



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



Université Echahid Cheikh Larbi Tébessi De Tébessa

Faculté Des Sciences Exactes Et Sciences De La
Nature Et De La Vie

Mémoire de fin d'étude

Pour l'obtention du diplôme de MASTER

Domaine : Informatique

Option : Réseaux et sécurité informatique

Thème

***Une approche intelligente pour La prédiction
des problèmes cardiaque***

Réalisé Par :

Goufi Hadil

Devant le jury :

Dr. Thabet Khawla	MCB	Université Larbi Tébessi	Président
Mme. Bouakkaz Fatima	MAA	Université Larbi Tébessi	Examineur
Dr. Merzoug Soltane	MCA	Université Larbi Tébessi	Encadreur

Date de soutenance : 10 juin 2024

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

وَمَا تَوْفِيقِي إِلَّا بِاللَّهِ

Remerciements

Je souhaite avant tout exprimer ma gratitude envers Dieu, le Tout-Miséricordieux, pour m'avoir accordé la volonté, le courage, la patience, ainsi que la force morale et physique nécessaires pour mener à bien ce travail.

*Je tiens à remercier mon encadrant **Dr. Merzoug Soltane**, maître de conférences classe A à l'université de Tébessa, pour son rôle de direction et d'orientation dans ce travail. Je le remercie sincèrement pour sa disponibilité, son suivi assidu, ses précieux conseils et son aide précieuse. Merci infiniment.*

Je tiens à remercier s'adressent également à tous les membres du Jury, qui ont accepté de juger mon travail

Je remercie vont aussi à tous les enseignants du département d'Informatique.

Merci à mon père, ma mère, mon frère et mes sœurs pour leur compréhension, leur patience et leur soutien inconditionnel. Je suis extrêmement reconnaissant envers ma famille qui a toujours été la source de mon succès, et je ne saurai jamais comment les remercier suffisamment. Je prie Dieu pour qu'Il les préserve et les garde près de moi le plus longtemps possible.

Enfin, je tiens à remercier toutes les personnes qui de près ou de loin ont contribué à la réalisation de ce travail.

Dédicace

A mes chers parents,

Je dédie ce modeste travail à la lumière de ma vie, la boussole de mon chemin et mon soutien moral, mon amour ma chère maman. Et celui à qui je dois mon respect et mon amour est mon cher père qui m'a donné courage, Aucune dédicace ne peut véritablement exprimer ma dette envers eux, pour leur bienveillance, leur affection et leur soutien. Ils sont de véritables trésors de bonté, de générosité et de tendresse, et je leur témoigne mon amour profond et ma gratitude immense. Que Dieu les préserve.

A mon chère frère (Oussama), A mes chères sœurs (Hidaya, Ghada),

Je suis profondément reconnaissant que vous soyez présents dans ma vie, car vous êtes le secret de ma joie et la source de mon bonheur. Merci à chacun de vous, Oussama, Hidaya et Ghada, pour votre soutien précieux. Vous êtes l'amour de mon cœur. Que Dieu vous protège de tout mal.

À toute ma famille,

À tous ceux dont le nom peut être oublié, mais dont le lien reste gravé dans mon cœur.

Hadil

Résumé

La détection précoce des maladies cardiaques est un facteur essentiel à la réussite des soins de santé. Ces dernières années, le domaine médical a vu émerger diverses méthodes permettant de prédire les maladies cardiaques avant qu'elles ne surviennent, basées sur l'apprentissage automatique et l'apprentissage profond. En effet, les maladies cardiaques restent une cause majeure de décès dans le monde, notamment la maladie coronarienne, qui est l'une des maladies les plus dangereuses. Elle est responsable d'une grande partie des cas de patients et des décès liés aux maladies cardiovasculaires, étant donné que cette maladie est latente et n'apparaît pas cliniquement. Il est donc nécessaire de la diagnostiquer et de la traiter le plus tôt possible. L'objectif de ce travail est de développer un modèle de prédiction de la maladie coronarienne capable de détecter l'apparition précoce d'une maladie coronarienne, souvent mortelle.

Notre projet de recherche vise à prédire la maladie coronarienne à l'aide de techniques d'apprentissage profond. Nous avons utilisé un réseau de neurones convolutifs pour atteindre cet objectif. On pense que cela contribuera à améliorer les taux de diagnostic précoce et à réduire les complications associées à cette maladie.

Mots clés : prédiction, apprentissage automatique, apprentissage profond, maladies cardiovasculaires, maladie coronarienne, réseau neuronal convolutif

Abstract

Early detection of heart disease is a critical factor in achieving successful health care. In recent years, the medical field has witnessed the emergence of various methods to predict heart diseases before they occur based on machine learning and deep learning. Indeed, heart disease is still a major cause of death in the world, especially coronary artery disease, which is one of the most dangerous diseases. It is responsible for a large portion of patient cases and deaths related to cardiovascular disease, given that this disease is latent and does not appear clinically. Therefore, it is necessary to diagnose it. And treat it as soon as possible. The goal of this work is to develop a prediction model for coronary artery disease capable of detecting the early onset of coronary artery disease, which can often be fatal.

Our research project aims to achieve prediction of coronary artery disease using deep learning techniques. We used a convolutional neural network to achieve this goal. It is believed that this will contribute to improving early diagnosis rates and reducing complications associated with this disease

Keywords : prediction, machine learning, deep learning, cardiovascular disease, coronary artery disease, convolutional neural network

الملخص

إن الكشف المبكر عن أمراض القلب يُعتبر عاملاً حاسماً في تحقيق الرعاية الصحية الناجحة. في السنوات الأخيرة شهد المجال الطبي ظهور طرق مختلفة للتنبؤ بأمراض القلب قبل حدوثها بناء على التعلم الآلي والتعلم العميق. فعلا، لا تزال أمراض القلب سببا رئيسيا للوفيات في العالم، وخاصة مرض الشريان التاجي الذي يعدد من أخطر الأمراض فهو مسؤول عن جزء كبير من حالات المرضى والوفيات المرتبطة بأمراض القلب والأوعية الدموية نظرا إلى أن هذا المرض كامن ولا يظهر سريريا، لذلك، من الضروري تشخيصه وعلاجه في أسرع وقت ممكن. هدف هذا العمل هو تطوير نموذج للتنبؤ بمرض الشريان التاجي قادر على كشف الظهور المبكر لمرض الشريان التاجي الذي يمكن أن يكون في اغلب الاحيان قاتلا

مشروع بحثنا يستهدف تحقيق التنبؤ بمرض الشريان التاجي باستخدام تقنيات التعلم العميق. قمنا باستخدام الشبكة العصبية التلافيفية لتحقيق هذا الهدف. يُعتقد أن ذلك سيساهم في تحسين معدلات التشخيص المبكر والحد من المضاعفات المرتبطة بهذا المرض

الكلمات المفتاحية: التنبؤ، التعلم الآلي، التعلم العميق، أمراض القلب والأوعية الدموية، مرض الشريان التاجي، الشبكة العصبية التلافيفية

Table des matières

Liste des figures	vii
Liste des tableaux	ix
Liste des abréviations	x
Introduction générale	1
Chapitre 01 : Les maladies cardiaques et L'intelligence artificielle dans le domaine médical ...	4
1. Introduction	5
2. Les maladies cardiaques	5
2.1 Définition	5
2.2 Types des maladies cardiaques	6
2.3 Cas d'étude la maladie coronarienne	6
2.3.1 Définition	7
2.3.2 Causes	7
2.3.3 Les catégories et les complications principales	8
2.3.4 Les facteurs de risque	8
2.3.5 Les Symptômes	9
2.3.6 Diagnostic	9
3. L'intelligence artificielle	10
3.1 L'intelligence artificielle dans le domaine médical	10
3.2 Applications de l'intelligence artificielle dans le domien médicales	10
3.2.1 Accélérer le diagnostic des maladies rares	10
3.2.2 Favoriser la médecine de précision avec des traitements personnalisés	11
3.2.3 Des robots intelligents pour assister les médecins	11
3.2.4 Développer la chirurgie assistée par ordinateur	11
3.2.5 Accélérer la prévention des épidémies et effets secondaires des médicaments	11
3.3 Exemples d'applications de l'intelligence artificielle en médicales	12

3.3.1	Dans le domaine des maladies cardiovasculaire	12
3.3.2	En cancérologie	12
3.3.3	Dans le domaine de la transplantation	12
3.3.4	Appui au diagnostic en gynécologie-obstétrique	13
3.3.5	La prévention en santé mentale	13
4.	Conclusion	13
Chapitre 02 : État de l'art		14
1.	Introduction	15
2.	L'intelligence artificielle (IA)	15
2.1	L'apprentissage Automatique : Machine Learning	16
2.1.1	Définition	16
2.1.2	Les Méthodes et Les principaux types d'algorithmes d'Apprentissage Automatique...16	
2.2	L'apprentissage Profond : Deep learning	20
2.2.1	Définition	20
2.2.2	Réseau de neurones profonds « DNN »	20
2.2.3	Les grandes classes de réseaux de neurones profonds	20
3.	Les travaux connexes	22
3.1	Tableaux de comparaison	28
3.2	Synthèse	32
4.	Conclusion	33
Chapitre 03 : Contribution		35
1.	Introduction	36
2.	Problématique	36
3.	Architecture du système	37
4.	Dataset	39
5.	Le prétraitement de données	44
5.1	Nettoyage des données	45
5.2	Normalisation des données	45
5.3	Standardisation des données	46
5.4	Extraction des caractéristiques	46
5.5	Division des données	48

6. Classification avec Le Réseau Neuronal Convolutif (CNN)	48
7. Entraînement le modèle	49
8. Conclusion	52
Chapitre 04 : Implémentation	53
1. Introduction	54
2. Les outils de développement utilisés	54
2.1 Langage de programmation	54
2.2 Les Logiciels	54
2.3 Le matériel	56
3. Mesures de performance	56
4. Chargement des données	58
5. Prétraitement des données	58
6. Entraîner le modèle	61
7. Test du modèle	62
8. Discussion	65
9. Conclusion	65
Conclusion générale	66
Bibliographie	68

Liste des figures

Figure 1.1 : le muscle cardiaque et leur vaisseaux sanguins	6
Figure 1.2 : Effets de la maladie coronarienne	7
Figure 1.3 : Effets de plaque d'athérosclérose à l'artère coronaire	7
Figure 1.4 : Les symptômes d'une maladie coronarienne	9
Figure 2.1 : La relation entre IA, ML et DL	15
Figure 2.2 : Les méthodes d'apprentissage automatique et leur fonctionnement	17
Figure 2.3 : Les trois grandes classes de réseaux de neurones profonds	21
Figure 3.1 : Architecture du système de prédiction de La maladie coronarienne	38
Figure 3.2 : Description de dataset	39
Figure 3.3 : Graphe illustrant les deux classes	41
Figure 3.4 : Visualisation relation Age /maladie cardiaque	42
Figure 3.5 : Distribution de Dataset selon Le sexe	43
Figure 3.6 : Visualisation relation sexe /maladie cardiaque	43
Figure 3.7 : Les valeurs manquantes	44
Figure 3.8 : Phase de prétraitement	45
Figure 3.9 : les attributs sélectionnés	48
Figure 3.10 : Division des données	48
Figure 3.11 : Architecture CNN Proposé	49
Figure 3.12 : Les résultats de L'entraînement	49
Figure 3.13: Train Accuracy du CNN	50
Figure 3.14 : Train Precision du CNN	50
Figure 3.15 : Train Recall du CNN	51

Figure 3.16 : Train F1 Score du CNN	51
Figure 4.1 : La structure de l'ensemble de données	58
Figure 4.2 : l'état des valeurs après le Prétraitement	59
Figure 4.3 : l'étape d'extraction des caractéristiques	59
Figure 4.4 : Normalisation des caractéristiques	59
Figure 4.5 : division des données	60
Figure 4.6 : Convertissez les données en une matrice à deux colonnes (One-Hot Encoding)	60
Figure 4.7 : Standardisation des caractéristiques	61
Figure 4.8 : L'entraînement du modèle CNN	62
Figure 4.9 : Test Accuracy du CNN	62
Figure 4.10 : Test Precision du CNN	63
Figure 4.11 : Test Recall du CNN	63
Figure 4.12 : Test F1 Score du CNN	64

Liste des tableaux

Tableau 1 : Les catégories et les complications principales de La maladie coronarienne	8
Tableau 2 : Les facteurs de risque de la maladie coronarienne	8
Tableau 3 : les déférente méthodes de l'Apprentissage Automatique et leur type d'Algorithmes de chaque méthode et les Algorithme commun	19
Tableau 4 : Récapitulation des travaux connexes	31
Tableau 5 : Description de dataset	39
Tableau 6 : Description des attributs	40
Tableau 7 : Les facteurs de risque /Les attributs de dataset	47
Tableau 8 : Matrice de confusion	57
Tableau 9 : Les mesures de performance/L'équation	57
Tableau 10 : Résultats obtenus avec le model CNN	64

Liste des abréviations

MCV : maladies cardiovasculaires

IA : intelligence artificielle

ML : machine learning

DL : deep learning

DNN : Deep Neural Network

CHD : Coronary Heart Disease (CAD : Coronary Artery Disease)

MCC : Matthews Correlation Coefficient

AUC : Area Under the Receiver Operating Characteristic Curve

KNN : K-Nearest Neighbors

MLP : Multilayer Perceptron

SVM : Support Vector Machine

CNN : Convolutional Neural Network

BN : Batch Normalization

ReLU : Rectified Linear Unit

ADAM : Optimiseur ADAM (Adaptive Moment Estimation)

ROC : Receiver Operating Characteristic

AUROC : Area Under the Receiver Operating Characteristic Curve

DT : Decision Tree

LR : Logistic Regression

NB : Naive Bayes

LASSO : Least Absolute Shrinkage and Selection Operator

RF : Random Forest

XGB : Extreme Gradient Boosting

NN : Neural Network

RBF : Radial Basis Function

GBT : Gradient Boosting

VGG-16 : VGG-16 Convolutional Neural Network

1D CNN : 1D Convolutional Neural Network

ResNet18 : Residual Network-18

GRU : Gated Recurrent Unit

RNN : Recurrent Neural Network

ANN : Artificial Neural Network

Introduction générale

Introduction générale

Actuellement, les maladies cardiaques sont l'un des problèmes les plus importants qui menacent la vie humaine. Elles se positionnent en tant que principale responsable des décès à l'échelle mondiale, entraînant le décès d'environ 17,9 millions de personnes annuellement, Parmi ces décès, on estime que 7,4 millions sont dus à une cardiopathie coronarienne [3], cette maladie est la première cause de décès prématurés dans les pays développés elle peuvent durer longtemps sous une forme latente, sans se manifester cliniquement, C'est ce qui fait de la maladie coronarienne un problème de santé mondial très important, car elle est responsable d'une grande partie des cas de patients et des décès liés aux maladies cardiovasculaires, pour L'identification des personnes à risque de développer une maladie coronarienne est primordiale pour une intervention rapide et des efforts de prévention, ce qui pourrait éviter des décès prématurés, Le processus de diagnostic de ce maladie nécessite une batterie étendue de tests, tels que la mesure de la pression artérielle, du taux de glucose, des signes vitaux, des douleurs thoraciques, des électrocardiogrammes et de la fréquence cardiaque maximale, entre autres. En complément de ces analyses, des études cliniques, des antécédents médicaux des patients et des réponses à leurs questions sont également prises en compte. Cependant, la réalisation de tous ces diagnostics est souvent sujette à des erreurs ou à des retards dans les tests. De plus, cette démarche engendre des coûts élevés, et sa prend du temps pour les évaluations.

Dans ce contexte, d'importants travaux de recherche ont été effectués pour prédire les maladies cardiaques en mettant l'accent sur la maladie coronarienne en utilisant des techniques d'intelligence artificielle, notamment l'apprentissage automatique et l'apprentissage profond. Ces méthodes s'appuient sur les données du patient, ses antécédents médicaux et une série d'analyses pour établir des modèles prédictifs.

Dans le cadre de ce projet, notre principal objectif est de développer une approche basée sur l'apprentissage profond pour prédire la maladie coronarienne en utilisant les données des patients. Pour atteindre cet objectif, nous avons mis en place une méthode spécifique en implémentant un modèle. Notre travail est organisé en quatre chapitres principaux, qui sont les suivants :

Chapitre 01 : Les maladies cardiaques et L'intelligence artificielle dans le domaine médical

Présente les maladies cardiaques avec leurs types, Se concentrer sur la maladie coronarienne et présenter tout ce qui s'y rapporte leurs Définition, Causes, Les catégories et les complications principales, Les facteurs de risque, Les Symptômes, le Diagnostic, et Discuter de la contribution de

l'intelligence artificielle dans le domaine médical et de ses applications les plus importantes dans diverses spécialités médicales.

Chapitre 02 : État de l'art

Il offre au lecteur un aperçu rapide et complet du domaine L'intelligence artificielle et leur sous-domaines L'apprentissage Automatique et L'apprentissage Profond y compris Les concepts de chaque domaine, il propose également une étude exploratoire et analytique des travaux connexes.

Chapitre 03 : Contribution

Une proposition de modèle de prédiction de La maladie coronarienne est présentée, Pour ce faire, un algorithme de Réseau Neuronal Convolutif (CNN) est utilisé pour la classification ainsi que la présentation des premiers résultats obtenus grâce à sa entraînement le modèle.

Chapitre 04 : Implémentation

Les différents outils et langages de développement qui permettent l'utilisation d'algorithme ont été présentés, ainsi que les mesures de performance utilisées pour évaluer notre modèle. De plus, les résultats de cette proposition ont été présentés.

Enfin, la conclusion, il convient de souligner l'importance de cette démarche dans le domaine de la prédiction des maladies cardiaques en mettant l'accent sur la maladie coronarienne. Cependant, il est également essentiel de reconnaître de ces limites. La conclusion met en évidence les principaux résultats obtenus au cours de cette thèse et propose quelques perspectives de recherche future basées sur ces résultats.

Chapitre 01 :

*Les maladies cardiaques et
L'intelligence artificielle dans le
domaine médical*

1. Introduction :

Les maladies cardiaques sont un groupe de troubles qui affectent la structure ou la fonction du cœur et comprennent un large éventail de problèmes de santé liés au cœur et aux vaisseaux sanguins. Les maladies cardiaques sont les maladies les plus courantes qui entraînent les taux de mortalité les plus élevés dans le monde, en particulier la maladie coronarienne, qui arrive en tête de liste des maladies cardiaques mortelles si elle n'est pas prédictive.

Dans ce chapitre, nous parlerons des maladies cardiaques, de leurs types, causes et symptômes, ainsi que de la contribution de l'intelligence artificielle dans le domaine médical en prédisant les maladies cardiaques, ce qui conduit à son tour à secourir la vie des patients cardiaques.

2. Les maladies cardiaques :

2.1 Définition :

Les maladies cardiovasculaires (MCV) regroupent divers troubles affectant le système circulatoire, tels que les maladies congénitales, ischémiques ou coronariennes, les maladies cérébrovasculaires, les affections vasculaires périphériques, l'insuffisance cardiaque et le rhumatisme cardiaque. La maladie ischémique ou coronarienne est la forme la plus répandue de MCV, affectant les vaisseaux sanguins qui alimentent le muscle cardiaque. Les maladies cérébrovasculaires se caractérisent par des problèmes de circulation sanguine dans les vaisseaux cérébraux, tandis que les affections vasculaires périphériques touchent principalement les vaisseaux qui irriguent les bras et les jambes. L'insuffisance cardiaque survient lorsque le cœur ne pompe pas suffisamment de sang pour répondre aux besoins énergétiques du corps. Le rhumatisme cardiaque est une maladie infectieuse affectant les articulations et les valves cardiaques, tandis que les maladies congénitales sont des anomalies cardiaques présentes dès la naissance [1].

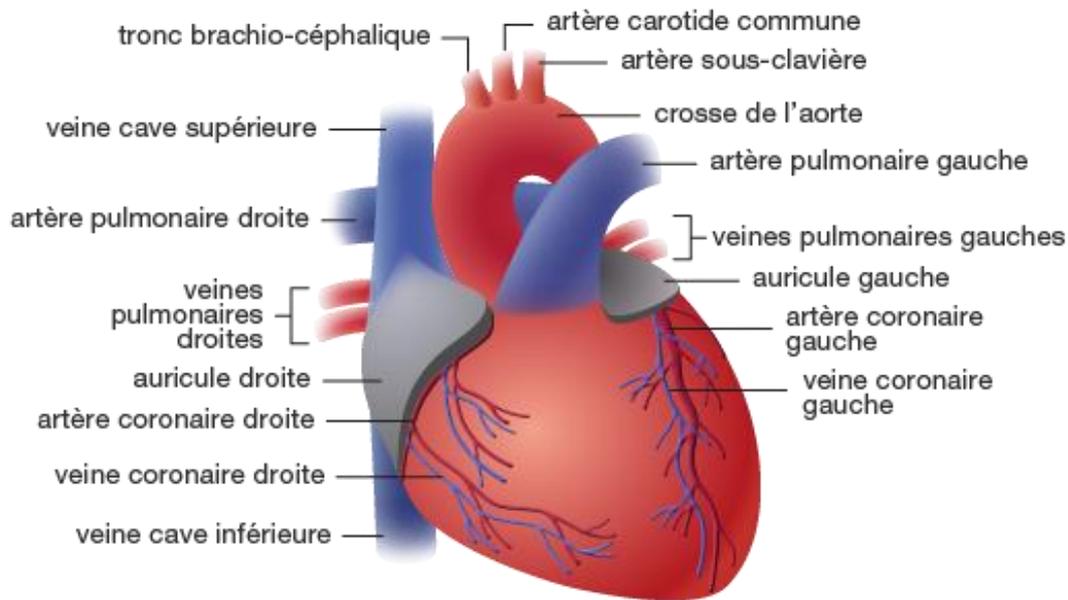


Figure 1.1 : le muscle cardiaque et leur vaisseaux sanguins [2].

2.2 Types des maladies cardiaques :

Les maladies cardiovasculaires sont un groupe d'affections qui affectent le cœur et les vaisseaux sanguins, notamment [3] :

- Maladie coronarienne (affecte les vaisseaux sanguins qui irriguent le muscle cardiaque)
- Maladie cérébro-vasculaires (affecte les vaisseaux sanguins irriguant le cerveau)
- Maladie artérielle périphérique (affecte les vaisseaux sanguins irriguant les bras et les jambes)
- Cardiopathie rhumatismale, qui affecte le muscle cardiaque et les valvules et est causée par un rhumatisme articulaire aigu provoqué par la bactérie streptocoque.
- Malformations cardiaques congénitales (anomalies structurelles du cœur présentes à la naissance)
- Thrombose veineuse profonde et embolie pulmonaire (un caillot sanguin bloquant une veine de la jambe qui peut se rompre et se déplacer vers le cœur ou les poumons).

2.3 Cas d'étude la maladie coronarienne :

La maladie coronarienne est la première cause de mortalité dans les pays développés (7,4 millions dans le monde (organisation mondiale de la santé)), est une pathologie des artères du cœur appelées artères coronaires, qui nourrissent le cœur qui nourrit à son tour tous les organes du corps.

2.3.1 Définition :

La maladie coronarienne est une maladie cardio-vasculaire qui touche les artères ayant pour fonction d'alimenter le cœur en sang (artères coronaires). Elle est souvent causée par l'athérosclérose, une accumulation de plaques à l'intérieur de la paroi des artères. Cette accumulation rétrécit peu à peu l'intérieur des artères et ralentit le flot de sang [4].

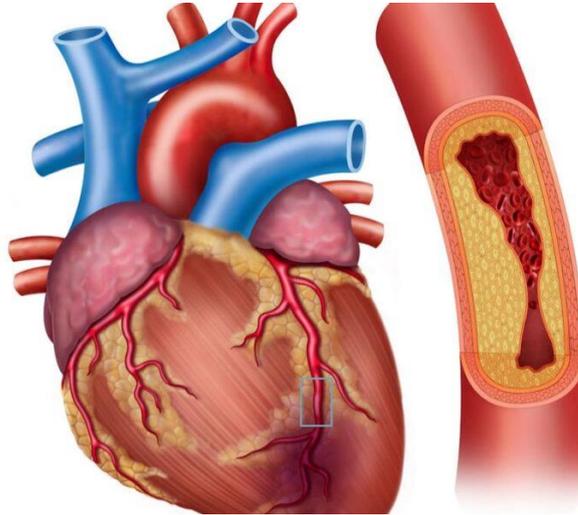


Figure 1.2 : Effets de la maladie coronarienne [5].

2.3.2 Causes :

La maladie des artères coronaires est presque toujours due à l'accumulation graduelle de cholestérol et d'autres matériels lipidiques (appelée athérome ou plaque d'athérosclérose) dans la paroi de l'artère coronaire. Ce processus, appelé athérosclérose, peut affecter de nombreuses artères et pas seulement celles du cœur [6].

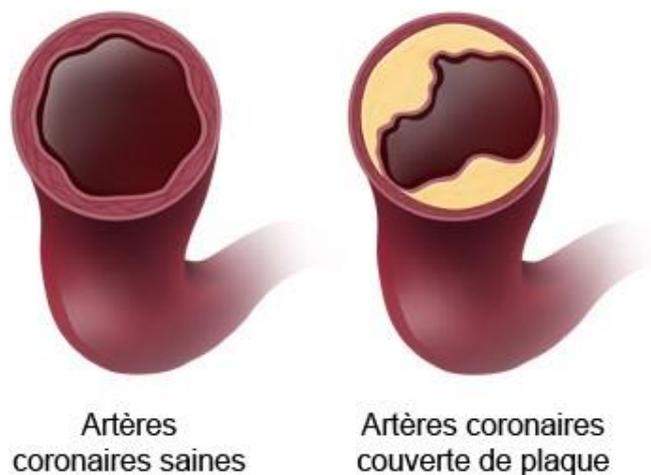


Figure 1.3 : Effets de plaque d'athérosclérose à l'artère coronaire [4].

2.3.3 Les catégories et les complications principales :

Vous trouverez ci-dessous un tableau 1 résumant les catégories ou divisions de La maladie coronarienne, ainsi que les principales complications de cette maladie.

Les catégories	Les complications principales
<ul style="list-style-type: none">• Angine de poitrine (ou angor stable)• Syndrome coronarien aigu (ou angor instable)	<ul style="list-style-type: none">• La mort subite• L'insuffisance cardiaque• Plusieurs types d'infarctus du myocarde• Crise cardiaque

Tableau 1 : Les catégories et les complications principales de La maladie coronarienne.

2.3.4 Les facteurs de risque :

Il y a plusieurs facteurs de risque associés à la maladie coronarienne, certains étant non modifiables tandis que d'autres peuvent être modifiés, le Tableau 2 résume ces facteurs

Pas modifiables	Modifiables
<ul style="list-style-type: none">• Âge avancé• Sexe masculin• Hérité	<ul style="list-style-type: none">• Alimentation trop riche en graisse saturée• Hypercholestérolémie• Tabagisme• Hypertension artérielle• Diabète• Sédentarité• Obésité abdominale• Troubles psycho-sociaux (dont stress et dépression)• Abus d'alcool.

Tableau 2 : Les facteurs de risque de la maladie coronarienne.

2.3.5 Les Symptômes :

Les symptômes d'une maladie coronarienne ont généralement besoin de plusieurs années pour se manifester. La maladie continue de progresser et se manifeste finalement sous la forme d'une angine de poitrine ou d'une crise cardiaque [4]. Les symptômes suivants peuvent survenir aussi bien pendant l'effort physique que pendant le repos :

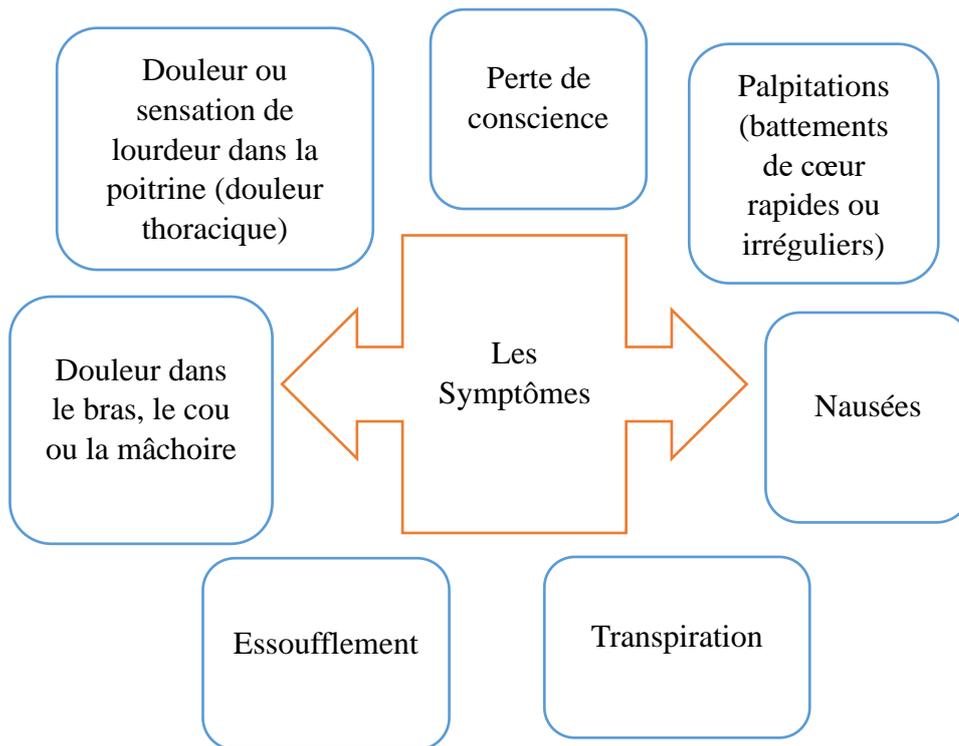


Figure 1.4 : Les symptômes d'une maladie coronarienne.

2.3.6 Diagnostic :

Les examens utilisés couramment pour diagnostiquer une maladie coronarienne comprennent [4] :

- L'électrocardiographie : pour déceler les problèmes de rythme cardiaque ou les signes d'une crise cardiaque
- L'épreuve d'effort sur tapis roulant : pour évaluer la fonction cardiaque lorsque le cœur doit fournir un effort plus important que d'habitude (durant une séance d'exercice)
- L'imagerie nucléaire de perfusion : pour repérer les régions du cœur qui reçoivent moins de sang

- L'échocardiographie : pour déterminer le volume de sang pompé par le cœur. Cet examen peut être réalisé au cours d'une séance d'exercice ou après l'administration d'un médicament qui stimule le cœur.
- La coronarographie par tomodensitométrie : pour repérer les blocages dans les artères du cœur
- Le cathétérisme cardiaque : pour repérer les artères obstruées ou rétrécies

3. L'intelligence artificielle :

3.1 L'intelligence artificielle dans le domaine médical :

L'intelligence artificielle n'appartient pas qu'au futur, elle est déjà largement présente et imprègne dans le domaine de la santé. Sous diverses formes, allant des simples algorithmes aux robots chirurgicaux, l'IA peut apporter une aide quotidienne aux professionnels de la santé en prévenant les effets indésirables, en accélérant les diagnostics et en perfectionnant la médecine de précision, Alors qu'elle devient de plus en plus indispensable dans le domaine de la santé, l'intelligence artificielle reste un concept souvent mal défini, ce qui peut susciter à la fois des fantasmes et des inquiétudes chez les patients et les professionnels de santé. Bien que les premiers algorithmes aient été développés dès les années 1960, ce n'est que récemment que nous avons observé une multiplication de ses applications dans tous les domaines de la médecine. De la détection des maladies rares à l'ophtalmologie en passant par la prévention des épidémies, l'intelligence artificielle a pour objectif d'automatiser certaines tâches et de compiler puis analyser de vastes quantités d'informations. Elle peut élargir les capacités de détection des pathologies en exploitant des téraoctets de données, ou encore assister les professionnels de santé dans des procédures chirurgicales complexes grâce à l'utilisation de robots médicaux [7].

3.2 Applications de l'intelligence artificielle dans le domaine médicales :

3.2.1 Accélérer le diagnostic des maladies rares :

L'utilisation du machine learning ou du deep learning dans le domaine de l'intelligence artificielle permet d'améliorer et d'enrichir les connaissances des professionnels de la santé sur diverses maladies. Il s'agit d'un outil technologique particulièrement précieux pour la détection des maladies rares, qui affectent plus de 3 millions de personnes en France, soit environ 4,5% de la population. Actuellement, plus de 6 000 maladies rares sont répertoriées, et un diagnostic précoce est essentiel pour un traitement efficace. L'objectif est donc de lutter contre l'errance diagnostique en fournissant aux médecins des

bases de données intelligentes, tout en accélérant la prise en charge des patients afin qu'ils puissent bénéficier rapidement d'un parcours de soins adapté à leurs besoins [7].

3.2.2 Favoriser la médecine de précision avec des traitements personnalisés :

L'objectif de la médecine de précision, également appelée médecine personnalisée, est de fournir un traitement adapté à chaque patient, en prenant en compte ses symptômes spécifiques. L'intelligence artificielle joue un rôle crucial en analysant les données de santé du patient, ce qui permet de détecter les variations d'une maladie d'un individu à un autre. Cette analyse des variations peut être déterminante dans le choix du meilleur traitement, surtout pour les maladies complexes [7].

3.2.3 Des robots intelligents pour assister les médecins :

L'intelligence artificielle trouve déjà des applications concrètes dans le domaine de la médecine. Sous forme de robots et de technologies d'automatisation intelligente, elle assiste les chirurgiens, rationalise la logistique médicale, automatise des tâches en laboratoire et assure des tâches telles que le nettoyage des chambres et la distribution de médicaments dans les hôpitaux. Cette technologie en plein essor transforme déjà la pratique médicale en offrant des solutions plus précises, efficaces et autonomes [7].

3.2.4 Développer la chirurgie assistée par ordinateur :

La chirurgie assistée par ordinateur utilise des outils basés sur l'intelligence artificielle pour aider les chirurgiens à planifier et réaliser des interventions chirurgicales. Cela englobe la simulation, la localisation des instruments, le traitement d'images et de données médicales, ainsi que l'utilisation de robots chirurgicaux. L'objectif est d'offrir des interventions adaptées aux besoins des patients, tout en améliorant la précision et l'efficacité des chirurgiens. Un exemple concret est le projet TherA-image du CHU de Rennes, qui permet d'effectuer des interventions chirurgicales réelles dans un environnement innovant développé en collaboration avec le LTSI [7].

3.2.5 Accélérer la prévention des épidémies et effets secondaires des médicaments :

L'intelligence artificielle a été largement utilisée pendant la pandémie de Covid-19, notamment dans la recherche de vaccins, le suivi de l'évolution de l'épidémie et la pharmacovigilance. Elle a permis d'accélérer le séquençage du génome, de réaliser des diagnostics rapides, d'analyser les scanners et d'utiliser des robots pour des tâches de maintenance et de livraison. Des logiciels ont été développés pour observer les tendances liées à la Covid-19, tandis que l'intelligence artificielle a été intégrée dans des applications de pharmacovigilance pour détecter les effets secondaires des médicaments et des vaccins [7].

3.3 Exemples d'applications de l'intelligence artificielle en médicales :

3.3.1 Dans le domaine des maladies cardiovasculaire :

L'intelligence artificielle trouve également des applications dans le domaine de la santé cardiovasculaire. Par exemple, une équipe de l'University College of London développe un algorithme qui utilise des données d'imagerie par résonance magnétique cardiaque pour prédire le risque de décès après un infarctus du myocarde ou un accident vasculaire cérébral. En France, le Samu travaille sur l'adaptation d'un programme d'IA danois capable d'aider les opérateurs du 15 à détecter un arrêt cardiaque en analysant en temps réel des signaux verbaux (mots-clés) et non verbaux (intonation de la voix, rythme de la respiration, etc.) [8], et en plus de Ultromics est une entreprise technologique mondiale qui développe des solutions SaaS révolutionnaires dans le domaine de l'échocardiographie. Leur produit phare, EchoGo, est le premier service d'échocardiographie autonome au monde, utilisant l'intelligence artificielle basée sur le cloud pour faciliter le diagnostic des maladies cardiovasculaires. Les professionnels de la santé peuvent ainsi prendre des décisions rapides et précises lors de l'interprétation des images cardiaques [9].

3.3.2 En cancérologie :

L'intelligence artificielle est de plus en plus utilisée en cancérologie pour faciliter la gestion des données, améliorer le diagnostic et guider la prise en charge thérapeutique. Des projets tels que ConSoRe permettent de collecter et analyser les données massives dans le domaine de la cancérologie, offrant aux médecins la possibilité de trouver des patients répondant à des critères spécifiques et de visualiser l'évolution de leur maladie. L'IA peut détecter certains cancers avec une précision supérieure à celle des médecins, même dans les cas atypiques. Des recherches en bioinformatique explorent les liens entre les données génomiques et les manifestations cliniques des cancers, ouvrant la voie à des diagnostics génétiques personnalisés et à de nouvelles options thérapeutiques. De plus, des algorithmes aident à prédire la réponse des patients à certains traitements, permettant d'éviter des interventions chirurgicales inutiles. Des programmes d'aide à la décision thérapeutique (le projet européen Desiree, auquel ont participé l'AP-HP et le laboratoire LIMICS) ont également été développés pour le cancer du sein, basés sur les meilleures pratiques et l'expérience des réunions de concertation pluridisciplinaires [8].

3.3.3 Dans le domaine de la transplantation :

Une équipe française a coordonné une étude internationale dans le domaine de la transplantation pour valider un algorithme universel capable de prédire le risque de rejet d'un rein transplanté chez un

patient. Cette avancée permet d'améliorer le suivi des patients transplantés et d'optimiser le développement de nouveaux traitements immunosuppresseurs [8].

3.3.4 Appui au diagnostic en gynécologie-obstétrique :

Le projet SUOG (Smart Ultrasound in Obstetrics and Gynecology) utilise l'intelligence artificielle pour améliorer le suivi de la grossesse en gynécologie-obstétrique. Il intègre un programme d'IA à l'appareil d'échographie, ce qui permet au praticien de recevoir en temps réel des suggestions d'images à réaliser pour poser un diagnostic en cas de pathologie suspectée. Ce projet, mené par des équipes de Sorbonne-Université, l'Inserm et l'AP-HP en France, vise à optimiser la prise en charge des patientes [8].

3.3.5 La prévention en santé mentale :

Le projet PsyCARE vise à développer une intelligence artificielle (IA) pour la détection précoce de la schizophrénie et de la psychose chronique. Dirigé par Marie-Odile Krebs de l'Université de Paris, le projet vise à mettre en place une prise en charge psychiatrique préventive et personnalisée. L'IA utilisera la modélisation des connaissances et des biomarqueurs pour identifier les personnes à risque [8].

4. Conclusion :

Nous avons discuté dans ce chapitre à tout ce qui concerne les maladies cardiovasculaires, y compris La maladie coronarienne est l'une des maladies les plus dangereuses qui entraînent la mort, leurs causes, leurs signes, leurs caractéristiques, les étapes de diagnostic et les caractéristiques de ces maladies cardiaques distinctives, ainsi que les installations et le soutien apportés par l'intelligence artificielle. Dans le domaine médical, avec son contributeur au développement et à la prospérité de ce domaine. Avec la présence d'exemples d'application, tout cela a conduit à l'avancée technologique dans ce domaine, ce qui a contribué en réduisant le nombre de décès.

Dans le prochain chapitre, nous discuterons des techniques d'intelligence artificielle et Nous découvrirons également les techniques utilisées par les chercheurs dans les derniers articles dans ce domaine.

Chapitre 02 :
État de l'art

1. Introduction :

L'intelligence artificielle (IA) est un domaine de recherche en pleine expansion. Ses utilisations ont pénétré divers domaines de notre vie quotidienne. Ces dernières années, l'intelligence artificielle est passée d'une promesse d'avenir à une norme d'innovation. Grâce à ses applications qui concernent tous les êtres humains activités, elle permet notamment d'améliorer la qualité des soins.

Dans ce chapitre, nous parlerons les deux des sous-domaines de l'intelligence artificielle : machine Learning, Deep Learning et leur méthode, technique et algorithmes, La contribution de ces deux domaines dans la prédiction des maladies cardiaques en particulier la maladie coronarienne à travers de l'étude des travaux connexes dans le domaine de prédiction des maladies cardiaques, Ce sera le cœur de notre sujet dans ce chapitre.

2. L'intelligence artificielle (IA) :

Le Deep learning (L'apprentissage Profond) et le machine learning (L'apprentissage Automatique) ayant tendance être utilisés de manière interchangeable, mais il est important de noter les nuances entre les deux, le deep learning et le machine learning sont tous deux des sous-domaines de l'intelligence artificielle, et le deep learning est en fait un sous-domaine du machine learning [10].

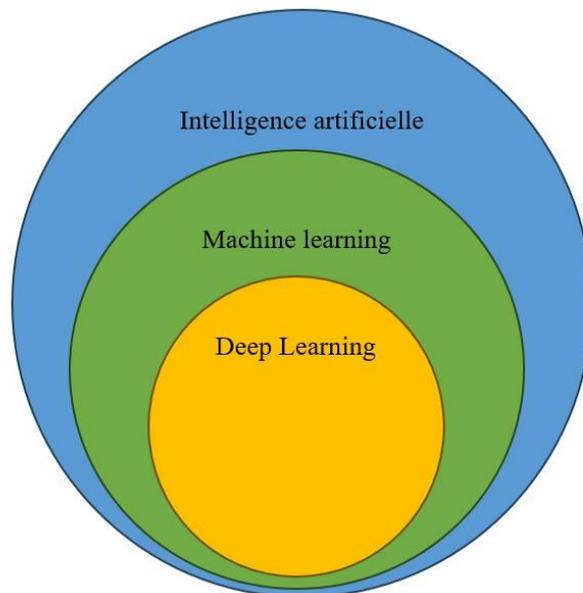


Figure 2.1 : La relation entre IA, ML et DL.

2.1 L'apprentissage Automatique : Machine Learning

2.1.1 Définition :

« L'apprentissage automatique (en anglais machine learning, littéralement « apprentissage machine »), apprentissage artificiel ou apprentissage statistique est un champ d'étude de l'intelligence artificielle qui se fonde sur des approches mathématiques et statistiques pour donner aux ordinateurs la capacité d'« apprendre » à partir de données, c'est-à-dire d'améliorer leurs performances à résoudre des tâches sans être explicitement programmés pour chacune. Plus largement, il concerne la conception, l'analyse, l'optimisation, le développement et l'implémentation de telles méthodes. On parle d'apprentissage statistique car l'apprentissage consiste à créer un modèle dont l'erreur statistique moyenne est la plus faible possible. » [11].

2.1.2 Les Méthodes et Les principaux types d'algorithmes d'Apprentissage Automatique :

Différentes approches d'apprentissage automatique sont utilisées en fonction des problèmes à résoudre et des données disponibles. La Figure 2.2 présente un organigramme expliquant Les méthodes d'apprentissage automatique et leur fonctionnement.

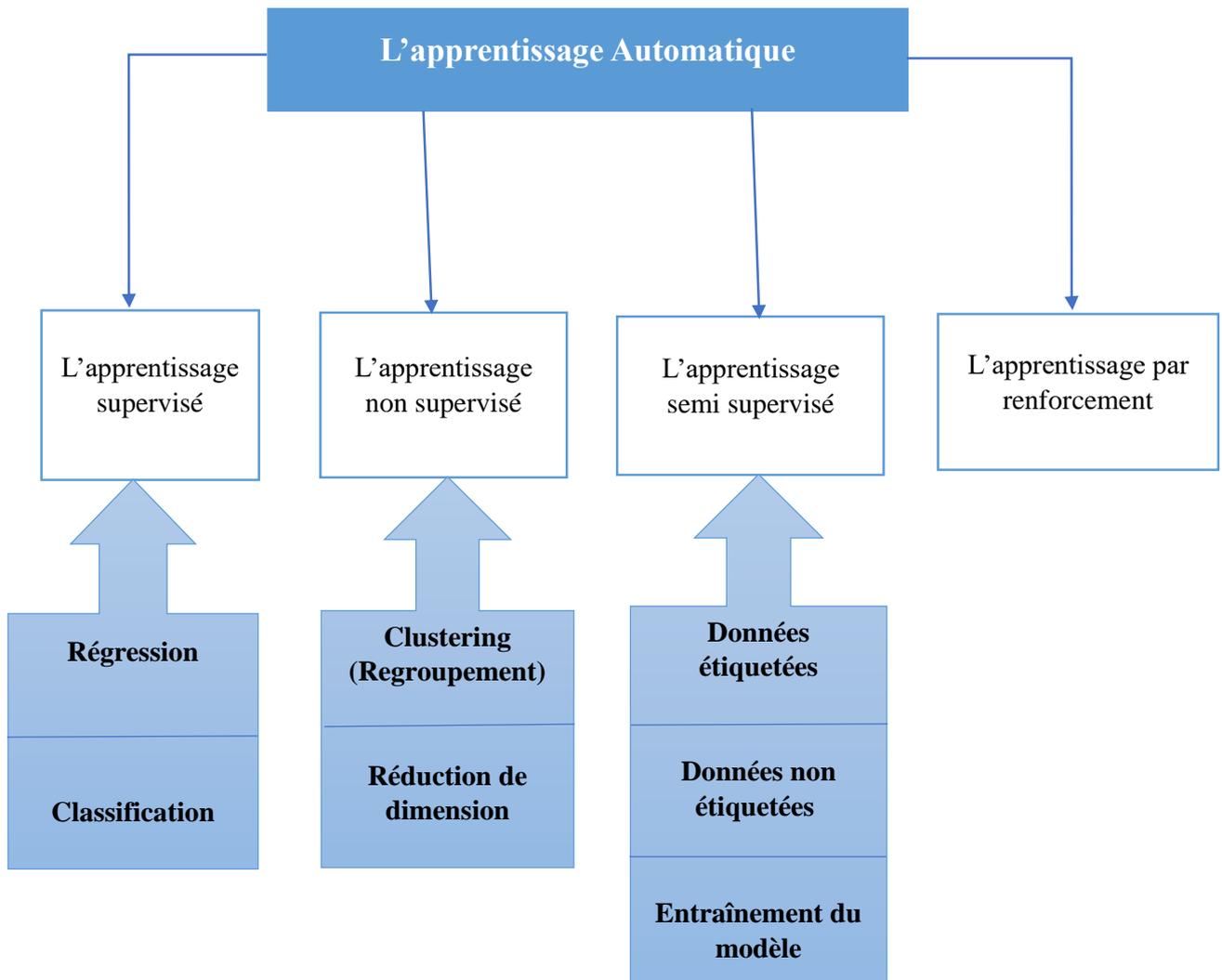


Figure 2.2 : Les méthodes d'apprentissage automatique et leur fonctionnement.

Le Tableau 3 présente un résumé des types/méthodes et Les principaux algorithmes d'apprentissage automatique les plus connus, il clarifie et explique les éléments présentés dans l'organigramme ci-dessus (Figure 2.2).

Types d'apprentissage	Définition	Types d'Algorithmes	Algorithme commun
Apprentissage supervisé	Parmi les quatre types d'algorithmes de machine learning, les algorithmes d'apprentissage supervisé sont les plus simples. Ces algorithmes requièrent une supervision directe de la part du développeur du modèle. Le développeur étiquette le corpus de données d'exemples et définit des limites strictes sur lesquelles l'algorithme fonctionnera.	<p>Régression : est le processus d'identification des modèles et de calcul des prédictions des résultats continus</p> <p>Classification : est le processus qui consiste à apprendre à partir d'échantillons de données passés et à entraîner manuellement le modèle pour prédire les résultats essentiellement binaires (oui / non, vrai / faux, 0/1).</p>	<ul style="list-style-type: none"> • Machines à vecteurs de support (SVM) • Régression linéaire • Régression logistique • Bayes naïf • Analyse discriminante linéaire • Arbres de décision • Algorithme des K plus proches voisins • Réseaux de neurones (Perceptron multicouche) • Boosting • Forêts aléatoires
Apprentissage non supervisé	Lorsque le système ou l'opérateur dispose uniquement d'exemples de données sans étiquettes, et que le nombre de classes et leur nature n'ont pas été prédéterminées, on parle d'apprentissage non supervisé ou clustering en anglais. Aucun expert n'est requis. L'algorithme doit découvrir par lui-même la structure plus ou moins cachée des données. Le partitionnement de données, data clustering en anglais, est un algorithme d'apprentissage non supervisé.	<p>Clustering : Identifier des groupes similaires dans les données</p> <p>Réduction de dimensionnalité : Réduire la complexité des données tout en préservant les informations importantes</p>	<ul style="list-style-type: none"> • K-signifie clustering • t-SNE (incorporation de voisin stochastique distribué en t) • PCA (analyse en composantes principales) • Règle d'association.
Apprentissage semi-supervisé	Les algorithmes d'apprentissage semi-supervisé représentent un juste milieu entre les algorithmes supervisés et non supervisés, offrant un compromis entre les	<p>Fonctionnement des algorithmes semi-supervisés :</p> <p>Données étiquetées : Ces données sont associées à des sorties connues</p>	<ul style="list-style-type: none"> • LabelPropagation • LabelSpreading • Co-training • Self-Training • Multi-view learning

	deux approches. Ces modèles semi-supervisés combinent certains aspects des deux méthodes pour former une approche distincte.	<p>Données non étiquetées : Ces données n'ont pas de sorties associées, mais elles sont essentielles pour enrichir notre modèle</p> <p>Entraînement du modèle : L'idée est que le modèle apprendra à partir des données étiquetées et généralisera ensuite ces connaissances aux données non étiquetées</p>	
Apprentissage par renforcement	L'apprentissage par renforcement implique le développement d'un système autonome qui s'améliore au fil du temps grâce à des séquences contiguës d'essais et d'échecs, Ce processus combine l'utilisation de données étiquetées avec des interactions adaptatives aux données d'entrée pour renforcer les performances du système.		<ul style="list-style-type: none"> • Q-Learning • Deep Q Networks (DQN) • Policy Gradient Methods • Différence temporelle (TD) • Recherche arborescente Monte-Carlo (SCTM) • Agents critiques d'acteur asynchrones (A3C)

Tableau 3 : les différentes méthodes de l'Apprentissage Automatique et leur type d'Algorithmes de chaque méthode et les Algorithmes communs.

2.2 L'apprentissage Profond : Deep learning

2.2.1 Définition :

Le Deep Learning, ou apprentissage profond, est un sous-ensemble du Machine Learning, ou apprentissage automatique, basé sur des réseaux neuronaux artificiels. Le processus d'apprentissage est qualifié de profond parce que la structure des réseaux neuronaux artificiels se compose de plusieurs couches d'entrée, de sortie et masquées. Chaque couche contient des unités qui transforment les données d'entrée en informations que la couche suivante peut utiliser une tâche prédictive spécifique. Grâce à cette structure, une machine est capable d'apprendre au travers de son propre traitement de données [12].

2.2.2 Réseau de neurones profonds « DNN » :

Le Deep Neural Network, également connu sous le nom de réseau de neurones profond, est une technologie basée sur des algorithmes inspirés du fonctionnement du cerveau humain. Son objectif est de simuler l'activité cérébrale, en particulier la reconnaissance de motifs et la transmission d'informations entre les différentes couches de connexions neuronales. Ce qui le distingue, c'est qu'il est composé d'au moins deux couches, ce qui lui permet de traiter les données de manière complexe en utilisant des modèles mathématiques avancés. [13].

2.2.3 Les grandes classes de réseaux de neurones profonds :

Les réseaux de neurones profonds sont des architectures puissantes largement utilisées dans le domaine de l'apprentissage profond. Il y a trois grandes classes [14] :

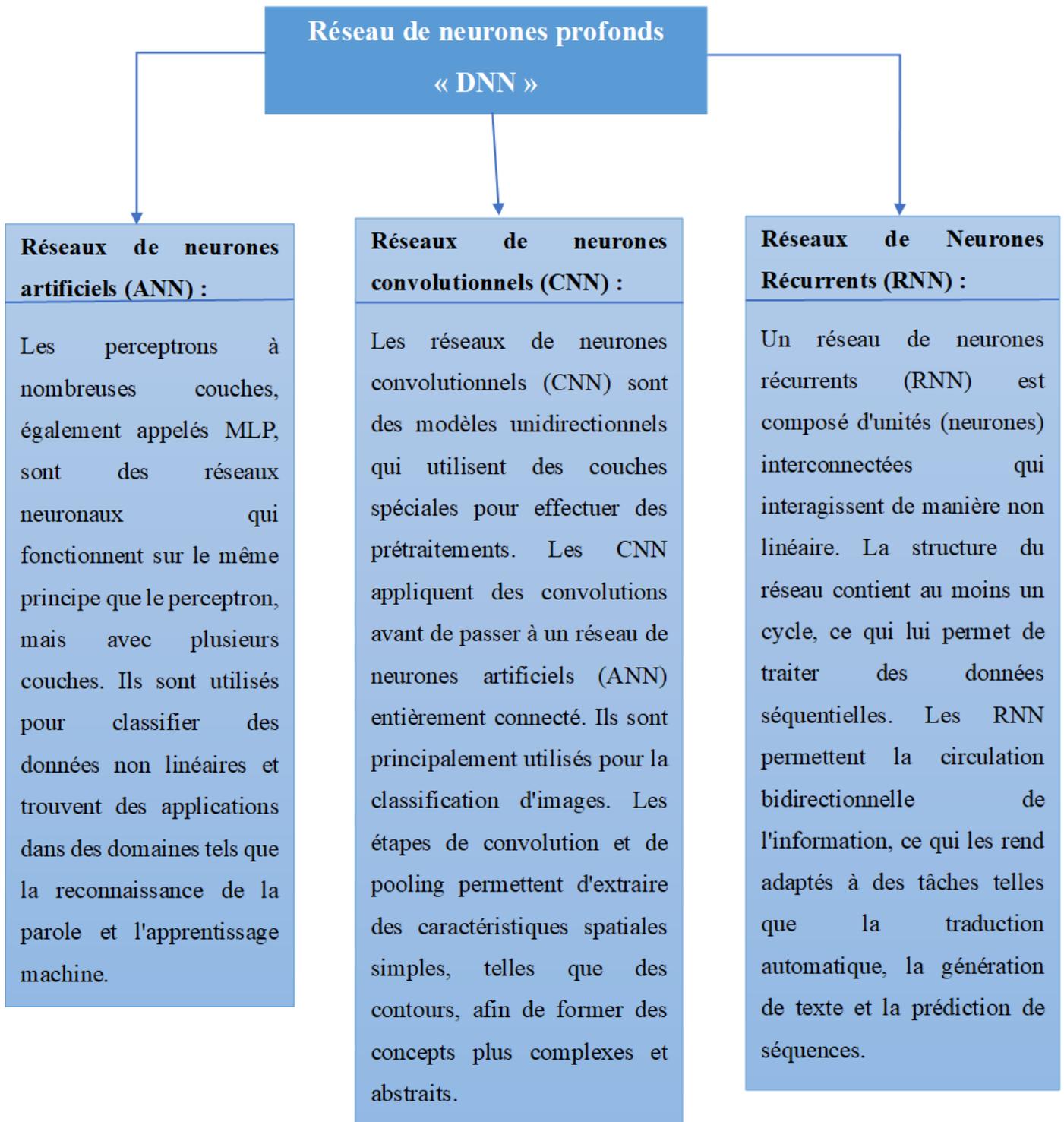


Figure 2.3 : Les trois grandes classes de réseaux de neurones profonds.

3. Les travaux connexes :

1) Machine Learning and Deep Neural Network Techniques for Heart Disease Prediction

les auteurs ont fourni une étude approfondie sur la prédiction des maladies cardiaques en utilisant des techniques de machine learning (ML) et des réseaux de neurones profonds (DNN) (DL), et compare les performances de différents modèles de classification, dont des techniques de machine learning (ML) et des réseaux de neurones profonds (DNN) , Elle utilise un ensemble de données de référence provenant de la base de données UCI sur les maladies cardiaques, comprenant 14 caractéristiques liées aux maladies cardiaques, Le modèle de classification Extreme Gradient Boosting a obtenu la plus haute précision parmi les modèles de ML (81,10 %), par contre Le réseau de neurones profonds à trois couches (DNN3) a fourni la meilleure précision (85,41 %) en utilisant des caractéristiques sélectionnées en entrée [15].

2) Prediction of Heart Disease Using Machine Learning : A Systematic Literature Review

les auteurs ont fourni une revue systématique des travaux de recherche qui traitent de l'utilisation de l'apprentissage automatique (ML) et de l'apprentissage profond (DL) pour prédire les maladies cardiaques (Le regroupement de 43 articles analysés), Les pays tels que l'Inde, la Chine et le Pakistan ont mené le plus grand nombre d'études sur l'utilisation de ML et de DL pour prédire les maladies cardiaques , Les techniques les plus utilisées étaient Random Forest, SVM et Logistic Regression, parmi lesquelles XGBoost, Ensemble Deep learning et Stacking ont obtenu les meilleurs résultats en termes de précision entre 90 % et 100 % , et Python a été considéré comme l'outil le plus performant, Les métriques les plus couramment utilisées étaient l'exactitude, la précision et le score F1, et le type de maladie appliqué était la maladie coronarienne, et les algorithmes de sélection étaient Kernel et Information Gain [16].

3) Intelligent Methods for Early Prediction of Heart Disease

L'auteur met en évidence l'importance de l'apprentissage automatique (ML) pour prédire avec précision les maladies cardiaques. En utilisant une variété d'algorithmes de ML et un ensemble de données soigneusement collecté, plusieurs classificateurs pour le diagnostic des maladies cardiaques ont été développés et évalués. Les classificateurs développés ont été évalués à l'aide d'un nouvel ensemble de données provenant du Médical Help Center, un hôpital privé et un centre cardiaque à Erbil, en Irak, Les techniques utilisées ont atteint un niveau de précision supérieur à 92 %, Le

classificateur de boosting de gradient a obtenu la meilleure performance, avec une précision de 98,51 % lors de la phase de test [17].

4) Heart disease prediction using distinct artificial intelligence techniques : performance analysis and comparison

Les auteurs ont examiné l'utilisation de diverses techniques d'intelligence artificielle (ML) pour la prédiction des maladies cardiaques ,en utilisant la technique de sélection de sous-ensemble de caractéristiques basée sur la corrélation avec la recherche Best First, Les facteurs les plus significatifs identifiés comprennent l'âge, le sexe, le tabagisme, l'obésité, l'alimentation, l'activité physique, le stress, le type de douleur thoracique, la douleur thoracique antérieure, la pression artérielle diastolique, le diabète, la troponine, l'ECG ,Plusieurs techniques d'intelligence artificielle ont été appliquées et comparées, et La forêt aléatoire utilisant des caractéristiques sélectionnées a atteint le taux de précision le plus élevé (90 %) par rapport à l'utilisation de toutes les caractéristiques d'entrée et d'autres techniques d'intelligence artificielle , L'approche proposée pourrait être utilisée comme un cadre d'assistance pour prédire les maladies cardiaques à un stade précoce [18].

5) Deep Learning Predicts Heart Failure With Preserved, Mid-Range, and Reduced Left Ventricular Ejection Fraction From Patient Clinical Profiles

Les auteurs ont concentré sur l'utilisation de techniques de deep learning pour prédire l'insuffisance cardiaque en fonction de la fraction d'éjection ventriculaire gauche (FEVG) des patients, La FEVG est un indicateur clé pour évaluer l'insuffisance cardiaque chez les patients atteints de maladie coronarienne, Elle aide à catégoriser l'insuffisance cardiaque en trois types : préservée (HFpEF), intermédiaire (HFmEF) et réduite (HFrEF),Les données de 303 patients atteints de maladie coronarienne ont été obtenues et catégorisées selon les directives ASE/EACVI en HFpEF (FEVG > 55%), HFmEF ($50\% \leq \text{FEVG} \leq 55\%$) et HFrEF (FEVG < 50%),et Les profils cliniques comprenaient 13 marqueurs démographiques et cliniques regroupés en facteurs de risque cardiovasculaire, médicaments et antécédents ,et Les techniques de deep learning employées ont atteint un niveau de précision 90,43% [19].

6) Predicting cardiovascular events with deep learning approach in the context of the internet of things

Les auteurs ont exploré l'utilisation de techniques de deep learning pour prédire les événements cardiovasculaires, Pour identifier et prévenir rapidement ces événements, le réseau de neurones Long Short-Term Memory (LSTM) a été utilisé en raison de sa capacité à apprendre des dépendances à long terme, et Un Deep Belief Network (DBN) a également été utilisé pour représenter et sélectionner les caractéristiques les plus efficaces et pertinentes de l'ensemble de données enregistré, Les résultats expérimentaux ont montré que le LSTM-DBN proposé avait une précision moyenne de 88,42 %, ce qui est significativement meilleur par rapport à toutes les autres approches de deep learning et classifications traditionnelles [20].

7) Prediction of Heart Disease Using a Collection of Machine and Deep Learning Algorithms

Les auteurs ont examiné l'utilisation d'une variété d'algorithmes d'apprentissage automatique et de deep learning pour la prédiction des maladies cardiaques, et L'objectif est de déterminer quels patients sont plus susceptibles de souffrir de maladies cardiaques en se basant sur un certain nombre de caractéristiques médicales (Kaggle),en utiliser De nombreux algorithmes différents d'apprentissage automatique et Deep Learning pour prédire et classer les patients atteints de maladies cardiaques, La force du modèle proposé était très satisfaisante et a pu prédire la présence d'une maladie cardiaque chez une personne en utilisant Deep Learning et Random Forest Classifier, qui ont montré une bonne précision 92.23% par rapport aux autres classificateurs utilisés [21].

8) Heart Disease Detection using Hybrid Machine Learning and IoT (Software Based)

Les auteurs ont abordé la détection des maladies cardiaques en utilisant une combinaison d'approches de machine learning et de l'Internet des Objets (IoT) ,utilise Un modèle hybride de machine learning est proposé, combinant des approches de forêt aléatoire (Random Forest) avec des méthodes linéaires pour prédire les maladies cardiaques, Les données sont collectées via des dispositifs IoT et traitées à l'aide de ce modèle pour identifier les signes précurseurs de maladies cardiaques, le modèle de prédiction atteint un niveau de performance supérieur avec un taux de précision de 88,7 % pour prédire les maladies cardiaques, La forêt aléatoire hybride avec modèle linéaire (HRFLM) s'est avérée particulièrement efficace pour atteindre ces résultats [22].

9) Detection of Heart Disease Using Supervised Machine Learning

Les auteurs présentent une étude sur la détection des maladies cardiaques en utilisant des techniques d'apprentissage supervisé, Le modèle proposé inclut plusieurs techniques de machine learning pour obtenir des prédictions précises des maladies cardiaques, L'ensemble de données utilisé dans cette étude provient du Référentiel d'apprentissage automatique de l'Université de Californie à UCI ,il est composé de 13 variables mesurées sur 271 individus, Diverses méthodes d'évaluation, telles que la sensibilité, la précision, la spécificité, le MCC, la matrice de confusion et la précision, sont utilisées pour l'évaluation des performances du modèle, L'analyse comparative a montré que l'arbre de décision offre la plus haute précision de 85,21 % avec toutes les caractéristiques [23].

10) Prediction of Coronary Heart Disease using Supervised Machine Learning Algorithms

Les auteurs ont concentré sur la prédiction de la maladie coronarienne en utilisant des algorithmes d'apprentissage supervisé, le modèle de prédiction pour les maladies coronariennes (CHD) est proposé, impliquant le remplacement des valeurs nulles, le rééchantillonnage, la standardisation, la normalisation, la classification et la prédiction, L'approche vise à prédire le risque de CHD en utilisant des algorithmes d'apprentissage automatique tels que Random Forest, Decision Trees et K-Nearest Neighbours , La validation croisée K-fold est utilisée pour générer de l'aléatoire dans les données, Ces algorithmes sont testés sur le jeu de données "Framingham Heart Study", qui comprend 4240 enregistrements ,Dans l'analyse expérimentale, Random Forest, Decision Tree et K-Nearest Neighbour ont atteint une précision de 96,8 %, 92,7 % et 92,89 % respectivement [24].

11) Predicting Coronary Heart Disease Using an Improved LightGBM Model : Performance Analysis and Comparison

Les auteurs présentent une étude sur l'utilisation d'un modèle LightGBM optimisé pour la prédiction de la maladie coronarienne (CHD), Pour optimiser le classificateur LightGBM, les hyperparamètres du modèle LightGBM ont été ajustés, sa fonction de perte a été améliorée, et le modèle a été formé en utilisant ces hyperparamètres ajustés, Le cadre d'optimisation des hyperparamètres le plus avancé, OPTUNA, a été appliqué pour optimiser les hyperparamètres du modèle de prédiction. La fonction de perte améliorée est appelée la perte focale (FL), Le modèle a été évalué en utilisant des données CHD de l'Institut Framingham Heart. Pour évaluer la performance du modèle de prédiction, diverses métriques ont été utilisées, y compris la précision, le rappel, le score F, l'exactitude, le MCC, la

sensibilité, la spécificité et l'AUC. La valeur AUC du modèle proposé était de 97,8 %, ce qui était mieux que celle des autres modèles comparatifs [25].

12) Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers

Les auteurs ont exploré l'utilisation de divers classificateurs d'apprentissage automatique pour prédire efficacement la présence de la maladie coronarienne (CHD). La maladie coronarienne est l'une des principales causes de décès dans le monde, Prédire la CHD est un défi majeur dans l'analyse des données cliniques, Onze classificateurs d'apprentissage automatique ont été utilisés pour introduire le modèle de prédiction, en utilisant diverses combinaisons de caractéristiques et des algorithmes de classification, Les arbres boostés par gradient et le perceptron multicouche ont atteint une précision de 95% dans le modèle de prédiction de la CHD. Le Random Forest a donné une meilleure performance avec un niveau de précision de 96,28%, une spécificité et une sensibilité de 0.9628 et 0.9537 respectivement [26].

13) Deep learning of heart-sound signals for efficient prediction of obstructive coronary artery disease

Les auteurs ont concentré sur l'utilisation de l'apprentissage profond pour analyser les signaux sonores du cœur afin de prédire efficacement la maladie coronarienne obstructive, L'étude a inclus 320 sujets suspectés de CAD qui ont subi une CAG. Des techniques de filtrage avancées et des modèles d'apprentissage profond de pointe (VGG-16, CNN 1D et ResNet18) ont été utilisés pour analyser les signaux sonores du cœur et identifier la CAD obstructive (définie comme au moins une sténose ≥ 50 %), Dans l'ensemble de test, VGG-16 a montré les meilleures performances avec une aire sous la courbe ROC (AUC) de 0,834, tandis que ResNet-18 et CNN-7 ont obtenu des AUC de seulement 0,755 et 0,652 respectivement. VGG-16 a démontré une sensibilité de 80,4 % et une spécificité de 86,2 % dans l'ensemble de test. Le modèle diagnostique combiné de VGG et de scores DF a atteint une AUC de 0,915, et l'AUC pour VGG combiné avec des scores PTP était de 0,908. La sensibilité et la spécificité de VGG-16 ont dépassé 0,85 chez les patients avec occlusion des artères coronaires et ceux avec 3 lésions vasculaires [27].

14) A hybrid deep neural net learning model for predicting Coronary Heart Disease using Randomized Search Cross-Validation Optimization

Les auteurs ont exploré une méthode avancée pour la prédiction des maladies cardiaques coronariennes en utilisant un modèle d'apprentissage profond hybride, a proposé un modèle hybride d'apprentissage de réseau neuronal profond pour prédire CHD à l'aide de l'ensemble de données BRFSS-2015, Ce modèle combine plusieurs techniques de machine learning et de deep learning, tel que BiLSTM, GRU, RSCV, LSTM, BiLSTM-GRU, Le modèle a atteint une précision impressionnante de 93,3% dans la prédiction des maladies cardiaques coronariennes, ce qui démontre l'efficacité de l'apprentissage profond pour cette tâche et souligne le potentiel d'utilisation de techniques avancées dans la prédiction des maladies cardiaques pour améliorer les résultats de santé [28].

15) Prediction of Coronary Artery Disease Using Machine Learning Techniques with Iris Analysis

Les auteurs ont proposé une méthode novatrice pour le diagnostic non invasif de la maladie coronarienne (CAD) en utilisant l'analyse de l'iris, L'iridologie, une méthode d'analyse de l'iris pour diagnostiquer les conditions de santé, a été combinée avec des techniques de traitement d'image pour détecter la maladie chez un total de 198 volontaires, dont 94 avec CAD et 104 sans, Le modèle proposé à un taux de précision de 93 % pour prédire la CAD en utilisant le classificateur de machine à vecteurs de support (SVM). Avec la méthode proposée, la CAD peut être diagnostiquée de manière préliminaire par analyse de l'iris sans nécessiter d'électrocardiographie, d'échocardiographie et de tests d'effort. De plus, la méthode proposée peut être facilement utilisée pour soutenir les applications de télémédecine pour la CAD dans les systèmes de télémédecine intégrés [29].

3.1 Tableaux de comparaison :

Référence	Techniques	Évaluation Paramètres	Datasets	Accuracy	Conclusion / travaux futurs
[15]	XGBoost, Ada Boost, Light Gradient Boosting Machine, CatBoost, Random Forest, Ridge, Decision Tree, Logistic Regression, K Neighbors, SVM-Linear Kernel, Naive Bayes et des réseaux de neurones profonds (DNN3 et DNN4).	Accuracy, Precision, Recall, F-1 score, et AUC curve	UCI	ML (Extreme Gradient Boosting) 81.10% DL (DNN3) 85.41%	
[16]	Random Forest, SVM et Logistic Regression, XGBoost, Ensemble Deep learning et Stacking	accuracy, precision, et F1 score	UCI	Entre 90 % et 100 %	<ul style="list-style-type: none"> • L'exploration de différentes pathologies cardiaques • Amélioration de la qualité des données • Mise à jour continue des modèles • Interprétation des résultats pour les professionnels de la santé
[17]	ML algorithms		Dataset local	ML techniques 92 % The gradient-boosting classifier 98.51 %	
[18]	Logistic regression, Naïve Bayes, KNN, SVM, decision tree, random forest, MLP.	Accuracy, Precision, Recall, F1-score, ROC-AUC Score	UCI	Random forest 90%	<ul style="list-style-type: none"> • l'amélioration des modèles de prédiction et l'exploration de nouvelles techniques d'intelligence artificielle pour améliorer les résultats cliniques

[19]	CNN, BN, ReLu, ADAM, ROC, AUROC.	Accuracy	Dataset local	deep learning 90.43%	<ul style="list-style-type: none"> • ajouter des données démographiques et cliniques supplémentaires.
[20]	LSTM-DBN, LSTM, DBN, simple RNN, GRU, CNN, Ensemble, MLP, SVM, Logistic Regression, Random Forest, PCA, AutoEncoder.	Precision, Sensitivity, F-measure, Accuracy, Fall-out, Miss rate, Specificity	Dataset local (IoT)	LSTM-DBN 88.42%	<ul style="list-style-type: none"> • la nécessité de recherches plus approfondies pour aborder la complexité et l'efficacité de la méthode proposée et pour combler le fossé entre l'utilisation académique et pratique des approches d'apprentissage profond et la collecte de données.
[21]	Gaussian Mixture, Nearest Centroid, MultinomialNB, Logistic RegressionCV, Linear SVC, Linear Discriminant Analysis, SGD Classifier, Extra Tree Classifier, Calibrated ClassifierCV, Quadratic Discriminant Analysis, GaussianNB, Random Forest Classifier, ComplementNB, MLP Classifier, BernoulliNB, Bagging Classifier, LGBM Classifier, Ada Boost Classifier, K Neighbors Classifier, Logistic Regression, Gradient Boosting Classifier,	Accuracy, Precision, Recall, F1_score	Kaggle	Deep Learning, Random Forest 92.23%	<ul style="list-style-type: none"> • l'exactitude de L'apprentissage profond est le plus élevé parmi les collections d'algorithmes que utilisés
[22]	The hybrid random forest linear model (HRFLM)		Dataset local collecté avec (IoT)	88.7%	<ul style="list-style-type: none"> • La coordination de l'IoT dans les services médicaux a donné aux patients des stratégies indolores pour respecter leurs limites en matière de bien-être, ce qui Il s'agit d'une énorme

					amélioration par rapport aux techniques traditionnelles
[23]	SVM, DT, LR, KNN, NB, Relief, LASSO	Accuracy, Sensitivity, Specificity, Precision, MCC	UCI	Decision tree 85.21%	<ul style="list-style-type: none"> •Généraliser davantage le modèle à le rendre compatible avec d'autres techniques de sélection de fonctionnalités et plus résistant aux ensembles de données contenant des quantités importantes de données manquantes •Une autre stratégie future consiste à utiliser le Deep Learning algorithmes.
[24]	Random Forest, Decision Trees, K-Nearest Neighbours	Accuracy, Precision, Recall Specificity, F1 Score, ROC	Framingham heart study dataset (Kaggle)	Random Forest 96,8 %	<ul style="list-style-type: none"> •Random Forest est une bonne sélection Pour obtenir un modèle de prédiction robuste •Un ensemble de données plus important est nécessaire pour obtenir un meilleur modèle de formation, et concentrer sur l'importance de prétraitement
[25]	HY_OptGBM (LightGBM, loss function, OPTUNA)	Precision, Recall, F-score, AUROC, AUPRC and MCC, Accuracy, Sensitivity, Specificity	Framingham heart study dataset (Kaggle)	97.8%	<ul style="list-style-type: none"> •Les ensembles de données augmentent, il faudra obtenir un succès résultat en ajustant les paramètres par défaut présentés , En plus d'utiliser un modèle unique pour prédire les maladies coronariennes, alternativement, on peut envisager de construire un modèle de prédiction en combinant plusieurs modèles.
[26]	LR, SVM, NB, RF, XGB, DT, NN, RBF, KNN, GBT, MLP	Accuracy, Precision (specificity) , Recall	Heart Disease (UCI Repository)	Random Forest 96.28%	<ul style="list-style-type: none"> •Des ensembles de données supplémentaires pour tenter d'obtenir des conclusions plus fiables

		(sensitivity), F-Measure			•Optimiser les paramètres des classificateurs ML et des méthodes d'apprentissage en profondeur à l'aide de techniques métaheuristiques
[27]	VGG-16, 1D CNN, ResNet18	Sensitivity, specificity, AUC, F1-Score, Accuracy	Dataset local	VGG-16 95 %	•les résultats introduisent un nouveau concept selon lequel l'utilisation d'algorithmes d'apprentissage profond avec des signaux sonores cardiaques bruts peut faciliter la détection de la coronaropathie obstructive.
[28]	BiLSTM, GRU, RSCV, LSTM, BiLSTM-GRU	Positive predicated value (PPV), Recall or Sensitivity, Specificity, Negative predicted Value(NPV), F1-Score, Accuracy	The BRFSS-2015 Dataset	BiLSTM-GRU 98.28%	•Ce modèle BiLSTM-GRU proposé sera appliqué à d'autres nouveaux ensembles de données d'enquête BRFSS sur les maladies cardiaques ainsi qu'à d'autres prévisions de maladies.
[29]	SVM, statistical analysis, Gray-Level Co-Occurrence Matrix (GLCM), et Gray Level Run Length Matrix (GLRLM)	Sensitivity, specificity, F1-Score, Accuracy, Geometric Mean, Precision	Dataset local	SVM 93 %	•Cette étude fournit une référence pour détecter la CAD à partir d'images d'iris. Dans les études futures, cela peut être connu la relation entre diverses maladies cardiaques, et l'amélioration des performances peut être obtenue en essayant différentes extractions de caractéristiques détectant diverses maladies à l'aide de réseaux neuronaux convolutifs.

Tableau 4 : Récapitulation des travaux connexes.

3.2 Synthèse :

Les auteurs des articles présentés portent sur l'utilisation de techniques d'apprentissage automatique (Machine Learning) et d'apprentissage profond (Deep Learning) pour la prédiction des maladies cardiaques, plus de 55 techniques ont été utilisées dans cette étude, où l'apprentissage automatique était le plus largement utilisé. Jusqu'à présent, la précision de l'apprentissage profond était la plus élevée d'apprentissage automatique surtout dans les articles qui comparent l'utilisation de ces deux sous-domaines de l'intelligence artificielle: d'apprentissage automatique (Machine Learning) et d'apprentissage profond (Deep Learning). Différents modèles et algorithmes ont été utilisés dans ces études et les techniques de Machine Learning ont atteint un taux de précision plus élevé est : Extreme Gradient Boosting (XGBoost), Random Forest, Support Vector Machine (SVM), Logistic Regression, decision tree, Multilayer Perceptron (MLP), Hybrid Random Forest Linear Model (HRFLM), et Les techniques de Deep Learning ont atteint un taux de précision plus élevé est : DNN3 (3-layer network), Convolutional Neural Network (CNN), Batch Normalization (BN), Rectified Linear Unit (ReLU), Adaptive Moment Estimation (ADAM), Receiver Operating Characteristic (ROC), Area Under the Receiver Operating Characteristic Curve (AUROC), Long Short-Term Memory - Deep Belief Network (LSTM-DBN), Visual Geometry Group (VGG). Les performances des modèles varient, mais la plupart des approches ont obtenu des taux de précision élevés, allant de 81,10 % à 100 %. Certains articles ont également utilisé des techniques de sélection de caractéristiques pour améliorer les performances des modèles (LASSO, PCA). L'utilisation de techniques de deep learning, telles que les réseaux de neurones LSTM et les réseaux de croyance profonde, a montré de bons résultats dans la prédiction des maladies cardiaques. Des métriques couramment utilisées pour évaluer les performances des modèles incluent l'exactitude, la précision, le score F1 et la sensibilité. Les ensembles de données utilisés proviennent de diverses sources, notamment la base de données UCI sur les maladies cardiaques et d'autres centres médicaux. Les articles mettent également en évidence l'importance d'une détection précoce des maladies cardiaques et soulignent le potentiel de l'apprentissage automatique pour aider à identifier les facteurs de risque et les signes précurseurs. Certains articles ont également exploré l'utilisation de l'Internet des Objets (IoT) pour collecter des données médicales et améliorer la détection des maladies cardiaques. Des pays ont également été mentionnés tel que L'Inde, la Chine et le Pakistan étaient les Pays avec le plus grand nombre d'études sur l'utilisation du ML (Machine Learning) et du DL (Deep Learning) pour prédire les maladies cardiaques, ils ont également souligné la maladie coronarienne Cette maladie étant la plus mortelle des maladies cardiaques en raison du taux de mortalité très élevé qu'elle provoque tout au long de l'année, Chaque article présente une approche différente pour prédire la maladie et évalue les performances des

modèles proposés tel que [24] utilise des algorithmes d'apprentissage supervisé tels que Random Forest, Decision Trees et K-Nearest Neighbours pour prédire le risque de maladie coronarienne. Les résultats montrent une précision élevée pour ces algorithmes, [25] se concentre sur l'utilisation d'un modèle amélioré LightGBM pour la prédiction de la maladie coronarienne. Les hyperparamètres du modèle sont ajustés à l'aide d'OPTUNA, et une fonction de perte améliorée est utilisée. Le modèle obtient une valeur AUC élevée, dépassant les autres modèles comparatifs, [26] explore l'utilisation de divers classificateurs d'apprentissage automatique pour prédire la présence de la maladie coronarienne. Les arbres boostés par gradient et le perceptron multicouche obtiennent de bons résultats, mais le Random Forest atteint la meilleure performance avec une précision élevée, [27] se concentre sur l'utilisation de l'apprentissage profond pour analyser les signaux sonores du cœur et prédire la maladie coronarienne obstructive. VGG-16 est identifié comme le meilleur modèle avec une performance supérieure aux autres modèles, [28] propose un modèle d'apprentissage profond hybride pour prédire la maladie coronarienne en combinant plusieurs techniques de machine learning et de deep learning. Le modèle atteint une précision impressionnante dans la prédiction des maladies cardiaques coronariennes, et [29] présente une méthode novatrice utilisant l'analyse de l'iris pour le diagnostic non invasif de la maladie coronarienne. Le modèle proposé, utilisant le classificateur SVM, obtient un taux de précision élevé.

Dans l'ensemble, ces articles démontrent l'efficacité des techniques d'apprentissage automatique et d'apprentissage profond pour prédire les maladies cardiaques en particulier la maladie coronarienne. Chaque approche présente ses propres avantages et performances, contribuant ainsi à l'avancement de la prédiction de cette maladie grave, à une meilleure compréhension des facteurs de risque et à l'amélioration des soins cardiovasculaires.

4. Conclusion :

Nous avons discuté dans ce chapitre à tout ce qui concerne L'intelligence artificielle (IA) en abordant les deux des sous-domaines machine learning, deep learning et son influence et sa contribution dans le domaines de prédiction des maladies cardiovasculaires grâce à ses technologies qui permettent aux machines d'imiter l'intelligence humaine, à travers de l'étude des travaux connexes qui a montré une grande précision concernant les technique utilisées et intégrées de ML et DL ,cette étude a également mis en lumière la prédiction de la maladie coronarienne en raison de la gravité de cette maladie cardiovasculaire et les techniques utilisé pour obtenir une grande prédiction.

Dans le chapitre suivant, nous verrons la présentation de l'architecture globale du système, En s'attaquant aux facteurs de risque les plus importants qui contribuent à l'émergence de la maladie coronarienne.

Chapitre 03 :

Contribution

1. Introduction :

Dans le chapitre précédent, nous avons présenté les différentes techniques utilisées pour prédire les maladies cardiaques, notamment la maladie coronarienne, en discutant de certains travaux menés dans ce contexte compte tenu du grand développement observé ces dernières années dans le domaine de l'apprentissage profond. Et sa technologie qui imite l'esprit humain, il est toujours intéressant de découvrir de nouvelles solutions à notre problème qui est de prédire la maladie coronarienne en identifiant les facteurs de risque. Ce chapitre sera consacré à la présentation de la structure de notre système.

2. Problématique :

La mort subite est devenue à l'heure actuelle l'un des cas de décès les plus courants. Sans avertissement préalable, le patient se retrouve parmi les morts sans même savoir à l'avance qu'il est un patient cardiaque. C'est essentiellement ce qui distingue la maladie coronarienne. L'Organisation Mondiale de la Santé a classé le type de cette maladie cardiovasculaire le plus dangereux et le plus mortel, cela est dû à son taux de mortalité très élevé. Aujourd'hui, cette maladie est devenue la cause principale de décès prématuré dans les pays développés et dans la plupart des pays du monde.

Cela est probablement dû au fait que le diagnostic de ce type de maladie par des experts prend beaucoup de temps et coïncide avec le manque d'experts ayant les connaissances nécessaires. En fin de compte, il s'agit d'une maladie aux symptômes complexes, difficiles à expliquer par le patient et à diagnostiquer par le médecin.

On peut dire de cette maladie qu'elle est latente et n'apparaissent pas cliniquement. Il existe une cause principale à cette maladie, qui est l'accumulation de graisses dans le sang (c'est-à-dire le cholestérol LDL). Ce processus, c'est-à-dire l'accumulation de graisse, commence à l'adolescence, ce qui provoque au fil du temps un durcissement des artères, c'est-à-dire des artères coronaires qui nourrissent le cœur, ce qui fournisse en sang. Cela se produit souvent en raison de l'accumulation de plaque dans les parois des artères, ce qui réduit progressivement leur diamètre, ralentit le flux sanguin et peut également conduire à leur blocage complet.

Les méthodes automatiques telles que l'apprentissage profond peuvent résoudre ce problème en entraînant un modèle pour faire des prédictions précoces sur les personnes à risque de développer une maladie coronarienne, évitant ainsi une grande partie des souffrances de la maladie ainsi que la possibilité d'une mort certaine due à cette maladie en identifiant les facteurs de risque tel que : Âge avancé, Sexe masculin, Hérité, Alimentation trop riche en graisse saturée, Hypercholestérolémie

(cholestérol LDL), Tabagisme, Hypertension artérielle, Diabète, Sédentarité, Obésité abdominale (BMI), Troubles psycho-sociaux (dont stress et dépression), Abus d'alcool

L'objectif de cette partie est de développer un modèle de classification capable de prédire s'il existe ou non une possibilité d'exposition à une maladie coronarienne dans le futur, en utilisant une base de données : Ensemble de données de l'étude cardiaque de Framingham (Framingham heart study dataset) et les algorithmes qui seront utilisés à des fins de classification sont CNN.

3. Architecture du système :

Le diagramme ci-dessous (Figure 3.1) représente la structure du modèle de prédiction, à travers lequel nous tenterons de donner un aperçu de notre système qui prédit si un patient présente un risque à 10 ans de développer une maladie coronarienne dans le futur. Cela se fait en soumettant ces données aux étapes nécessaires qui permettent de les traiter, d'extraire les caractéristiques les plus importantes et de les diviser en données d'entraînement qui permettent d'entraîner le modèle et de données de test qui permettent de tester le modèle à l'aide de l'algorithme pour la classification CNN.

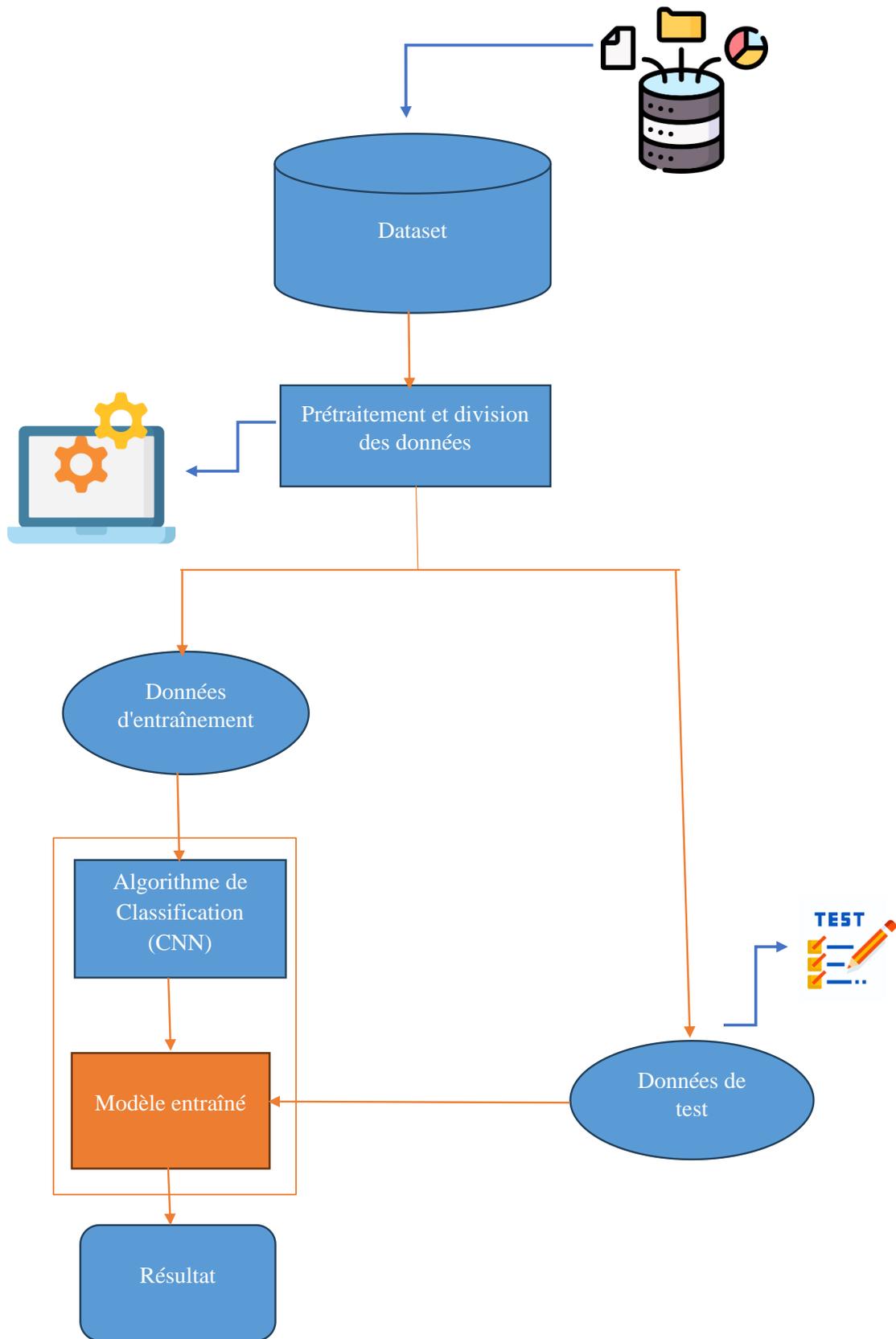


Figure 3.1 : Architecture du système de prédiction de La maladie coronarienne.

4. Dataset :

L'ensemble de données sur les maladies cardiaques « Framingham » comprend plus de 4 240 enregistrements, 16 colonnes et 15 attributs. L'objectif de l'ensemble de données est de prédire si le patient présente un risque de maladie coronarienne future (CHD) sur 10 ans [30].

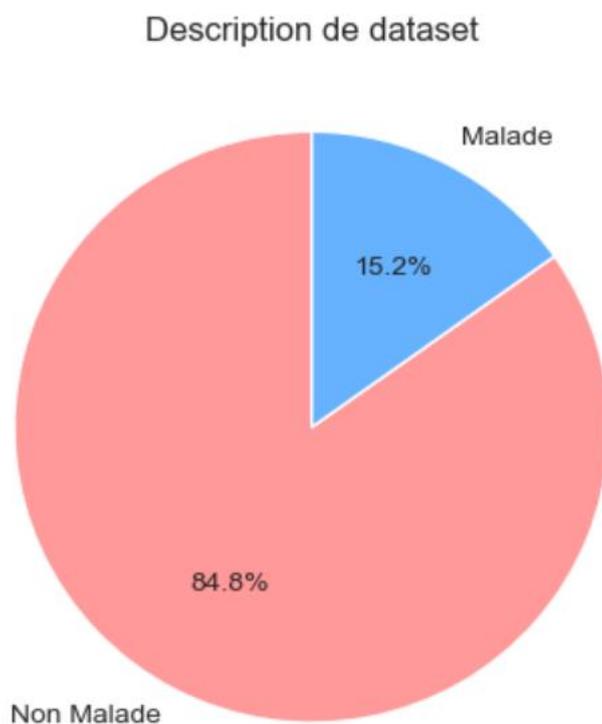


Figure 3.2 : Description de dataset.

	Classe	Instant
1	Non Malade	3596
2	Malade	644
	Totale	4 240

Tableau 5 : Description de dataset.

Dataset contient les attributs mentionnés dans le tableau 6 :

ID	Attribut	
	Abréviations	Détail
01	male	Représentant le genre masculin (1) ou féminin (0)
02	age	Âge en ans (32 ans -70 ans)
03	education	Éducation obtenue (1 : secondaire supérieur,2 : obtention du diplôme ,3 : post-diplôme ,4 : doctorat)
04	currentSmoker	Si le patient est ou non un fumeur actuel (oui ="1" ou non="0")
05	cigsPerDay	Le nombre de cigarettes que la personne a fumées en moyenne au cours d'une journée.
06	BPMeds	Si le patient prenait ou non des médicaments contre l'hypertension (oui ="1" ou non="0")
07	prevalentStroke	Si le patient a déjà eu un accident vasculaire cérébral (oui ="1" ou non="0")
08	prevalentHyp	Si le patient était hypertendu ou non (oui ="1" ou non="0")
09	diabetes	Si le patient était diabétique ou non (oui ="1" ou non="0")
10	totChol	Taux de cholestérol total
11	sysBP	Tension artérielle systolique
12	diaBP	Tension artérielle diastolique
13	BMI	Indice de masse corporelle
14	heartRate	Fréquence cardiaque
15	glucose	Niveau de glucose
16	TenYearCHD	Risque de maladie coronarienne sur 10 ans (binaire : « 1 » signifie « Oui », « 0 » signifie « Non »)

Tableau 6 : Description des attributs.

La Figure 3.3 représente les deux classes contenues dans la base de données, où 0 représente la classe de patients ne souffrant pas de maladie cardiaque, et 1 représente la classe de patients souffrant d'une maladie cardiaque, il y'a Un écart important entre le nombre des patients qui ne souffrent pas de maladie cardiaque et parmi les patients souffrant de maladies cardiaques.

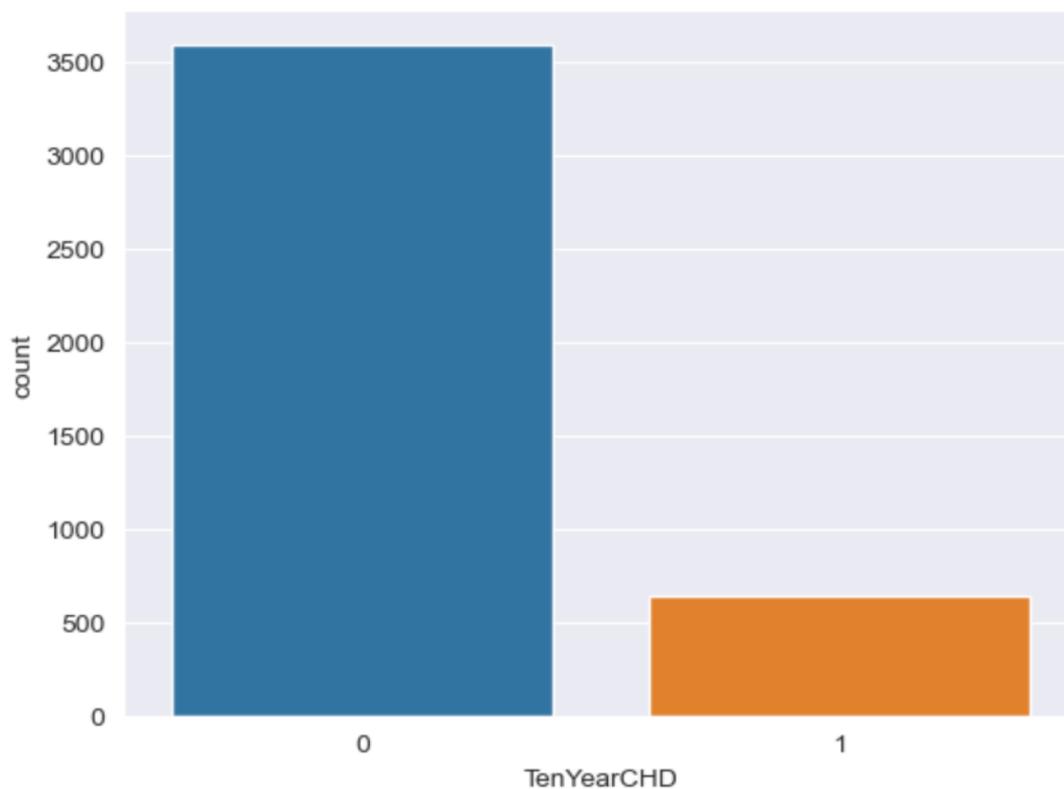


Figure 3.3 : Graphe illustrant les deux classes.

L'âge est l'un des indicateurs qui fournit des informations importantes dans la détermination de l'identification de la maladie. La Figure 3.4 présente la répartition des patients en fonction de leur âge, qui varie entre 32 et 70 ans.

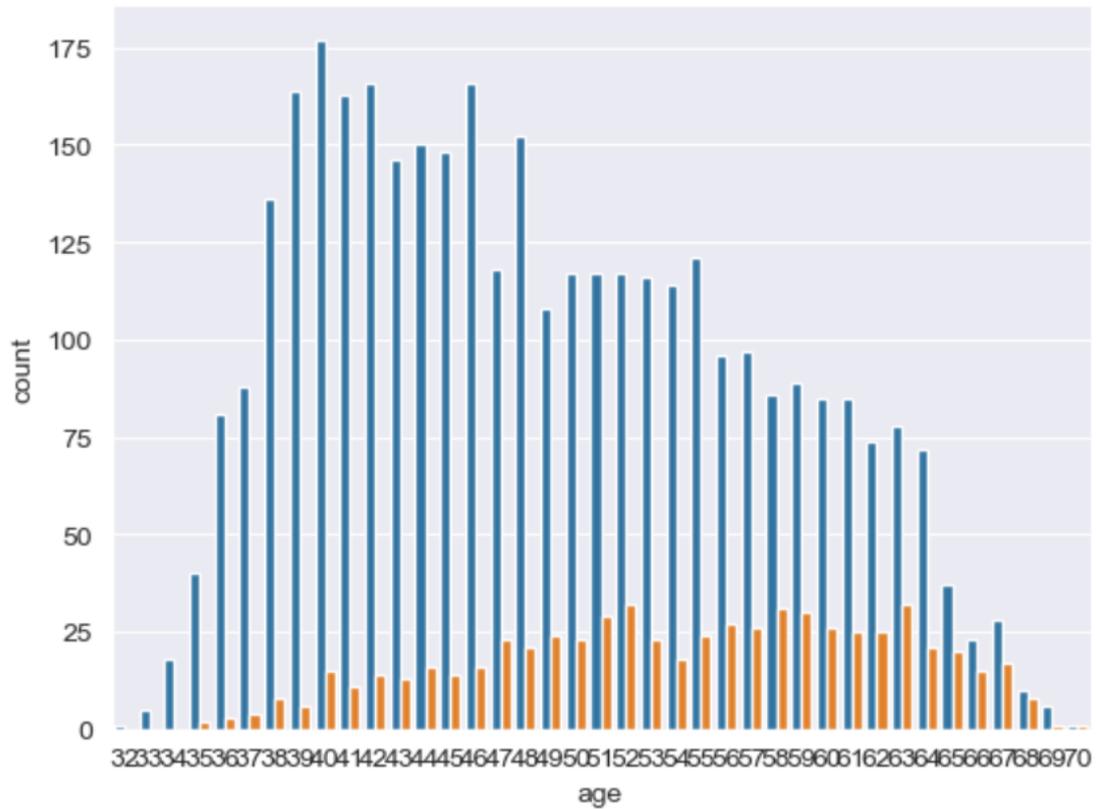


Figure 3.4 : Visualisation relation Age /maladie cardiaque.

Le sexe est également l'un des facteurs de risque directement affectés dans la détermination de la maladie, car le sexe masculin est le plus susceptible à la maladie. La Figure 3.5 représente le nombre de chaque sexe, car le nombre de femmes est supérieur au nombre d'hommes, tandis que La Figure 3.6 représente le nombre de femmes malades et non malades, ainsi que le nombre d'hommes malades et non malades, car il apparaît que le nombre d'hommes malades est supérieur au nombre de femmes malades.

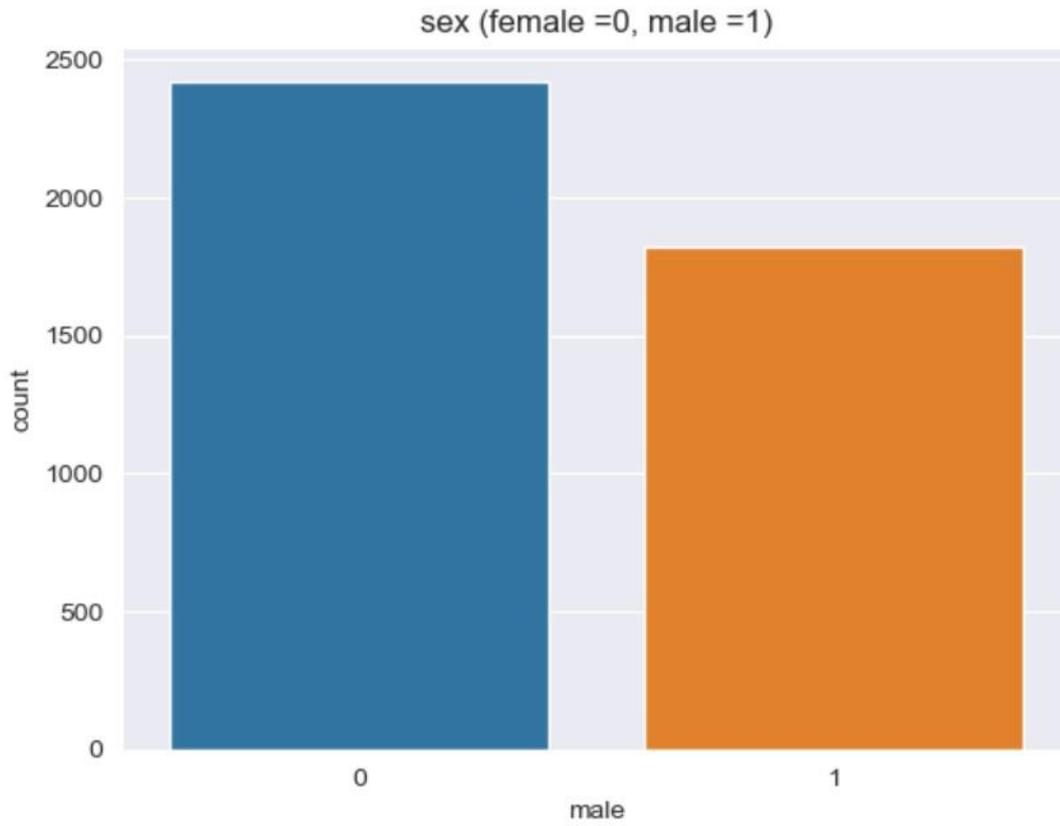


Figure 3.5 : Distribution de Dataset selon Le sexe.

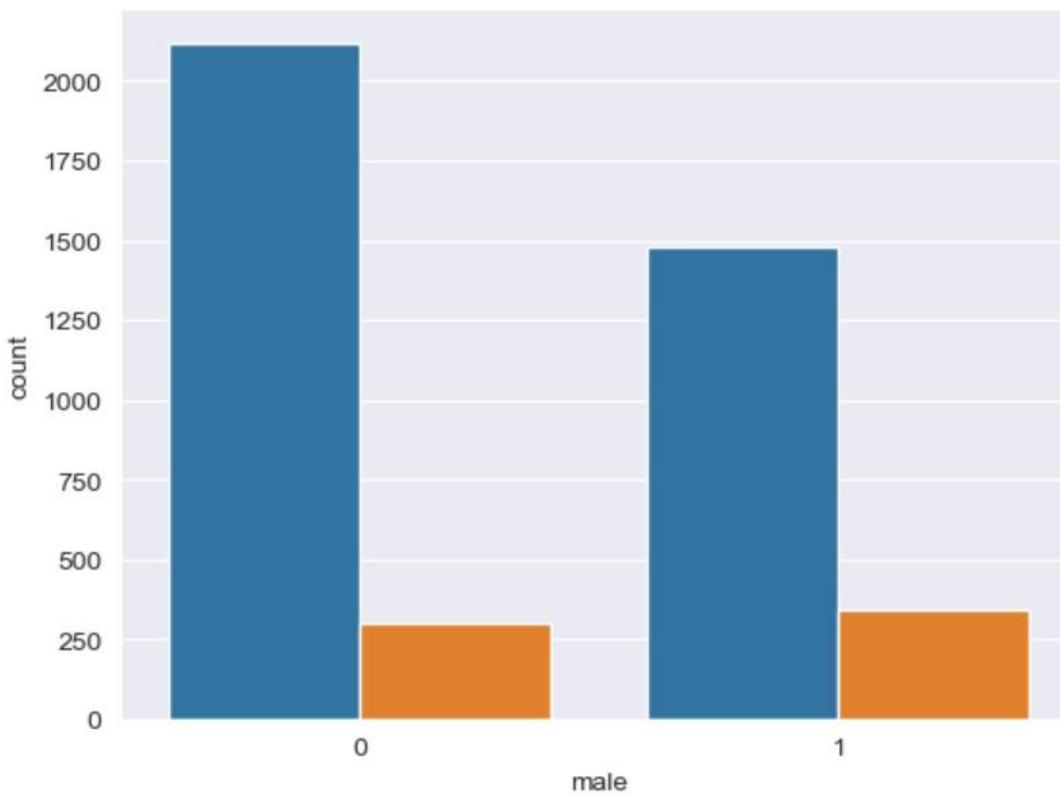


Figure 3.6 : Visualisation relation sexe /maladie cardiaque.

5. Le prétraitement de données :

Le prétraitement des données est une étape fondamentale dans le domaine de l'analyse de données, Son objectif est de transformer les données brutes en un format compréhensible et adapté à l'analyse, Tout d'abord, il est essentiel de traiter les valeurs manquantes dans le jeu de données (voir la figure 3.7).

```
df.isnull().sum()
male          0
age           0
education     105
currentSmoker 0
cigsPerDay    29
BPMeds        53
prevalentStroke 0
prevalentHyp  0
diabetes      0
totChol       50
sysBP         0
diaBP         0
BMI           19
heartRate     1
glucose       388
TenYearCHD   0
dtype: int64
```

Figure 3.7 : Les valeurs manquantes.

Dans le cadre du prétraitement de données, il est souvent nécessaire de convertir les variables catégorielles en variables numériques en utilisant des techniques telles que le codage one-hot. De plus, il est recommandé de normaliser ou de standardiser les caractéristiques numériques afin de les mettre à la même échelle. Enfin, il est essentiel de diviser l'ensemble de données en ensembles d'entraînement et de test afin de pouvoir évaluer les performances du modèle. La proportion de division peut varier en fonction de la taille du jeu de données et des exigences spécifiques du problème. Cette étape de prétraitement des données implique plusieurs étapes, dont certaines sont illustrées dans la figure suivante.

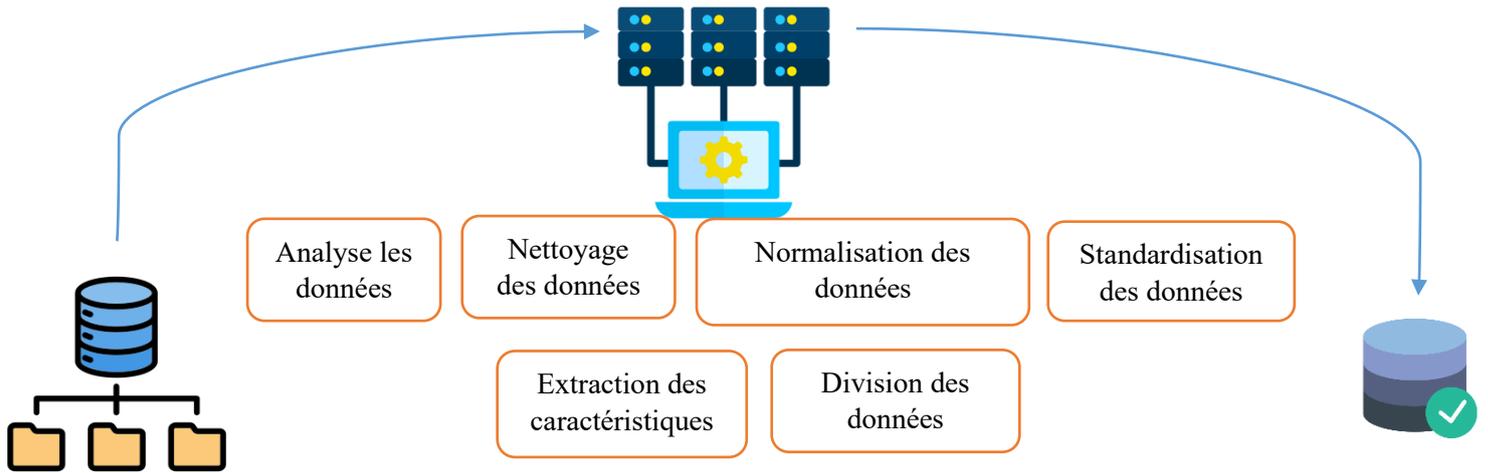


Figure 3.8 : Phase de prétraitement.

5.1 Nettoyage des données :

Au cours de cette étape, les données incomplètes, altérées ou inexactes sont identifiées et corrigées ou supprimées afin d'améliorer les critères de qualité. Le nettoyage des données est une étape cruciale dans l'analyse du jeu de données de l'étude Framingham Heart Study, car les données réelles sont rarement parfaites. Elles peuvent comporter des valeurs manquantes, des erreurs, des doublons, des incohérences et d'autres problèmes.

5.2 Normalisation des données :

La normalisation, également connue sous le nom d'échelle Min-Max, est une technique de mise à l'échelle des valeurs où celles-ci sont décalées et réajustées pour se situer dans une plage de [0, 1].

Et sa formule s'écrit comme ci-dessous : $X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$

Si on l'applique, par exemple, à l'indicateur d'âge, on trouve :

Exemple : $X=39$

$$X' = \frac{39-32}{70-32} \quad \text{On a } X' = 0.18 \in [0, 1]$$

5.3 Standardisation des données :

La standardisation est une technique de mise à l'échelle des valeurs centrées autour de la moyenne avec un écart type unitaire. Elle permet de rendre les données comparables et cohérentes, facilitant ainsi leur analyse. Cependant, avant de standardiser les données, il est important de prendre en compte certaines variables catégorielles qui nécessitent une conversion spécifique. Ces variables, telles que : diabetes, male, BPMeds..., diffèrent des autres variables indicatives comme : totChol, BMI.... Il faut donc convertir ces variables, et puis nous continuons le processus de standardisation, selon la formule suivante :

$$X' = \frac{X - \mu}{\sigma}$$

Où

X : la valeur d'origine

μ : la moyenne des valeurs des caractéristiques.

σ : l'écart type des valeurs des caractéristiques.

X' : la valeur standardisée

Une fois que la formule a été appliquée à l'ensemble des données, nous obtenons des données standardisées prêtes pour la prochaine étape.

5.4 Extraction des caractéristiques :

L'extraction des caractéristiques est une étape cruciale du prétraitement des données. Elle implique la collecte, la préparation et la sélection des caractéristiques pertinentes, ainsi que l'extraction et l'évaluation de ces caractéristiques pour améliorer les performances de prédiction. Cette étape permet de simplifier les ressources nécessaires et de déterminer un sous-ensemble de caractéristiques extraites essentielles à l'analyse des données.

Dans cette étude, nous concentrerons sur l'identification des facteurs de risque de maladie coronarienne, que nous avons mentionnés précédemment. Le tableau 7 représente les facteurs de risque de maladie coronarienne, qui ont été soigneusement sélectionnés en fonction de leur impact direct sur l'émergence et l'exacerbation de cette maladie, avec la définition et le détail de chaque facteur séparément, ainsi que l'attribut qui lui correspond À partir de dataset utilisée.

Les facteurs de risque	Les attributs
Âge avancé : L'âge est un facteur de risque continu qui augmente progressivement l'incidence de maladie coronarienne. Ce risque devient significatif à partir de 50 ans chez les hommes et à partir de 60 ans chez les femmes.	age
Sexe masculin : Avant l'âge de 70 ans, les hommes représentent les deux tiers des cas d'infarctus, Cependant, après la ménopause, cette disparité diminue chez les femmes et disparaît complètement après l'âge de 75 ans.	male
Hypercholestérolémie : Caractérisées par des anomalies des lipides circulants, présentent un facteur de risque majeur pour les maladies cardiovasculaires. Est l'élévation du Low Density Lipoprotein-cholestérol ou cholestérol lié aux lipoprotéines de faible densité > 1,60 g/L (4,1 mmol/L). Le LDL-cholestérol a un rôle direct sur l'accroissement des plaques d'athérome et sur leur rupture par instabilité.	totChol
Obésité abdominale : L'obésité est un facteur de risque cardiovasculaire, et ce risque est corrélé à l'indice de masse corporelle (IMC), en particulier lorsque l'obésité est de type androïde, par prépondérance de graisses intraabdominales ; très souvent associée à d'autres facteurs de risque (HTA, diabète)	BMI
Hypertension artérielle : L'hypertension artérielle, qu'elle soit permanente, paroxystique, traitée ou non, constitue un facteur de risque pour la maladie coronarienne. En effet, l'hypertension artérielle multiplie par trois le risque de développer cette maladie.	BPMeds, heartRate, sysBP, diaBP, prevalentHyp, prevalentStroke
Tabagisme : Le tabagisme aggrave les lésions athéromateuses en altérant la fonction endothéliale, perturbant ainsi la vasomotricité et activant l'agrégation plaquettaire. Il est athérogène et prothrombotique. Il multiplie par 5 le risque d'infarctus du myocarde.	currentSmoker, cigsPerDay
Diabète : Le diabète de type 1 ou de type 2 est associé à une augmentation du risque de maladies cardiovasculaires. Les complications cardiovasculaires peuvent survenir dès l'âge de 30 ans dans le cas du diabète de type 1, mais l'incidence croissante du diabète de type 2 en fait un facteur de risque très préoccupant	diabetes, glucose

Tableau 7 : Les facteurs de risque /Les attributs de dataset.

Maintenant, nous supprimons tous les attributs similaires qui jouent le même rôle, de sorte que l'ensemble d'attributs sur lequel nous allons travailler soit :

```
feature_cols = ['age', 'male', 'totChol', 'BMI', 'prevalentHyp', 'heartRate', 'prevalentStroke', 'BPMeds', 'currentSmoker', 'diabetes']
```

Figure 3.9 : les attributs sélectionnés.

5.5 Division des données :

Au cours de cette étape, l'ensemble de données est divisé en deux ensembles distincts : un ensemble d'apprentissage représentant 80 % des données et un ensemble de test représentant 20 % des données (Figure 3.10). L'ensemble d'apprentissage est utilisé pour entraîner le modèle, tandis que l'ensemble de test est utilisé pour évaluer les performances du modèle.

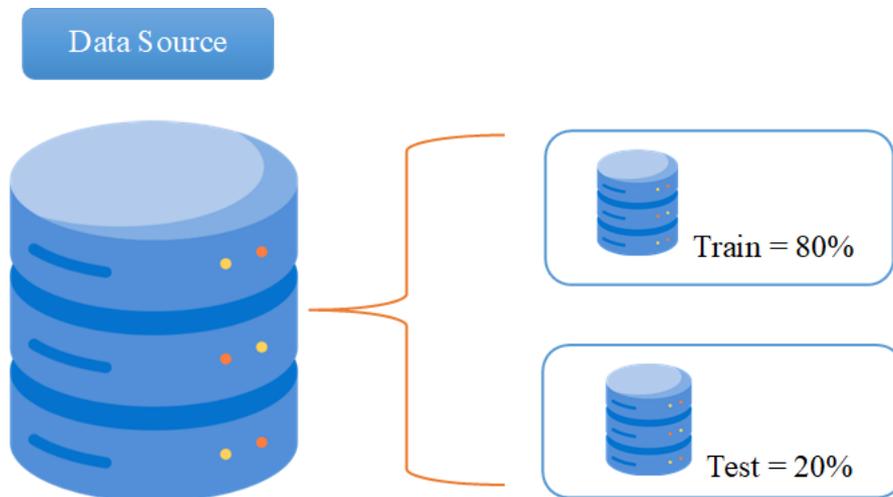


Figure 3.10 : Division des données.

6. Classification avec Le Réseau Neuronal Convolutif (CNN) :

Dans le cadre de la détection de la maladie coronarienne, l'architecture CNN (Convolutional Neural Network) est utilisée pour apprendre les schémas et les relations entre les attributs du cœur et la maladie coronarienne. Cette architecture, illustrée dans la Figure 3.11, est composée de couches d'entrée, de couches cachées et de couches de sortie. Chaque couche est constituée de neurones qui effectuent des calculs et des transformations sur les données d'entrée pour produire une sortie. Étant couramment utilisée pour des tâches de classification, l'architecture CNN est un choix approprié pour la détection de la maladie coronarienne.

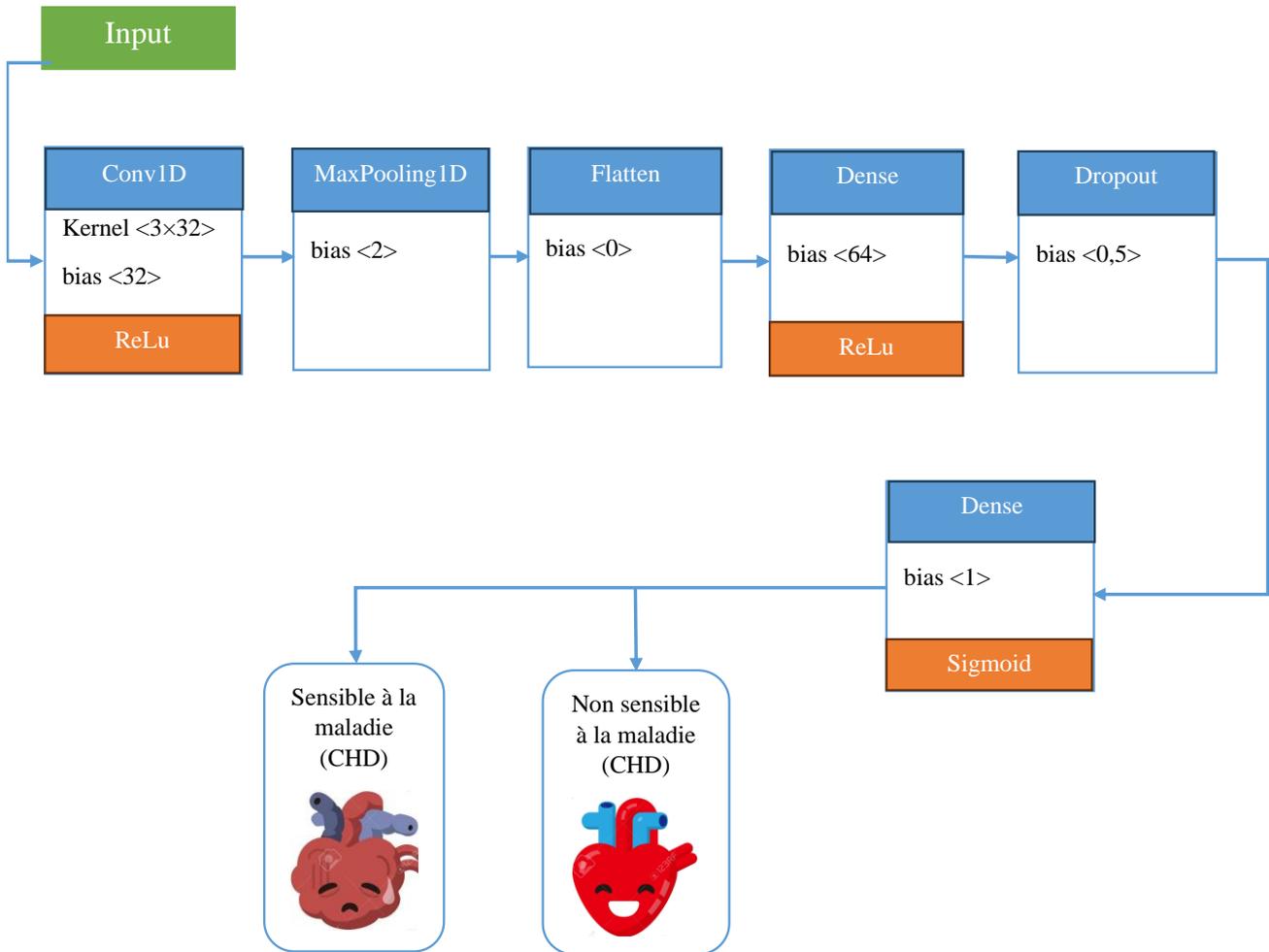


Figure 3.11 : Architecture CNN Proposé.

7. Entraînement le modèle :

Après avoir effectué toutes les opérations de traitement des données et choisi l'algorithme, nous entraînerons ensuite le modèle et verrons les résultats de cette formation, qui nous montre les résultats de toutes les étapes préalables que nous avons suivies depuis les étapes de prétraitement jusqu'au choix de l'algorithme et Couches CNN. Les résultats sont présentés dans les Figures ci-dessous.

Train Accuracy: 0.8573113207547169
 Train Precision: 0.8580757252215575
 Train Recall: 0.8573113207547169
 Train F1 Score: 0.8040871371669795

Figure 3.12 : Les résultats de L'entraînement.

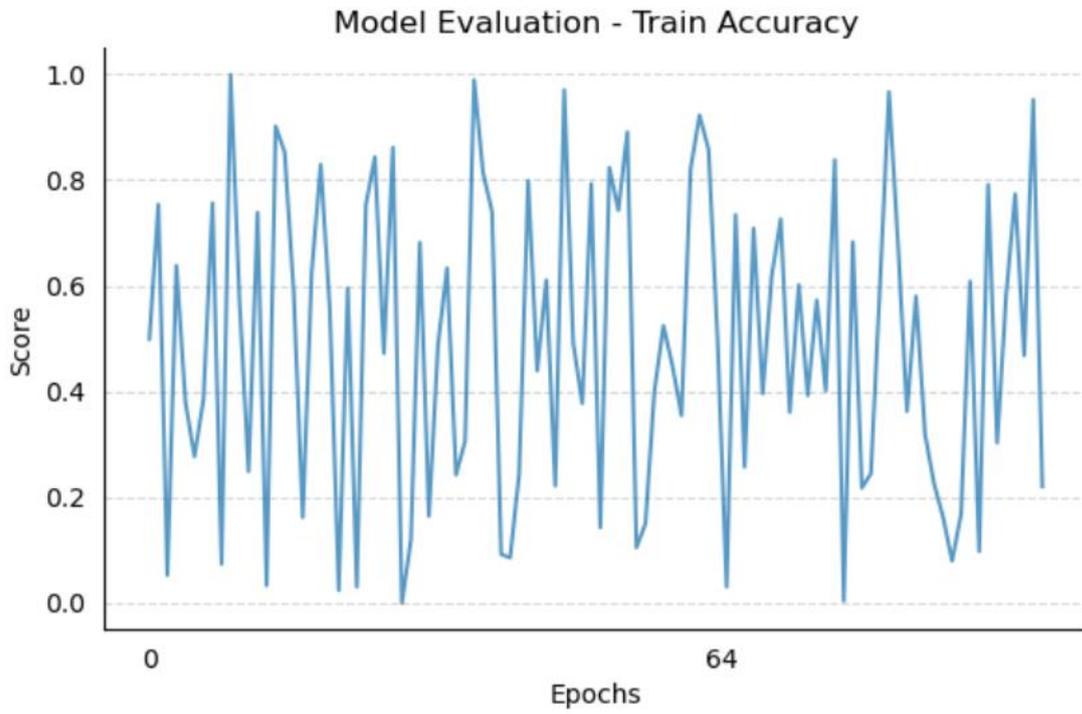


Figure 3.13 : Train Accuracy du CNN

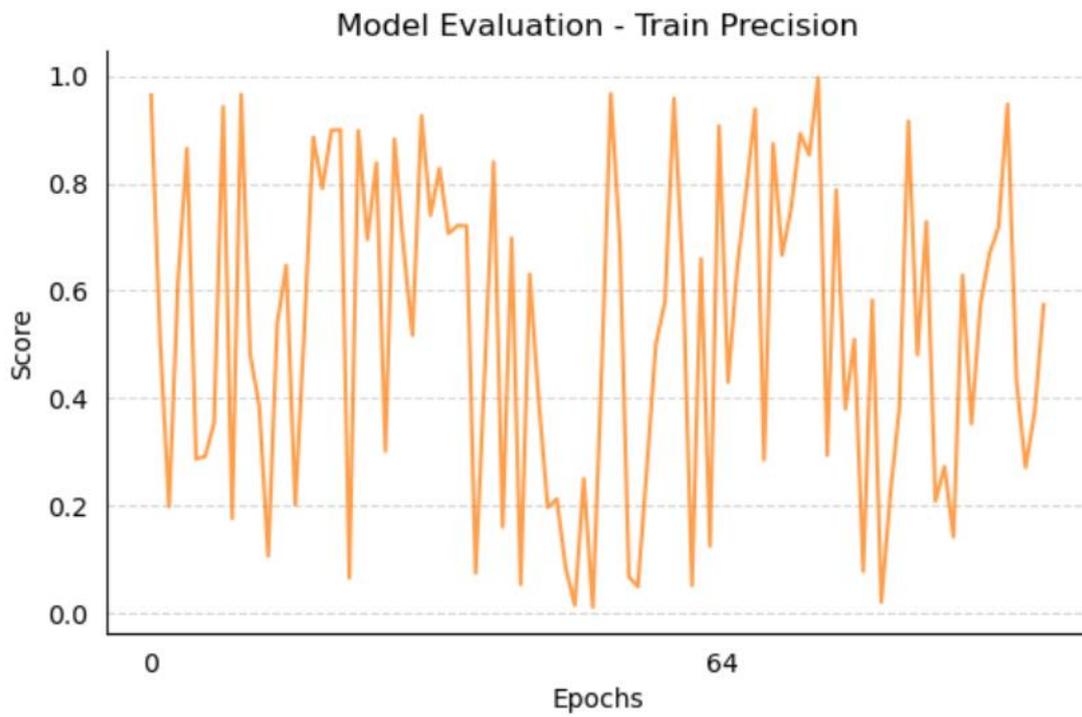


Figure 3.14 : Train Precision du CNN

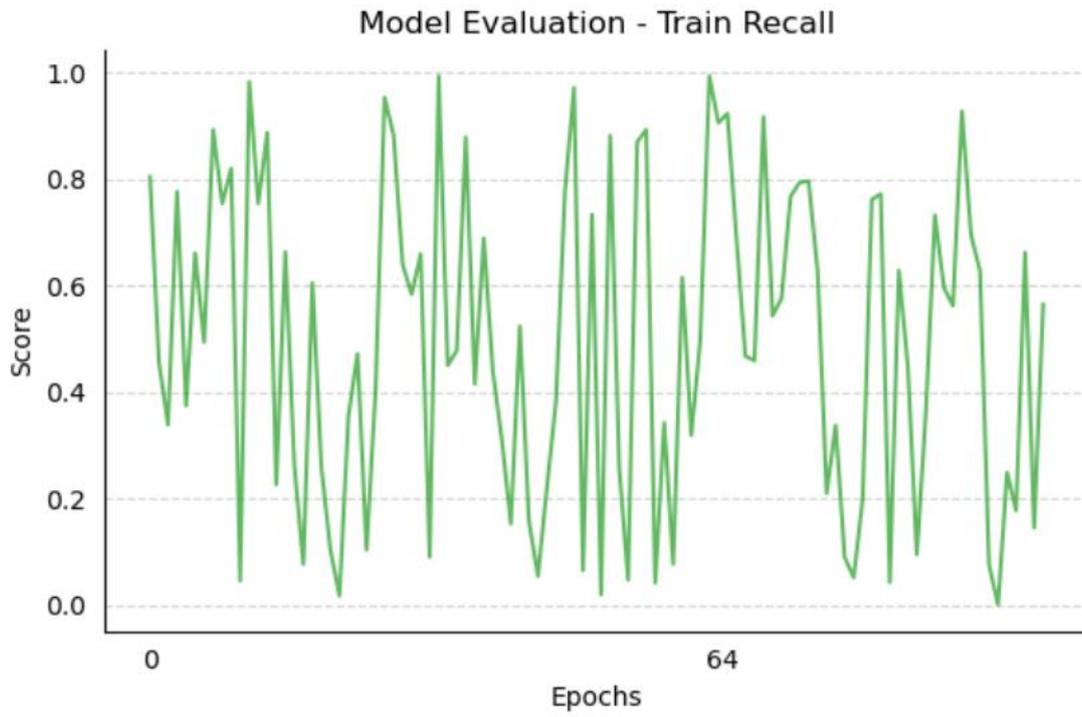


Figure 3.15 : Train Recall du CNN

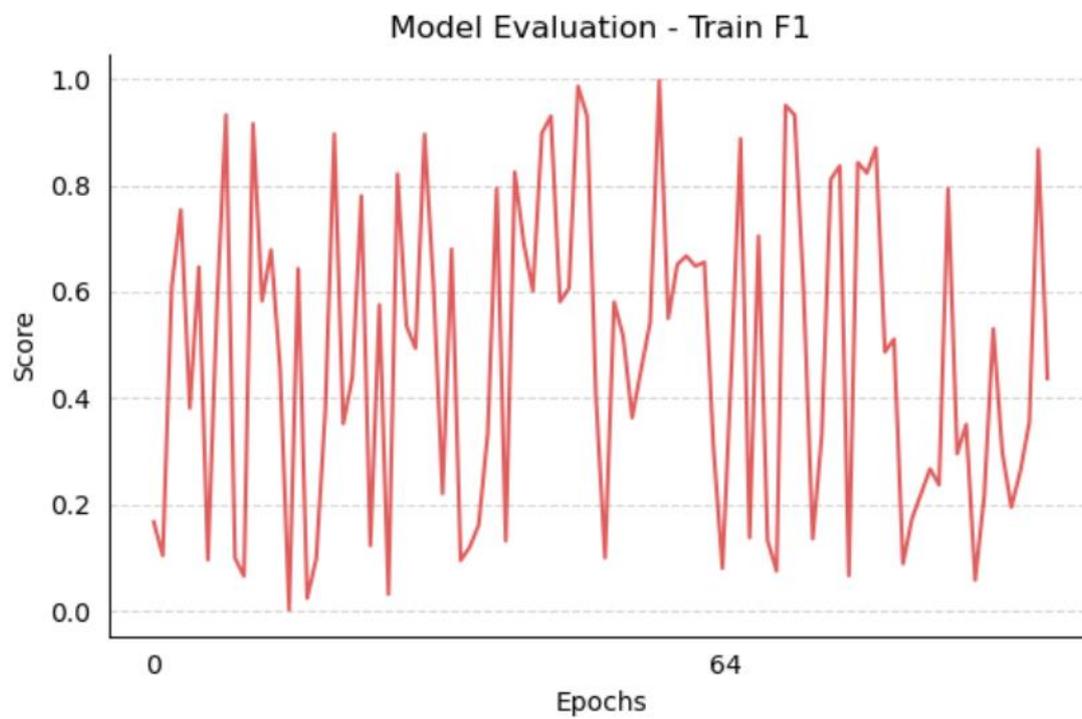


Figure 3.16 : Train F1 Score du CNN

8. Conclusion :

Prédire l'état du patient est un processus fondamental et très important, car il permet de sauver des vies. Ce chapitre est donc consacré à la présentation du système que nous proposons pour prédire la maladie cardiovasculaire la plus grave, la maladie coronarienne. Le système proposé utilise l'algorithme CNN, Ceci après avoir découvert la base de données, présenté les étapes pour la traiter, extrait les caractéristiques les plus importantes et les diviser, ainsi que pris connaissance du schéma de couches CNN proposé pour ce modèle.

Dans le prochain chapitre nous devons réaliser des tests et évaluer les performances ainsi que lister les différents outils qui nous ont permis de réaliser nos expérimentations et tests.

Chapitre 04 :

Implémentation

1. Introduction :

Dans ce chapitre, nous présenterons les outils utilisés dans cette étude, et nous testerons également le modèle que nous avons proposé dans le chapitre précédent, en plus des méthodes appliquées pour atteindre notre objectif. Après cela, nous présenterons les résultats obtenus.

2. Les outils de développement utilisés :

Dans cette partie, nous présenterons dans un premier temps le langage de programmation Python utilisé pour implémenter le modèle, Lors de la préparation de l'environnement d'implémentation, nous avons installé la distribution Anaconda pour Windows 10, en plus d'outils logiciels tels que des bibliothèques importées : Keras, Tensorflow, Matplotlib, Numpy, Sklearn, Pandas, Seaborn

2.1 Langage de programmation :

Python 3.11.7



Python est un langage de programmation interprété, multiparadigme et multiplateformes. Il favorise la programmation impérative structurée, fonctionnelle et orientée objet. Il est doté d'un typage dynamique fort, d'une gestion automatique de la mémoire par ramasse-miettes et d'un système de gestion d'exceptions [31].

2.2 Les Logiciels :

Keras 3.3.1



Keras est une bibliothèque qui permet d'interagir avec les algorithmes de réseaux de neurones profonds et d'apprentissage automatique [32].



Tensorflow 2.16

TensorFlow est un outil open source d'apprentissage automatique développé par Google. Le code source a été ouvert le 9 novembre 2015 par Google et publié sous licence Apache [33].



NumPy 1.26.4

NumPy est une bibliothèque pour langage de programmation Python, destinée à manipuler des matrices ou tableaux multidimensionnels ainsi que des fonctions mathématiques opérant sur ces tableaux [34].



Matplotlib 3.8.3

Matplotlib est une bibliothèque du langage de programmation Python destinée à tracer et visualiser des données sous forme de graphiques. Elle peut être combinée avec les bibliothèques python [35].



Pandas 2.2.2

Pandas est une bibliothèque écrite pour le langage de programmation Python permettant la manipulation et l'analyse de données. Elle propose en particulier des structures de données et des opérations de manipulation de tableaux numériques et de séries temporelles [36].



Seaborn 0.13.2

Seaborn est une bibliothèque de visualisation de données Python basée sur matplotlib. Il fournit une interface de haut niveau pour dessiner des graphiques statistiques attrayants et informatifs [37].



Scikit-learn 1.4.2

Scikit-learn est une bibliothèque libre Python destinée à l'apprentissage automatique. Elle est développée par de nombreux contributeurs notamment dans le monde académique par des instituts français d'enseignement supérieur et de recherche comme Inria [38].

2.3 Le matériel :



Jupyter Notebook

Jupyter Notebook est une application de création de blocs-notes, sous l'égide du projet Jupyter. Construit sur la puissance du format notebook informatique, Jupyter Notebook offre de nouvelles façons rapides et interactives de prototyper et d'expliquer votre code, d'explorer et de visualiser vos données et de partager vos idées avec d'autres [39].

3. Mesures de performance :

La matrice de confusion est utilisée comme mesure de performance pour évaluer la qualité de correspondance entre les valeurs prédites et les valeurs réelles obtenues par un modèle. Elle est particulièrement utile dans le contexte de la maladie par exemple, où elle peut être représentée de la manière suivante :

		Si l'état de santé du patient est réellement malade ou non	
		Est malade	N'est pas malade
La prédiction fournie par le modèle	Est malade	Vrai positif (VP) Se produit lorsque les valeurs réelles et prédites sont positives et identiques.	Faux positif (FP) Se produit lorsque les valeurs réelles et prédites sont différentes. Cela signifie que le modèle prédit que le patient est malade alors qu'en réalité, il ne l'est pas.
	N'est pas malade	Faux négatif (FN) Se produit lorsque le modèle prédit à tort que le patient n'est pas malade, alors qu'en réalité, il l'est.	Vrai négatif (VN) Se produit lorsque les valeurs réelles et prédites sont négatives et identiques.

Tableau 8 : Matrice de confusion.

Ces valeurs permettent de comprendre les mesures utilisées pour comparer ou valider les modèles, et voici quelques-unes de ces mesures :

Les mesures de performance	L'équation
Accuracy : La précision est l'une des principales mesures de performance pour la classification. Elle est définie comme la proportion des classifications correctes par rapport à l'ensemble total des échantillons.	$Accuracy = \frac{VP + VN}{VP + VN + FP + FN}$
Rappel : C'est la petite proportion des individus par rapport à la quantité globale des individus applicables.	$Rappel = \frac{VP}{VP + FN}$
Précision : C'est la proportion des individus qui sont correctement identifiées par le modèle.	$Précision = \frac{VP}{VP + FP}$
F1-score : C'est la moyenne entre la précision et le rappel	$F1 - score = 2 \times \frac{Précision \times Rappel}{Précision + Rappel}$

Tableau 9 : Les mesures de performance/L'équation.

4. Chargement des données :

Une fois les bibliothèques nécessaires appelées, la première étape consiste à charger la base de données dans l'environnement Jupyter Notebook. La Figure 4.1 ci-dessous présente le résultat du chargement des données, illustrant la structure de l'ensemble global de données avec 16 colonnes et 4240 instances.

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	diabetes	totChol	sysBP	diaBP	BMI	heartRate	glucose	TenYearC
0	1	39	4.0	0	0.0	0.0	0	0	0	195.0	106.0	70.0	26.97	80.0	77.0	
1	0	46	2.0	0	0.0	0.0	0	0	0	250.0	121.0	81.0	28.73	95.0	76.0	
2	1	48	1.0	1	20.0	0.0	0	0	0	245.0	127.5	80.0	25.34	75.0	70.0	
3	0	61	3.0	1	30.0	0.0	0	1	0	225.0	150.0	95.0	28.58	65.0	103.0	
4	0	46	3.0	1	23.0	0.0	0	0	0	285.0	130.0	84.0	23.10	85.0	85.0	
...
4235	0	48	2.0	1	20.0	NaN	0	0	0	248.0	131.0	72.0	22.00	84.0	86.0	
4236	0	44	1.0	1	15.0	0.0	0	0	0	210.0	126.5	87.0	19.16	86.0	NaN	
4237	0	52	2.0	0	0.0	0.0	0	0	0	269.0	133.5	83.0	21.47	80.0	107.0	
4238	1	40	3.0	0	0.0	0.0	0	1	0	185.0	141.0	98.0	25.60	67.0	72.0	
4239	0	39	3.0	1	30.0	0.0	0	0	0	196.0	133.0	86.0	20.91	85.0	80.0	

4240 rows x 16 columns

Figure 4.1 : La structure de l'ensemble de données.

5. Prétraitement des données :

Cette base de données ou dataset contient des valeurs manquantes, comme nous l'avons mentionné dans le chapitre précédent. Pour résoudre ce problème, nous devons remplir les valeurs manquantes dans les colonnes désignées, La figure 4.2 montre que le nombre de valeurs manquantes dans chaque colonne est de 0, ce qui signifie qu'après le traitement que nous avons effectué, nous n'avons plus de valeurs manquantes.

```
[17]: df.isnull().sum()

[17]: male          0
      age          0
      education    0
      currentSmoker 0
      cigsPerDay    0
      BPMeds        0
      prevalentStroke 0
      prevalentHyp  0
      diabetes      0
      totChol       0
      sysBP         0
      diaBP         0
      BMI           0
      heartRate     0
      glucose       0
      TenYearCHD    0
      dtype: int64
```

Figure 4.2 : l'état des valeurs après le Prétraitement.

Une fois que nous avons rempli les valeurs manquantes dans les colonnes spécifiées, nous procédons à l'étape d'extraction des caractéristiques, telle que présentée dans le chapitre précédent. Cette étape est importante car elle inclut les caractéristiques sélectionnées ainsi que la variable cible sur laquelle nous allons travailler.

```
feature_cols = ['age', 'male', 'totChol', 'BMI', 'prevalentHyp', 'heartRate', 'prevalentStroke', 'BPMeds', 'currentSmoker', 'diabetes']
X = df[feature_cols]
Y = df['TenYearCHD']
```

Figure 4.3 : l'étape d'extraction des caractéristiques.

Après cela on applique la formule de Normalisation que nous avons présenté dans le chapitre précédent à l'ensemble globale de données, nous obtenons des données normalisées prêtes pour la prochaine étape, comme le montre la figure 4.4

```
[21]:
```

	age	male	totChol	BMI	prevalentHyp	heartRate	prevalentStroke	BPMeds	currentSmoker	diabetes
0	0.184211	1.0	0.149406	0.277024	0.0	0.363636	0.0	0.0	0.0	0.0
1	0.368421	0.0	0.242784	0.319680	0.0	0.515152	0.0	0.0	0.0	0.0
2	0.421053	1.0	0.234295	0.237518	0.0	0.313131	0.0	0.0	1.0	0.0
3	0.763158	0.0	0.200340	0.316045	1.0	0.212121	0.0	0.0	1.0	0.0
4	0.368421	0.0	0.302207	0.183228	0.0	0.414141	0.0	0.0	1.0	0.0

Figure 4.4 : Normalisation des caractéristiques.

La division des données est une étape essentielle en science des données, Elle consiste à séparer l'ensemble de données en deux parties distinctes ou plus, Dans notre cas, nous avons divisé notre ensemble de données en deux parties : une première partie qui représente 80 % des données et est utilisée pour entraîner le modèle, et une deuxième partie qui représente 20 % des données et est utilisée pour évaluer ou tester le modèle, Le code ci-dessous fournit une description de la solution, et la Figure 4.5 illustre cette étape.

```
from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, Y_test = train_test_split(X_norm, Y, test_size=0.2, random_state=42)

print("X_test :" ,X_test.shape[0])
print("X_train :" ,X_train.shape[0])

X_test : 848
X_train : 3392
```

Figure 4.5 : division des données.

Le codage "One-Hot Encoding" est une technique de conversion des variables catégorielles en une représentation numérique. De nouvelles variables binaires sont créées pour chaque catégorie de la variable d'origine. La valeur de la variable est 1 (ou False) si le patient est malade et 0 (ou True) si le patient n'est pas malade. Le codage "One-Hot Encoding" est important dans l'analyse de données et l'apprentissage profond, car il permet d'appliquer des modèles d'apprentissage profond aux variables catégorielles, d'éviter les biais et d'uniformiser le traitement des différentes catégories.

```
Y_test
array([[False,  True],
       [ True, False],
       [ True, False],
       ...,
       [False,  True],
       [ True, False],
       [ True, False]])
```

Figure 4.6 : Convertissez les données en une matrice à deux colonnes (One-Hot Encoding).

La standardisation des variables nous aide à éliminer les différences de niveaux, à uniformiser les effets et à améliorer les performances des modèles. C'est une étape fondamentale dans l'analyse de données et l'apprentissage profond pour garantir des résultats précis et fiables, ce code dans la Figure 4.7 applique la transformation de standardisation aux variables indépendantes des données d'entraînement et de test. Cela permet de mettre les variables à la même échelle et de faciliter l'entraînement et la comparaison des modèles.

```
scaler = StandardScaler()  
X_train_scaled = scaler.fit_transform(X_train)  
X_test_scaled = scaler.transform(X_test)
```

```
X_train_scaled:  
[[ 0.39974402 -0.86045318 -0.37209143 ... -0.1802175  0.98769364  
 -0.16789985]  
 [ 1.68620673 -0.86045318  0.04148891 ... -0.1802175  0.98769364  
 -0.16789985]  
 [-1.35452332 -0.86045318 -1.19925209 ... -0.1802175 -1.01245969  
 -0.16789985]  
 ...  
 [ 1.21840211 -0.86045318  1.39711333 ... -0.1802175 -1.01245969  
 -0.16789985]  
 [-1.23757217  1.16217828 -0.50995154 ... -0.1802175  0.98769364  
 -0.16789985]  
 [-1.70537679 -0.86045318  0.24827907 ... -0.1802175 -1.01245969  
 -0.16789985]]  
X_test_scaled:  
[[-0.06806061 -0.86045318  0.5239993 ... -0.1802175  0.98769364  
 -0.16789985]  
 [-0.76976754  1.16217828 -0.83162513 ... -0.1802175  0.98769364  
 -0.16789985]  
 [-0.53586523  1.16217828  0.91460294 ... -0.1802175  0.98769364  
 -0.16789985]  
 ...  
 [ 1.56925558 -0.86045318  1.58092681 ... -0.1802175 -1.01245969  
 -0.16789985]  
 [-0.53586523 -0.86045318  1.21329985 ... -0.1802175 -1.01245969  
 -0.16789985]  
 [-1.23757217  1.16217828 -0.30316137 ... -0.1802175 -1.01245969  
 -0.16789985]]
```

Figure 4.7 : Standardisation des caractéristiques.

6. Entraîner le modèle :

Le modèle de réseau de neurones convolutionnel (Convolutional Neural Network, CNN) est utilisé pour la classification binaire en utilisant des couches convolutives pour extraire les caractéristiques des données d'entrée, suivies de couches denses pour apprendre des relations complexes et d'une couche de sortie pour la classification finale. Le modèle est entraîné en utilisant l'optimiseur "adam" et la fonction de perte "binary_crossentropy".

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv1D, MaxPooling1D, Flatten, Dense, Dropout

model = Sequential()

model.add(Conv1D(32, kernel_size=3, activation="relu", input_shape=(X_train_resaped.shape[1], 1)))
model.add(MaxPooling1D(pool_size=2))
model.add(Flatten())

model.add(Dense(64, activation="relu"))
model.add(Dropout(0.5))

model.add(Dense(Y_train_binary.shape[1], activation="sigmoid"))

model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
model.summary()
```

Figure 4.8 : L'entrainement du modèle CNN.

7. Test du modèle :

Les résultats ont été obtenus en appliquant l'algorithme de classification (CNN) à l'ensemble de données utilisé, Les résultats de classification de ce modèle sont présentés par les Figures et le Tableau 10 :

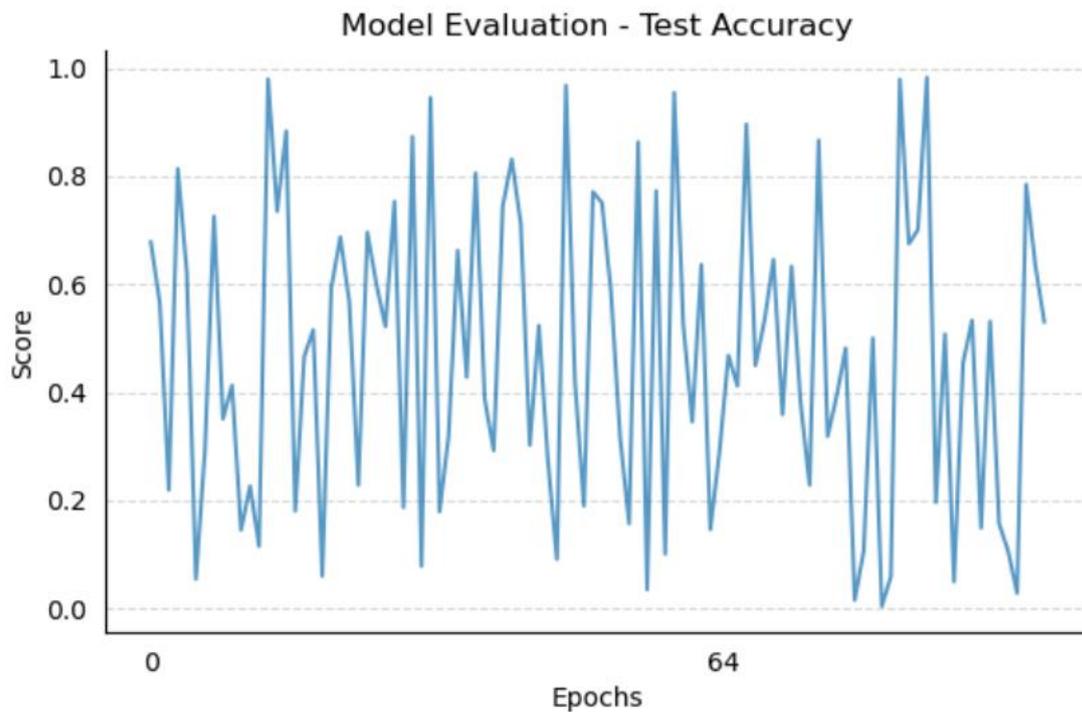


Figure 4.9 : Test Accuracy du CNN

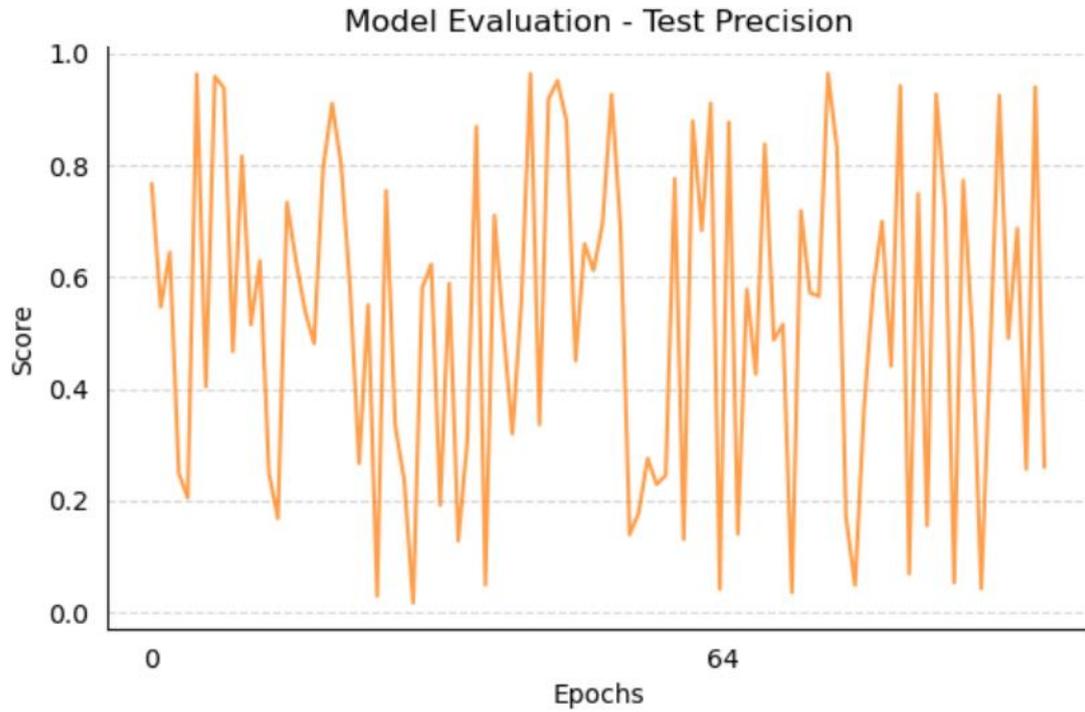


Figure 4.10 : Test Precision du CNN

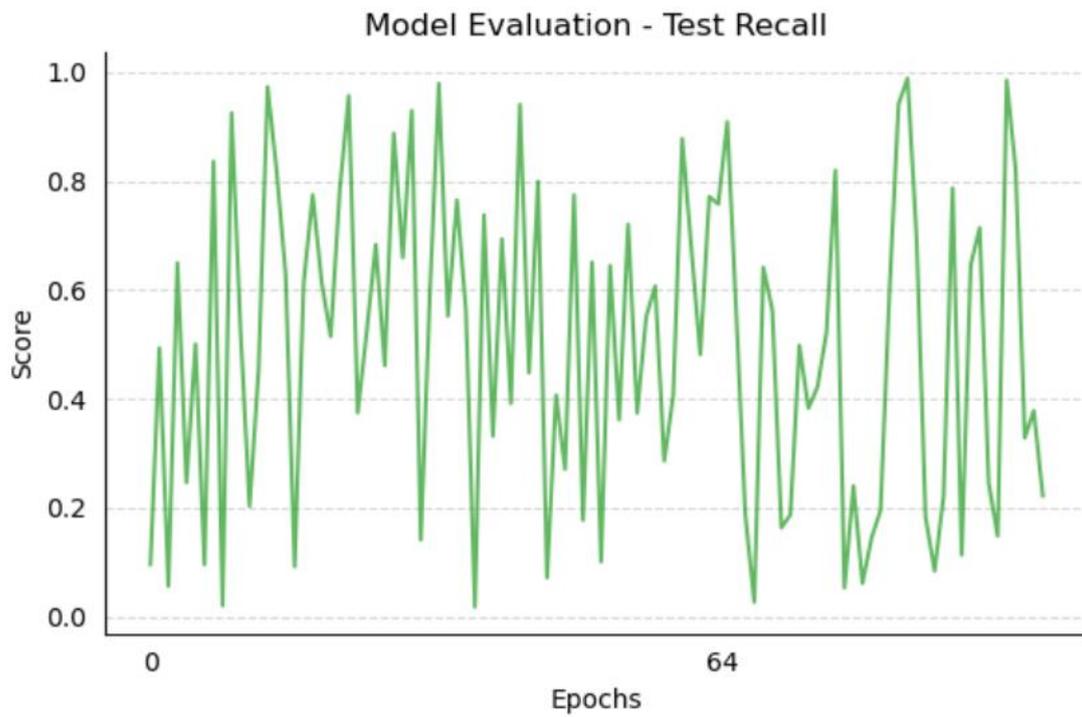


Figure 4.11 : Test Recall du CNN

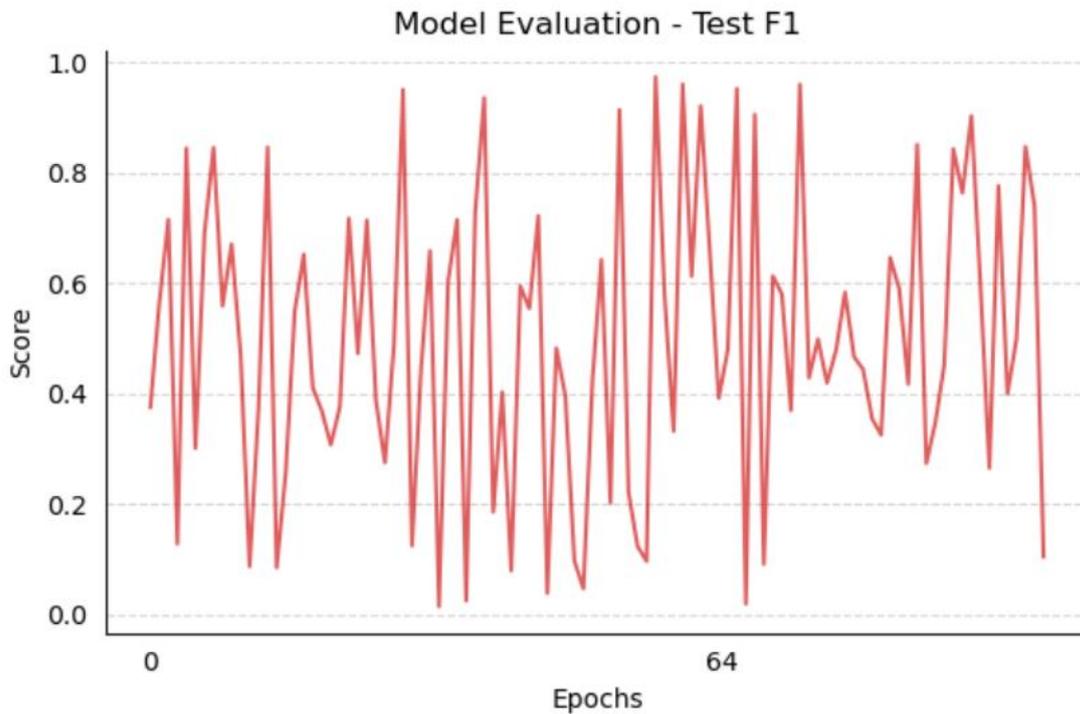


Figure 4.12 : Test F1 Score du CNN

Les résultats de modèle sont présentés dans le Tableau 10 :

Train				Test			
Acc (%)	Prec (%)	Rec (%)	F1 Score (%)	Acc (%)	Prec (%)	Rec (%)	F1 Score (%)
85.73	85.80	85.73	80.40	86.08	84.38	86.08	80.81

Tableau 10 : Résultats obtenus avec le model CNN.

Les résultats du tableau indiquent les performances du modèle (CNN) sur les ensembles d'entraînement et de test, suggèrent que le modèle a de bonnes performances, avec des scores élevés pour la précision, le rappel et le score F1 sur les ensembles d'entraînement et de test.

8. Discussion :

Les avancées dans le domaine de l'e-santé, de la télémédecine et du traitement des données massives transforment considérablement la prestation des soins de santé. Un aspect crucial de cette transformation consiste à rendre les données massives accessibles via des systèmes informatisés, afin d'assister les médecins et les professionnels de la santé dans leurs diagnostics, en les rendant plus rapides et plus précis.

Notre objectif est de développer un outil d'aide à la prise de décision pour la prédiction des maladies cardiovasculaires, en particulier la maladie coronarienne. Au cours de nos recherches, nous avons constaté que la méthode de deep learning utilisant la technique CNN est une option prometteuse pour atteindre cet objectif. Il est important de souligner que cette technique n'a pas pour but de remplacer le médecin, mais plutôt de servir d'outil intelligent permettant d'exploiter les vastes quantités de données disponibles sur des patients ayant déjà reçu un diagnostic précis.

9. Conclusion :

Dans ce chapitre, nous avons présenté les différentes étapes que nous avons suivies pour concevoir et implémenter notre modèle de prédiction des maladies cardiaques, en mettant l'accent sur la maladie coronarienne. Nous avons décrit les différents outils et langages de développement utilisés pour ce projet. Ensuite, nous avons examiné en détail les résultats de performance de l'algorithmes employés (CNN), en utilisant différentes mesures de performance.

Conclusion générale

Conclusion générale

L'utilisation des techniques d'apprentissage profond pour prédire les maladies représente un domaine prometteur de la médecine moderne. Ces techniques permettent d'analyser les données des patients, d'apprendre les modèles associés à différentes maladies et de prédire la probabilité de survenue future d'une maladie spécifique. Les applications de l'apprentissage profond dans le domaine médical constituent une avancée majeure en matière de soins de santé. Elles facilitent la détection précoce des maladies, ainsi qu'une amélioration du diagnostic et du traitement individualisés. Les algorithmes d'apprentissage profond peuvent aider à détecter les symptômes précoces des maladies et à prédire les risques de développer certaines affections. Cela permet aux médecins et aux professionnels de la santé d'intervenir plus rapidement pour diagnostiquer et traiter les patients, ce qui se traduit par une amélioration des résultats thérapeutiques.

Dans le cadre de ce projet, nous avons examiné les recherches récentes dans le domaine de la prédiction des maladies cardiaques, en mettant l'accent sur la maladie coronarienne. À partir de l'analyse des études examinées, nous avons constaté que certaines recherches proposaient une approche basée sur la comparaison des performances des algorithmes pour choisir le plus efficace. Cependant, cette méthode s'est révélée moins efficace que les systèmes hybrides utilisant plusieurs algorithmes. De plus, la plupart de ces études étaient centrées sur l'apprentissage automatique, bien que les recherches suggèrent souvent l'utilisation d'algorithmes d'apprentissage profond pour améliorer les résultats. C'est pourquoi nous avons opté pour cette approche en utilisant l'algorithme de CNN, ce qui nous a permis d'obtenir une précision de prédiction de 86.08%.

La précision dans de tels cas revêt une importance cruciale, car elle a un impact direct sur la vie des individus concernés. Il est essentiel de comprendre que les chiffres ne représentent pas seulement des données, mais également des vies humaines. Ainsi, toute marge d'erreur dans la précision des résultats reste préoccupante.

En continuant ce travail à l'avenir, nous avons le potentiel de contribuer de manière significative à l'avancement des connaissances et à l'amélioration des soins de santé dans le domaine des maladies cardiovasculaires, en particulier la maladie coronarienne, qui est à la fois dangereuse et mortelle. À l'avenir, nous espérons que le futur est d'accroître la précision de la prédiction de la maladie coronarienne afin de réduire autant que possible les taux d'erreur. Nous aspirons à garantir un niveau de confiance plus élevé, assurant ainsi une sécurité maximale pour la vie des individus concernés.

Bibliographie

Bibliographie

- [1] : Cynthia Khanji. (2018). Évaluation de la qualité des soins et des services préventifs cardiovasculaires en première ligne. Université de Montréal, 342.
- [2] : Fonctionnement du cœur. (s. d.). CHUV. <https://www.chuv.ch/fr/cardiologie/car-home/patients-et-famille/fonctionnement-du-coeur>
- [3] : World Health Organization : WHO. (2021, 11 juin). Maladies cardiovasculaires. [https://www.who.int/fr/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)#:~:text=Principaux%20faits&text=On%20estime%20%C3%A0%2017%2C7,de%20la%20mortalit%C3%A9%20mondiale%20totale](https://www.who.int/fr/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)#:~:text=Principaux%20faits&text=On%20estime%20%C3%A0%2017%2C7,de%20la%20mortalit%C3%A9%20mondiale%20totale).
- [4] : Maladie coronarienne (Athérosclérose). (2024, 22 mars). Institut de Cardiologie de L'Université D'Ottawa. <https://www.ottawaheart.ca/fr/maladie-du-c%C5%93ur/maladie-coronarienne-atherosclerose>
- [5] : Carthage_Medical_Ref. (2022, 17 octobre). Maladies des artères coronaires. MedCarthage. <https://www.carthage-medical.com/blog/maladies-des-arteres-coronaires/>
- [6] : Sweis, R. N., Jivan, A., & Msd, M. (2022, 7 juin). Présentation de la maladie des artères coronaires (MAC). Manuels MSD Pour le Grand Public. <https://www.msmanuals.com/fr/accueil/troubles-cardiaques-et-vasculaires/maladie-des-art%C3%A8res-coronaires/pr%C3%A9sentation-de-la-maladie-des-art%C3%A8res-coronaires-mac>
- [7] : Mousseseye. (2023, 12 décembre). Comment l'intelligence artificielle peut-elle aider les professionnels de santé ? MSD Connect. <https://www.msconnect.fr/comment-lintelligence-artificielle-peut-elle-aider-les-professionnels-de-sante/>
- [8] : les applications de l'intelligence artificielle en médicales <https://www.frm.org/recherches-financees/intelligence-artificielle/tout-savoir-ia>
- [9] : Next-generation Echocardiography is here. (s. d.). Ultromics. <https://www.ultromics.com/>
- [10] : Qu'est-ce que l'intelligence artificielle (IA) ? | IBM. (s. d.). <https://www.ibm.com/fr-fr/topics/artificial-intelligence>
- [11] : Frédéric Sur. (2023). Introduction à l'apprentissage automatique. Tronc commun scientifique FICM 2A École des Mines de Nancy,178.

- [12] : Lgayhardt. (2024, 19 janvier). Apprentissage profond et apprentissage automatique - Azure Machine Learning. Microsoft Learn. <https://learn.microsoft.com/fr-fr/azure/machine-learning/concept-deep-learning-vs-machine-learning?view=azureml-api-2>
- [13] : Kassel, R. (2023, 9 novembre). Deep Neural Network : qu'est-ce qu'est-ce qu'un réseau de neurones profond ? Formation Data Science | DataScientest.com. <https://datascientest.com/deep-neural-network>
- [14] : Réseaux de neurones profonds (DNN). (s. d.). Techniques de L'Ingénieur. <https://www.techniques-ingenieur.fr/base-documentaire/technologies-de-l-information-th9/intelligence-artificielle-42679210/introduction-aux-reseaux-de-neurones-h3730/reseaux-de-neurones-profonds-dnn-h3730niv10004.html>
- [15] : Rahman, S., Hasan, M. M., & Sarkar, A. K. (2022). Machine Learning and Deep Neural Network Techniques for Heart Disease Prediction. IEEE. <https://doi.org/10.1109/iccit57492.2022.10055902>
- [16] : Prediction of Heart Disease Using Machine Learning : A Systematic Literature Review. (2023). Journal Of System And Management Sciences, 13(6). <https://doi.org/10.33168/jsms.2023.0603>
- [17] : Hamdi A. Al-Jamimi. (2023). Intelligent Methods for Early Prediction of Heart Disease. IEEE. <https://doi.org/10.1109/codit58514.2023.10284226>
- [18] : Md. Imam Hossain, Mehadi Hasan Maruf, Md. Ashikur Rahman Khan, Farida Siddiqi Prity, Sharmin Fatema, Md. Sabbir Ejaz, Md. Ahnaf Sad Khan. (2023). Heart disease prediction using distinct artificial intelligence techniques : performance analysis and comparison. Iran Journal of Computer Science, 6 :397–417.
- [19] : Alkhodari, M., Jelinek, H. F., Karlas, A., Soulaïdopoulos, S., Arsenos, P., Doundoulakis, I., Gatzoulis, K. A., Tsioufis, K., Hadjileontiadis, L. J., & Khandoker, A. H. (2021). Deep Learning Predicts Heart Failure With Preserved, Mid-Range, and Reduced Left Ventricular Ejection Fraction From Patient Clinical Profiles. Frontiers In Cardiovascular Medicine, 8. <https://doi.org/10.3389/fcvm.2021.755968>
- [20] : Dami, S., & Yahaghizadeh, M. (2021). Predicting cardiovascular events with deep learning approach in the context of the internet of things. Neural Computing & Applications, 33(13), 7979-7996. <https://doi.org/10.1007/s00521-020-05542-x>

- [21] : Ali M. A. Barhoom, Abdelbaset Almasri, Bassem S. Abu-Nasser, Samy S. Abu-Naser.(2022). Prediction of Heart Disease Using a Collection of Machine and Deep Learning Algorithms. International Journal of Engineering and Information Systems, Vol. 6 Issue 4, Pages :1-13.
- [22] : Salunke, T., Jagade, P., Pawar, S., Rathod, P., & Ghawate, N. (2023). Heart Disease Detection using Hybrid Machine Learning and IoT (Software Based). Journal Of Artificial Intelligence, Machine Learning And Neural Network, 33, 10-13. <https://doi.org/10.55529/jaimltn.33.10.13>
- [23] : Amna Kanwal, Dr.Khawaja Tehseen Ahmad,Mr. Muhammad Kamran Abid, Dr.Naeem Aslam4. (2022). Detection of Heart Disease Using Supervised Machine Learning. VFAST Transactions on Software Engineering, Volume 10, Number 3, pp : 58-70.
- [24] : Divya Krishnani, Anjali Kumari, Akash Dewangan, Aditya Singh, Nenavath Srinivas Naik. (2019). Prediction of Coronary Heart Disease using Supervised Machine Learning Algorithms. IEEE Region 10 Conference (TENCON 2019), pp : 367-372.
- [25] : Yang, H., Chen, Z., Yang, H., & Tian, M. (2023). Predicting Coronary Heart Disease Using an Improved LightGBM Model : Performance Analysis and Comparison. IEEE Access, 11, 23366-23380. <https://doi.org/10.1109/access.2023.3253885>
- [26] : Hassan, C. A. U., Iqbal, J., Irfan, R., Hussain, S., Algarni, A. D., Bukhari, S. S. H., Alturki, N., & Ullah, S. S. (2022). Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers. Sensors, 22(19), 7227. <https://doi.org/10.3390/s22197227>
- [27] : Ainiwaer, A., Hou, W. Q., Qi, Q., Kadier, K., Qin, L., Rehemuding, R., Mei, M., Wang, D., Ma, X., Dai, J. G., & Tong, Y., MA. (2024). Deep learning of heart-sound signals for efficient prediction of obstructive coronary artery disease. Heliyon, 10(1), e23354. <https://doi.org/10.1016/j.heliyon.2023.e23354>
- [28] : Sharma, N., Malviya, L., Jadhav, A., & Lalwani, P. (2023). A hybrid deep neural net learning model for predicting Coronary Heart Disease using Randomized Search Cross-Validation Optimization. Decision Analytics Journal, 9, 100331. <https://doi.org/10.1016/j.dajour.2023.100331>
- [29] : Özbilgin, F., Kurnaz, Ç., & Aydın, E. (2023). Prediction of Coronary Artery Disease Using Machine Learning Techniques with Iris Analysis. Diagnostics, 13(6), 1081. <https://doi.org/10.3390/diagnostics13061081>
- [30] : Framingham heart study dataset. (2022, 19 avril). Kaggle. <https://www.kaggle.com/datasets/aasheesh200/framingham-heart-study-dataset/data>

Bibliographie

- [31]: Contributeurs aux projets Wikimedia. (2024, 15 mai). Python (langage). [https://fr.wikipedia.org/wiki/Python_\(langage\)](https://fr.wikipedia.org/wiki/Python_(langage))
- [32]: Contributeurs aux projets Wikimedia. (2024a, mai 9). Keras. <https://fr.wikipedia.org/wiki/Keras>
- [33]: Contributeurs aux projets Wikimedia. (2024a, avril 29). TensorFlow. <https://fr.wikipedia.org/wiki/TensorFlow>
- [34]: Contributeurs aux projets Wikimedia. (2024a, février 12). NumPy. <https://fr.wikipedia.org/wiki/NumPy>
- [35]: Contributeurs aux projets Wikimedia. (2024a, février 12). Matplotlib. <https://fr.wikipedia.org/wiki/Matplotlib>
- [36]: Contributeurs aux projets Wikimedia. (2024c, avril 26). Pandas. <https://fr.wikipedia.org/wiki/Pandas>
- [37]: seaborn : statistical data visualization — seaborn 0.13.2 documentation. (s. d.). <https://seaborn.pydata.org/>
- [38]: Contributeurs aux projets Wikimedia. (2024c, avril 23). Scikit-learn. <https://fr.wikipedia.org/wiki/Scikit-learn>
- [39]: The Jupyter Notebook — Jupyter Notebook 7.2.0 documentation. (s. d.). <https://jupyter-notebook.readthedocs.io/en/stable/notebook.html>