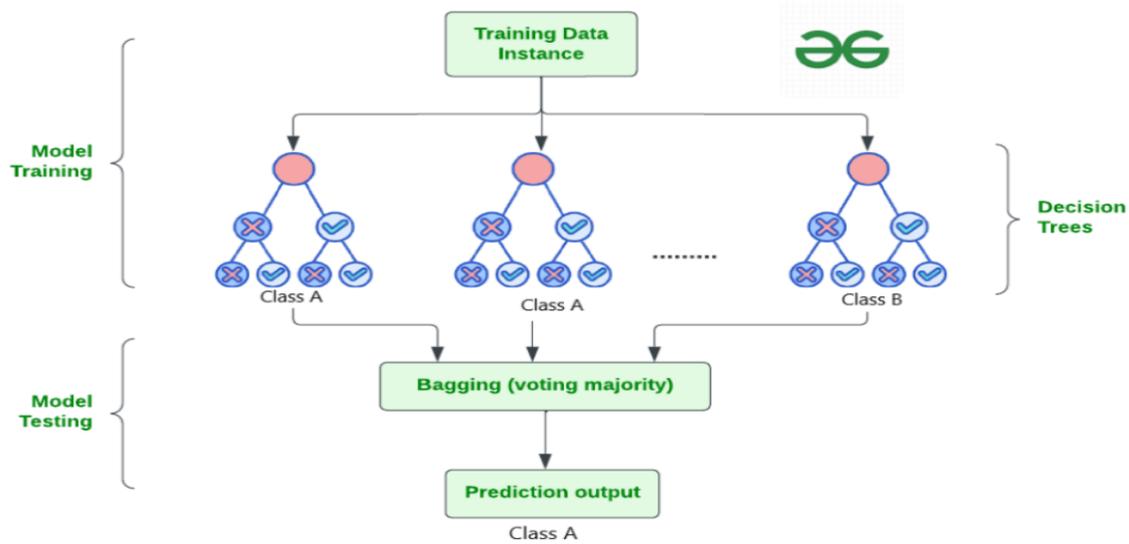


FIGURE 3.15 – Schéma de fonctionnement de Random Forest [67].

### 3.6.2.6 GridSearchCv

GridSearchCV, également connu sous le nom de Grid Search Cross-Validation, est une méthode employée dans le domaine de l'apprentissage automatique afin d'améliorer les hyperparamètres. C'est une méthode systématique qui vise à déterminer la combinaison optimale des hyperparamètres pour un modèle spécifique.

GridSearchCV fonctionne par 5 étapes :

1. Définition de la grille d'hyper paramètres : Pour débuter, il est nécessaire de définir une grille d'hyper paramètres avec leurs valeurs correspondantes. Par exemple, lors de la configuration des paramètres d'une machine à vecteurs de support (SVM), il est possible de créer une grille en fonction de paramètres tels que le type de noyau, la valeur C et le gamma.
2. Recherche exhaustive : Suite à cela, GridSearchCV effectue une recherche approfondie sur toutes les combinaisons envisageables de ces hyperparamètres. Par exemple, si vous disposez de trois hyperparamètres avec des valeurs possibles de 3, 4 et 2, GridSearchCV évaluera le modèle en utilisant une combinaison de  $2 \times 3 \times 4 = 24$  combinaisons différentes.
3. Validation croisée : Pour chaque ensemble des hyperparamètres, GridSearchCV emploie une méthode connue sous le nom de validation croisée afin d'évaluer les performances du modèle. Il subdivise les données d'entraînement en différents sous-ensembles, entraîne le modèle sur certains d'entre eux, puis l'évalue sur le reste. Ce processus est répété à plusieurs reprises, avec des sous-ensembles distincts utilisés pour l'entraînement et les tests à chaque fois.

4. Évaluation des performances : Un score de performance moyen est calculé par GridSearchCV après chaque tour de validation croisée en fonction de la métrique d'évaluation sélectionnée (par exemple, précision, erreur quadratique moyenne). Ce score témoigne de l'efficacité du modèle en utilisant une combinaison particulière des hyperparamètres.
5. Sélection de la meilleure combinaison : Après l'évaluation de toutes les combinaisons, GridSearchCV détermine tous les hyperparamètres qui ont obtenu le meilleur score de performance.

GridSearchCV est utilisé, car elle automatise le processus d'essai de différentes combinaisons des hyperparamètres et elle élimine le risque de sur apprentissage [68].

### 3.6.3 Évaluation des résultats

Dans cette section, nous allons évaluer les résultats avec et sans GridSearchCv.

#### 3.6.3.1 Sans GridSearchCv

##### 1. Évaluation selon les métriques de performances

Évaluer les performances d'un modèle de classification est essentiel pour mesurer son efficacité et prendre des décisions éclairées. Voici une brève description des quatre métriques de performances qu'on a utilisées pour évaluer les performances des modèles [69].

- **Accuracy** La précision est la métrique la plus couramment utilisée dans la conversation quotidienne. La précision répond à la question : "Parmi toutes les prédictions que nous avons faites, combien étaient vraies?"

$$Accuracy = \frac{truepositives + truenegatives}{truepositives + truenegatives + falsenegatives + falsepositives}$$

- **Recall** Le rappel se concentre sur la capacité du modèle à trouver tous les positifs. Le rappel est également appelé taux de vrais positifs et répond à la question "Sur tous les points de données qui devraient être prédits comme vrais, combien avons-nous correctement prédit comme vrais?"

$$Recall = \frac{truepositives}{truepositives + falsenegatives}$$

- **F1 Score** Le score F1 est une mesure qui combine le rappel et la précision. Comme nous l'avons vu, il y a un compromis entre la précision et le rappel, le score F1 peut donc

être utilisé pour mesurer à quel point nos modèles font ce compromis de manière efficace.

$$F1 = 2 * \frac{precision * recall}{precision + recall}$$

- **Précision** : La précision est une mesure qui vous donne la proportion de vrais positifs par rapport à la quantité totale de positifs que le modèle prédit. Elle répond à la question "Sur toutes les prédictions positives que nous avons faites, combien étaient vraies ?"

$$Precision = \frac{truepositives}{truspositives + falsenegatives}$$

Voici un tableau qui illustre les métriques de performances

	<b>Précision(%)</b>	<b>Accuracy (%)</b>	<b>Recall (%)</b>	<b>F1_Score(%)</b>
<b>Logistique Régression</b>	87.8788	88.5246	90.625	89.2308
<b>KNN</b>	93.3333	90.1639	87.500	90.3226
<b>Random Forest Classifier</b>	90.000	86.8852	84.375	87.0968
<b>Decision Tree Classifier</b>	77.4194	75.4098	75.000	76.1905
<b>SVC</b>	90.3226	88.5246	87.500	88.8889

TABLEAU 3.1 – Comparaison des Performances de Divers modèles de Classification

Voici un graphique comparatif qui montre les métriques de performances comme illustré dans la Figure 3.16 .

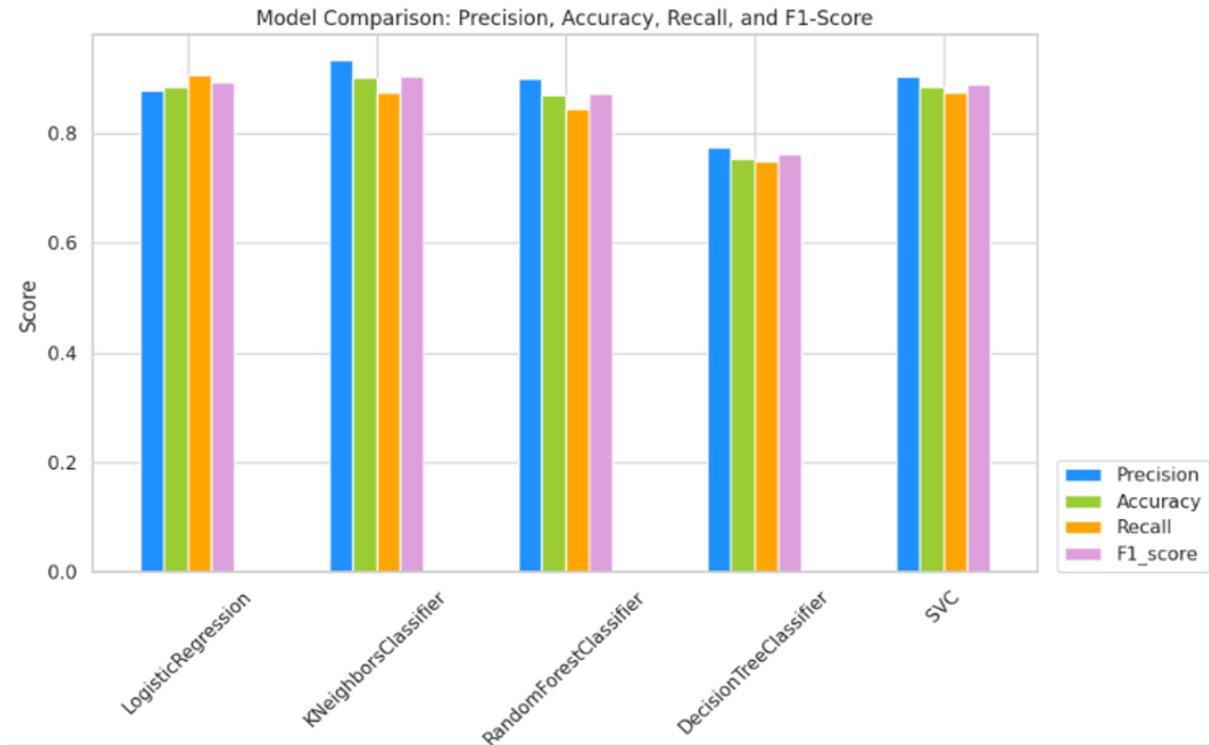


FIGURE 3.16 – Évaluation des performances des modèles.

## 2. Évaluation selon les métriques d’erreurs

La validation des modèles de machine learning est rendue possible grâce à plusieurs métriques erreur.. Pour évaluer les performances de nos modèles, nous utiliserons les trois métriques suivantes : l’erreur moyenne absolue (MAE), l’erreur quadratique moyenne (MSE) et l’explication de la variance du coefficient ( $R^2$ ).[70]

**MAE** : L’erreur moyenne absolue représente la moyenne de la différence absolue entre de réelles valeurs prédites dans le jeu de données. Elle mesure la moyenne des résidus dans le jeu de données.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}|$$

Where,  $\hat{y}$  — la valeur prédit de y.

$y_i$  — la valeur réelle de y.

**MSE** : L’erreur quadratique moyenne représente la moyenne des différences au carré entre les valeurs originales et prédites dans l’ensemble de données. Elle mesure la variance des résidus.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2$$

**R2 SCORE** : Le coefficient de détermination ou R carré représente la proportion de la variance dans la variable dépendante qui est expliquée par le modèle de régression linéaire. C'est un score sans échelle, c'est-à-dire que, indépendamment des valeurs étant petites ou grandes, la valeur de R carré sera inférieure à un.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2}$$

Voici un tableau qui illustre les métriques d'erreurs des modèles utilisés.

	MAE	MSE	R2_SCORE
Logistique Régression	0.114754	0.114754	0.539871
KNN	0.098361	0.098361	0.605603
Random Forest Classifier	0.131148	0.131148	0.474138
Decision Tree Classifier	0.245902	0.245902	0.014009
SVC	0.114754	0.114754	0.539871

TABLEAU 3.2 – Évaluation des performances des modèles basée sur les métriques d'erreur.

La Figure 3.17 suivante présente un graphique de bar comparatif qui illustre les métriques d'erreurs

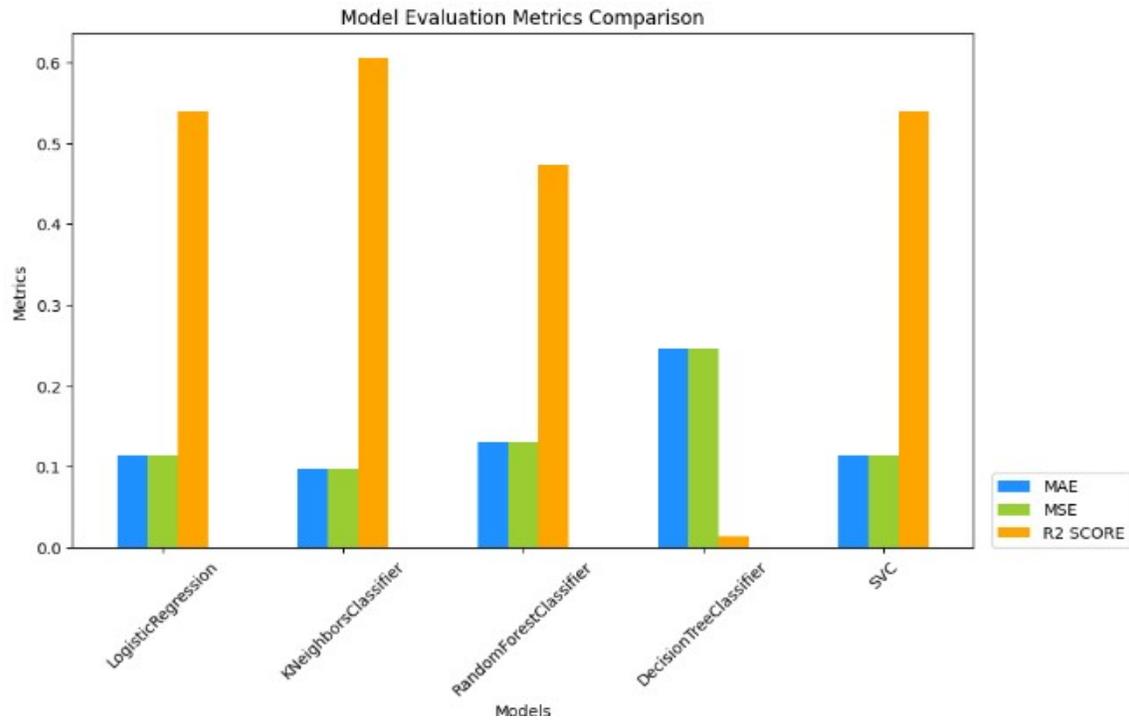


FIGURE 3.17 – Évaluation des métriques des modèles.

### 3 Selon le temps d'exécution

Ce tableau montre le temps d'exécution de chaque modèle de machine learning :

Modèle	Temps d'execution (en seconds)
LogisticRegression	0.008705
KNeighborsClassifier	0.002500
RandomForestClassifier	0.105064
DecisionTreeClassifier	0.002155
SVC	0.002918

TABLEAU 3.3 – Évaluation des performances des modèles Selon le temps d'exécution.

Le temps d'exécution de chaque algorithme est illustré dans le graphique comparatif suivant :

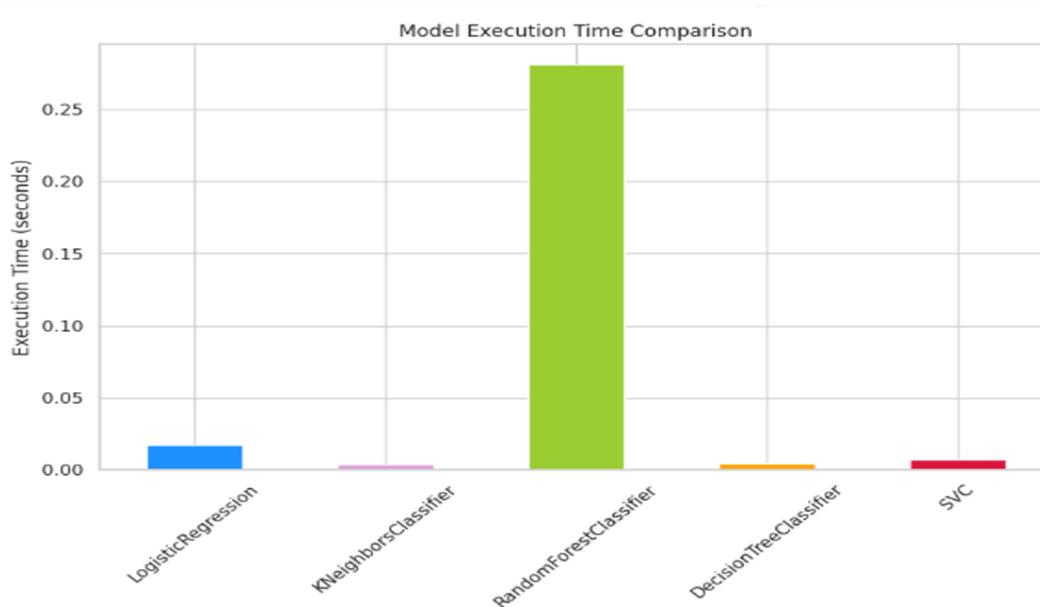


FIGURE 3.18 – Comparaison du temps d'exécution des modèles utilisés.

### 3.6.3.2 Avec GridSearchCv

Pour chaque modèle, nous avons utilisé plusieurs hyperparamètres.

- **La Régression Logistique** : Nous avons exploré différentes valeurs pour le paramètre de régularisation  $C$ .
- **KNN** : nous avons testé plusieurs nombres de voisins "K".
- **Random Forest Classifier** : Nous avons varié le nombre d'arbres  $n_{estimators}$  et les caractéristiques maximales ( $max_{featurer}$ ).
- **Decision Tree** : Différentes profondeurs d'arbre  $max_{depth}$  et seuils de division minimal  $min_{samplesplit}$  ont été étudiés.

- **Support Vector Machine (SVC)** : Nous avons effectué des tests sur diverses valeurs de C et les différents noyaux.

La Figure 3.19 ci-dessus montre le morceau de code implémenté basé sur GridSearchCv.

```

classifiers = {
    'LogisticRegression': (LogisticRegression(), {'C': [0.01, 0.1, 1, 10, 100]}),
    'KNeighborsClassifier': (KNeighborsClassifier(), {'n_neighbors': [3, 5, 7, 9]}),
    'RandomForestClassifier': (RandomForestClassifier(), {'n_estimators': [50, 100, 200], 'max_features': ['auto', 'sqrt', 'log2']}),
    'DecisionTreeClassifier': (DecisionTreeClassifier(), {'max_depth': [None, 10, 20, 30], 'min_samples_split': [2, 10, 20]}),
    'SVC': (SVC(), {'C': [0.1, 1, 10, 100], 'kernel': ['linear', 'rbf']})
}

# Calculer les métriques
for model_name, (model, params) in classifiers.items():
    grid_search = GridSearchCV(model, params, cv=5, scoring='precision')
    grid_search.fit(X_train, y_train)
    best_model = grid_search.best_estimator_
    predictions = best_model.predict(X_test)
    
```

FIGURE 3.19 – Implémentation de GridSearchCv.

### 1. Évaluation selon les métriques de performances

Le tableau 6 illustre les métriques de performances :

Modèle	Précision(%)	Accuracy(%)	RECALL(%)	F1 SCORE(%)
LogisticRegression	92.5926	85.2459	78.125	84.7458
KNeighborsClassifier	87.0968	85.2459	84.375	85.7143
RandomForestClassifier	96.4286	90.1639	84.375	90.000
DecisionTreeClassifier	75.8621	72.1311	68.750	72.1311
SVC	90.000	86.8852	84.375	87.0968

TABEAU 3.4 – Comparaison des Performances de Divers modèles de Classification.

Le graphique suivant montre les métriques de performances comme illustré dans la Figure 3.20 .

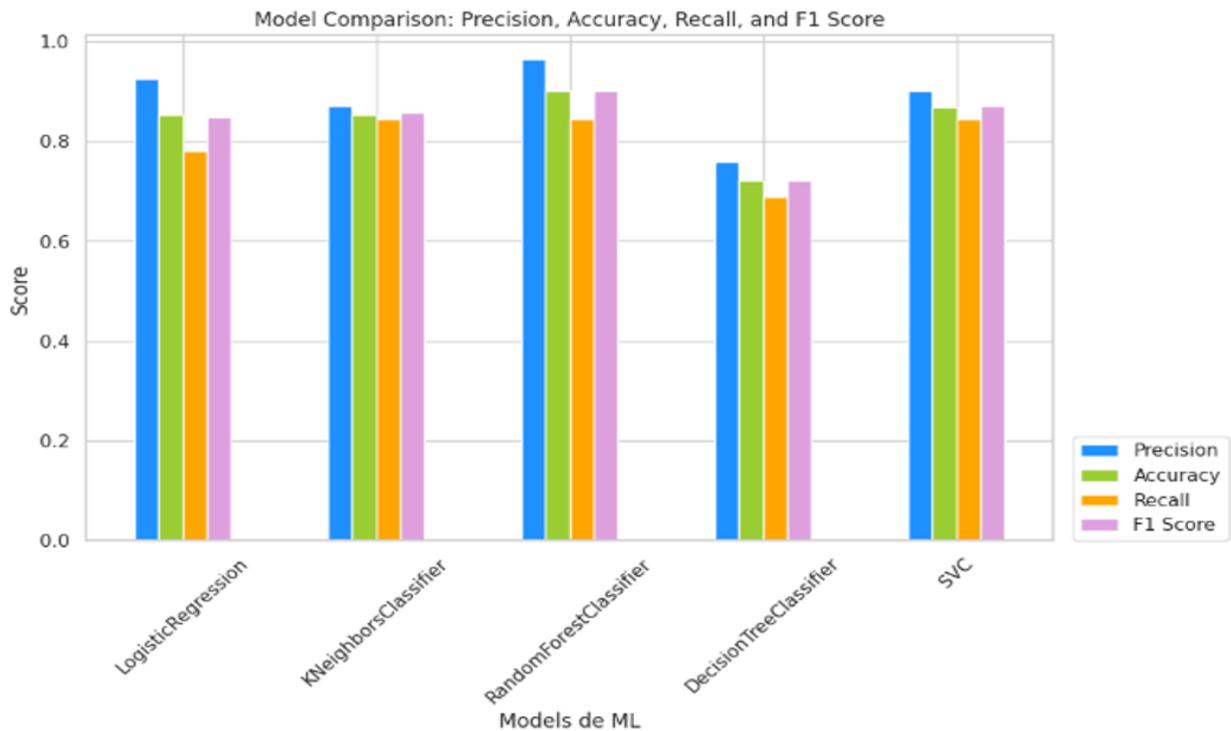


FIGURE 3.20 – Comparaison des métriques de performance avec GridSearchCv.

## 2. Évaluation selon les métriques d’erreurs

Le Tableau 3.5 présente les métriques d’erreurs des différents modèles utilisés.

	MAE	MSE	R2_SCORE
<b>Logistique Régression</b>	0.147541	0.147541	0.408405
<b>KNN</b>	0.147541	0.147541	0.408405
<b>Random Forest Classifier</b>	0.098361	0.098361	0.605603
<b>Decision Tree Classifier</b>	0.278689	0.278689	-0.117457
<b>SVC</b>	0.131148	0.131148	0.474138

TABLEAU 3.5 – Analyse comparative des modèles selon les métriques d’erreurs.

La figure ci-dessus présente un graphique comparatif qui illustre les métriques d’erreurs

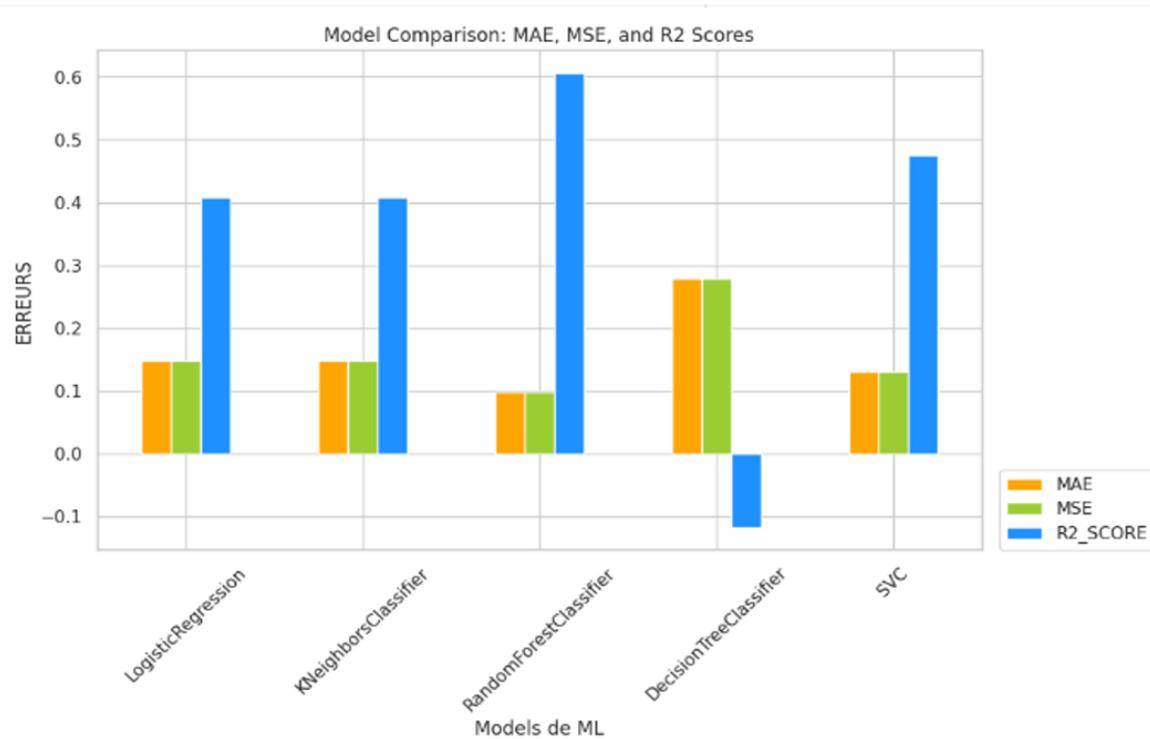


FIGURE 3.21 – Analyse des erreurs des modèles suite à l’utilisation de GridSearchCv.

### 3. Selon le temps d’exécution

La durée d’exécution de chaque modèle d’apprentissage automatique est indiqué dans ce tableau

Modèle	Execution Time (seconds)
LogisticRegression	0.006768
KNeighborsClassifier	0.001566
RandomForestClassifier	0.111436
DecisionTreeClassifier	0.002153
SVC	0.002782

TABEAU 3.6 – Évaluation des résultats des modèles En fonction de la durée d’exécution.

Le temps d’exécution de chaque algorithme est illustré dans la figure suivante :

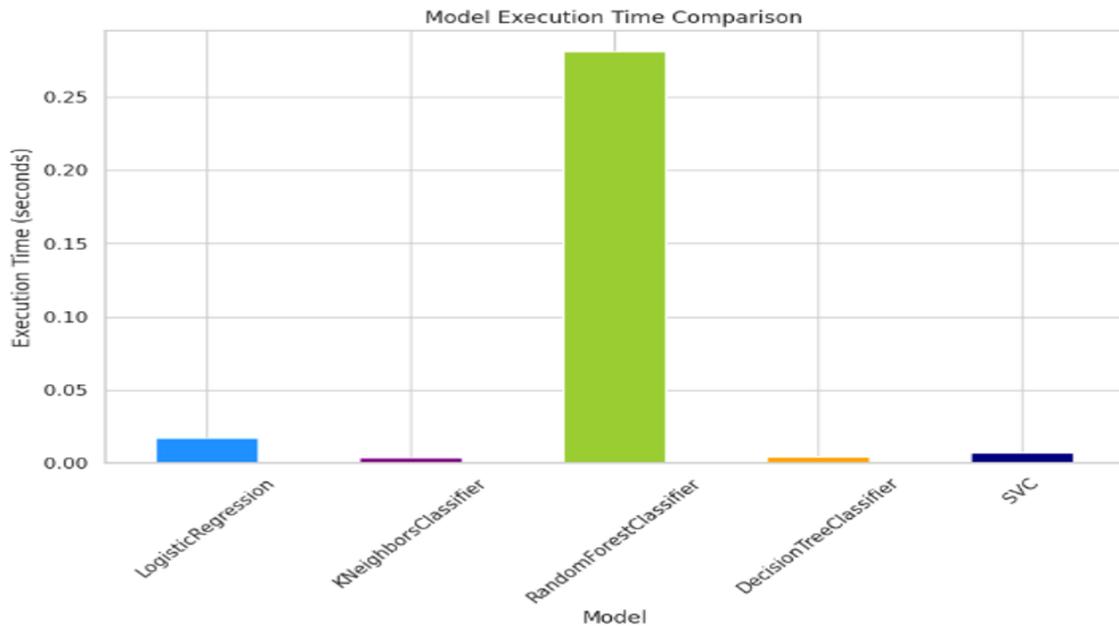


FIGURE 3.22 – Analyse de la durée d’exécution des modèles employés.

## 3.7 Interprétation des résultats et discussion

À cette étape, nous allons interpréter et échanger sur les résultats obtenus.

### 3.7.1 Sans l’utilisation de GridSearchCV

#### 3.7.1.1 Comparaison des Performances des Algorithmes

Après avoir analysé les résultats obtenus sans l’utilisation de GridSearchCv, plusieurs observations peuvent être faites. En termes de précision ,d’accuracy et F1 Score, le KNN a eu le score le plus élevé par rapport aux autres algorithmes avec 93.33%,90.16% et 90.32% respectivement.

En termes rappel, la régression logistique a le plus score élevé avec 90.62%. Le modèle Decision Tree ses performances étaient les plus médiocres parmi les modèles testés.

#### 3.7.1.2 Taux d’Erreur

En ce qui concerne le MSE, MAE, Le KNN a les plus faibles taux d’erreurs et le meilleur en R2 SCORE avec 60.56%, ce qui montre sa performance et sa précision parmi les autres modèles en termes d’erreurs.

En revanche, Decision Tree a les taux d’erreurs les plus élevés, ce qui montre une performance inférieure.

### 3.7.1.3 Temps d'Exécution

L'analyse des temps d'exécution montre que les modèles Decision Tree , SVC et le KNN sont les plus rapides. En revanche, RandomForest prend beaucoup de temps pour son exécution en raison de sa complexité et du nombre d'arbres générés.

### 3.7.1.4 Évaluation Globale

En résumé, sans l'utilisation de l'optimiseur des hyperparamètres GridSearchCv le modèle KNN est le plus performant, car il a obtenu le score le plus élevé en termes de précision, accuracy et F1 Score. Il présente les taux d'erreurs les plus faibles en MSE, MAE et le meilleure en R2 Score. La régression logistique a obtenu le meilleur score en rappel.

En revanche , RandomForest a donné des bonnes performances, mais nécessite un temps d'exécution plus long. Le modèle Decision Tree a les performances les plus faibles sur toutes les métriques évaluées, mais c'est le plus rapide en termes de temps d'exécution.

## 3.7.2 Avec l'utilisation de GridSearchCv

Avec l'utilisation de l'optimiseur des hyper paramètres GridSearchCV, On a démontré les résultats des améliorations des métriques de performance pour plusieurs algorithmes.

### 3.7.2.1 Comparaison des Performances des Algorithmes

Les modèles RandomForest et régression logistique ont montré une amélioration précise en termes de précision avec 96.42% et 92.59% respectivement.

En termes d'accuracy, le RandomForest et le SVC ont des résultats satisfaisantsobtenus avec 90.16% et 86.88% respectivement.

Le Random Forest a été le modèle le plus élevé en F1 score avec 90%. Et en ce qui concerne le rappel, les modèles KNN, SVC et Random Forest ont obtenu un rappelélevé de 84.37%.

Le modèle Decision Tree, ses performances étaient les plus faibles parmi les modèles testés.

### 3.7.2.2 Taux d'Erreur

Les taux d'erreur les plus bas (MAE de 0.098361 et MSE de 0.098361) et le meilleur R2 SCORE (0.605603) ont été obtenus par le RandomForest,

### 3.7.2.3 Temps d'Exécution

Les modèles KNN, Decision Tree et SVC sont restés les plus rapides à s'exécuter même avec l'utilisation de l'optimiseur de GridSearchCv.

Random Forest prend beaucoup de temps pour s'exécuter en raison de sa complexité accrue de l'optimisation des hyper paramètres.

### 3.7.2.4 Évaluation Globale

Avec GridSearchCv, le RandomForest a obtenu les meilleurs résultats en précision, accuracy, F1 Score et en R2 Score, avec les taux d'erreurs les plus bas.

La régression logistique a également montré une amélioration notable en précision.

En termes de temps d'exécution, les modèles KNN, SVC et Decision Tree sont restés les plus rapides malgré l'utilisation de GridSearchCv.

## 3.7.3 Prédiction Des résultats

### 3.7.3.1 Prédiction sans utiliser GridSearchCv

La prédiction avec le meilleur algorithme KNN en termes de précision sont mentionnés dans la Figure 3.23 pour 30 personnes des données de test.

	Patient	Réelle	Prédite
0	1	En bonne santé	Maladie cardiovasculaire
1	2	Maladie cardiovasculaire	Maladie cardiovasculaire
2	3	Maladie cardiovasculaire	En bonne santé
3	4	Maladie cardiovasculaire	En bonne santé
4	5	Maladie cardiovasculaire	Maladie cardiovasculaire
5	6	Maladie cardiovasculaire	Maladie cardiovasculaire
6	7	Maladie cardiovasculaire	Maladie cardiovasculaire
7	8	Maladie cardiovasculaire	Maladie cardiovasculaire
8	9	Maladie cardiovasculaire	En bonne santé
9	10	En bonne santé	En bonne santé
10	11	En bonne santé	En bonne santé
11	12	En bonne santé	En bonne santé
12	13	Maladie cardiovasculaire	Maladie cardiovasculaire
13	14	Maladie cardiovasculaire	Maladie cardiovasculaire
14	15	Maladie cardiovasculaire	Maladie cardiovasculaire
15	16	En bonne santé	En bonne santé
16	17	En bonne santé	En bonne santé
17	18	Maladie cardiovasculaire	Maladie cardiovasculaire
18	19	Maladie cardiovasculaire	Maladie cardiovasculaire
19	20	En bonne santé	En bonne santé
20	21	En bonne santé	En bonne santé
21	22	En bonne santé	En bonne santé
22	23	Maladie cardiovasculaire	Maladie cardiovasculaire
23	24	En bonne santé	En bonne santé
24	25	Maladie cardiovasculaire	Maladie cardiovasculaire
25	26	En bonne santé	En bonne santé
26	27	En bonne santé	En bonne santé
27	28	Maladie cardiovasculaire	Maladie cardiovasculaire
28	29	Maladie cardiovasculaire	Maladie cardiovasculaire
29	30	Maladie cardiovasculaire	Maladie cardiovasculaire
30	31	En bonne santé	En bonne santé

FIGURE 3.23 – Prédiction avec le KNN sans utiliser GridSearchCV.

Les résultats montrent que le KNN est capable de prédire correctement la majorité des cas des personnes atteintes des maladies cardiovasculaires et pour les personnes en bonne santé.

### 3.7.3.2 Prédiction avec l'utilisation de GridSearchCv

Avec l'optimiseur par GridSearchCV, nous avons déterminé que Le RandomForest est l'algorithme le plus précis. Les prédictions effectuées sur un échantillon de 30 personnes des données de test sont présentées dans la Figure 3.24.

	Réelle	Prédite
0	En bonne santé	En bonne santé
1	Maladie cardiovasculaire	Maladie cardiovasculaire
2	Maladie cardiovasculaire	Maladie cardiovasculaire
3	Maladie cardiovasculaire	En bonne santé
4	Maladie cardiovasculaire	Maladie cardiovasculaire
5	Maladie cardiovasculaire	Maladie cardiovasculaire
6	Maladie cardiovasculaire	Maladie cardiovasculaire
7	Maladie cardiovasculaire	Maladie cardiovasculaire
8	Maladie cardiovasculaire	En bonne santé
9	En bonne santé	Maladie cardiovasculaire
10	En bonne santé	En bonne santé
11	En bonne santé	En bonne santé
12	Maladie cardiovasculaire	Maladie cardiovasculaire
13	Maladie cardiovasculaire	Maladie cardiovasculaire
14	Maladie cardiovasculaire	Maladie cardiovasculaire
15	En bonne santé	En bonne santé
16	En bonne santé	En bonne santé
17	Maladie cardiovasculaire	Maladie cardiovasculaire
18	Maladie cardiovasculaire	Maladie cardiovasculaire
19	En bonne santé	En bonne santé
20	En bonne santé	En bonne santé
21	En bonne santé	En bonne santé
22	Maladie cardiovasculaire	Maladie cardiovasculaire
23	En bonne santé	En bonne santé
24	Maladie cardiovasculaire	Maladie cardiovasculaire
25	En bonne santé	En bonne santé
26	En bonne santé	En bonne santé
27	Maladie cardiovasculaire	Maladie cardiovasculaire
28	Maladie cardiovasculaire	Maladie cardiovasculaire
29	Maladie cardiovasculaire	Maladie cardiovasculaire
30	En bonne santé	En bonne santé
31	Maladie cardiovasculaire	Maladie cardiovasculaire

FIGURE 3.24 – Prédiction avec le RandomForest avec l'utilisation de GridSearchCV.

Les résultats indiquent que le RandomForest optimisé avec GridSearchCv est très performant avec une grande précision dans la prédiction. Les erreurs de la prédiction sont minimales.

Lorsqu'on compare les résultats de prédiction sans utiliser GridSearchCv et avec l'utilisation de GridSearchCv, on remarque que les résultats obtenus sont similaires.

## 3.8 Conclusion

Dans ce chapitre, nous avons proposé notre architecture et ses différentes étapes : la collecte, stockage et traitement des données, ainsi que la visualisation des données.

Nous avons ensuite détaillé la partie Machine Learning, en commençant par le prétraitement, la normalisation, visualisation des données du dataset Cleveland.

Nous avons créé différents modèles de ML et évalué leurs performances avec et sans l'utilisation de l'optimiseur d'hyper paramètres GridSearchCV. Les résultats obtenus ont montré que l'optimiseur GridSearchCV a permis d'augmenter les performances de certains modèles de machine Learning tels que Random Forest et la régression logistique. Les comparaisons effectuées selon les métriques de performances, erreurs et le temps d'exécution de chaque modèle utilisé ont montré les forces et les faiblesses de chaque modèle.

Enfin, nous avons interprété les résultats obtenus et proposer des suggestions pour des travaux futurs comme l'exploration d'autres modèles de Machine Learning et de Deep Learning. L'objectif de ces propositions est d'améliorer l'efficacité des modèles pour une meilleure prédiction et un suivi optimal des personnes atteintes des maladies cardiovasculaires.

# Conclusion et perspectives

L'évolution rapide des technologies d'Intelligence Artificielle et de l'Internet des Objets crée des opportunités dans de nombreux secteurs, y compris la santé.

Tout au début de notre travail, nous nous sommes intéressés à l'étude approfondie des technologies de l'Intelligence Artificielle, du Machine Learning, de l'Internet des Objets et du Cloud Computing. Ces innovations technologiques offrent des perspectives prometteuses, en particulier pour la surveillance et la prévention des maladies cardiovasculaires. À travers trois chapitres principaux, ce travail explore la révolution technologique en cours, l'état de l'art de l'application de l'IoT et de l'IA dans les soins de santé, et propose une architecture innovante pour la surveillance continue des patients cardiaques.

La thématique principale de cette étude est de combiner l'intelligence artificielle et l'Internet des objets afin de concevoir des systèmes de surveillance intelligents. La capacité de ces systèmes à traiter les données des patients en temps réel facilite la détection précoce des anomalies et permet des interventions rapides, réduisant ainsi le nombre de décès liés à ces affections.

Dans ce contexte, nous avons ensuite implémenté les diverses formes d'apprentissage automatique et les algorithmes les plus pertinents pour la prédiction des maladies cardiaques. Nous avons mis en évidence l'importance de tester et d'évaluer plusieurs algorithmes pour déterminer celui qui convient le mieux à notre problématique spécifique selon une haute précision. Les résultats de notre étude ont montré que l'utilisation de l'optimiseur d'hyperparamètres GridSearchCV améliore considérablement les performances des modèles de machine learning, tels que le Random Forest et la régression logistique. Cette amélioration se traduit par une précision accrue, une réduction des erreurs et une exécution plus rapide des modèles, ce qui est important pour des applications en temps réel dans le domaine de la santé.

En outre, notre étude ouvre la voie à des innovations futures. Les recherches futures devraient se concentrer sur l'amélioration des algorithmes, l'intégration de nouvelles sources de données, et le développement de systèmes encore plus robustes et sécurisés. Pour avoir des résultats plus performants et améliorer l'efficacité des modèles de machine learning, on propose les suggestions suivantes :

- **Augmentation du Dataset** : Augmenter la taille du dataset aide à améliorer la généralisation des modèles.
- **Utiliser d'autres algorithmes de machine learning** : Utiliser d'autres algorithmes de machine learning comme les algorithmes d'ensemble learning comme XGBoost ou Gradient Boosting, et de deep learning, comme les réseaux de neurones profonds et les techniques de l'apprentissage non supervisé comme le clustering (Kmeans). La combinaison de plusieurs modèles peut augmenter les performances des résultats.

En conclusion, ce mémoire met en lumière le potentiel énorme de l'IA et de l'IoT pour transformer la gestion des maladies cardiovasculaires. En offrant des outils pour une surveillance précise et continue, ces technologies peuvent non seulement améliorer les soins aux patients, mais aussi réduire les taux de mortalité et les coûts de santé. Il est essentiel de continuer à explorer ces technologies tout en tenant compte des défis éthiques et pratiques pour maximiser leurs bénéfices pour la société.

# Bibliographie

- [1] IEEE. "Internet of Things (IoT) Definitions." *IEEE Internet of Things*. Publié le 27 May 2015. <https://iot.ieee.org/definition.html>. (Consulter Le 22 mars 2024).
- [2] Challal, Y.: "Sécurité de l'internet des objets : vers une approche cognitive et systémique," *Thèse de doctorat en Informatique*. Université de Technologie de Compiègne. 2012.
- [3] Bulletin, IEEE IoT: "Three Major Challenges Facing IoT" March 2017. <https://www.ques10.com/p/30825/enlist-and-explain-various-service-model-and-deplo/>. (Consulter Le 22 mars 2024).
- [4] <https://www.iot-valley.fr/fr/transformer/quest-ce-que-linternet-des-objets>, (Consulter Le 24 Mars 2024).
- [5] N, Soumyalatha: "Study of IoT : Understanding IoT Architecture, Applications, Issues and Challenges." *Conference Paper, Mai 2016*. [https://www.researchgate.net/publication/330501274\\_Study\\_of\\_IoT\\_Understanding\\_oT\\_Architecture\\_Applications\\_Issues\\_and\\_Challenges](https://www.researchgate.net/publication/330501274_Study_of_IoT_Understanding_oT_Architecture_Applications_Issues_and_Challenges).
- [6] Mell.P et Grance.T . Sp 800-145: *the nist definition of cloud computing*. Technical report, Gaithersburg, MD, United States. 2011.
- [7] Verma, M: " Artificial intelligence and its scope in different areas with special reference to the field of education ". Volume 3, Issue 1 ; January 2018 ; Page No." 05-10".
- [8] Ertel, W: "Introduction to Artificial Intelligence". Springer nternational Publishing :Berlin/Heidelberg. Germany, 2018, p.17-.
- [9] Woolf, B P: 'Machine Learning Building Intelligent Interactive Tutors '. Dans Interac-tive, chap. 7, p. 3. 2009.
- [10] Jordan M.I, Mitchell T.M: "Machine learning : Trends, perspectives, and prospects". Science, vol. 349, 2015, pp. 255–260.

- [11] Bendiab, G.: *Sécurité des applications métiers au niveau du Cloud Computing :Contrôle d'accès au niveau des APIs du Cloud Computing. PhD thesis*. Université Abdelhamid Mehri – Constantine 2, 2015.
- [12] *Organisation Mondiale de la Santé. (Mai 2017). Maladies cardiovasculaires. Récupéré sur.* [https://www.who.int/fr/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/fr/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)). (Consulter le 1 avril 2024).
- [13] Buyya, R . Broberg, J & Goscinski A.: *Cloud computing : Principles and paradigms*. volume 87. John Wiley Sons, Hoboken, New Jersey, 2011.
- [14] Edgar, T. W., & Manz D. O.: (2017) . *Machine Learning. Research Methods for Cyber-security*. (pp. 153-173).
- [15] Batta, M: '*Machine Learning Algorithms - A Review*'. 'International Journal of Science and Research (IJSR)', ISSN : 2319-7064, (2018).
- [16] Wittek, P: "*Unsupervised Learning.*" In *Quantum Machine Learning, 2014*. <https://www.sciencedirect.com/topics/computer-science/unsupervised-learning>. (Consulté Le 15 mars 2024.).
- [17] Reddy B.E, Reddy Y, Viswanath P: "*Titre de l'article.*". International Journal of Engineering Technology, February 2018.
- [18] 'Seyed S ', 'Michael S ', ' Enda H': '*Deep Reinforcement Learning : An Overview* '. The College of Engineering and Informatics National University of Ireland, Galway, Republic of Ireland, 23 june 2018.
- [19] ' Chahal A', 'Preeti G': '*Machine Learning and Deep Learning* '. International Journal of Innovative Technology and Exploring Engineering (IJITEE) Volume-8 Issue-12, October 2019, ISSN : 2278-3075 (Online).
- [20] <https://addepto.com/blog/deep-learning-architecture/>,  
EdwinLisowski, 'Deep Learning Architecture',dernier accès 'le 16 mars 2024'.
- [21] Madhavan S, M. Tim J: '*Deep learning rchitectures*' . *IBM Developer*. <https://developer.ibm.com/articles/cc-machine-learning-deep-learning-architectures/>. (Consulter le 16 mars 2024).
- [22] ' Yamashita R', ' Nishio M', ' R K G Do' ' K Togashi': '*Convolutional neural networks : an overview and application in radiology*'. 22 June 2018.
- [23] ' Bergmann D', ' Stryker C'. <https://www.ibm.com/topics/autoencoder>. Dernier accès : '16/03/2024.

- [24] Valavanidis, A: ' *Artificial Intelligence (AI) Applications. he most important technology we ever develop and we must ensure it is safe and beneficial to human civilization I* '. 'April 2023'.
- [25] Janiesch, C., Zschech P. et Heinrich K: " *Apprentissage automatique et apprentissage profond. Marchés électroniques 31 , 685-695* ". . April 2021.
- [26] Hilley, D.: *Cloud computing : A taxonomy of platform and infrastructure-level offerings*. Technical report, Georgia Institute of Technology, 1 April 2009.
- [27] Stewart, E: ' *Deep learning rchitectures* ' . *EM360 Tech*. <https://em360tech.com/tech-article/deep-learning-vs-machine-learning>. (Consulter le 26 mars 2024).
- [28] Khwaja, F: " *AI, Cloud, Internet Of Things, Combined* " .*QuickStart*. <https://www.quickstart.com/artificial-intelligence/ai-cloud-internet-of-things-combined/>. (Consulter le 27 mars 2024).
- [29] Kushida KE, Murray J et Zysman J: *Cloud computing : From scarcity to abundance*. Journal of Industry, Competition and Trade, 2015.
- [30] Frissou, Y: *Challenges. Dans Réplication des données dans le Cloud Computing*. (p. 12). Mémoire de Master Recherche en Informatique, Université A/Mira de Béjaïa, 2017 p. 12.
- [31] Howell, J.: " *Top IoT Technologies.* " *101 Blockchains*. <https://101blockchains.com/top-iot-technologies/>. (Consulter le 14 avril 2024).
- [32] Peters, L: *Cloud Computing Trends for 2019 Récupéré de Network Unlimited*. <https://www.networksunlimited.com/cloud-computing-trends-for-2019/>. Publié le 22 février 2019.
- [33] Chaurasia, N: " *Cloud Service Models.* " *Sprintzeal Blog*. Publié le 27 avril 2023. *Disponible sur* :. <https://www.sprintzeal.com/blog/cloud-service-models>. (Consulté le 28 mars 2024).
- [34] Fu, Arron: *7 Types de Structures de Cloud Computing*. *UniPrint.net* Publié le 03 mars 2017. *Disponible sur* :. <https://www.uniprint.net/fr/7-types-cloud-computing-structures/>. (Consulter le 27 mars 2024).
- [35] Mattioli, J: " *Vers un déploiement industriel de l'intelligence artificielle* " *AFAS*. Publié le 6 octobre 2020. *Disponible sur* :. <https://www.afas.fr/vers-un-deploiement-industriel-de-lintelligence-artificielle/>, (Consulté le 28 mars 2024).

- [36] Karunathilake, S P et Gamage U G: '*Review Article Secondary Prevention of Cardiovascular Diseases and Application of Technology for Early Diagnosis*'. , Consulté le 20 avril 2024.
- [37] Modi, J: "*6 Most Common Causes of Heart Disease.*" *BuzzRx Blog*. Disponible sur : <https://www.buzzrx.com/blog/6-most-common-causes-of-heart-disease>, Consulté le 20 avril 2024.
- [38] Shenton, Parkway: "*Symptoms and Causes of Coronary Artery Disease.*" Disponible sur : <https://www.parkwayshenton.com.sg/conditions-diseases/coronary-artery-disease/symptoms-causes>, Consulté le 18 avril 2024.
- [39] Today, Medical News: "*Coronary Artery Disease : Symptoms, Causes, and Treatment.*" Disponible sur : <https://www.medicalnewstoday.com/articles/184601#outlook>, Consulté le 18 avril 2024.
- [40] Today, Medical News: "*National Cancer Centre Singapore (NCCS). "Adult Congenital Heart Disease (ACHD)." Disponible sur : <https://www.nccs.com.sg/patient-care/conditions-treatments/adult-congenital-heart-disease-achd/>, Consulté le 18 avril 2024.*
- [41] *National Heart Lung, and Blood Institute (NHLBI)*. "*Causes of Peripheral Artery Disease.*" Disponible sur : <https://www.nhlbi.nih.gov/health/peripheral-artery-disease/causes>, Consulté le 18 avril 2024.
- [42] Hôpitaux\_Universitaires\_de\_Genève: "*Facteurs de risque cardiovasculaire.*" Dernière mise à jour le 06 septembre 2023. Disponible sur : <https://www.hug.ch/coeur-vaisseaux-cerveau/facteurs-risque-cardiovasculaire>. Consulté le 18 avril 2024.
- [43] "*People Heart Health*" (2020, 28 octobre). *What Are the Non-Modifiable Risk Factors of Heart Disease? Know*. Disponible sur : <https://www.peoplehearthealth.com/blog/what-are-the-non-modifiable-risk-factors-of-heart-disease-know/>. Consulté le 18 avril 2024.
- [44] *British Heart Foundation*. (Janvier 2024). *Fiche d'information mondiale sur les maladies cardiaques et circulatoires*. Repéré à. <https://www.bhf.org.uk/-/media/files/for-professionals/research/heart-statistics/bhf-cvd-statistics-global-factsheet.pdf>, Consulté le 20 avril 2024.
- [45] *Agence de Presse Algérienne (APS)*. (31 Janvier 2021). "*Plus de 50% des décès en Algérie dus aux maladies chroniques*". Récupéré sur. <https://www.aps.dz/sante-science-technologie/>

- 116791-plus-de-50-des-deces-en-algerie-dus-aux-maladies-chroniques.  
Consulté le 19 avril 2024.
- [46] Debasish.S, Amrutanshu.P, Abhilash.P Bhimasen.M Adyasha.R Ghanashyam.S Manoranjan.D: “*Enhancing Cardiovascular Disease Prediction Based on AI and IoT Concepts*”, in *International Journal on Recent and Innovation Trends in Computing and Communication*, November 2023. Disponible sur : <http://dx.doi.org/10.17762/ijritcc.v11i10.8483>.
- [47] Ashfaq.Z, Mumtaz.R, Rafay.A. Zaidi.S.M.H. Saleem.H Mumtaz.S Shahid.A . Poorter.E.D. Moerman I.: “ *Embedded AI-Based Digi-Healthcare* ”, in *Appl. Sci.* 2022, 12, 519. Disponible sur : <https://www.mdpi.com/2076-3417/12/1/519>.
- [48] Alnaggar.M, Handosa.M, Medhat.T RashadM.Z: “ *An IoT-based Framework for Detecting Heart Conditions using Machine Learning*”. in (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 14, No. 4, 2023.
- [49] Ahamed J, Koli AM, Ahmad K Jamal MA Gupta BB: “*CDPS-IoT : Système de prédiction des maladies cardiovasculaires basé sur l'apprentissage automatique.*”. publié le 7 septembre 2021.
- [50] *L'interprétabilité en apprentissage machine. Un regard sur les réseaux de neurones artificiels profonds.* - *Scientific Figure on ResearchGate*. Disponible sur : [https://www.researchgate.net/figure/Reseau-de-neurones-recurrents\\_fig5\\_343135807](https://www.researchgate.net/figure/Reseau-de-neurones-recurrents_fig5_343135807), [consulté le 26 mai 2024].
- [51] Shah, S: (15 Mars 2022). “*Convolutional Neural Network : An Overview.*” *Analytics Vidhya*. Disponible sur : <https://www.analyticsvidhya.com/blog/2022/01/convolutional-neural-network-an-overview/>, [consulté le 26 mai 2024].
- [52] “*Autoencoders : A Comprehensive Guide.*” *V7 Labs Blog*. Disponible sur : <https://www.v7labs.com/blog/autoencoders-guide>. Consulté le 26 mai 2024.
- [53] “*Google Colab : Tout savoir.*” *DataScientest*. Disponible sur : <https://datascientest.com/google-colab-tout-savoir#:~:text=Google%20Colab%20permet%20d'acc%C3%A9der,stock%C3%A9s%20sur%20votre%20Google%20Drive>. Consulté le 3 juin 2024.
- [54] *Python Documentation*. Disponible sur : <https://docs.python.org/3/tutorial/>. Consulté le 3 juin 2024.
- [55] *Python Doctor*. Disponible sur : <https://python.doctor/>. Consulté le 3 juin 2024.

- [56] *Data Bird*. "Bibliothèque Python." Disponible sur : <https://www.data-bird.co/blog/bibliotheque-python>. Consulté le 3 juin 2024.
- [57] *Ritwikb3*. "Heart Disease Cleveland Dataset." Kaggle. Disponible sur : <https://www.kaggle.com/datasets/ritwikb3/heart-disease-cleveland>. Consulté le 3 juin 2024.
- [58] *Geek Python*. "How to Use Sklearn StandardScaler." Disponible sur : <https://geekpython.in/how-to-use-sklearn-standardscaler#:~:text=StandardScaler%20is%20used%20to%20standardize,to%20differences%20in%20feature%20scales>. Consulté le 3 juin 2024.
- [59] Nelson, D: "Quels sont les  $k$  plus proches voisins ?" *Unite.AI*. Disponible sur : <https://www.unite.ai/fr/quels-sont-k-voisins-les-plus-proches/>, Consulté le 4 juin 2024.
- [60] *Médium*: "Image de visualisation des données." Disponible sur : [https://miro.medium.com/v2/resize:fit:640/format:webp/0\\*2\\_qzcm2gSe9l67aI.png](https://miro.medium.com/v2/resize:fit:640/format:webp/0*2_qzcm2gSe9l67aI.png), Consulté le 4 juin 2024.
- [61] Kanade, Vijay: "What Is Logistic Regression?" *Spiceworks*. Disponible sur : <https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-logistic-regression/#:~:text=Regression%20Best%20Practices-,What%20Is%20Logistic%20Regression%3F>. Consulté le 4 juin 2024.
- [62] *Spiceworks*: "What Is Logistic Regression?" Disponible sur : <https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-logistic-regression/>. Consulté le 4 juin 2024.
- [63] *Journal du Net*. "Machine à vecteurs de support (SVM) : Définition et cas d'usage." Disponible sur : <https://www.journaldunet.fr/intelligence-artificielle/guide-de-l-intelligence-artificielle/1501879-machine-a-vecteurs-de-support-svm-definition-et-cas-d-usage>, Consulté le 4 juin 2024.
- [64] *JavaTpoint*: "Image de l'algorithme de machine learning Support Vector Machine (SVM)." Disponible sur : <https://static.javatpoint.com/tutorial/machine-learning/images/support-vector-machine-algorithm.png>, Consulté le 4 juin 2024.
- [65] *GeeksforGeeks*: "Decision Tree in Machine Learning." Disponible sur : <https://www.geeksforgeeks.org/decision-tree/>, Consulté le 9 juin 2024.

- [66] DataCamp: *"Decision Tree for Heart Attack Prevention."* Disponible sur : [https://images.datacamp.com/image/upload/v1677504957/decision\\_tree\\_for\\_heart\\_attack\\_prevention\\_2140bd762d.png](https://images.datacamp.com/image/upload/v1677504957/decision_tree_for_heart_attack_prevention_2140bd762d.png), Consulté le 9 juin 2024.
- [67] GeeksforGeeks: *"Random Forest Algorithm in Machine Learning."* Disponible sur : <https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/>, Consulté le 9 juin 2024.
- [68] Nashaat, Mohammed: *"Hyperparameter Tuning with GridSearchCV."* Medium. Disponible sur : <https://medium.com/@mohammednashaat29/hyperparameter-tuning-with-gridsearchcv-8724f215a383>, Consulté le 4 juin 2024.
- [69] AI, LabelF: *"What is Accuracy, Precision, Recall, and F1 Score?"* Disponible sur : <https://www.labelf.ai/blog/what-is-accuracy-precision-recall-and-f1-score>, Consulté le 4 juin 2024.
- [70] Vidhya, Analytics: *"MAE, MSE, RMSE, Coefficient of Determination (Adjusted R-squared) : Which Metric is Better?"* Medium. Disponible sur : <https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e>, Consulté le 4 juin 2024.

## *Résumé*

Le monde contemporain est témoin d'une révolution technologique majeure qui touche divers aspects de la vie quotidienne, telles que l'intelligence artificielle (IA), le Machine Learning et l'Internet des Objets (IoT) dans la surveillance des maladies cardiaques. À l'ère actuelle, ces technologies permettent de traiter de vastes quantités de données pour créer des modèles prédictifs et des systèmes intelligents, améliorant ainsi la prédiction, la classification et la prise de décision dans divers domaines, y compris la santé.

Notre étude propose une approche novatrice pour le suivi des patients cardiaques, exploitant les capteurs IoT pour collecter des données en temps réel et les techniques de machine learning pour analyser et prédire les crises cardiaques. Les résultats obtenus confirment l'efficacité de cette méthode pour améliorer la surveillance continue et la gestion des patients souffrant de maladies cardiovasculaires.

**Mots-Clés :** IOT ; ML ; IA ; Cloud Computing ; Santé ; cardiaques ; maladies cardiovasculaires.

## *Abstract*

Today's world is witnessing a major technological revolution affecting various aspects of everyday life, such as artificial intelligence (AI), machine learning and the Internet of Things (IoT) in the monitoring of heart disease. In the current era, these technologies enable vast amounts of data to be processed to create predictive models and intelligent systems, improving prediction, classification and decision-making in various fields, including healthcare.

Our study proposes an innovative approach for monitoring heart patients, exploiting IoT sensors to collect data in real time and machine learning techniques to analyse and predict heart attacks. The results confirm the effectiveness of this method in improving the ongoing monitoring and management of patients suffering from cardiovascular disease.

**Keywords :** IOT ; ML ; IA ; Cloud Computing ; Health ; Cardiac ; Cardiovascular disease.