



Ministère de l'Enseignement Supérieur et de la
Recherche Scientifique

Université Echahid Cheikh Larbi Tébessi – Tébessa

Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie



Département : *Mathématiques et Informatique*

Mémoire de fin d'étude

Pour l'obtention du diplôme de MASTER

Domaine : Mathématiques et Informatique

Filière : Informatique

Option : Systèmes d'information

Thème

***Système de surveillance intelligent basé sur l'IoT pour la
prédiction des maladies cardiaques***

Réalisé Par :

Raoued Chabou

Devant le jury :

- | | | | |
|------------------------------|-----|------------------------------------|--------------|
| ● Dr. Ahmed Zeggari | MCA | Université Larbi Tébessi- Tébessa | Président |
| ● Dr. Khaoula Tabet | MCB | Université Laarbi Tébessi- Tébessa | Examineur |
| ● Dr. Mohamed Yassine Aaouam | MCA | Université Larbi Tébessi- Tébessa | Encadreur |
| ● Dr. Issam Bendib | MCA | Université Larbi Tébessi- Tébessa | Co-Encadreur |

Date de soutenance: 05/06/2023

Thème

*Système de surveillance intelligent basé sur l'IoT
pour la prédiction des maladies cardiaques*

Raoued Chabou

Université de Larbi Tébessi

Dédicaces

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

C'est à ceux qui croient à l'expression de ces paroles que je dédie notre travail

*A mon très **cher père.***

*A ma très **chère mère.***

Qui sans eux, je ne serais pas là où j'en suis maintenant, pour tous leurs sacrifices, leurs prières, leurs amours et leurs soutiens tout au long de mon parcours académique.

A mes très chères sœurs et frères

A tous mes amis sans exception avec qui j'ai passé d'agréables moments.

REMERCIEMENT

Tout d'abord, je remercie Allah Tout Puissant, qui nous a donné la force et la patience pour mettre en œuvre ce projet.

Deuxièmement, je tiens à remercier chaleureusement le Dr. Haouam Mohamed Yassine et le Dr. Bendib Issam pour la qualité de leur encadrement, leur patience, leur rigueur et leur disponibilité lors de la préparation de ce mémoire.

Nous remercions les membres du jury d'avoir accepté de juger ce travail

Résumé

L'utilisation de l'IoT et de l'apprentissage automatique peut améliorer le diagnostic et la prédiction des maladies cardiaques. L'IoT permet une collecte de données en temps réel, une surveillance à distance et des interventions médicales à distance. Cependant, la sécurité et la confidentialité des données sont des défis importants. Différents algorithmes d'apprentissage automatique ont été testés, et le modèle Random Forest avec une normalisation Min-Max des données a donné les meilleures performances avec une précision de 94,70% et une exactitude de 94,75%.

Il est important de continuer à explorer d'autres modèles et techniques pour améliorer les performances du système.

Mots clés: IOT , Machine Learning, Prédiction , Maladies cardiaques

Abstract

The use of IoT and machine learning can improve the diagnosis and prediction of cardiac diseases. IoT enables real-time data collection, remote monitoring, and remote medical interventions. However, data security and confidentiality are significant challenges. Various machine learning algorithms have been tested, and the Random Forest model with Min-Max data normalization has shown the best performance with a precision of 94.70% and an accuracy of 94.75%.

It is important to continue exploring other models and techniques to enhance system performance.

Keywords: IoT, Machine Learning, Prediction, Cardiac Diseases.

ملخص

استخدام الإنترنت الأشياء والتعلم الآلي يمكن أن يحسن تشخيص وتنبؤ الأمراض القلبية. تسمح الإنترنت الأشياء بجمع البيانات في الوقت الحقيقي والمراقبة عن بُعد والتدخل الطبي عن بُعد. ومع ذلك، فإن الأمان وسرية البيانات تُعدُّ تحديات هامة. تم اختبار خوارزميات تعلم آلي مختلفة، وأظهر نموذج الغابة العشوائية مع تطبيع أعطى أفضل أداء مع دقة عالية تبلغ % 94.70 وضبط % 94.75

من المهم الاستمرار في استكشاف نماذج وتقنيات أخرى لتحسين أداء النظام.

كلمات مفتاحية: الإنترنت الأشياء، التعلم الآلي، التنبؤ، الأمراض القلبية.

TABLE DES MATIÈRES

Introduction générale	1
IOT pour la surveillance cardiaque	3
1.1 Introduction	3
1.2 Les Maladies Cardiaques	3
1.2.1 Anatomie du cœur	3
1.2.2 Type des maladies cardiaques	4
1.2.3 Causes des maladies cardiaques	5
1.2.4 Diagnostic des maladies cardiaque	6
1.3 Internet des Objets (IoT)	9
1.3.1 L'architecture de l'IoT:	10
1.3.2 Avantages et inconvénients de l'IoT	12
1.4 Internet of Medical Things (IoMT)	14
1.4.1 Applications de l'IoMT	14
1.4.2 Défis et considérations	15
1.5 Conclusion	16
Apprentissage automatique	17
2.1 Introduction	17
2.2 L'apprentissage automatique	18
2.3 Les différents types d'apprentissage	18
2.3.1 Apprentissage supervise (Supervised Learning)	18
2.3.2 Apprentissage non supervise (Unsupervised Learning)	19
2.3.3 Apprentissage par renforcement	20
2.4 Algorithmes de Machine Learning	20
2.4.1 Random Forest	20
2.4.2 Les Machines à Vecteurs de Support (SVM)	22
2.4.3 Les arbres de décision	23
2.4.4 Les k plus proches voisins	24
2.5 Domaines d'application de ML	24
2.6 Les métriques d'évaluations des systèmes de machine learning	25
2.7 Conclusion	25
Prédiction des maladies cardiaques: Méthodes et résultats récents	26
3.1 Introduction	26

3.2 Revue des travaux connexes	26
3.3 Synthèse des travaux connexes	29
3.4 Conclusion	30
Conception et réalisation du système	31
4.1 Introduction	31
4.2 Préparation des données	32
4.2.1 Collecte des données	32
4.2.2 Sélection des fonctionnalités	33
4.2.3 Normalisation des données	33
4.2.4 Séparation des ensembles de données	34
4.3 Réalisation du système	34
4.3.1 Outils et bibliothèques utilisés pour le développement du système proposé	34
4.3.2 Description du dataset	35
4.3.3 Exploration des données	36
4.3.4 Implémentation et résultats	38
4.5 Conclusion	43
Conclusion générale	44
Bibliographie	45

LISTE DES FIGURES

Figure 1.1– Schéma général du cœur.....	4
Figure 1.2–Les ondes du signal ECG.....	7
Figure 1.3–le fonctionnement d'une oxymétrie de pouls.....	8
Figure 1.4–Le stéthoscope électronique.....	9
Figure 1.5–Mesure de la tension artérielle.....	9
Figure 1.6–les 4 étapes de l'architecture IOT.....	11
Figure 2-1. La relation entre IA, ML et DL.....	17
Figure 2-2. L'apprentissage supervisé.....	19
Figure 2-3. L'apprentissage non supervisé.....	19
Figure 2-4. L'apprentissage par renforcement.....	20
Figure 2-5. Fonctionnement de Random Forest.....	21
Figure 2-6. –Exemple de marge maximum(hyperplan valide).....	23
Figure 2-7 – Arbre de décision.....	23
Figure 4.1- architecture globale de système.....	32
Figure 4-2. Extrait du dataset Cleveland.....	35
Figure 4-3– Graphe illustre les deux classes	37
Figure 4.4– Matrice de corrélation de l'ensemble de Train	38

LISTE DES TABLEAUX

Tableau 3-1. Synthèse des travaux connexes sur la prédiction des maladies cardiaques	29
Tableau 4-1 Évaluation des résultats par kernel.....	38
Tableau 4-2-Résultats SVC sans et avec normalisation.....	39
Tableau 4-3-Evaluation des résultats selon le nombre d'estimateurs.....	40
Tableau 4-4-Résultats RF sans et avec normalisation.....	40
Tableau 4-5- Evaluation de l'algorithme KNN selon l'hyperparamètre K.....	41
Tableau 4-6- Résultats RF sans et avec normalisation.....	41
Tableau 4-7-Evaluation de l'algorithme de l'arbre de décision.....	42
Tableau 4-8-Résultats AD sans et avec normalisation.....	42

INTRODUCTION GÉNÉRALE

L'internet des objets (IoT) et l'apprentissage automatique (machine learning) ont le potentiel de révolutionner le domaine de la santé, en particulier en ce qui concerne les maladies cardiaques. Grâce à l'IoT, il est désormais possible de collecter en temps réel une multitude de données sur la santé des individus, à l'aide de capteurs et de dispositifs portables. Ces données comprennent des informations telles que la fréquence cardiaque, la pression artérielle, l'activité physique et bien d'autres paramètres.

En utilisant des algorithmes de machine learning avancés, ces données peuvent être analysées pour détecter des schémas et des anomalies qui pourraient indiquer la présence de maladies cardiaques. Par exemple, en surveillant les variations de la fréquence cardiaque d'une personne sur une période prolongée, il est possible de détecter des signes précurseurs de problèmes cardiaques tels que des arythmies ou des battements irréguliers.

De plus, l'apprentissage automatique peut être utilisé pour développer des modèles prédictifs qui évaluent le risque individuel de développer une maladie cardiaque. En utilisant des ensembles de données massifs et des techniques d'apprentissage profond, ces modèles peuvent prendre en compte une multitude de facteurs tels que l'âge, le sexe, les antécédents familiaux, le mode de vie et les habitudes alimentaires pour fournir des estimations personnalisées de probabilité de développer une maladie cardiaque.

En intégrant l'IoT et le machine learning, les médecins et les professionnels de la santé peuvent bénéficier de données en temps réel et de modèles prédictifs précis pour diagnostiquer précocement les maladies cardiaques, optimiser les traitements et améliorer les résultats pour les patients. Cela ouvre la voie à une médecine plus proactive et personnalisée, offrant de nouvelles opportunités pour prévenir et gérer les maladies cardiaques de manière plus efficace.

L'objectif principal de ce travail est de mettre en œuvre un système d'IoT pour la prédiction des maladies cardiaques, en exploitant les avantages de la surveillance continue et de l'analyse en temps réel des données physiologiques. En fournissant aux individus des informations précieuses sur leur santé cardiaque et en les alertant en cas de risques potentiels, ce système peut jouer un rôle crucial dans la prévention, la gestion et le traitement des maladies cardiaques.

Le manuscrit est organisé en quatre chapitres:

- **Chapitre 1** : IoT pour la surveillance cardiaque
Dans ce premier chapitre, nous explorons l'application de l'Internet des objets (IoT) dans le domaine de la surveillance cardiaque.
- **Chapitre 2** : Apprentissage automatique
Dans ce deuxième chapitre, Nous présentons les différents types d'algorithmes utilisés, tels que les arbres de décision et les algorithmes de classification, et leur rôle dans l'analyse des données cardiaques pour la prédiction des maladies cardiaques.
- **Chapitre 3** : Travaux connexes.
Le troisième chapitre examine les travaux de recherche connexes et les avancées déjà réalisées dans le domaine de la surveillance cardiaque basée sur l'IoT et l'apprentissage automatique. Nous passons en revue les études et les projets existants, en mettant en évidence les approches innovantes, les résultats obtenus et les limitations éventuelles. Cela permet de situer notre travail dans le contexte de la recherche actuelle.
- **Chapitre 4** : Implémentation et résultats expérimentaux.
Dans ce dernier chapitre, nous détaillons la conception et l'implémentation de notre système de surveillance cardiaque basé sur l'IoT et l'apprentissage automatique. Nous décrivons l'architecture du système, les choix technologiques effectués et les étapes de développement. Nous présentons également les résultats expérimentaux obtenus en utilisant notre système, en évaluant sa précision et son efficacité dans la prédiction des maladies cardiaques.

En **conclusion**, nous résumons les principales contributions de notre travail et les perspectives futures.

CHAPITRE 1

IOT POUR LA SURVEILLANCE CARDIAQUE

1.1 Introduction

Les maladies cardiovasculaires regroupent différents troubles qui affectent le cœur et les vaisseaux sanguins. En raison de leur impact considérable sur la santé et la vie de nombreuses personnes à travers le monde, les chercheurs se sont concentrés sur la réduction des risques associés à ces affections ainsi que sur la capacité de prédire leur apparition.

Dans ce chapitre introductif, nous examinerons de manière approfondie les maladies cardiaques, en mettant l'accent sur leurs différents types, leurs causes et leurs symptômes. Nous aborderons également l'Internet des objets (IoT) et ses caractéristiques, ainsi que son architecture. Nous discuterons des domaines d'application de l'IoT en lien avec les maladies cardiovasculaires, en mettant en évidence comment cette technologie peut contribuer à la prévention, au diagnostic et au traitement de ces affections.

Enfin, nous concluons ce chapitre en récapitulant les points clés abordés concernant les maladies cardiaques et l'IoT, et en soulignant l'importance de ces domaines dans la lutte contre les maladies cardiovasculaires.

1.2 Les Maladies Cardiaques

Le cœur est la pompe vitale responsable de la circulation du sang dans le corps humain. Cependant, tout comme les autres organes, il est également susceptible de développer des maladies. Les maladies cardiovasculaires, qui affectent le cœur et les vaisseaux sanguins, constituent une préoccupation majeure en matière de santé publique..

1.2.1 Anatomie du cœur

Le cœur est un muscle situé entre les deux poumons au niveau du thorax,. Il est protégé par le sternum et les côtes. Sa position est légèrement décalée vers la gauche de la ligne médiane du corps. Il est entouré par le péricarde, un sac membranaire qui le maintient en place et le protège des frottements contre les autres organes environnants.

Le cœur se décompose en deux grandes régions (Figure 1.1):

- les oreillettes reçoivent les veines, leurs parties supérieures communiquent avec ces poches accessoires, les auricules, qui, comme leur nom l'indique, constituent des sortes d'oreilles ondulées à leur surface.
- les ventricules s'étendent vers la gauche et convergent vers la pointe du cœur.

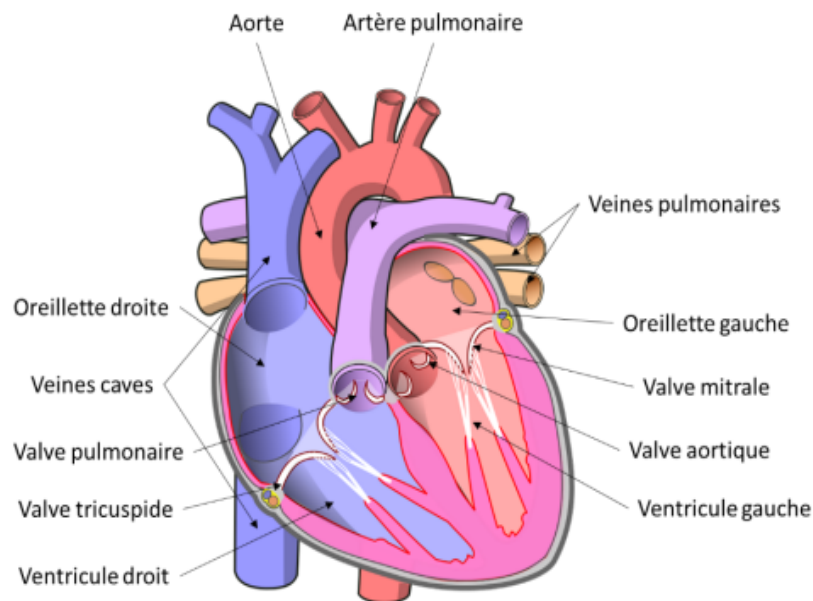


Figure 1.1– Schéma général du cœur.

1.2.2 Type des maladies cardiaques

Les maladies cardiovasculaires (MCV) englobent différents types de troubles touchant le système circulatoire, tels que [1]:

- **Maladie coronarienne:** une affection des vaisseaux sanguins qui alimentent le muscle cardiaque.
- **Maladie cérébrovasculaire:** une pathologie des vaisseaux sanguins qui alimentent le cerveau.
- **Maladie artérielle périphérique:** un trouble des vaisseaux sanguins qui alimentent les bras et les jambes.

- **Cardiopathie rhumatismale:** des dommages au muscle cardiaque et aux valves cardiaques dus au rhumatisme articulaire aigu, provoquée par des bactéries streptococciques.
- **Cardiopathie congénitale:** des malformations congénitales qui affectent le développement et le fonctionnement normaux du cœur en raison d'anomalies structurelles.
- **Thrombose veineuse profonde et embolie pulmonaire:** des caillots sanguins se forment dans les veines des jambes, qui peuvent se déloger et se déplacer vers le cœur et les poumons, entraînant ainsi une embolie pulmonaire.

1.2.3 Causes des maladies cardiaques

Outre les causes congénitales, les principales causes des maladies cardiovasculaires incluant l'hypertension artérielle, le diabète, les troubles des lipides sanguins, le tabagisme, une alimentation déséquilibrée, le manque d'activité physique et le surpoids [2].

- **Hypertension:** Le risque de MCV augmente lorsque la tension artérielle sur les parois des vaisseaux sanguins est trop haute. Le traitement de l'hypertension artérielle réduit l'incidence de la plupart des MCV et diminue le taux de mortalité toutes causes confondues.
- **Diabète :** Le diabète est étroitement lié aux maladies cardiovasculaires, augmentant significativement leur incidence. Les patients diabétiques sont plus susceptibles de développer une maladie cardiovasculaire en raison de l'association entre le diabète et de nombreux autres facteurs de risque cardiovasculaire. Par conséquent, le traitement du diabète joue un rôle essentiel dans la prévention des maladies cardiovasculaires.
- **Dyslipidémie :** La dyslipidémie, caractérisée par des taux anormaux de graisses dans le sang, est associée à un risque accru de maladies cardiovasculaires. La réduction du taux de cholestérol est l'une des stratégies les plus efficaces pour prévenir les maladies cardiovasculaires.
- **Tabagisme :** Le tabagisme constitue un facteur de risque majeur de maladies cardiovasculaires. Il existe une corrélation directe entre le nombre de cigarettes fumées par jour et le risque de maladie cardiovasculaire. Le tabagisme double le

taux de mortalité cardio-vasculaire. Arrêter de fumer est donc le moyen le plus efficace de prévenir les maladies cardiovasculaires.

- **Alimentation malsaine:** Le régime alimentaire joue un rôle important dans le risque de maladies cardiovasculaires en influençant le cholestérol, la tension artérielle, le poids et le diabète. Une consommation excessive de boissons gazeuses est déconseillée. De plus, une alimentation riche en sucre, en sel et en gras saturés est associée aux maladies cardiovasculaires, et il est recommandé de suivre une alimentation équilibrée selon les directives de prévention des maladies cardiovasculaires.

1.2.4 Diagnostic des maladies cardiaque

Le diagnostic des maladies cardiovasculaires implique généralement une évaluation approfondie des antécédents médicaux, des symptômes et des facteurs de risque, ainsi que des tests et des examens complémentaires. Voici quelques éléments clés à considérer lors du diagnostic des maladies cardiovasculaires:

Antécédents médicaux

Les médecins recueillent les informations des patients sur les antécédents familiaux de maladies cardiaques, les antécédents personnels de problèmes de santé, les habitudes de vie, le régime alimentaire, le niveau d'activité physique, etc. dans le but d'établir un diagnostic précis. Ces informations jouent un rôle crucial dans l'évaluation du risque de maladies cardiovasculaires et aident le médecin à comprendre le contexte global de la santé du patient. En recueillant ces informations, le médecin peut identifier les facteurs de risque potentiels, déterminer les éventuelles causes sous-jacentes des symptômes et formuler un plan de traitement adapté. Cela permet de garantir un diagnostic correct et de mettre en place des mesures de prévention et de gestion appropriées pour les maladies cardiovasculaires.

Électrocardiographie (ECG)

C'est une technique non invasive qui mesure les impulsions électriques émises par le cœur lors de sa contraction. Des électrodes sont placées sur la peau du patient pour enregistrer les variations électriques à différents endroits du corps. L'impulsion électrique initiale

provient du nœud sinusal et se propage à travers le cœur, provoquant la contraction des muscles cardiaques. L'ECG produit un graphique des variations du potentiel électrique du cœur au fil du temps. Il permet d'évaluer le rythme cardiaque, l'intervalle entre les battements et la présence de problèmes électriques ou structurels du cœur. C'est un outil essentiel pour diagnostiquer les troubles du rythme cardiaque, les anomalies de la conduction, les infarctus du myocarde, etc..

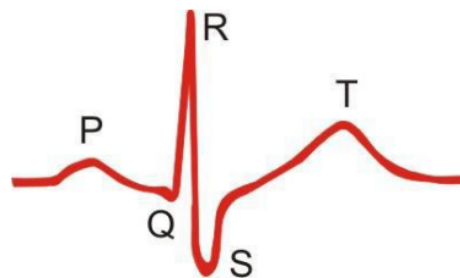


Figure 1.2–Les ondes du signal ECG.

L'onde P correspond à l'impulsion électrique qui engendre la contraction des oreillettes. L'impulsion est légèrement retardée et atteint ensuite les muscles ventriculaires, générant l'onde complexe QRS. L'onde T correspond à la repolarisation des cellules pendant la diastole ventriculaire, préparant ainsi le cœur à un nouveau battement.

Oxymétrie de pouls

L'oxymétrie de pouls, également connue sous le nom de saturation en oxygène, est une méthode de mesure non invasive utilisée pour évaluer la saturation en oxygène de l'hémoglobine dans les capillaires. Elle fournit une mesure appelée saturation pulsée en oxygène (SPO₂), qui est étroitement liée à la saturation artérielle en oxygène (SAO₂).

L'oxygène est présent dans le sang sous deux formes: sous forme dissoute (PaO₂) et sous forme combinée à l'hémoglobine (HbO₂). Cependant, la quantité d'oxygène dissous est considérablement plus faible que celle combinée à l'hémoglobine.

Le principe de fonctionnement des oxymètres de pouls repose sur l'absorption de la lumière rouge et infrarouge par le sang, en fonction de sa saturation en oxygène. Le capteur,

généralement placé à l'extrémité du doigt, comporte un émetteur et un récepteur de lumière (Figure 1-3).

L'émetteur envoie des faisceaux de lumière rouge et infrarouge à travers le tissu vasculaire. Une partie de cette lumière est absorbée par l'hémoglobine oxygénée et désoxygénée présente dans les capillaires. Le récepteur mesure ensuite l'intensité de la lumière transmise à travers le tissu.

En comparant l'absorption de la lumière rouge et infrarouge, l'oxymètre de pouls peut déterminer la proportion d'hémoglobine saturée en oxygène. La SPO2 est calculée en pourcentage, représentant la quantité d'hémoglobine oxygénée par rapport à l'hémoglobine totale.

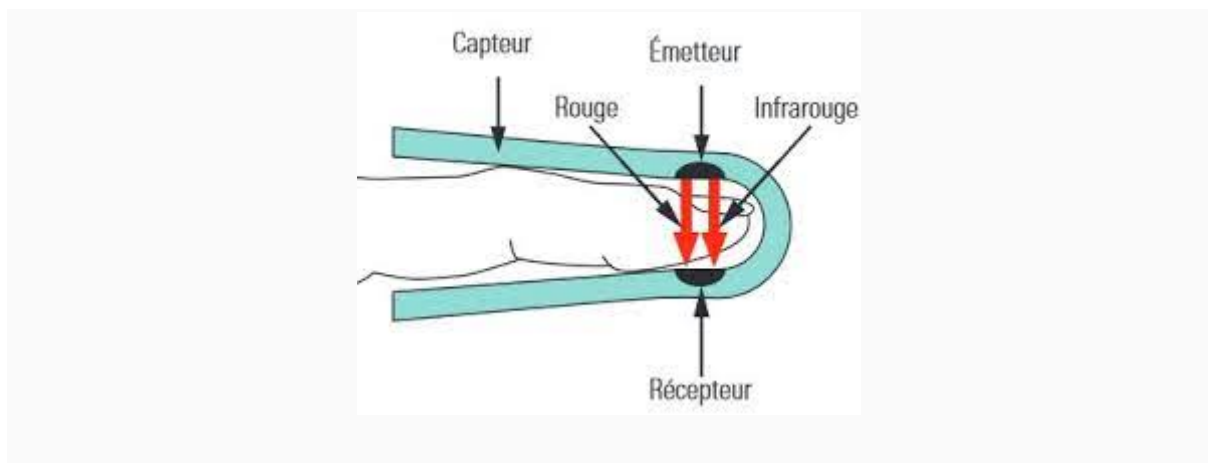


Figure 1-3 Fonctionnement d'une oxymétrie de pouls.

Le stéthoscope électronique

Le stéthoscope est un instrument de diagnostic médical utilisé pour l'auscultation, c'est-à-dire l'écoute des sons internes du corps. Il se compose d'un tube flexible avec des embouts auriculaires et un pavillon qui transmet et amplifie les sons par résonance.

L'auscultation est une technique essentielle utilisée par les médecins pour évaluer le fonctionnement des organes internes et détecter d'éventuelles anomalies.



Figure 1-4 Le stéthoscope électronique

Mesure de la tension artérielle

Les appareils de mesure de la tension artérielle (non invasifs), appelés tensiomètre reposent sur le principe de la détection des ondes de pression artérielle. La méthode utilisée est dite oscillométrique, Elle s'effectue sur les membres du patient.

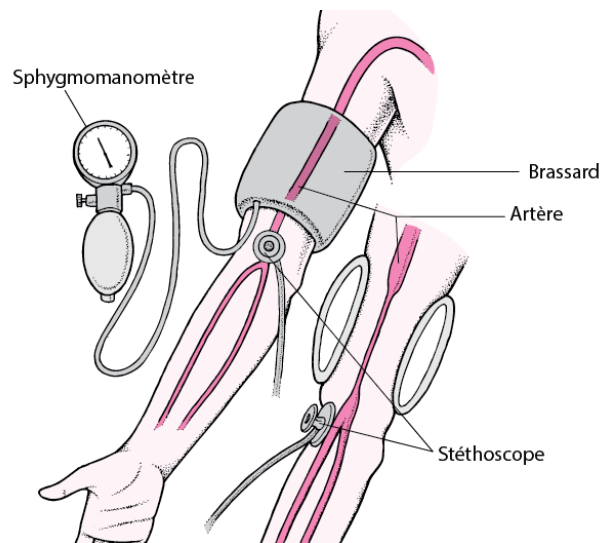


Figure 1-5 Mesure de la tension artérielle

1.3 Internet des Objets (IoT)

l'IoT (Internet des Objets) est l'acronyme de "Internet of Things" en anglais. Selon l'UIT (Union Internationale des Télécommunications), l'IoT est défini comme une infrastructure

mondiale pour la société de l'information, permettant de disposer de services évolués en interconnectant des objets, physiques ou virtuels, grâce aux technologies de l'information et de la communication interopérables existantes ou en évolution [3].

L'internet des objets est une technologie émergente qui vise à créer un réseau global de machines et d'appareils interconnectés, capables de communiquer et d'échanger des données entre eux par le biais d'Internet. Cela signifie que des objets ordinaires, tels que des appareils électroménagers, des véhicules, des capteurs, des dispositifs médicaux, des équipements industriels, peuvent être connectés à Internet et échanger des informations de manière autonome.

L'IoT permet de collecter des données provenant de ces objets connectés, de les analyser et de les utiliser pour prendre des décisions, améliorer les processus et optimiser les performances. Par exemple, dans le domaine de la santé, les dispositifs portables peuvent surveiller les signes vitaux d'un patient et envoyer ces données en temps réel à un système de suivi médical.

1.3.1 L'architecture de l'IoT:

L'architecture de l'IoT comprend plusieurs briques de systèmes connectés pour assurer que les données générées par les capteurs des objets sont collectées, stockées et traitées dans des entrepôts de données volumineuses (big data warehouses), et que les actionneurs des objets exécutent les commandes envoyées via une application utilisateur [4]. La figure 1.6 présente une architecture détaillée de l'IOT. Les fonctionnalités de chaque couche sont décrites ci-dessous:

Couche de perception

Cette couche représente les objets physiques équipés de capteurs qui collectent des données. Ces capteurs peuvent être intégrés dans différents types d'objets, tels que des véhicules, des équipements industriels, des appareils domestiques, des dispositifs portables, etc. Ils mesurent des paramètres tels que la température, l'humidité, la pression, la localisation, etc..

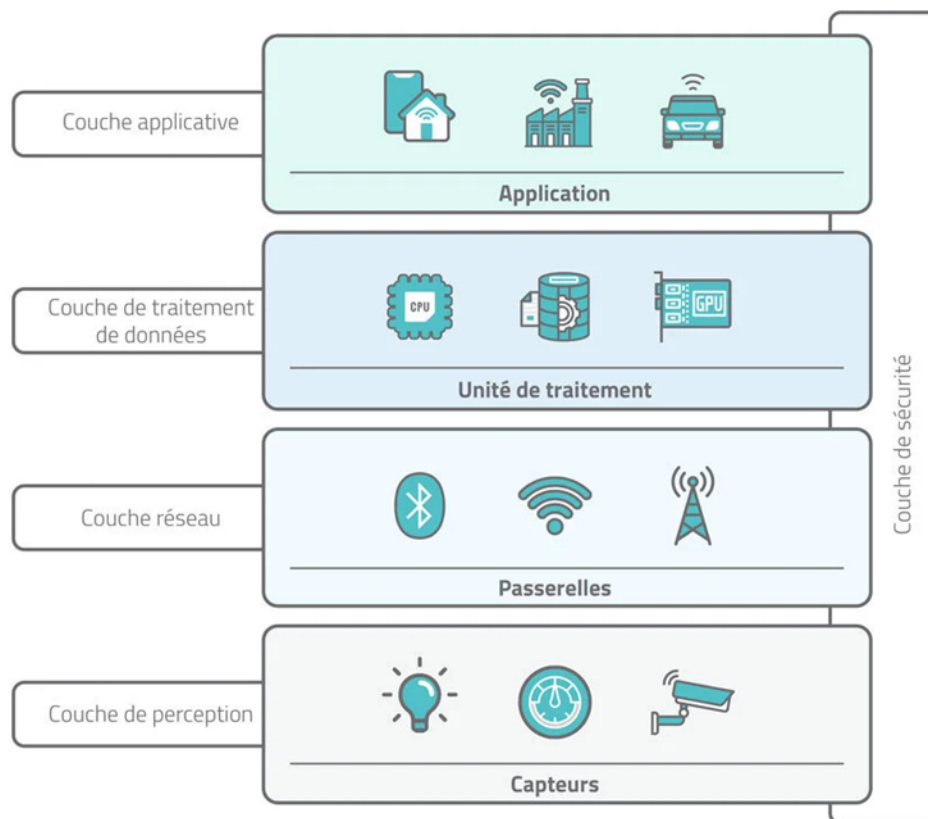


Figure 1-6 Architecture de l'IoT [24]

Couche réseau

La couche réseau joue un rôle crucial dans l'architecture de l'IoT. Elle est responsable de la connectivité entre les différents appareils et objets intelligents, les serveurs et les appareils réseau. Sa fonction principale est de permettre la transmission des données collectées par les appareils connectés.

La couche réseau utilise diverses technologies de communication pour établir des connexions et transmettre les données. Cela peut inclure des réseaux sans fil tels que le Wi-Fi, le Bluetooth, les réseaux cellulaires (comme la 4G ou la 5G), ou des connexions filaires comme l'Ethernet.[25]

Couche de traitement de données

Cette couche est responsable du stockage et du traitement des grandes quantités de données générées par les objets connectés. Elle utilise des technologies de stockage distribué et des outils d'analyse de données pour extraire des informations utiles à partir de ces données massives. Les données collectées peuvent être utilisées pour la prise de décision, la génération de rapports, l'apprentissage automatique (machine learning), etc.

Couche applicative

Cette couche représente les interfaces utilisateur et les applications qui permettent aux utilisateurs d'interagir avec les objets connectés et d'exploiter les données collectées. Cela peut inclure des applications mobiles, des tableaux de bord en ligne, des systèmes de surveillance, des applications industrielles, des systèmes de gestion énergétique, etc.

Actuellement, les applications peuvent être construites directement sur les plateformes IoT qui offrent une infrastructure de développement logiciel avec des outils prêts à l'emploi pour l'exploration de données, l'analyse avancée et la visualisation de données.

Les plateformes IoT fournissent une infrastructure de développement logiciel qui simplifie la création d'applications dans l'environnement IoT. Elles proposent des outils prêts à l'emploi pour l'exploration des données, l'analyse avancée et la visualisation des données. Ces plates-formes facilitent le déploiement rapide d'applications IoT et permettent aux développeurs de se concentrer sur la création de fonctionnalités spécifiques aux besoins des utilisateurs.

1.3.2 Avantages et inconvénients de l'IoT

L'IoT présente plusieurs avantages et inconvénients. Voici quelques-uns des principaux [26]:

Avantages de l'IoT :

- **Surveillance à distance:** L'IoT permet la surveillance à distance des patients, ce qui peut être particulièrement utile pour les patients atteints de maladies chroniques ou les personnes âgées qui ont besoin d'une surveillance continue.
- **Prévention des maladies:** Les dispositifs IoT peuvent aider à prévenir les maladies en surveillant les habitudes de vie et en fournissant des rappels et des conseils personnalisés pour maintenir une bonne santé. Par exemple, les montres connectées

peuvent suivre l'activité physique, la fréquence cardiaque et les calories brûlées, ce qui encourage les utilisateurs à adopter un mode de vie sain.

- **Amélioration des soins médicaux:** L'IoT permet une meilleure gestion des soins médicaux en fournissant aux professionnels de santé des données précises et en temps réel sur l'état des patients. Cela facilite le suivi des patients, la détection précoce des problèmes de santé et l'ajustement rapide des traitements.
- **Prise de décision basée sur les données:** L'IoT génère une quantité massive de données qui peuvent être utilisées pour prendre des décisions éclairées dans le domaine de la santé. L'analyse de ces données peut aider à identifier les tendances, les schémas et les facteurs de risque, ce qui peut contribuer à une meilleure prise en charge des patients et à des décisions médicales plus précises.

Inconvénients de l'IoT:

- **Sécurité:** L'un des principaux défis de l'IoT est la sécurité. Les dispositifs connectés peuvent être vulnérables aux attaques malveillantes, notamment le piratage, l'interception de données et les violations de la vie privée. Il est crucial de mettre en place des mesures de sécurité robustes pour protéger les données et les systèmes contre les cyberattaques.
- **Confidentialité des données :** L'IoT implique la collecte massive de données provenant de divers capteurs et appareils. Cela soulève des préoccupations quant à la confidentialité des données personnelles
- **Interopérabilité :** L'IoT implique souvent l'utilisation de différents appareils et technologies provenant de différents fabricants. Cela peut entraîner des problèmes d'interopérabilité, où les appareils ne peuvent pas communiquer efficacement entre eux. Des normes et des protocoles de communication communs doivent être établis pour assurer une interopérabilité fluide.
- **Complexité et coûts :** La mise en place d'un système IoT peut être complexe et coûteuse. Cela nécessite des investissements en matière d'infrastructure, de matériel, de logiciels et de maintenance.
- **Dépendance à la connectivité réseau :** L'IoT dépend d'une connectivité réseau stable et fiable pour fonctionner correctement. Tout problème de connectivité peut entraîner des interruptions de service et affecter les performances de l'IoT.

1.4 Internet of Medical Things (IoMT)

L'Internet of Medical Things (IoMT), également connu sous le nom d'Internet des objets médicaux, est une extension spécifique de l'Internet des objets (IoT) dans le domaine de la santé. Il se réfère à l'utilisation de dispositifs médicaux et d'objets connectés qui collectent, transmettent et analysent des données médicales pour améliorer les soins de santé.[27]

1.4.1 Applications de l'IoMT

L'IoMT englobe un large éventail d'applications et de technologies dans le domaine de la santé. Voici quelques exemples :

Dispositifs portables

Les dispositifs portables tels que les montres intelligentes, les bracelets et les capteurs peuvent collecter des données sur les signes vitaux tels que la fréquence cardiaque, la pression artérielle, la qualité du sommeil, etc. Ces données peuvent être transmises à des professionnels de la santé pour la surveillance à distance ou l'évaluation de l'état de santé d'un individu.

Implants médicaux connectés

Certains dispositifs médicaux implantables, tels que les stimulateurs cardiaques, les prothèses articulaires et les implants de surveillance, peuvent être connectés à l'IoMT. Cela permet de surveiller en temps réel le fonctionnement de ces implants, de recueillir des données sur leur performance et de faciliter leur maintenance ou leur ajustement si nécessaire.

Systèmes de santé connectés

Les systèmes de santé connectés intègrent différents appareils médicaux et dispositifs IoT pour faciliter la gestion des soins de santé. Cela peut inclure des dossiers médicaux électroniques, des systèmes de gestion de l'inventaire, des systèmes de surveillance des patients en temps réel, des dispositifs de suivi des médicaments, etc. Ces systèmes visent à améliorer l'efficacité, la précision et la sécurité des soins de santé.

Télésanté

L'IoMT facilite également la prestation de services de santé à distance, tels que la téléconsultation et la télésurveillance. Les patients peuvent consulter des médecins à

distance, partager des données médicales en temps réel et recevoir des recommandations ou des conseils médicaux, ce qui peut être particulièrement utile pour les personnes vivant dans des régions éloignées ou ayant des difficultés d'accès aux soins de santé.

En permettant le suivi à distance et en évitant les visites médicales fréquentes ou les hospitalisations prolongées, l'loMT peut réduire les coûts de soins de santé pour les patients, les fournisseurs de soins et les systèmes de santé en général.

Bien-être et soins préventifs

L'loMT offre des possibilités pour la promotion des soins préventifs. Les dispositifs compatibles loMT, tels que les appareils portables, les applications mobiles et les dispositifs de surveillance à domicile, permettent le suivi de l'alimentation, de l'activité physique et de la qualité de vie, contribuant ainsi à la prévention des maladies et à l'amélioration du bien-être.

1.4.2 Défis et considérations

L'adoption de l'Internet des Objets Médicaux (loMT) dans les soins de santé présente également des défis et des considérations. Voici certains des défis courants et des aspects à prendre en compte:

Sécurité et confidentialité des données

La protection des données de santé est une préoccupation majeure dans l'loMT. Des mesures de sécurité robustes doivent être mises en place pour prévenir les violations et assurer la confidentialité des informations médicales sensibles.

Interopérabilité des dispositifs

Les nombreux appareils loMT utilisent différents protocoles et normes de communication. L'interopérabilité entre les dispositifs est un point clé pour garantir une intégration transparente des données et une collaboration efficace entre les professionnels de santé.

Formation et adoption

L'intégration de l'IoMT dans les soins de santé nécessite une formation adéquate des professionnels de santé et une adoption généralisée. Il est essentiel d'impliquer les parties prenantes et de promouvoir une culture de confiance envers ces nouvelles technologies.

1.5 Conclusion

Dans ce chapitre nous avons exploré deux domaines importants : les maladies cardiaques et l'Internet des objets (IdO) appliqué à la santé. Nous avons examiné les différentes maladies cardiaques, leurs causes et leurs symptômes, mettant en évidence l'importance d'un diagnostic précis pour assurer des soins appropriés.

Ensuite, nous avons abordé l'IdO, qui est une technologie permettant la communication entre les objets physiques et virtuels. Nous avons souligné son utilisation dans le domaine de la santé pour améliorer les soins aux patients. L'IdO offre des avantages tels que la collecte de données en temps réel, la surveillance à distance et les interventions médicales à distance.

Cependant, nous avons également noté les défis associés à l'IdO, tels que la sécurité des données et la confidentialité. Il est crucial de mettre en place des mesures de sécurité appropriées pour protéger les informations médicales sensibles et garantir la confidentialité des patients.

En résumé, l'IdO présente un potentiel énorme pour améliorer la santé des patients en permettant une meilleure surveillance et des interventions rapides. Cependant, il est essentiel de trouver un équilibre entre les avantages offerts par l'IdO et les préoccupations liées à la sécurité et à la confidentialité des données.

CHAPITRE 02

APPRENTISSAGE AUTOMATIQUE

2.1 Introduction

L'apprentissage automatique (Machine Learning) et l'apprentissage profond (Deep Learning) sont des domaines de l'intelligence artificielle qui permettent aux ordinateurs d'apprendre à partir de données et de prendre des décisions intelligentes. Le domaine d'application de ces approches est très vaste tel que la reconnaissance vocale, la vision par ordinateur, le traitement du langage naturel, la détection précoce de maladies, l'analyse d'images médicales, etc. [6].

Le Machine Learning (ML) est une branche de l'intelligence artificielle qui vise à développer des modèles et des algorithmes capables d'apprendre à partir des exemples et des données pour effectuer des tâches spécifiques sans être explicitement programmés. Le Deep Learning (DL), quant à lui, est une approche avancée du ML qui utilise des réseaux de neurones artificiels profonds, inspirés du fonctionnement du cerveau humain, pour traiter des données complexes.

La relation entre les trois concepts IA, ML et DL est résumée dans la figure 2.1:

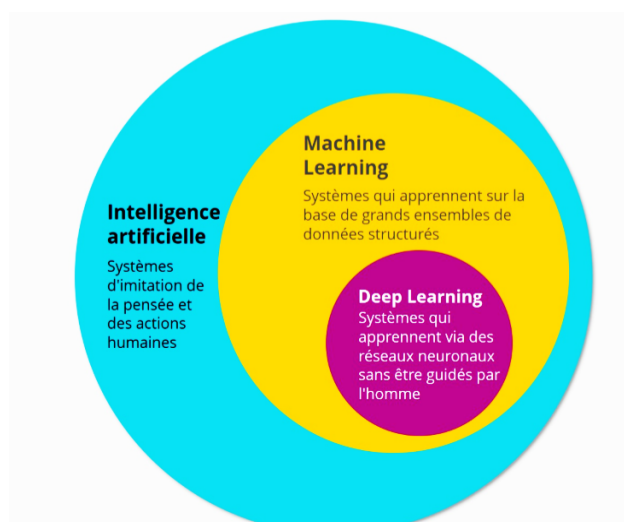


Figure 2-1. La relation entre IA, ML et DL .

2.2 L'apprentissage automatique

L'apprentissage automatique (ML) est un sous-domaine de l'intelligence artificielle, le terme faisant référence à la capacité des systèmes informatiques à trouver indépendamment des solutions aux problèmes en reconnaissant des modèles dans des bases de données. Il construit aussi un modèle mathématique basé sur des données d'échantillons appelé « données d'entraînement », afin de rendre prédictions ou décisions sans être explicitement programmé pour exécuter la tâche [7].

L'objectif du Machine Learning est de reconnaître parmi des données des structures souvent trop difficiles à détecter ou à mesurer manuellement. À partir de ces structures, on peut chercher à classer des individus, des objets, à prédire la valeur d'une variable à un certain horizon, à expliquer l'apparition ou non d'une caractéristique.

2.3 Les différents types d'apprentissage

L'apprentissage automatique comprend différents types d'algorithmes et d'approches qui peuvent être utilisés pour résoudre des problèmes et effectuer des tâches spécifiques. Il existe principalement trois types d'apprentissage automatique:

2.3.1 Apprentissage supervisé (Supervised Learning)

L'apprentissage supervisé (supervised learning en anglais) est certainement la forme de machine learning la plus répandue. Elle consiste à entraîner le programme sur des données étiquetées, où chaque exemple est associée à une étiquette ou une classe prédéfinie (Figure 2-2).

Le modèle est alimenté par un ensemble de données de formation, comprenant à la fois les observations (c'est-à-dire les entrées) et les résultats correspondants (c'est-à-dire les sorties). Ensuite, le modèle déduit mathématiquement la relation entre les caractéristiques (features) des données et leurs étiquettes. Le modèle construit est capable de prédire ou de classer de nouveaux exemples non vus auparavant.

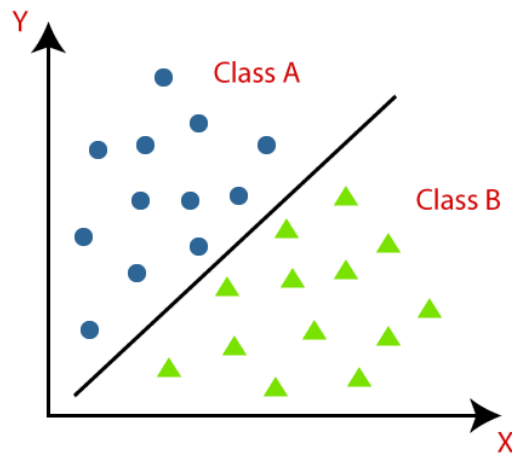


Figure 2-2. L'apprentissage supervisé.

Les algorithmes d'apprentissage supervisé courants comprennent les arbres de décision, les réseaux de neurones, les machines à vecteurs de support (SVM) et les méthodes de régression linéaire.

2.3.2 Apprentissage non supervisé (Unsupervised Learning)

Dans ce mode d'apprentissage, le modèle est entraîné sur un ensemble de données non étiquetées, où aucune classe ou étiquette n'est fournie. Il doit donc déterminer le nombre de catégories à créer, puis trier les données dans chaque catégorie. Ces algorithmes sont utilisés lorsqu'il n'y a pas d'échantillon disponible (Il est à noter que cette méthode peut être utilisée pour créer un échantillon et ensuite utiliser l'apprentissage supervisé pour classer de nouvelles données) [8].

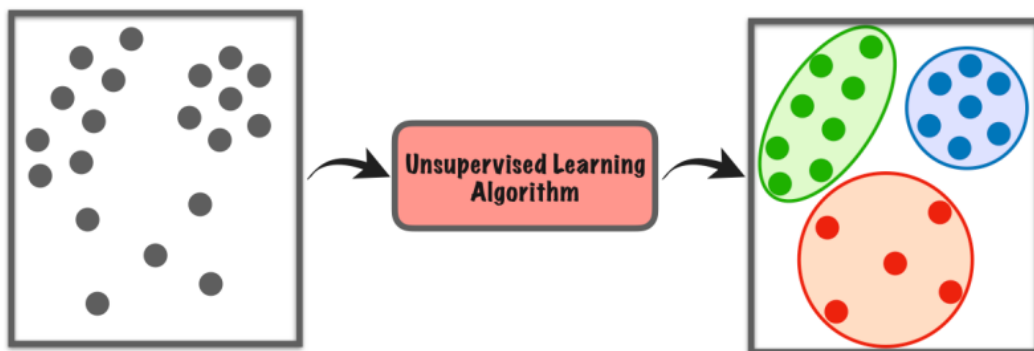


Figure 2-3. L'apprentissage non supervisé.

Les algorithmes d'apprentissage non-supervisé courants comprennent le regroupement (clustering), la réduction de dimension (dimensionality reduction) et l'association de règles.

2.3.3 Apprentissage par renforcement

L'apprentissage automatique par renforcement est basé sur la capacité de la machine à apprendre à partir de différentes situations de test et d'erreur dans divers contextes. Même si les résultats sont connus à l'avance ou non, la machine ne sait pas toujours quelles décisions prendre pour obtenir les meilleurs résultats. Dans ce type d'apprentissage, l'algorithme doit associer lui-même les schémas de réussite, et c'est leur répétition qui permet à l'algorithme d'obtenir constamment des résultats précis et fiables [23].

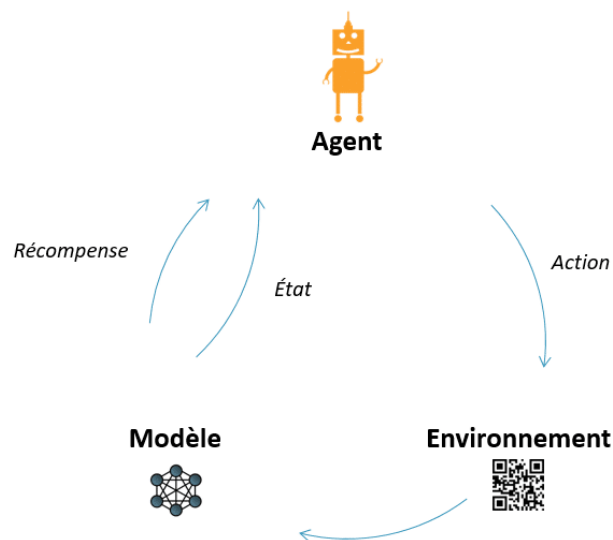


Figure 2-4. L'apprentissage par renforcement.

2.4 Algorithmes de Machine Learning

Il existe une grande variété d'algorithmes de Machine Learning, et certains sont plus couramment utilisés que d'autres, en particulier lorsqu'il s'agit de données étiquetées.

2.4.1 Random Forest

La Random Forest (RF) est un algorithme d'apprentissage supervisé polyvalent qui peut être utilisé à la fois pour la classification et la régression. Cette méthode est considérée comme l'une des plus flexibles et des plus faciles à utiliser. RF est composé d'arbres de décision, et plus il y a d'arbres, plus la forêt est robuste.

Les forêts aléatoires créent des arbres de décision sur des échantillons de données sélectionnés au hasard. Ensuite, elles obtiennent des prédictions à partir de chaque arbre et sélectionnent la meilleure solution par vote. De plus, les forêts aléatoires fournissent également un indicateur assez fiable de l'importance des caractéristiques (features) [8].

La figure 2-5 illustre son fonctionnement:

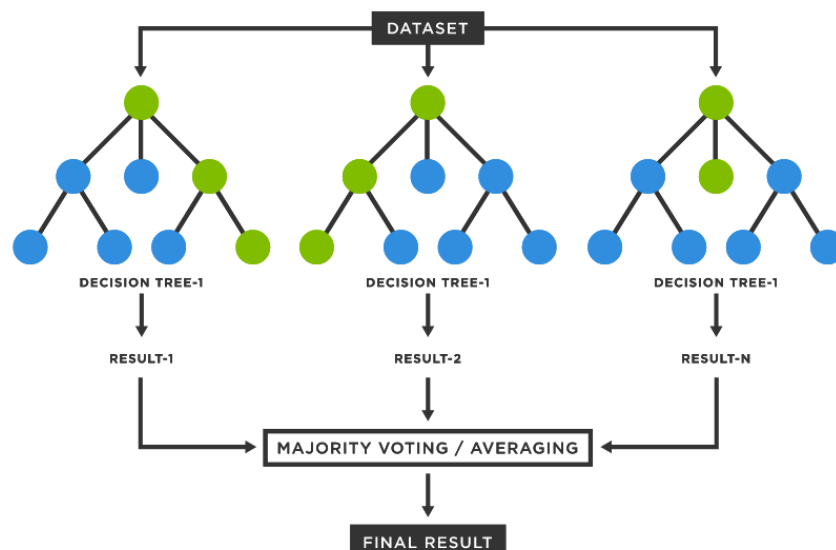


Figure 2-5. Fonctionnement de Random Forest.

L'algorithme RF offre plusieurs hyperparamètres qui peuvent être optimisés pour obtenir de meilleures performances.

- **n-estimators** : Il s'agit du nombre d'arbres dans la forêt..
- **max-depth** : C'est la profondeur maximale de chaque arbre dans la forêt..
- **min-samples-split**: C'est le nombre minimum d'échantillons requis pour diviser un nœud interne de l'arbre.
- **max-features**: Il s'agit du nombre maximum de variables que RF est autorisé à essayer dans un arbre individuel.

2.4.2 Les Machines à Vecteurs de Support (SVM)

Les SVMs sont des algorithmes d'apprentissage automatique supervisé puissants qui sont utilisés à la fois pour la classification et la régression. Mais généralement, ils sont utilisés dans les problèmes de classification où les données sont linéairement séparables.

Effectivement, les Machines à Vecteurs de Support (SVM) présentent plusieurs avantages :

- Les SVM peuvent traiter des ensembles de données avec un grand nombre de paramètres (également appelés dimensions), ce qui en fait un choix approprié pour des problèmes avec des espaces de caractéristiques complexes et de grande dimension.
- Les SVM utilisent un nombre relativement restreint d'hyper paramètres.
- Les SVM garantissent de bons résultats théoriques car ils sont basés sur des principes mathématiques solides.
- Ils peuvent égaler ou dépasser en performance les réseaux de neurones ou modèles gaussiens.

Les concepts principaux du SVM sont les suivants :

- **Vecteurs de support** : Les points de données les plus proches de l'hyperplan sont appelés vecteurs de support. La ligne de séparation sera définie à l'aide de ces points de données.
- **Hyperplan**: Un hyperplan est un plan de décision qui sépare un ensemble d'objets en différentes classes..
- **Marge**: Une marge est l'espace entre les deux lignes de séparation, qui est calculée comme la distance perpendiculaire entre l'hyperplan et les vecteurs de support les plus proches. Si la marge est plus grande entre les classes, alors elle est considérée comme une bonne marge, une plus petite marge est une mauvaise marge.

L'objectif principal du SVM est d'identifier l'hyperplan avec la marge maximale entre les vecteurs de support.

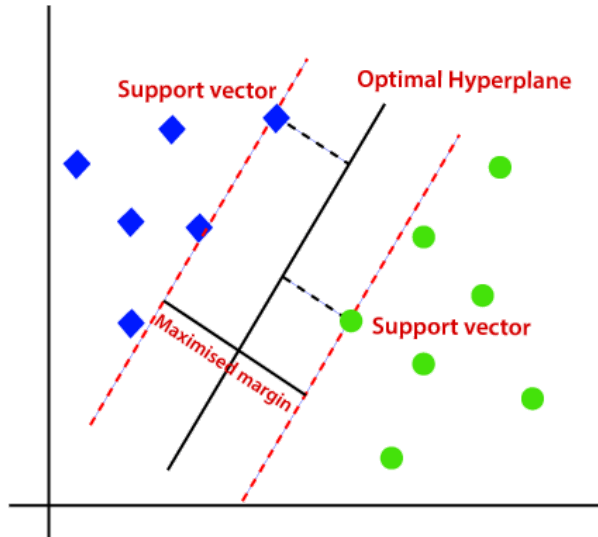


Figure 2-6. Exemple de marge maximale (hyperplan valide).

2.4.3 Les arbres de décision

L'apprentissage par arbre de décision est une approche permettant de construire des arbres qui serviront de modèle prédictif. Un arbre de décision est composé de nœuds, de branches et de feuilles. Le nœud de départ, appelé nœud racine, représente la caractéristique de décision initiale. On part du sommet de l'arbre et on descend en empruntant les feuilles jusqu'à arriver en bas de l'arbre ce qui va permettre de prendre une décision.



Figure 2-7. Arbre de décision

2.4.4 Les k plus proches voisins

L'algorithme des k plus proches voisins (K-NN) est un algorithme d'apprentissage supervisé: il est nécessaire d'avoir des données labellisées. À partir d'un ensemble de données étiquetées, il sera possible de classer (déterminer le label) d'une nouvelle donnée.

Pour sélectionner la valeur de k qui convient à vos données, nous exécutons plusieurs fois l'algorithme KNN avec différentes valeurs de k. Le choix du nombre k, c'est-à-dire du nombre de voisins à considérer, est un hyperparamètre important dans l'algorithme k-NN. Il peut avoir un impact sur la performance du modèle et sur sa capacité à généraliser. Un k trop petit peut conduire à une sensibilité au bruit ou au surapprentissage, tandis qu'un k trop grand peut conduire à une sous-représentation des classes minoritaires.

2.5 Domaines d'application de ML

Le Machine Learning (ML) est une discipline qui trouve des applications dans de nombreux domaines.

- **Vision par ordinateur** : Le ML est utilisé pour la détection d'objets, la reconnaissance faciale, la segmentation d'images, la classification d'images médicales, etc.
- **Énergie** : Le ML est utilisé pour la prévision de la consommation d'énergie, la gestion de l'énergie intelligente, la maintenance prédictive des équipements, etc.
- **Marketing et publicité** : Le ML est utilisé pour la recommandation de produits, la segmentation de clients, l'analyse de sentiments des consommateurs, la personnalisation des publicités, etc.
- **Industrie de la finance** : L'apprentissage automatique est de plus en plus utilisé dans le secteur financier. Les banques utilisent principalement le ML pour trouver des modèles dans les données, mais aussi pour prévenir la fraude.
- **Santé et médecine** : Le ML est utilisé pour la détection de maladies, le diagnostic médical, la prédiction de résultats cliniques, l'analyse d'images médicales, la découverte de médicaments, etc.

2.6 Les métriques d'évaluations des systèmes de machine learning

Evaluer les performances d'un modèle de classification est essentiel pour mesurer son efficacité et prendre des décisions éclairées. Voici quelques-unes des métriques les plus couramment utilisées dans les systèmes de classification et de prédiction

- **Précision:** Elle représente le rapport entre les vrais positifs parmi les exemples prédits comme positifs.

$$Precision = \frac{TP}{TP+FP}$$

- **Le rappel:** Le rappel ou la sensibilité est le rapport entre les exemples positifs correctement classés et tous les exemples de la classe réelle.

$$Rappel = \frac{TP}{TP+FN}$$

- **Le F1-Score:** Le F1-score est une métrique qui combine les deux métriques précédentes. Elle est la moyenne pondérée de la précision et du rappel. Par conséquent, ce score prend en compte à la fois les faux positifs et les faux négatifs.

$$F1\ Score = \frac{2 \times Precision \times Rappel}{Precision + Rappel}$$

- **Accuracy:** Il s'agit de la fraction des exemples bien classés par rapport à toutes les prédictions

$$Exactitude = \frac{TP+TN}{TP+TN+FP+FN}$$

2.7 Conclusion

Dans ce chapitre, nous avons présenté l'apprentissage automatique dans ses différentes formes, à savoir l'apprentissage supervisé, non supervisé et avec renforcement.

Nous avons aussi abordé les domaines d'application variés de l'apprentissage automatique. De la vision par ordinateur, où les machines peuvent comprendre et interpréter des informations visuelles, à la santé, où elles peuvent aider au diagnostic et à la prédiction de résultats cliniques et des applications dans de nombreux secteurs.

Nous avons également exploré différents algorithmes d'apprentissage automatique, tels que les arbres de décision, les machines à vecteurs de support (SVM), les k plus proches voisins

(k-NN), etc. Chaque algorithme a ses propres caractéristiques et est adapté à des types spécifiques de problèmes.

CHAPITRE 03

PRÉDICTION DES MALADIES CARDIAQUES: MÉTHODES ET RÉSULTATS RÉCENTS

3.1 Introduction

Dans le but de bien mener et aborder notre sujet, dans ce chapitre nous présentons les travaux connexes réalisés pour les systèmes de prédiction des maladies cardiaques, nous avons débuté par effectuer une recherche et une lecture approfondies des travaux les plus pertinents portant sur la classification des maladies cardiaques en utilisant des méthodes de Machine Learning.

3.2 Revue des travaux connexes

Au cours des dernières années, de nombreuses recherches ont été proposées pour la prédiction des maladies cardiaques. Dans cette section, nous présenterons certaines des solutions récentes qui ont été étudiées. Le tableau 3.1 résume les résultats obtenus par ces différentes solutions.

- Dans l'étude menée par Muktevi Srivenkatesh [21], un jeu de données provenant de Kaggle a été utilisé sans nécessiter de prétraitement des informations. Les performances du modèle de diagnostic ont été évaluées en utilisant des métriques telles que la précision, la sensibilité et la spécificité.

Cette étude a proposé un modèle de prédiction pour déterminer si une personne est atteinte d'une maladie cardiovasculaire ou non, et pour fournir une sensibilisation ou un diagnostic à cet égard. Pour cela, les précisions des algorithmes tels que le Support Vector Machine, le Random Forest, le classifieur Naive Bayes et la régression logistique ont été comparées sur un ensemble de données prélevé dans une région afin de présenter un modèle précis de prédiction des maladies cardiovasculaires.

Les algorithmes d'apprentissage automatique étudiés ont réussi à prédire les maladies cardiovasculaires chez les patients avec une précision variant entre 58,71 % et 77,06 %. Il a été démontré que la régression logistique présente une meilleure précision (77,06 %) par rapport aux différents algorithmes d'apprentissage automatique.

- Le travail présenté par Chintan M. Bhatt et al. [18], explore l'utilisation des techniques d'apprentissage automatique pour prédire la probabilité de maladies cardiaques. Ceci peut être bénéfique pour les professionnels de la santé et les patients. Ces techniques incluent Random Forest, Arbre de décision, Perceptron multicouche et le classificateur XGBoost.

Le dataset utilisé dans le travail de Chintan M. Bhatt et al. comprend 70 000 dossiers de patients avec 12 caractéristiques distinctes. Ce dataset est disponible publiquement sur Kaggle, ce qui facilite la reproductibilité et la vérification des résultats par d'autres chercheurs. Cette étude est basée sur l'utilisation des échantillons de petite taille, ce qui peut affecter la généralisabilité des résultats. Pour surmonter cette limitation, l'étude proposée a utilisé un ensemble de données plus vaste et plus diversifié.

Pour évaluer les performances des techniques d'apprentissage automatique utilisées, les auteurs ont mesuré plusieurs métriques, notamment la précision (Precision), le rappel (Recall), l'exactitude (Accuracy) et le F1-Score. Les résultats obtenus ont montré que le modèle MLP (Perceptron multicouche) a donné le meilleur score avec une Accuracy de 87.28 %, un Recall de 84.85%, une Precision de 88.70% et un F1-Score de 86.71%.

- Une autre étude a été présentée par Julker Nayeem et al. [19]. Différents types d'algorithmes d'apprentissage automatique supervisés ont été utilisés pour prédire la présence de maladies cardiaques chez les patients.

Dans ce travail, la technique d'imputation de la valeur moyenne a été appliquée pour gérer les valeurs manquantes présentes dans l'ensemble de données. Cela permet de remplacer les valeurs manquantes par la moyenne des autres valeurs présentes dans

la même colonne. Pour évaluer l'exactitude des prévisions, les algorithmes KNN, Naive Bayes et Random Forest ont été appliqués à l'ensemble de données sur les maladies cardiaques. La figure 3-2 illustre le système global de classification utilisé dans ce travail.

Chacune des méthodes de classification utilisées a bien fonctionné lorsqu'elle a été appliquée aux observations du dataset contenant des valeurs manquantes. Le classificateur Random Forest a été le plus performant parmi les trois algorithmes étudiés. Il a obtenu un score d'Accuracy de 95,63%, une précision de 95,28%, un rappel de 88,89% et un F1-score de 87,36%.

- Dans l'étude menée par Rubini PE et al. [20], une méthodologie a été présentée pour prédire les maladies cardiovasculaires en utilisant quatre algorithmes différents: Random Forest, Régression logistique, Algorithme de Bayes naïf et SVM.

La comparaison des résultats obtenus par les différents algorithmes a révélé que l'algorithme de l'arbre de décision était le plus précis. Les mesures d'exactitude, de précision et de F-score ont confirmé cette conclusion en montrant que l'algorithme de l'arbre de décision produisait les prédictions les plus précises par rapport aux autres algorithmes avec une exactitude de 99.01 %, une précision de 99.01% et un F1-Score de 99.50%. Cependant, il est intéressant de noter que les mesures de rappel et de spécificité n'ont pas pu fournir d'informations utiles pour distinguer les algorithmes.

Les auteurs ont observé que, dans le jeu de données utilisé, l'algorithme SVM n'a pas réussi à surpasser les autres algorithmes, même s'il est considéré comme l'un des meilleurs algorithmes de classification pour les ensembles de données médicaux. Cette observation souligne que les performances des algorithmes peuvent varier selon le contexte spécifique du jeu de données et les caractéristiques des problèmes de classification.

Les résultats obtenus par les différentes solutions étudiées pour la prédiction des maladies cardiaques sont récapitulés dans le tableau ci-dessous.

Référence	Date	Technique	Accuracy	Dataset utilisé
Muktevi Srivenkatesh [21]	2020	KNN	58.71%	Kaggle
		SVM	71.55%	
		Logistic Regression	77.06%	
		Naive Bayes	69.72%	
		Random Forest	74.31%	
Chintan M. Bhatt et al. [18]	2023	Random forest	86.92%	Kaggle
		Arbre de décision	53.00%	
		XGBoost	87.02%	
Julker Nayeem et al [19]	2022	KNN	87.36%	Kaggle
		Naives Bayes	88.89%	
		Random Forest	95.63%	
Rubini PE et al. [20]	2021	Arbre de décision	99.01%	PIDD
		Naives Bayes	78.88%	
		Discriminant	83.50%	
		Random Forest	93.40%	
		SVM	76.57%	

Tableau 3-1. Résultats des travaux connexes sur la prédiction des maladies cardiaques

3.3 Synthèse des travaux connexes

Les études présentées ont utilisé diverses méthodes d'apprentissage automatique, notamment les arbres de décision, les régressions logistiques, les algorithmes de Bayes naïfs, les machines à vecteurs de support (SVM) et les réseaux de neurones.

Les résultats ont montré que les méthodes d'apprentissage automatique étaient capables de prédire avec précision la présence de maladies cardiaques chez les patients. L'algorithme de

l'arbre de décision a souvent été identifié comme étant le plus précis, produisant des scores élevés en termes d'exactitude, de précision et de F1-score. Cependant, il est important de noter que les mesures de rappel et de spécificité n'ont pas toujours été des indicateurs fiables pour distinguer les performances des algorithmes.

Il est important de noter que ces études ont été menées sur des ensembles de données spécifiques et que les performances des algorithmes peuvent varier en fonction du contexte et des caractéristiques des données. Il est donc essentiel de considérer ces résultats dans le cadre spécifique de chaque étude et de prendre en compte d'autres facteurs tels que la taille de l'échantillon, la diversité des données et les caractéristiques spécifiques du problème de prédiction des maladies cardiaques.

3.4 Conclusion

Dans ce chapitre, nous avons examiné les travaux connexes récents portant sur les méthodes de prédiction des maladies cardiaques. Pour répondre à notre problématique, nous nous inspirerons de ces études en testant plusieurs algorithmes d'apprentissage automatique et en sélectionnant celui qui présente les meilleures performances pour notre jeu de données.

Dans ce contexte, nous chercherons à appliquer les algorithmes les plus pertinents et les plus efficaces dans la prédiction des problèmes de santé et des maladies. En utilisant un jeu de données spécifique, nous évaluerons les performances de ces algorithmes et déterminerons celui qui est le plus approprié pour notre problématique.

CHAPITRE 4

CONCEPTION ET REALISATION DU SYSTEME

4.1 Introduction

Dans les chapitres précédents, nous avons examiné en détail les maladies cardiovasculaires, ainsi que les différents algorithmes d'apprentissage automatique utilisés pour leur prédiction. Dans ce chapitre, nous présentons notre système proposé pour prédire ces maladies, en mettant l'accent sur son architecture générale, l'environnement de développement utilisé et les résultats obtenus.

Nous commençons par présenter l'architecture générale de notre système. Cette architecture englobe les différentes étapes, de la collecte des données médicales à la construction du modèle de prédiction. Nous mettons également en évidence l'importance du prétraitement des données pour garantir des résultats fiables et cohérents.

Ensuite, nous décrivons l'environnement de développement que nous avons utilisé pour la mise en œuvre de notre système. Cet environnement comprend des outils et des bibliothèques spécifiques pour l'apprentissage automatique, qui nous ont permis de développer et d'entraîner notre modèle de prédiction.

Nous présentons ensuite les résultats obtenus à l'aide de notre système de prédiction des maladies cardiaques. Nous évaluons les performances de notre modèle en utilisant différentes mesures telles que la précision, le rappel et la spécificité. Ces résultats nous permettent d'évaluer l'efficacité de notre système et de déterminer sa capacité à fournir des prédictions précises.

Enfin, nous analysons et discutons les résultats obtenus. Nous examinons les forces et les faiblesses de notre système, en identifiant les aspects qui ont contribué à ses performances ainsi que ceux qui pourraient nécessiter des améliorations.

L'architecture générale de notre système est illustrée dans la figure 4-1.

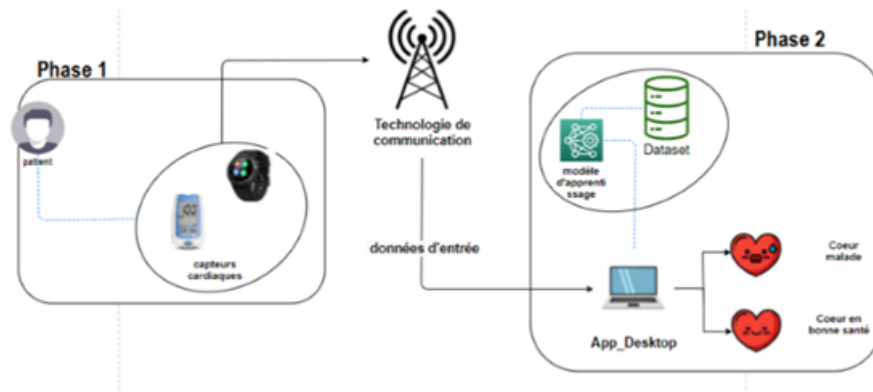


Figure 4.1- architecture globale de système

4.2 Préparation des données

La préparation des données est une étape cruciale pour entraîner un algorithme d'apprentissage automatique. Elle vise à transformer les données brutes en un format approprié pour l'entraînement du modèle. Voici les principales étapes de préparation des données :

4.2.1 Collecte des données

Pour prédire les problèmes cardiaques, nous pouvons classer les données en trois niveaux différents:

- Au premier niveau, nous trouvons les données personnelles fixes, telles que l'âge et le sexe du patient. Ces informations fournissent une base initiale pour évaluer les risques cardiovasculaires.
- Le deuxième niveau comprend des données périodiques collectées à des intervalles réguliers, comme le taux de sucre dans le sang et le taux de cholestérol sérique.
- Le troisième niveau concerne les données en temps réel, telles que la pression artérielle et le rythme cardiaque. Ces données fournissent des informations précises sur l'état actuel du patient et permettent une surveillance continue de sa santé cardiovasculaire.

Pour collecter les données du premier niveau, telles que l'âge et le sexe, notre système fournit des interfaces utilisateur où les utilisateurs peuvent saisir ces informations manuellement. Cela peut se faire lors de l'inscription ou lors de la création d'un profil dans notre plateforme.

Pour collecter les données du deuxième et du troisième niveau, nous utilisons des dispositifs portables et des dispositifs de surveillance de la santé à domicile (IoT). Ces dispositifs permettent de collecter des données périodiques sur la santé cardiaque du patient. Parmi ces dispositifs nous pouvons citer à titre d'exemple: le tensiomètre, le moniteur de cholestérol, le glucomètre et le moniteur ECG.

4.2.2 Sélection des fonctionnalités

La sélection des fonctionnalités consiste à choisir les variables ou les caractéristiques qui seront utilisées pour entraîner le modèle. Il est important de sélectionner les fonctionnalités les plus pertinentes et informatives pour éviter les problèmes de sur apprentissage et améliorer les performances du modèle. Des techniques telles que l'analyse de corrélation et les méthodes de sélection de fonctionnalités peuvent être utilisées pour effectuer cette étape. Dans cette étude, nous avons choisi d'utiliser les 13 attributs définis dans le dataset pour entraîner notre modèle de prédiction des maladies cardiovasculaires.

4.2.3 Normalisation des données

Cette étape permet de normaliser ou de mettre à l'échelle les données pour les ramener à une même échelle. Cela peut être réalisé à l'aide de techniques telles que la normalisation z-score ou la mise à l'échelle min-max. Cela garantit que toutes les variables ont une influence équitable sur le modèle.

La normalisation MinMax redimensionne les données de manière à ce qu'elles soient comprises dans une plage spécifique, généralement entre 0 et 1. Cela permet d'égaliser l'importance des différentes caractéristiques et d'éviter que certaines caractéristiques dominent les autres en raison de leurs plages initiales. La formule de normalisation est donnée par:

$$X_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

La standardization transforme les valeurs d'une caractéristique pour qu'elles aient une moyenne $\mu = 0$ et une variance $\sigma = 1$. La formule de standardization est donnée par:

$$x_{standard} = \frac{x - \mu}{\sigma}$$

4.2.4 Séparation des ensembles de données

Les données sont divisées en ensembles d'entraînement, de validation et de test. L'ensemble d'entraînement est utilisé pour entraîner le modèle, l'ensemble de validation est utilisé pour ajuster les hyperparamètres du modèle et 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 test.

4.3 Réalisation du système

L'objectif de cette phase est de construire un modèle d'apprentissage automatique capable de prédire avec précision les maladies cardiaques en utilisant les données pertinentes collectées et les techniques d'apprentissage automatique appropriées.

4.3.1 Outils et bibliothèques utilisés pour le développement du système proposé

Pour le développement du système proposé, nous avons utilisé les outils et les bibliothèques suivants:

- **Google Colab:** Nous avons utilisé Google Colab comme environnement de développement. Google Colab est une plateforme en ligne qui permet d'exécuter du code Python, notamment pour l'apprentissage automatique. Il fournit un accès gratuit aux ressources de calcul, notamment aux GPU, ce qui est avantageux pour l'entraînement de modèles d'apprentissage automatique.
- **Python:** Python est un langage de programmation largement utilisé dans le domaine de l'apprentissage automatique en raison de sa simplicité, de sa flexibilité et de sa richesse en bibliothèques dédiées à l'apprentissage automatique.
- **Tensorflow:** Une bibliothèque open-source pour l'apprentissage automatique qui fournit des outils pour la création et l'exécution de modèles d'apprentissage automatique, y compris les réseaux de neurones.
- **Keras:** Une bibliothèque open source écrite en Python pour l'apprentissage profond et est facile à utiliser, intuitive et rapide. Il peut être utilisé pour travailler à un

niveau élevé de certains frameworks tels que TensorFlow ou Knowledge Toolkit de Microsoft.

- **Matplotlib** : Une bibliothèque de langage de programmation Python qui permet de créer des graphiques et des visualisations pour analyser les données.. Elle peut être combinée avec les bibliothèques python de calcul scientifique NumPy et SciPy. Matplotlib est distribuée librement et gratuitement sous une licence de style BSD.
- **Pandas**: Une bibliothèque conçue pour la manipulation et l'analyse de données en langage Python. Elle est à la fois performante, flexible et simple d'utilisation..
- **Sklearn**: Scikit-learn est la bibliothèque la plus utile et la plus robuste pour l'apprentissage automatique en Python. Il fournit une sélection d'outils efficaces pour l'apprentissage automatique et la modélisation statistique, y compris la classification, la régression, le clustering et la réduction de la dimensionnalité via une interface de cohérence en Python.

4.3.2 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 (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 signifie aucune présence de maladies cardiaques et 1 signifie présence de maladies cardiaques.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

Figure 4-2. Extrait du dataset Cleveland.

- **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; 1 = angine de poitrine typique, 2 = angine de poitrine atypique, 3 = non-douleur angineuse, 4 = asymptomatique.
- **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 thalassemia; 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.

4.3.3 Exploration des donnée

1. **Visualisation des données:** Les deux classes de la base de données sont représentées dans la Figure 4-3. La classe 0 représente la classe des individus non malades. La classe 1 représente la classe des individus malades.



Figure 4-3. Graphe illustre les deux classes.

2. **Analyse de corrélation:** 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 4-4 représente la matrice de corrélation de notre jeu de données.

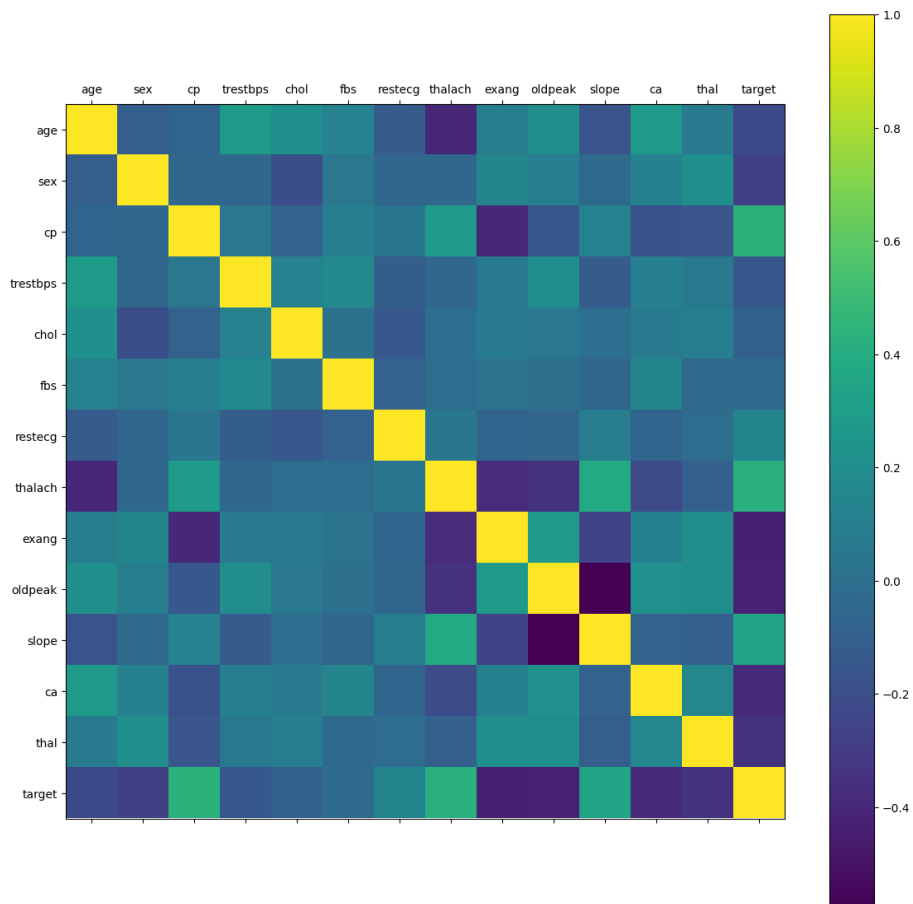


Figure 4-4. Matrice de corrélation de l'ensemble de Train .

3. **Nettoyage des données:** Cette étape de nettoyage des données vise à identifier, corriger ou supprimer les données incomplètes, altérées ou inexactes afin d'améliorer la qualité des données utilisées. Dans notre cas, grâce à la source intermédiaire [22] qui a déjà passé par l'étape de nettoyage des données, nous avons pu éviter ce traitement.

4.3.4 Implémentation et résultats

Pour implémenter notre modèle de prédiction, nous avons utilisé différents algorithmes d'apprentissage sans et avec normalisation des données.

1. Algorithme SVM:

L'implémentation de l'algorithme SVM (Support Vector Machine) dans la bibliothèque scikit-learn (sklearn) se fait en utilisant la classe "svm.SVC". L'algorithme SVM est utilisé pour la classification et la régression, et la classe "svm.SVC" est spécifiquement conçue pour la classification.

L'hyperparamètre clé pour l'algorithme SVM est le paramètre "kernel". Il spécifie le type de fonction de noyau à utiliser pour transformer les données d'entrée dans un espace de dimension supérieure. Les noyaux les plus couramment utilisés sont les noyaux linéaires, les noyaux polynomiaux, les noyaux de base radiale (RBF) et le noyau sigmoïde.

Les résultats obtenus pour chacun des noyaux utilisés sont présentés dans le tableau 4-1.

	linear	poly	rbf	sigmoid
Précision (%)	82.38	74.70	71.7	82.65
Rappel (%)	81.85	72.75	68.85	86.85
F1-score(%)	81.85	72.75	86.85	67.41
Exactitude(%)	82.87	74.75	70.81	72.41

Tableau 4-1. Evaluation des résultats par kernel.

D'après ces résultats, le noyau "sigmoid" a donné la meilleure précision (82.65%) parmi les différents noyaux testés.

Le noyau RBF" a une précision de 71.70%, c'est la plus faible des autres noyaux. Cependant, il présente un F1-score élevé de 86.85%, indiquant une bonne harmonie entre la précision et le rappel pour cette classe.

Le noyau "linear" a obtenu une précision de 82.38%, ce qui est légèrement inférieur à celle du noyau sigmoid". Cependant, il présente une précision solide et est proche de l'exactitude (82.87%).

Dans notre cas, nous choisissons le noyau "Linear" avec la meilleure exactitude (82.87%).

Après normalisation des données, les résultats obtenus sont comme suit:

kernel="linear"	sans normalisation	minmax normalisation	standard normalisation
Précision (%)	82.38	84.70	82.70
Rappel (%)	81.85	82.75	85.85
F1-score(%)	81.85	82.75	86.85
Exactitude(%)	82.87	84.75	83.81

Tableau 4-2. Résultats SVC sans et avec normalisation

Selon ces résultats, nous constatons que la normalisation a amélioré la performance de notre modèle avec une précision de 84.70% et une exactitude de 84.75% dans la cas de la normalization minmax et un rappel de 85.85% et un F1-score de 86.86% dans le cas de standardisation.

2. L'algorithme Random forest:

Le deuxième algorithme testé est Random Forest. Le nombre d'estimateurs (n_estimators) est en effet l'un des hyperparamètres clés de l'algorithme Random Forest. Il correspond au nombre d'arbres de décision indépendants qui seront construits dans la forêt.

Les résultats obtenus en sélectionnant différentes valeurs pour le nombre d'estimateurs sont illustrés dans le tableau 4-2.

Nb estimateurs	10	200	500	1000
Précision (%)	89.38	83.65	85.61	85.65
Rappel (%)	88.85	83.65	84.85	84.85
F1-score(%)	86.85	83.65	85.85	85.51
Exactitude(%)	87.87	84.75	85.81	85.81

Tableau 4-3. . Évaluation des résultats selon le nombre d'estimateurs.

Il est clair que parmi les différentes valeurs testées pour le nombre d'estimateurs, 10 est la valeur qui a donné les meilleures performances.

Après normalisation des données, les résultats obtenus sont comme suit:

estimateur =10	sans normalisation	minmax normalisation	standard normalisation
Précision (%)	89.38	94.70	92.7
Rappel (%)	88.85	92.75	94.85
F1-score(%)	86.85	92.75	96.85
Exactitude(%)	87.87	94.75	93.81

Tableau 4-4. Résultats RF sans et avec normalisation.

Selon ces résultats, nous constatons que la normalisation a amélioré la performance de notre modèle avec une précision de 94.70 % et une exactitude de 94.75% dans la cas de la normalisation minmax et un rappel de 94.85% et un F1-score de 96.86 dans le cas de la standardisation.

3. Algorithme KNN:

L'algorithme des k plus proches voisins (KNN) est un algorithme de classification qui repose sur le principe de trouver les k voisins les plus proches d'un point d'intérêt dans un espace

multidimensionnel. Dans le contexte de la prédiction des maladies cardiaques, les caractéristiques des patients, telles que l'âge, le sexe, les antécédents médicaux, les résultats des tests sanguins, etc., sont utilisées pour déterminer les voisins les plus proches.

Une fois que les voisins les plus proches ont été identifiés, le modèle KNN attribue une étiquette au point d'intérêt en se basant sur la classe majoritaire parmi ses voisins. Par exemple, si la majorité des k voisins sont étiquetés comme "malades", alors le modèle prédit que le patient est malade. La valeur de K est l'un des principaux hyperparamètres utilisés pour ajuster le modèle KNN.

Les résultats obtenus en sélectionnant différentes valeurs de K sont illustrés dans le tableau 4-3.

	10	11	12	13	14	16
Précision (%)	69.61	70.72	61.50	73.81	67.75	72.95
Rappel (%)	71.75	67.65	72.78	73.95	72.85	71.65
F1-score(%)	67.75	61.75	63.61	73.75	71.61	69.75
Exactitude(%)	70.71	70.71	71.71	74.75	71.91	72.54

Tableau 4-5. Evaluation de l'algorithme KNN selon l'hyperparamètre K

Les résultats obtenus montrent que le meilleur score est celui avec une exactitude de 74.75% en prenant en compte les 13 voisins les plus proches lors de la classification.

Après normalisation des données, les résultats obtenus sont comme suit:

k=13	sans normalisation	minmax normalisation	standard normalisation
Précision (%)	73.81	84.70	82.71
Rappel (%)	73.95	81.71	84.65
F1-score(%)	73.75	83.70	85.75
Exactitude(%)	74.75	84.75	83.81

Tableau 4-6. Résultats RF sans et avec normalisation.

Selon ces résultats, nous constatons que la normalisation a amélioré la performance de notre modèle avec une précision de 84.70 % et une exactitude de 84.75% dans la cas de la normalization minmax et un rappel de 84.65% et un F1-score de 85.75% dans le cas de la standardisation.

4. Algorithme d'arbre de décision:

L'hyperparametre utilisé pour ajuster cet algorithme est la profondeur maximale de l'arbre (max_depth).

	36	38	42	48
Précision (%)	76.75	76.5	78.61	76.65
Rappel (%)	77.85	77.65	78.85	77.81
F1-score(%)	77.85	77.65	78.85	76.61
Exactitude(%)	77.87	77.75	79.81	75.81

Tableau 4-7. Evaluation de l'algorithme de l'arbre de décision

En utilisant une profondeur maximale 42, la méthode de l'arbre de décision offre une exactitude de 79.81% et une precision de 78.61%. Cela signifie que le peut classifier les individus atteints ou non de maladies cardiaques avec une précision élevée.

Après normalisation des données, les résultat obtenus sont comme suit:

max_depth = 48	sans normalisation	minmax normalisation	standard normalisation
Précision (%)	73.81	75.70	79.71
Rappel (%)	73.95	75.71	81.76
F1-score(%)	73.75	75.70	76.75
Exactitude(%)	74.75	77.75	77.78

Tableau 4-8. Résultats AD sans et avec normalisation.

Selon ces résultats, nous constatons que la normalisation a amélioré la performance de notre modèle avec une précision de 79.71%, un rappel de 81.76%, un F1-score de 76.75% et une exactitude de 77.78% dans la cas de la standardisation.

Après avoir exploré plusieurs algorithmes de ML pour la classification binaire des maladies cardiaques avec et sans normalisation des données, que le modèle Random Forest avec une normalisation Min-Max des données a produit des résultats si prometteurs dans la classification binaire des maladies cardiaques.

Une précision de 94,70% et une exactitude de 94,75% sont des performances exceptionnelles. Cela signifie que notre modèle est capable de classer avec une grande précision les individus atteints ou non de maladies cardiaques, ce qui est essentiel dans le domaine médical où des décisions critiques sont prises en fonction des prédictions.

4.5 Conclusion

Dans ce chapitre, nous avons présenté l'architecture générale de notre système basé sur l'IoT et le ML pour la prédiction des maladies cardiaques. Cette architecture comprend deux phases essentielles: la préparation des données et l'implémentation du modèle de prédiction.

Dans la phase de préparation des données, nous avons décrit les différentes étapes, notamment la collecte des données, la sélection des caractéristiques, la normalisation et la séparation des données en ensembles d'entraînement, de validation et de test.

Pour l'implémentation, nous avons utilisé un environnement de développement dédié au ML, qui nous a permis de tester les modèles les plus pertinents tels que le SVM, le Random Forest, le KNN et l'arbre de décision. Nous avons effectué des expérimentations en utilisant ces modèles et présenté les résultats obtenus.

Les résultats ont montré que le modèle Random Forest avec une normalisation Min-Max des données a donné les meilleurs résultats pour notre ensemble de données. Il a obtenu 94,70% et une exactitude de 94,75%, ce qui indique une capacité de classification précise des individus atteints ou non de maladies cardiaques.

Il est important de souligner que les résultats peuvent varier en fonction des caractéristiques spécifiques de l'ensemble de données et du problème étudié. Par conséquent, nous pouvons explorer d'autres modèles et techniques, d'ajuster les hyperparamètres et de

considérer d'autres mesures d'évaluation pour affiner encore davantage les performances du système.

CONCLUSION GÉNÉRALE

Les maladies cardiaques sont un problème de santé majeur dans le monde entier, et l'utilisation de l'IoT et de l'apprentissage automatique peut contribuer à améliorer le diagnostic et la prédiction des maladies cardiaques, ce qui peut conduire à une prise en charge plus efficace des patients et à de meilleurs résultats cliniques.

Nous avons présenté l'IoT comme une technologie prometteuse dans le domaine de la santé, permettant une collecte de données en temps réel, une surveillance à distance et des interventions médicales à distance. Cependant, nous avons également mis en évidence les défis de sécurité et de confidentialité associés à l'IoT, qui doivent être pris en compte pour garantir la protection des données médicales sensibles.

Ensuite, nous avons exploré les différentes 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.

En utilisant différents algorithmes d'apprentissage automatique tels que le SVM, le Random Forest, le KNN et l'arbre de décision, nous avons effectué des expérimentations et évalué les performances de chaque modèle. Les résultats ont montré que le modèle Random Forest avec une normalisation Min-Max des données a donné les meilleures performances, avec une précision de 94,70% et une exactitude de 94,75%.

Cependant, il est important de noter que les résultats peuvent varier en fonction des caractéristiques spécifiques de l'ensemble de données et du problème étudié. Il est donc essentiel de continuer à explorer d'autres modèles, d'ajuster les hyperparamètres et de considérer d'autres mesures d'évaluation pour améliorer encore les performances du système.

BIBLIOGRAPHIE

- [1] w. h. organisation, «world health organisation,» 09 02 2022. . Available: <https://www.who.int/health-topics/cardiovascular-diseases#tab=tab 1>.
- [2] C. Khanji, "Évaluation de la qualité des soins et des services préventifs cardiovasculaires en première ligne", *International Journal for Quality in Health Care*, 2019
- [3] K. K. Patel, S. M. Patel et al., "Internet of things-IOT : definition, characteristics, enabling technologies, application & future challenges", *International journal of engineering science and computing*, t. 6, no 5, 2020.
- [4] L. Zhang, H. Yuan, S.-H. Chang et A. Lam, "Research on the overall architecture of Internet of Things middleware for intelligent industrial parks", *The International Journal of Advanced Manufacturing Technology*, p. 1-9, 2021.
- [5] S. Madakam, V. Lake, V. Lake, V. Lake et al., "Internet of Things (IoT) : A literature review", *International Journal of Computer and Communications*, t. 3, no 05, p. 164, 2022.
- [6] Chlo-Agathe Azencott, "Introduction au Machine Learning", *International Journal of Machine Learning Research*, 2021.
- [7] I. E. Naqa et M. J. Murphy, *Machine Learning in Radiation Oncology*, Springer International Publishing 2019.
- [8] A.-J. M. I., Q. M. H. et H. Mohammad, «Machine Learning Classification Techniques for Heart Disease Prediction: A Review,» *International Journal of Engineering & Technology*, pp. 5373-5379, 2020.
- [9] P. Mahé, "Noyaux pour graphes et Support Vector Machines pour le criblage virtuel de molécules". *Rapport de stage, DEA MVA 2020/2021, Septembre 2020*.
- [10] Haohan Wang, Bhiksha Raj, "On the Origin of Deep Learning", *Journal of Machine Learning Research*, 2021.
- [16] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural

- networks,"*International Conference on Machine Learning ICML(3)*, vol. 28, pp. 1310–1318, 2019.
- [17] Reza Bosagh Zadeh, Bharath Ramsundar, "TensorFlow for Deep Learning", 2020.
- [18]Chintan M. Bhatt 1, , Parth Patel 1, Tarang Ghetia 1 and Pier Luigi Mazzeo. "Effective Heart Disease Prediction Using Machine Learning Techniques",*International journal Artificial Intelligence Algorithms for Healthcare 2023*.
- [19] Md. Julker Nayeem, Sohel Rana, and Md. Rabiul Islam , "Prediction of Heart Disease Using Machine Learning Algorithms" , *European Journal of Artificial Intelligence and Machine*,2022.
- [20]Rubini PE1, Dr.C.A.Subasini, Dr.A.Vanitha Katharine, V.Kumaresan, S.GowdhamKumar, A Cardiovascular Disease Prediction using Machine Learning Algorithms,*journal of the Romanian Society for Cell Biology 2021* .
- [21] Muktevi Srivenkatesh." Prediction of Cardiovascular Disease using Machine Learning Algorithms", *International Journal of Engineering and Advanced Technology (IJEAT)*, Volume-9 Issue-3, February 2020.
- [22]D.LAPP,2022.[Enligne].Available:<https://www.kaggle.com/datasets/johnsmith88/heartdisease-dataset>.
- [23]Mohammed, M., M.B. Khan, and E.B.M. Bashier, Machine learning: algorithms and applications2016: Crc Press.
- [24] faker skandrani (2021). L'Internet des objets architecture de l'ioT.
- [25] M. Han and H. Zhang, 2019, "Business intelligence architecture based on internet of things " *Journal of Theoretical & Applied Information Technology*, vol. 50, no. 1 pp. 90-95.
- [26] A. F. A. Rahman, M. Daud et M. Z. Mohamad, "Securing sensor to cloud ecosystem using internet of things (iot) security framework", in Proceedings of the *International Conference on Internet of things and Cloud Computing*, 2018, p. 1-5.
- [27] Genestier, P., Letondeur, L., Zouarhi, S., Prola, A., & Temerson, J. M. (2017)." Blockchains et Smart Contracts: des perspectives pour l'Internet des objets (IoT) et pour l'e-santé". *In Annales des MinesRealites industrielles* (No. 3, pp. 70-73). FFE.

