

الجمهورية الجزائرية الديمقراطية الشعبية

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE**

وزارة التعليم العالي والبحث العلمي

**Ministère de l'Enseignement Supérieur et de la Recherche Scientifique**

جامعة أبي بكر بلقايد - تلمسان

Université Aboubakr Belkaïd – Tlemcen –

Faculté de TECHNOLOGIE



## **MEMOIRE**

Présenté pour l'obtention du **diplôme** de **MASTER**

**En** : Télécommunications

**Spécialité** : Réseaux et Télécommunications

**Par** : Belabdelli Fatima Zohra & Zeriouh Asmae Nour El Houda

### **Sujet**

**Evaluation des maladies cardiaques à l'aide de techniques de  
Machine Learning et Deep Learning**

Soutenu publiquement, le 06/06/2023, devant le jury composé de :

M. IRID Sidi Mohammed Hadj	MCA	Université de Tlemcen	Président
Mme SEDJELMACI Amina	MCB	Université de Tlemcen	Examinateur
M. HADJILA Mourad	MCA	Université de Tlemcen	Encadreur

# Dédicaces

*Je remercie **ALLAH** le tout-puissant de m'avoir accordé, ainsi que mon amie **Fatima Zohra**, l'opportunité de compléter ce mémoire.*

*Je tiens c'est avec grande plaisir que je dédie ce modeste travail*

*A mes chers parents ma mère et mon père pour leur patience, leur amour, leur soutien et leurs encouragements pendant tout mon parcours universitaire.*

*A ma chère famille **Kawther, Farah et Mohammed** qui m'ont toujours encouragé, et à qui je souhaite plus de succès.*

*Asmae Nour El Houda*

# Dédicaces

*J'ai l'honneur de dédier ce modeste travail :*

*A ma chère mère et mon précieux père pour tout leur soutien inconditionnel, à la fois moral et économique et leurs amour et sacrifices innombrables tout au long de ma vie.  
Qu'Allah leurs accorde santé, bonheur et longue vie.*

*A l'âme de mes chers, grand-père et grand-mère, **Miloud** et **Kheira** qui ont signifié et continuent de représenter tant de choses pour moi, bien qu'ils ne soient plus de ce monde, leurs souvenirs continuent de faire partie de ma vie. Qu'Allah leurs accorde son vaste paradis.*

*A mes chers frères et sœurs pour leurs encouragements permanents, et leur soutien moral à qui je souhaite beaucoup de bonheur et réussite dans leurs vies.*

*A ma chère amie et binôme **Asmae Nour El Houda** pour son aide et sa patience durant la préparation de ce projet et à mes fidèles amies et en particulier **Dounia** et **Khawla** avec qui j'ai passé les plus beaux et les plus agréables moments tout au long de mon cursus universitaire.*

**Fatima Zohra**

# Remerciements

*Nous voudrions tous d'abord remercier **ALLAH** le tout puissant, qui nous a donné la volonté, le courage et la patience pour développer et réaliser ce modeste travail.*

*Toute nos reconnaissances s'adressent à notre directeur de mémoire **Mr. HADJILA Mourad** pour la confiance qui nous a accordée ainsi que sa disponibilité et ses judicieux conseils qui nous a énormément aidé et contribué à alimenter nos réflexions.  
Nous avons eu beaucoup de plaisir de travailler à votre côté.*

*Nos sincères remerciements destinés aussi à **Mlle. FERHI Wafaa** pour sa rigueur, son orientation et son accompagnement tout au long de ce mémoire. Ce travail porte également votre empreinte.*

*Nous tenons également à remercier chaleureusement les membres du jury, **M. IRID Sidi Mohammed Hadj** pour avoir accepté de présider le jury de soutenance et **Mme SEDJELMACI Amina** pour avoir examiné notre travail.*

*Enfin, nos remerciements ont tous ceux qui sont présent de près ou de loin tout au long de notre parcours universitaire.*

# Résumé

Les maladies cardiaques constituent la première cause de décès dans le monde entier. La détection précoce de ces pathologies est d'une importance majeure pour assurer une intervention médicale préventive et réduire les complications potentielles. Ces dernières années, les évolutions technologiques notamment l'intelligence artificielle ont permis d'ouvrir de nouvelles perspectives intéressantes pour prédire les maladies cardiaques avec une meilleure précision. Les deux approches de classification binaire proposées dans ce mémoire ont été testées sur le dataset « Personal Key Indicator Heart Disease ». La première approche est une classification déséquilibrée alors que la seconde est une classification équilibrée. En premier lieu, les données de dataset seront prétraitées et nettoyées, puis les deux algorithmes d'apprentissage, le Deep Learning et le Machine Learning, ont été utilisés afin de construire des modèles performants en termes d'exactitude (Accuracy), de précision, de recall et de F1-score. Les résultats trouvés ont montré que la balance des données influe considérablement sur ces mesures d'évaluation pour les algorithmes de DL et que l'accuracy reste mal améliorée, néanmoins les algorithmes de ML étaient plus faciles à mettre en œuvre.

**Mots Clés :** Maladies cardiovasculaires, Intelligence artificielle, Apprentissage profond, Apprentissage automatique, Réseaux de neurones artificiels, Personal Key Indicators of Heart Disease dataset.

# Abstract

Heart diseases are the leading cause of death around the world. Early detection of these diseases is of major importance to ensure preventive medical intervention and reduce potential complications. In recent years, technological developments including artificial intelligence have opened up interesting new perspectives for predicting heart disease with greater precision. The two binary classification approaches proposed in this master thesis have been tested on the "Personal Key Indicator Heart Disease" dataset. The first approach is an unbalanced classification while the second is a balanced classification. First, the dataset will be preprocessed and cleaned, then the two learning algorithms, Deep Learning and Machine Learning, were used to build models that perform well in terms of accuracy, precision, recall and F1-score. The results found showed that the balance of the data significantly affects these evaluation measures for DL algorithms and accuracy remains poorly improved, however ML algorithms were easier to apply.

**Keywords :** Cardiovascular Disease, Artificial Intelligence, Deep Learning, Machine Learning, Artificial Neural Networks, Personal Key Indicators of Heart Disease dataset.

# Table des matières

<b>Dédicaces .....</b>	<b>I</b>
<b>Dédicaces .....</b>	<b>II</b>
<b>Remerciements .....</b>	<b>III</b>
<b>Résumé .....</b>	<b>IV</b>
<b>Abstract .....</b>	<b>V</b>
<b>Liste des figures .....</b>	<b>X</b>
<b>Liste des tableaux .....</b>	<b>XII</b>
<b>Acronymes.....</b>	<b>XIII</b>
<b>Introduction générale.....</b>	<b>1</b>
<b>Chapitre 1 Les maladies cardiaques.....</b>	<b>3</b>
1.1 Introduction .....	4
1.2 Le cœur .....	4
1.3 L'anatomie et le fonctionnement du cœur .....	5
1.3.1 Ventricule droit .....	6
1.3.2 Ventricule gauche.....	6
1.3.3 Oreillette droite .....	7
1.3.4 Oreillette gauche .....	7
1.4 Circulation sanguine .....	7
1.4.1 Circulation sanguine côté droit .....	7
1.4.2 Circulation sanguine côté gauche.....	8
1.5 Le battement du cœur .....	8
1.6 Les maladies cardiovasculaires .....	9
1.6.1 Les maladies vasculaires .....	9
1.6.2 Les maladies coronariennes.....	10
1.6.3 L'insuffisance cardiaque .....	10
1.6.4 L'angine de poitrine .....	10
1.6.5 L'infarctus du myocarde .....	10
1.6.6 Asthme cardiaque .....	11
1.6.7 L'arrêt cardiaque .....	11
1.7 Différence entre arrêt cardiaque, crise cardiaque et l'insuffisance cardiaque .....	11
1.7.1 L'arrêt cardiaque .....	11

1.7.2	La crise cardiaque.....	12
1.7.3	L'insuffisance cardiaque .....	12
1.8	Symptômes des maladies cardiaques.....	13
1.9	Diagnostic des maladies cardiaques .....	14
1.9.1	Tomodensitométrie (TDM) du cœur .....	14
1.9.2	Les analyses du sang .....	14
1.9.3	L'électrocardiographie .....	14
1.9.4	Echocardiographie.....	15
1.9.5	Epreuve d'efforts.....	16
1.9.6	Imagerie par résonance magnétique (IRM) du cœur.....	16
1.9.7	Angiographie coronarienne et cathétérisme cardiaque .....	17
1.10	Facteurs de risques.....	18
1.11	Prévention cardiovasculaire.....	19
1.11.1	Prévenir une crise cardiaque.....	19
1.12	Conclusion .....	20
<b>Chapitre 2 Machine Learning &amp; Deep Learning .....</b>		<b>21</b>
2.1	Introduction .....	22
2.2	Machine Learning.....	23
2.2.1	L'apprentissage non supervisé.....	23
2.2.2	L'apprentissage par renforcement.....	23
2.2.3	L'apprentissage supervisé.....	23
2.2.3.1	Dataset .....	23
2.2.3.2	Le modèle et ses paramètres .....	24
2.2.3.3	La Fonction Coût .....	24
2.2.3.4	L'Algorithme d'apprentissage .....	25
2.2.3.5	Les applications d'apprentissage supervisé .....	26
2.3	Types d'algorithmes en Machine Learning.....	26
2.3.1	K Nearest Neighbour Algorithm (KNN).....	26
2.3.2	Support Vector Machine (SVM) .....	27
2.3.3	Decision Tree Algorithm (DT).....	28
2.3.4	Random Forest Classification (RF).....	29
2.3.5	Naïve Bayes Algorithme (NB).....	30
2.4	Les applications du Machine Learning.....	31
2.4.1	Détection de fraude .....	31
2.4.2	Assistance médicale .....	31



2.4.3	Jeux et éducation .....	31
2.5	Deep Learning .....	32
2.5.1	Définition de l'apprentissage profond.....	32
2.5.2	Types d'apprentissage profond .....	32
2.5.2.1	Apprentissage supervisé .....	32
2.5.2.2	Apprentissage non supervisé .....	33
2.5.3	Les applications du Deep Learning .....	33
2.5.3.1	Détection des retards de développement chez les enfants .....	33
2.5.3.2	Marketing personnalisé.....	34
2.5.3.3	Détection de fausses nouvelles .....	34
2.6	Les réseaux de neurones artificiels .....	34
2.6.1	Introduction sur les réseaux de neurones .....	34
2.6.2	Fonctionnement des réseaux de neurones artificiels .....	35
2.6.3	Types des réseaux de neurones artificiels. ....	36
2.6.3.1	Réseaux neuronaux à propagation directe (FeedForward).....	36
2.6.3.2	Réseaux de neurones récurrents RNN .....	37
2.6.3.3	Réseaux de mémoire à long terme (LSTM) .....	37
2.6.3.4	Réseaux de neurones convolutifs (CNN) .....	39
2.6.4	Les fonctions d'activations.....	40
2.6.4.1	Introduction sur les fonctions d'activations.....	40
2.6.4.2	Fonction d'activation linéaire .....	40
2.6.4.3	Fonction d'activation non linéaire .....	41
2.6.4.4	Fonction d'activation en escalier .....	44
2.6.5	Fonction de perte (Loss Function) .....	44
2.6.5.1	Regression loss .....	44
2.6.5.2	Classifications loss .....	45
2.7	Comparaison entre Machine Learning et Deep Learning.....	45
2.8	Conclusion .....	47
<b>Chapitre 3 Conception d'un modèle intelligent : résultats et interprétations .....</b>		<b>48</b>
3.1	Introduction .....	49
3.2	Environnement d'exécution.....	49
3.2.1	Logiciels .....	49
3.2.2	Bibliothèques utilisées.....	50
3.3	Implémentation .....	51
3.4	Dataset .....	53

3.5	Description du dataset .....	53
3.6	Prétraitement de dataset .....	55
3.7	Nettoyage de dataset .....	55
3.8	Création du modèle de détection de la maladie cardiaque .....	56
3.8.1	Première approche (Imbalanced data) .....	56
3.8.1.1	Entraînement de modèle .....	57
3.8.1.2	Métriques d'évaluation du modèle .....	58
3.8.1.3	Implémentation des algorithmes de Machine Learning .....	60
3.8.1.4	Evaluation des résultats .....	60
3.8.1.5	Hyperparamètres de Deep Learning .....	61
3.8.1.6	Evaluation des résultats .....	62
3.8.2	Deuxième approche (balanced dataset) .....	65
3.8.2.1	Méthodes de data Sampling .....	65
3.8.2.2	Entraînement du modèle .....	66
3.8.2.3	Evaluation des résultats .....	66
3.9	Etude comparative .....	70
3.10	Conclusion .....	70
	<b>Conclusion générale .....</b>	<b>71</b>
	<b>Bibliographie.....</b>	<b>73</b>

# Liste des figures

<b>Figure 1.1</b> : Position du cœur. ....	4
<b>Figure 1.2</b> : Anatomie du cœur. ....	6
<b>Figure 1.3</b> : Circulation sanguine. ....	8
<b>Figure 1.4</b> : La conduction électrique dans le cœur. ....	9
<b>Figure 1.5</b> : Électrocardiogramme normal. ....	15
<b>Figure 1.6</b> : Echocardiographie du cœur. ....	16
<b>Figure 1.7</b> : Cathétérisme cardiaque et coronarographie. ....	17
<b>Figure 2.1</b> : Diagramme de Venn représentant la relation entre AI, ML et DL. ....	22
<b>Figure 2.2</b> : Modèle linéaire et modèle non linéaire. ....	24
<b>Figure 2.3</b> : Fonction coût. ....	25
<b>Figure 2.4</b> : Bon modèle (graphe a) vs Mauvais modèle (graphe b). ....	25
<b>Figure 2.5</b> : Les deux types de ML. ....	26
<b>Figure 2.6</b> : Classification d'un nouveau point avec KNN. ....	27
<b>Figure 2.7</b> : Modèle SVM. ....	28
<b>Figure 2.8</b> : Arbre de décision. ....	29
<b>Figure 2.9</b> : Forêt aléatoire. ....	30
<b>Figure 2.10</b> : Algorithme supervisé. ....	32
<b>Figure 2.11</b> : Algorithme non supervisé. ....	33
<b>Figure 2.12</b> : Un neurone humain. ....	35
<b>Figure 2.13</b> : Un réseau artificiel. ....	36
<b>Figure 2.14</b> : Le réseau FeedForward. ....	36
<b>Figure 2.15</b> : Réseaux de neurones récurrents (RNN). ....	37
<b>Figure 2.16</b> : Structure d'un réseau LSTM. ....	38
<b>Figure 2.17</b> : Un réseau de neurones convolutif (CNN). ....	40
<b>Figure 2.18</b> : Fonction linéaire. ....	41
<b>Figure 2.19</b> : Fonction Sigmoidé. ....	42
<b>Figure 2.20</b> : Fonction Tanh. ....	42
<b>Figure 2.21</b> : Fonction ReLU. ....	43
<b>Figure 2.22</b> : Fonction Softmax. ....	44
<b>Figure 3.1</b> : Approche de détection des maladies cardiaques. ....	52

<b>Figure 3.2 :</b> Visualisation de la classe Target.....	57
<b>Figure 3.3 :</b> Training et test split .....	57
<b>Figure 3.4 :</b> Diagramme de comparaison entre les algorithmes de ML. ....	61
<b>Figure 3.5 :</b> Résumé du modèle.....	63
<b>Figure 3.6 :</b> Model accuracy.....	63
<b>Figure 3.7 :</b> Model loss.....	64
<b>Figure 3.8 :</b> Matrice de confusion. ....	65
<b>Figure 3.9 :</b> Matrice de confusion de méthode Random Oversampling.....	67
<b>Figure 3.10 :</b> Matrice de confusion de méthode SMOTE. ....	68
<b>Figure 3.11 :</b> Matrice de confusion de méthode Random Undersampling.....	69

# Liste des tableaux

<b>Tableau 2.1:</b> Différence entre ML et DL. ....	46
<b>Tableau 3.1 :</b> Les paramètres d'environnement kaggle. ....	50
<b>Tableau 3.2 :</b> Bibliothèques utilisées. ....	51
<b>Tableau 3.3 :</b> Description des attributs de dataset.....	54
<b>Tableau 3.4 :</b> Caractéristiques générales du dataset.....	55
<b>Tableau 3.5 :</b> Changement de types des features. ....	56
<b>Tableau 3.6 :</b> Matrice de confusion.....	58
<b>Tableau 3.7 :</b> Comparaison des performances des algorithmes de ML. ....	60
<b>Tableau 3.8 :</b> Métriques d'évaluation pour le modèle DL. ....	64
<b>Tableau 3.9 :</b> Résultats des métriques de méthode Random Oversampling. ....	67
<b>Tableau 3.10 :</b> Résultats des métriques de méthode SMOTE.....	68
<b>Tableau 3.11 :</b> Résultats des métriques de méthode Random Undersampling. ....	69

# Acronymes

<b>AB</b>	Asthme Bronchique
<b>AC</b>	Arrêt Cardiaque
<b>ANN</b>	Artificial Neural Networks
<b>AP</b>	Angine de Poitrine
<b>AVC</b>	Accident Vasculaire Cérébral
<b>BNP</b>	Brain Natriuretic Peptide
<b>CDC</b>	Center for Disease Control
<b>CNN</b>	Convolution Neural Networks
<b>CT</b>	Coronaire par Tomodensitométrie
<b>CV</b>	Cardiovasculaire
<b>DEA</b>	Défibrillateur Externe Automatisé
<b>DL</b>	Deep Learning
<b>DT</b>	Decision Tree
<b>ECG</b>	Electrocardiogramme
<b>FDR CV</b>	Facteur De Risque Cardiovasculaire
<b>FN</b>	False Negative
<b>FP</b>	False Positive
<b>FP</b>	Forward Propagation
<b>HDL</b>	Lipoprotéines de Haute Densité
<b>IA</b>	Intelligence Artificielle
<b>IDE</b>	Environnements de Développement Intégré
<b>IM</b>	Infarctus Du Myocarde
<b>IRM</b>	Imagerie par Résonance Magnétique
<b>KNN</b>	K-Nearest Neighbour
<b>LDL</b>	Lipoprotéines de Faible Densité
<b>LSTM</b>	Réseaux de Mémoire à Long Terme
<b>MCV</b>	Maladies Cardiovasculaires

**ML** Machine Learning  
**NB** Naïve Bayes  
**NT-proBNP** N-Terminal proBNP  
**OD** Oreillette Droit  
**OG** Oreillette Gauche  
**PKIHD** Personal Key Indicators of Heart Disease  
**RCP** Réanimation CardioPulmonaire  
**ReLU** Rectified Linear Unit  
**RF** Random Forest  
**RNN** Recurrent Neural Networks  
**SVM** Support Vector Machine  
**TDM** TomoDensitoMétrie  
**TN** True Negative  
**TP** True Positive  
**VD** Ventricule Droit  
**VG** Ventricule Gauch

# Introduction générale

Avec la croissance de la population mondiale et les changements récents des modes de vie, le nombre de personnes souffrant de problèmes médicaux complexes est de plus en plus élevé.

Les maladies cardiaques constituent la première cause de mortalité de l'être humain dans le monde entier en raison d'une vie malsaine, citant le stress, le manque d'activité physique, l'obésité et d'autres facteurs qui risquent d'avoir des troubles cardiovasculaires qui influent négativement sur le corps et donc peuvent provoquer une mort subite. Pour éviter de contracter cette maladie, il est conseillé de déterminer et traiter très tôt ce défi, qui est un point complexe en particulier dans les pays en développement, en raison du manque d'équipements de diagnostic, de médecins et d'autres ressources qui affectent la prédiction et le traitement appropriés des patients cardiaques.

Il est donc nécessaire de disposer de systèmes de soins de santé capables de relever ces problèmes. Récemment, les progrès réalisés dans les domaines de l'électronique et de la médecine ont fait l'objet d'une attention croissante, de même que les applications intéressantes que ces technologies offrent, principalement pour le diagnostic, le suivi et l'analyse de la santé.

De nos jours, l'intelligence artificielle est l'un des progrès qui est devenue le seul moyen pourras faire un saut de développement dans le domaine médical grâce au Machine Learning et au Deep Learning. Ces derniers sont des techniques importantes dans cette branche utilisée pour établir des modèles qui peuvent être formés pour les prédictions de pandémie. Les dossiers médicaux peuvent également être transformés et analysés plus en profondeur pour de meilleures prédictions.

Le data mining est utile pour étudier et comprendre une grande quantité de dataset. Il permet d'extraire des données et de prendre des décisions en vue d'applications ultérieures. Les techniques les plus courantes couvertes par l'exploration de données sont le regroupement, l'extraction de règles d'association et les classifications. De nombreux algorithmes sont disponibles pour mettre en œuvre ces techniques de data mining.

Nous avons choisi pour réussir à la réalisation de ce mémoire le dataset "Personal Key Indicators of Heart Disease (PKIHD)" pour prédire la présence de la maladie cardiaque chez les patients en utilisant les deux techniques d'apprentissage ML et DL pour des résultats rapides et précis.

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



Le premier chapitre présente des généralités sur le cœur, l'organe le plus important dans le corps humain en citant leur anatomie et fonctionnement en se basant sur les symptômes et les facteurs de risques qui provoquent des maladies cardiovasculaires et comment diagnostiquer et parvenir contre ces maladies.

Le deuxième chapitre consiste à couvrir le domaine d'intelligence artificielle, en développant leurs deux sous-systèmes Machine Learning et Deep Learning avec leurs types, les applications et les algorithmes utilisés.

Le dernier chapitre est consacré à la mise en évidence des résultats obtenus tout en détaillant la méthodologie utilisée et les paramètres pris en compte.

Nous terminons par une conclusion générale qui résume nos différentes contributions et présente les perspectives de recherche de notre travail.

# **Chapitre 1**

## **Les maladies cardiaques**

## 1.1 Introduction

Les maladies cardiaques sont composées de multiples troubles qui affectent le cœur. Aujourd'hui, ces maladies sont la principale cause de décès dans le monde, avec 17.9 millions de décès par an, selon l'organisation mondiale de la santé. Diverses activités nocives sont responsables de l'augmentation de l'incidence des maladies cardiaques, notamment l'hypertension, l'obésité, l'augmentation des triglycérides etc. Un diagnostic et un traitement précoces de la maladie peuvent entraîner une réduction du nombre de décès. Pour comprendre et prédire les maladies cardiaques, nous devons avoir une compréhension approfondie du cœur et son fonctionnement. Nous vous présenterons également les causes et les effets des maladies cardiaques qui affectent négativement la psychologie, ce qui provoque à son tour un arrêt cardiaque et une mort subite.

## 1.2 Le cœur

Le cœur humain est une structure musculaire creuse qui ressemble à un cône, avec la base en haut et les pointes (apex) en bas et à gauche situés dans la cavité thoracique entre les poumons. Il a à peu près une taille d'un poing fermé et pèse environ 250 à 300 grammes pour un adulte. Il charge de pomper le sang environs de 8000 litres grâce à environ 100000 battements quotidiens au cours d'une journée typique pour soutenir tout le corps par l'acheminement du sang oxygéné vers les organes, les tissus et les cellules, et du retour du sang désoxygéné vers les poumons pour la ré-oxygénation (voir la figure 1.1).



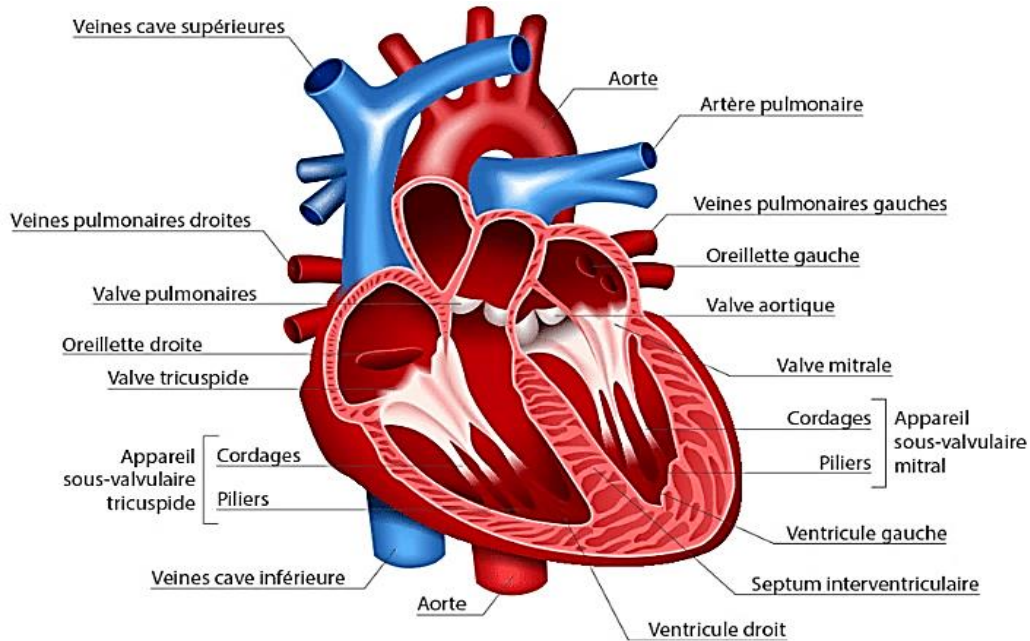
**Figure 1.1:** Position du cœur.

### 1.3 L'anatomie et le fonctionnement du cœur

Le cœur est enveloppé et protégé par le péricarde qui forme une sorte de sac qui recouvre complètement le cœur et s'étend jusqu'aux racines des grands vaisseaux. Il a deux parties, le péricarde séreux et le péricarde fibreux. Il est entouré d'une membrane appelée péricarde composé de deux couches. La couche externe du péricarde est attachée à la colonne vertébrale, au diaphragme et d'autres parties du corps via des ligaments. De plus, il entoure les vaisseaux sanguins des principaux organes du cœur. La couche interne du péricarde est reliée au muscle cardiaque, et une couche liquide sépare les deux couches de membrane, cela permet au cœur de bouger pendant un battement. La paroi du cœur est composée de trois couches : l'épicarde externe (couche fine), le myocarde moyen (couche épaisse) et l'endocarde interne (couche fine). Le myocarde est composé de fibres musculaires cardiaques, et ces fibres sont responsables de la contraction et de la relaxation qui font pomper le cœur [1].

Le cœur est divisé en deux côtés gauche et droit et il est composé de quatre chambres, dont deux sont situées en haut et deux en bas. Les deux chambres plus petites sont appelées le ventricule droit et ventricule gauche. Ils forcent le sang à quitter le cœur. Une paroi appelée septum interventriculaire les sépare, et chaque ventricule comporte deux valvules, une valvule d'entrée et l'autre pour la sortie. Les deux chambres supérieures sont l'oreillette droite et l'oreillette gauche. Elles acceptent le sang qui entre dans le cœur. La paroi qui les sépare s'appelle le septum inter auriculaire [2].

Pour le côté droit du cœur, il est responsable au pompage du sang vers les poumons, où il est enrichi d'oxygène et épuré du dioxyde de carbone. Le deuxième côté pompe le sang vers le reste du corps où l'oxygène et les substances nutritives sont fournis aux tissus et les déchets comme le dioxyde de carbone sont transférés au sang pour être éliminés par d'autres organes comme les poumons et les reins (Voir la figure 1.2) [2].



**Figure 1.2 :** Anatomie du cœur.

### 1.3.1 Ventricule droit

Le ventricule droit (VD) est la chambre inférieure droite du cœur qui a une forme conique et il est plus petit par rapport au ventricule gauche (VG). Leurs dimensions sont variées en fonction de l'âge, du sexe et de la taille de la personne. Chez un adulte en bonne santé, la taille moyenne du VD est d'environ 3.5cm de diamètre et 5cm de longueur.

Le VD reçoit le sang pauvre en oxygène de l'oreillette droite via la valvule tricuspide (valvule d'entrée) et le pompe vers les artères pulmonaires afin d'être réoxygéné dans les poumons par la valvule pulmonaire (valvule de sortie). Sa fonction dépend du système nerveux autonome et les hormones afin de s'adapter aux besoins du corps en oxygène [3].

### 1.3.2 Ventricule gauche

Le ventricule gauche est situé dans la partie inférieure gauche du cœur entouré par les autres structures cardiaques et il est plus gros que le VD avec une taille d'environ 4.5 cm de diamètre et 8cm de longueur chez un adulte en bonne santé. Il est divisé en trois parties principales : le muscle papillaire qui est responsable de la régulation de la pression sanguine dans le VG, le septum interventriculaire qui sépare le VD et le VG et la valvule mitrale qui permet au sang de circuler du VG vers l'aorte qui est la plus grande artère du corps qui transporte le sang oxygéné vers toutes les parties du corps. Le VG reçoit le sang oxygéné provenant des poumons par l'intermédiaire de l'oreillette gauche. Il se contracte ensuite pour pomper le sang dans l'aorte. Cette contraction du ventricule est déclenchée par un signal électrique qui vient du nœud sinusal [3].

### 1.3.3 Oreillette droite

L'oreillette droite (OD) est située dans la partie supérieure droite du cœur reliée au VD par une valve tricuspide, avec un format ovale et mesure environ 3 à 4 centimètres de longueur.

L'intérieur de l'OD est tapissé d'une membrane appelée endocarde qui est composée de cellules endothéliales et de tissu conjonctif. Cette membrane permet aux valves cardiaques de s'ouvrir et de se fermer correctement pour réguler le flux sanguin. Elle a un rôle important dans la circulation sanguine en pompant le sang veineux vers le ventricule droit qui est riche en dioxyde de carbone et pauvre en oxygène, et provient de toutes les parties du corps.

Une fois que le sang veineux entre dans OD, il est poussé à travers la valve tricuspide et dans le VD. De là, il est pompé vers les poumons où il est oxygéné avant le revenir au cœur pour être distribué dans le corps [3].

### 1.3.4 Oreillette gauche

L'oreillette gauche est une petite chambre à paroi mince qui mesure environ 4 cm sur 3cm.

Elle est isolée du VG par la valve mitrale, également appelée valve auriculo-ventriculaire gauche. OG est reliée aux veines pulmonaires, qui transportent le sang oxygéné des poumons vers le cœur. Elle fonctionne comme une pompe qui permet de remplir le VG avec du sang oxygéné. Lorsque le cœur se contracte, les muscles de l'OG se contractent également, ce qui fait que le sang est poussé dans le VG. Une fois le contenu est vidé, il se remplit à nouveau avec du sang oxygéné provenant des poumons [3].

## 1.4 Circulation sanguine

Les côtés droit et gauche du cœur travaillent ensemble pour assurer la circulation du sang dans tout le corps grâce à une série de systoles (contraction) et diastoles (relâchement) des oreillettes et ventricules. Ces étapes se déroulent en l'espace d'un battement de cœur, c'est-à-dire en une ou deux secondes [4].

### 1.4.1 Circulation sanguine côté droit

La circulation sanguine côté droit s'effectue comme suit [4] :

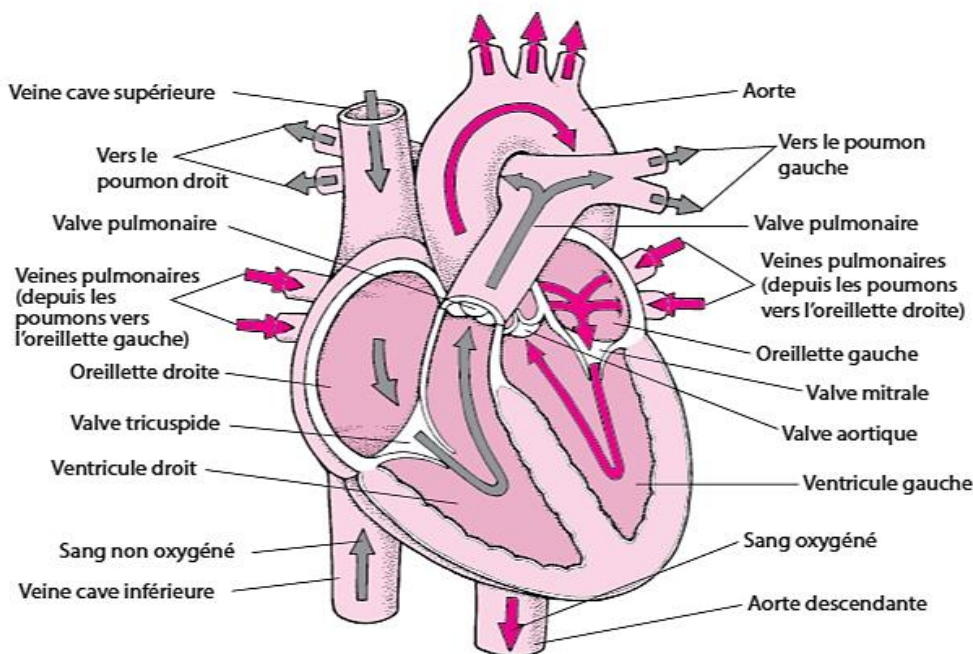
- Le sang pauvre en oxygène provenant de l'ensemble du corps pénètre dans l'oreillette droite par deux grandes veines inférieure et supérieure.
- La valve tricuspide s'ouvre pour permettre au sang de passer de LD au VD.
- Lorsque le VD est plein, il se comprime, ce qui ferme la valve tricuspide et ouvre la valve pulmonaire.
- Le sang circule dans l'artère pulmonaire jusqu'aux poumons, où il reçoit de l'oxygène.

## 1.4.2 Circulation sanguine côté gauche

La circulation sanguine côté gauche s'effectue comme suit [4] :

- Le sang riche en oxygène passe des poumons à LG par de grosses veines appelées veines pulmonaires.
- La valve mitrale s'ouvre pour envoyer le sang de LG vers le VG.
- Lorsque le VG est plein, il se comprime, ce qui ferme la valve mitrale et ouvre la valve aortique.
- Le cœur envoie le sang par la valve aortique vers l'aorte, où il s'écoule vers le reste du corps.

La figure 1.3 montre la circulation sanguine du cœur :



**Figure 1.3 :** Circulation sanguine.

## 1.5 Le battement du cœur

Un signal électrique est généré par le nœud sinusal (nommé aussi nœud sino-auriculaire ou nœud SA). Il s'agit d'une petite masse de tissu spécialisé située dans la chambre supérieure droite (oreillettes) du cœur. Le nœud sinusal génère un courant électrique régulièrement.

Les oreillettes sont alors activées. Le message électrique descend par les voies de conduction et provoque la contraction des ventricules du cœur et l'expulsion du sang. Les deux cavités supérieures du cœur (oreillettes) sont activées en premier et se contractent pendant un court intervalle de temps avant les deux cavités inférieures du cœur (ventricules).

L'impulsion électrique passe du nœud sinusal au nœud auriculo-ventriculaire (également appelé nœud AV). Là les impulsions sont ralenties pendant une très courte période, puis poursuivent leur chemin dans la voie de conduction via le faisceau de His jusqu'aux ventricules. Le faisceau de His se divise en deux voies droite et gauche, appelées branches du faisceau, pour activer les ventricules droite et gauche.

Normalement, au repos, lorsque l'impulsion électrique se déplace dans le cœur, ce dernier se contracte environ 60 à 100 fois par minute, en fonction de l'âge de la personne.

Chaque contraction des ventricules représente un battement de cœur. Les oreillettes se contractent une fraction de seconde avant les ventricules, de sorte que leur sang se décharge dans les ventricules avant que leur contraction se produise (voir la figure 1.4) [5].

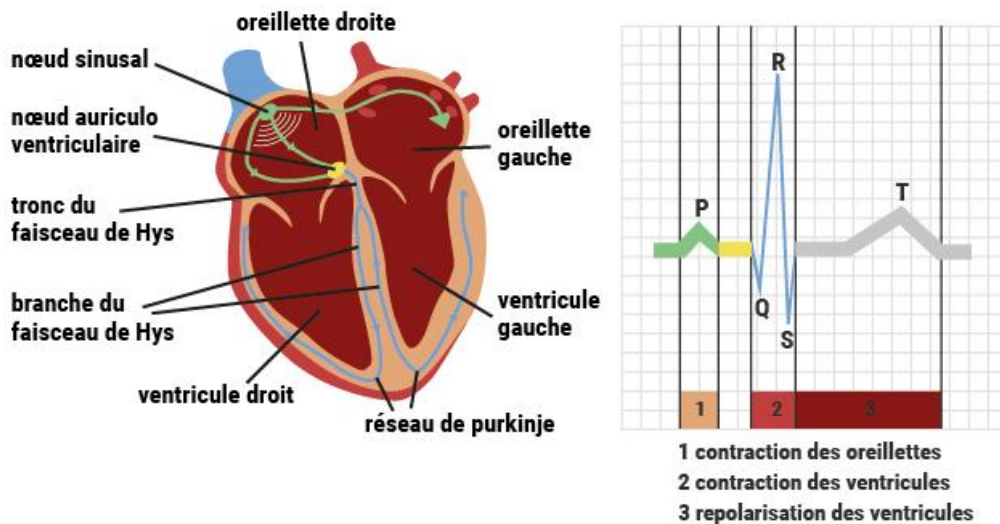


Figure 1.4 : La conduction électrique dans le cœur.

## 1.6 Les maladies cardiovasculaires

Les maladies cardiovasculaires (MCV) sont la première cause de mortalité dans le monde, avec environ 17.9 millions de décès par an [6]. Il existe de nombreux types de MCV, et chacun d'entre eux présente ses propres symptômes et traitement. Pour certains, des changements de mode de vie et des médicaments peuvent faire une grande différence dans l'amélioration de leur état de santé, et pour d'autres une intervention chirurgicale peut s'avérer nécessaire pour établir le bon fonctionnement du cœur [7]. Dans ce qui suit, nous aborderons les maladies cardiaques les plus courantes.

### 1.6.1 Les maladies vasculaires

Les maladies vasculaires ce sont des affections qui touchent le système circulatoire, notamment les artères et les veines. Elles peuvent être causées par une accumulation de plaque dans les vaisseaux sanguins, une inflammation ou une blessure. Ces maladies peuvent



avoir des conséquences graves sur la santé humaine comme l'infarctus du myocarde, l'accident vasculaire cérébral et l'insuffisance rénale. Les facteurs de risque incluent l'hypertension, le tabagisme, le diabète et l'obésité et une alimentation malsaine.

### **1.6.2 Les maladies coronariennes**

Ce sont un terme générique désignant une série d'affections qui influent sur la structure et le fonctionnement du cœur. Dans ce type de maladie, les artères du sang ne parviennent pas à fournir suffisamment de sang riche en oxygène au cœur. La cause de cette maladie dépend du type de la maladie. La coronaropathie est souvent causée par le cholestérol, une substance cireuse qui s'accumule à l'intérieur de la paroi des artères coronaires forme une plaque. Cette accumulation peut bloquer partiellement ou totalement la circulation sanguine dans les grandes artères du cœur. La maladie micro vasculaire coronarienne survient lorsque les minuscules vaisseaux sanguins du cœur ne fonctionnent pas normalement. Pour la plupart des gens, cette maladie peut être évitée grâce à un mode de vie favorable à la santé cardiaque [8].

### **1.6.3 L'insuffisance cardiaque**

L'insuffisance cardiaque est une affection de longue durée dans laquelle le cœur ne parvient pas à pomper le sang suffisamment bien pour répondre aux besoins de l'organisme. Le cœur fonctionne toujours, mais comme il ne peut pas gérer la quantité de sang qu'il en a besoin, le sang s'accumule dans d'autres parties de votre corps, la plupart du temps, il s'accumule dans les poumons, les jambes et les pieds. Cette maladie ne peut pas être soignée. Lorsque l'insuffisance cardiaque s'aggrave, votre muscle cardiaque pompe moins de sang vers vos organes et vous passez au stade suivant de l'insuffisance cardiaque, l'objectif du traitement est de vous empêcher d'avancer dans les niveaux ou de ralentir la maladie [4].

### **1.6.4 L'angine de poitrine**

L'angine de poitrine est une douleur thoracique qui résulte d'un manque de sang dans une partie du cœur. Elle peut ressembler à une crise cardiaque avec une pression ou un écrasement dans la poitrine. On le nomme souvent angine de poitrine ou douleur thoracique ischémique.

C'est un signe de maladie cardiaque qui se produit lorsque quelque chose bloque vos artères ou que le flux sanguin est insuffisant dans les artères qui amènent le sang riche en oxygène à votre cœur. L'angine de poitrine disparaît généralement rapidement. Cependant, elle peut être le signe d'un problème cardiaque potentiellement mortel. Il est important de savoir ce qui arrive et ce que vous pouvez faire pour éviter une crise cardiaque. En générale, les médicaments et les changements de mode de vie permettent de contrôler cette maladie. Si l'angine est plus grave, une intervention chirurgicale peut être nécessaire. Il est également possible que vous deviez porter un stent, un petit tube qui permet d'ouvrir les artères [7].

### **1.6.5 L'infarctus du myocarde**

L'infarctus du myocarde (IM) couramment appelé "crise cardiaque" est causé par une diminution ou un arrêt complet de la circulation sanguine dans une partie du myocarde. IM peut être silencieux et passer inaperçu, ou être un élément catastrophique conduisant à une diminution de l'hémodynamique et à une mort rapide. La plupart des IM sont dus à une

maladie coronarienne fondamentale. En cas de blocage de l'artère coronaire, le myocarde est privé d'oxygène. Une privation prolongée de l'apport d'oxygène au myocarde peut entraîner la mort des cellules myocardiques. Les patients peuvent présenter une douleur ou une pression thoracique qui peut remonter jusqu'au cou, à la mâchoire, à l'épaule ou au bras. De plus, l'ischémie myocardique peut être associée à des modifications de l'ECG et à une augmentation des indicateurs biochimiques tels que les troponines cardiaques [9].

### **1.6.6 Asthme cardiaque**

L'asthme cardiaque est différent de l'asthme bronchique (AB), mais il provoque des symptômes similaires. AB est déclenché par des allergies, des polluants, l'exercice physique, le stress ou une maladie pulmonaire. Les petites voies respiratoires des poumons sont irritées et enflammées à cause de ces déclencheurs. Cette inflammation se traduit par une respiration sifflante, une toux et un essoufflement. L'asthme cardiaque peut provoquer des symptômes similaires, mais il est causé par une accumulation de liquide dans le côté gauche du cœur. Cette accumulation de liquide peut être le résultat d'un cœur qui pompe faiblement, d'une valve qui fuit ou d'une malformation cardiaque présente à la naissance. Ces symptômes sont le signe d'une insuffisance cardiaque congestive. Les traitements dépendent de la cause (insuffisance cardiaque ou fuite de la valve) mais peuvent inclure des médicaments pour contrôler la pression artérielle et éliminer l'excès de liquide, un régime alimentaire approprié et des activités quotidiennes modifiées. Si la cause est une valve qui fuit ou une malformation cardiaque génétique, une intervention chirurgicale peut s'avérer nécessaire avec le temps [10].

### **1.6.7 L'arrêt cardiaque**

L'arrêt cardiaque soudain se produit soudainement et souvent sans avertissement. Il se produit lorsqu'un dysfonctionnement électrique du cœur provoque un rythme cardiaque irrégulier (arythmie). Son action de pompage étant perturbée, le cœur ne peut plus pomper le sang vers le cerveau, les poumons et les autres organes. Dans ce cas, la personne perd la connaissance et n'a plus de souffle. La mort se produit en quelques minutes si la victime n'est pas soignée [11].

## **1.7 Différence entre arrêt cardiaque, crise cardiaque et l'insuffisance cardiaque**

### **1.7.1 L'arrêt cardiaque**

On parle d'arrêt cardiaque lorsque le cœur cesse d'envoyer le sang et l'oxygène vers le cerveau et les autres organes et tissus. Il arrive parfois qu'une personne puisse être réanimée après un arrêt cardiaque, surtout si le traitement est entrepris immédiatement.

Cependant, plus le temps passe sans que du sang riche en oxygène soit envoyé au cerveau, moins il est probable que la personne puisse être réanimée et, si elle l'est, plus il est probable qu'elle ait des lésions cérébrales.

Des lésions cérébrales sont probables si l'arrêt cardiaque dure plus de 5 minutes sans l'intervention de premiers secours que constitue la réanimation cardiopulmonaire (RCP). Le décès est probable si l'arrêt cardiaque dure plus de 8 minutes. Ainsi, la RCP en cas d'arrêt cardiaque doit débiter le plus vite possible.

L'arrêt cardiaque peut être causé par tout ce qui fait que le cœur s'arrête de battre. L'une des causes fréquentes, surtout chez les adultes, est une anomalie du rythme cardiaque (arythmie). Une autre cause possible est l'arrêt respiratoire qui se produit lorsque les personnes se noient, sont atteintes d'une infection pulmonaire grave ou font une crise d'asthme sévère [2].

### 1.7.2 La crise cardiaque

Une crise cardiaque, également appelée infarctus du myocarde (IM), est une urgence médicale qui survient lorsque le flux sanguin vers une partie du muscle cardiaque est bloqué, généralement par un caillot de sang. Ce blocage prive le muscle cardiaque d'oxygène et de nutriments, entraînant la mort des cellules de la zone touchée. Les crises cardiaques provoquent généralement une douleur ou une gêne thoracique, ainsi que d'autres symptômes tels qu'un essoufflement, des nausées, des sueurs et des vertiges. Une prise en charge médicale rapide est essentielle pour minimiser les dommages au cœur et améliorer les chances de survie.

Les complications d'une crise cardiaque sont souvent dues à des lésions du muscle cardiaque, parmi ces complications [12] :

**Rythmes cardiaques irréguliers ou atypiques (arythmies) :** Les dommages causés par la crise cardiaque peuvent affecter la façon dont les signaux électriques circulent dans le cœur, entraînant des modifications du rythme cardiaque. Certaines peuvent être graves, voire mortelles.

**Choc cardiogénique :** Cette affection rare survient lorsque le cœur est soudainement et brusquement incapable de pomper le sang.

**Insuffisance cardiaque :** Des lésions importantes du tissu musculaire cardiaque peuvent rendre le cœur incapable de pomper le sang. L'insuffisance cardiaque peut être temporaire ou durable (chronique).

**Inflammation du tissu entourant le cœur (péricardite) :** Parfois, une crise cardiaque déclenche une réponse déficiente du système immunitaire. Cette affection peut être appelée syndrome de Dressler, syndrome post-infarctus du myocarde ou syndrome post-lésion cardiaque.

### 1.7.3 L'insuffisance cardiaque

L'insuffisance cardiaque est un trouble dans lequel le cœur est incapable de faire face aux demandes du corps, ce qui entraîne une réduction du flux sanguin, un retour (congestion) de sang dans les veines et les poumons et/ou d'autres changements susceptibles d'affaiblir ou de rigidifier davantage le cœur.

Une personne qui est atteinte d'insuffisance cardiaque présentera, dans certains cas, des symptômes d'essoufflement, de fatigue, d'œdème des pieds et des chevilles, ou de ballonnements. Même si on ne peut guérir l'insuffisance cardiaque, un diagnostic précoce, une modification des habitudes de vie et un traitement médicamenteux adapté aux besoins aident le patient à conserver un mode de vie actif [12].

## 1.8 Symptômes des maladies cardiaques

Il n'existe pas un symptôme très précis qui indique clairement un trouble cardiaque, mais un ensemble de symptômes qui permettent de prédire un diagnostic presque certain. Les médecins interrogeant la personne de ses antécédents médicaux, et pratiquent un examen clinique pour déduire les symptômes exacts. Cependant, une maladie cardiaque peut rester asymptomatique tant qu'elle n'a pas atteint une phase avancée. Des bilans biologiques et/ou radiologiques sont souvent réalisés pour confirmer le diagnostic et détecter une maladie cardiaque asymptomatique [2].

Les différents symptômes d'une maladie cardiaque sont [2] :

**La douleur thoracique :** est un trouble très fréquent, elle peut être aiguë ou atténuée. Certaines personnes peuvent ressentir une sensation de gêne, d'oppression, de pression, de gaz, ou de brûlure au niveau du thorax. Il arrive parfois que la douleur est ressentie dans le dos, la mâchoire, le cou, la partie supérieure de l'abdomen ou le bras. Autres symptômes peuvent accompagner cette douleur, comme les nausées, la toux, ou une difficulté à respirer.

**L'essoufflement (dyspnée comme un terme médical) :** est une sensation pénible et des difficultés à respirer. La fréquence et l'intensité de la respiration augmentent lors des activités physiques et aux altitudes élevées. La fréquence respiratoire peut également augmenter au repos chez des personnes atteintes de différentes maladies, comme les maladies pulmonaires ou autres ; par exemple, une fièvre peut faire augmenter la fréquence respiratoire.

**La fatigue :** est un besoin important de se reposer, ressenti par les personnes qui ont une énergie effondrée pour démarrer une activité et la poursuivre. Une fatigue après un effort physique, un stress prolongé et une privation de sommeil est considérée comme normale. Néanmoins, une fatigue qui augmente progressivement après des activités qui ne causaient pas un tel symptôme auparavant, peut être un symptôme révélateur d'un trouble cardiaque.

**Les palpitations :** sont la perception des battements cardiaques. C'est la sensation de vibrations, de coups, de course ou d'irrégularité dans les battements. Elles peuvent accompagner autres symptômes, comme un malaise thoracique ou un essoufflement.

**L'évanouissement (syncope comme un terme médical) :** est une perte de conscience brève et soudaine, durant laquelle, la personne tombe par terre puis redevient consciente. La personne peut avoir des signes avant de s'évanouir, comme un pouls faible, une respiration profonde, des nausées, des sueurs, une vision floue, des picotements des lèvres ou du bout des doigts, une douleur thoracique, ou des palpitations. Comme on peut s'évanouir sans aucun symptôme avertisseur.

**Le gonflement (œdème comme un terme médical) :** est un excès de liquide dans l'organisme qu'il s'agit principalement d'eau. Ce gonflement touche souvent les pieds et le bas des jambes. En général, le gonflement se produit lentement, et commence par une prise de poids, des yeux bouffis au lever le matin, et des chaussures serrées en fin de journée. Le patient peut ne pas remarquer ce gonflement, que quand il est considérable.

## 1.9 Diagnostic des maladies cardiaques

Quand un patient se présente chez le médecin, ce dernier réalise un examen clinique très minutieux, qui comporte un interrogatoire sur les antécédents médicaux du patient et examen physique. Parfois, l'examen clinique seul ne permet pas de poser un diagnostic certain d'une maladie cardiaque ou vasculaire, ainsi, des procédures diagnostiques spéciales sont souvent nécessaires pour confirmer le diagnostic, préciser l'extension et la gravité de la maladie, et contribuer au choix du traitement [2].

### 1.9.1 Tomodensitométrie (TDM) du cœur

Une tomodensitométrie (TDM) peut être utilisée pour détecter des anomalies structurelles du cœur, du sac qui l'enveloppe (péricarde), des vaisseaux sanguins principaux, des poumons et des structures de soutien dans le thorax.

La TDM multi détecteurs est un appareil de TDM très rapide qui peut prendre un cliché durant un seul battement cardiaque. Cette tomodensitométrie rapide (appelée angiographie coronarienne par tomodensitométrie) est parfois utilisée pour évaluer les artères qui irriguent le cœur (artères coronaires). Un agent de contraste (une substance visible sur les radiographies) est généralement injecté dans la veine du patient. On demande au patient de ne pas respirer pendant la procédure pour éviter que l'image ne soit floue [2].

### 1.9.2 Les analyses du sang

Les analyses de sang sont utiles pour dépister l'insuffisance cardiaque et rechercher des causes ou facteurs déclenchants possibles. De hauts niveaux de BNP (Brain Natriuretic Peptide) dans le sang ont été associés à l'insuffisance cardiaque et ce dosage est utile à la fois pour le diagnostic et la prise en charge thérapeutique. Le niveau de BNP ou de NT-proBNP (N-Terminal proBNP) est aujourd'hui mesuré systématiquement pour établir un diagnostic et déterminer la gravité de l'insuffisance cardiaque. De plus, les résultats sanguins sont obtenus rapidement. Un BNP normal chez un patient non traité pour insuffisance cardiaque laisse penser que les symptômes sont probablement dus à d'autres causes [13].

### 1.9.3 L'électrocardiographie

L'ECG est le résultat final d'une série complexe de processus physiologiques et technologiques.

Tout d'abord, des courants ioniques transmembranaires sont générés par des flux d'ions à travers les membranes cellulaires et entre les cellules adjacentes. Ces courants sont synchronisés par les séquences d'activation et de récupération cardiaques pour générer un fil électrique cardiaque à l'intérieur et autour du cœur, qui varie dans le temps au cours de l'activité cardiaque. Ce champ électrique passe à travers de nombreuses autres structures, notamment les poumons, le sang et les muscles squelettiques, qui perturbent le fil électrique cardiaque.

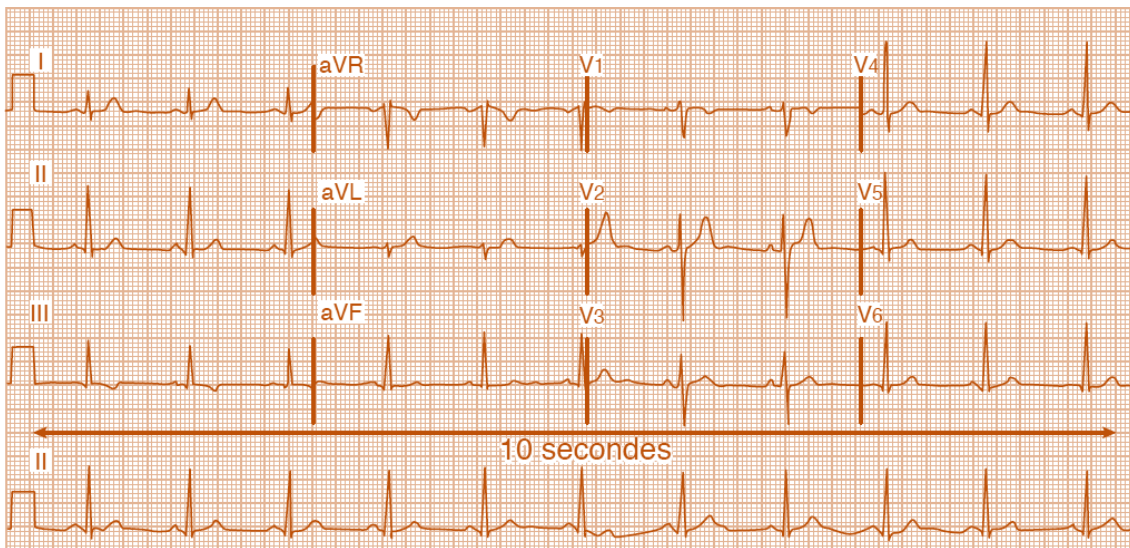
Les courants qui atteignent la peau sont ensuite détectés par des électrodes placées à des endroits spécifiques des extrémités et du torse, qui sont configurées pour produire des électrodes. Les sorties de ces électrodes sont amplifiées, filtrées, et affichées par divers

appareils pour produire un enregistrement électrocardiographique. Dans les systèmes informatisés, ces signaux sont numérisés, stockés, et traités par un logiciel de reconnaissance des formes.

Les critères de diagnostic sont ensuite appliqués, soit manuellement, soit à l'aide d'un ordinateur, afin de produire une interprétation.

Les tracés ECG enregistrés ou affichés sont enfin comparés à divers critères de diagnostic afin d'identifier des anomalies spécifiques. Dans certains cas, les critères électrocardiographiques sont dérivés de concepts physiologiques et constituent la seule base d'un diagnostic sans corrélation anatomique ou physiologique [14].

La figure 1.5 définit un exemple d'un ECG :



**Figure 1.5 :** Électrocardiogramme normal.

#### 1.9.4 Echocardiographie

L'échocardiographie est l'un des tests diagnostiques non invasifs les plus couramment présentant des maladies cardiovasculaires avérées ou suspectées (voir la figure 1.6).

L'échocardiographie permet une évaluation complète de la structure, de la fonction et de l'hémodynamique cardiovasculaires qui caractérisent les processus pathologiques. En outre, il n'y a pas d'effets secondaires connus associés à l'échocardiographie.

De plus, l'échocardiographie n'a pas d'effets secondaires connus, même en cas d'examen fréquents et répétés. Sa nature en temps réel, sa portabilité et son coût relativement faible rendent l'échocardiographie adaptable à la plupart des situations cliniques ou de recherche. L'échocardiographie a donc été utilisée avec succès pour fournir des informations mécanistes sur les résultats thérapeutiques et, dans certains cas, pour mesurer les changements fonctionnels et structurels considérés comme importants sur le plan thérapeutique [15].

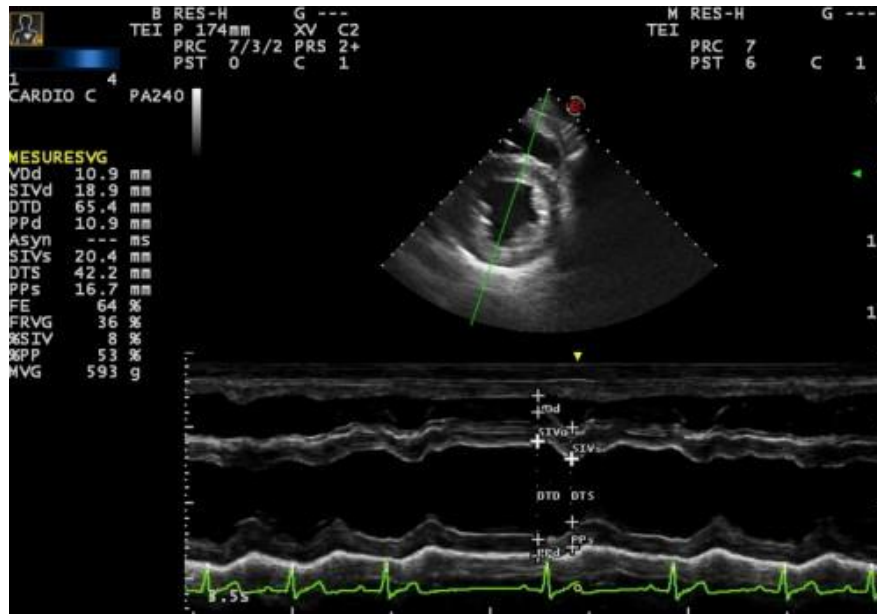


Figure 1.6 : Echocardiographie du cœur.

### 1.9.5 Epreuve d'efforts

L'entraînement à l'effort, avec la prévention secondaire, est la pierre angulaire des programmes de réadaptation pour les maladies cardiovasculaires athéromateuses. Il réduit la morbidité et la mortalité liées à la maladie, diminue le degré d'incapacité et joue un rôle dans le contrôle des facteurs de risque. Elle a un impact direct sur la réinsertion socioprofessionnelle. Il s'agit d'une thérapie à part entière dont l'approche (types d'exercices, intensité, fréquence, durée) doit être adaptée au patient. Compte tenu du poids financier des maladies cardiovasculaires sur le système de santé et de leur prévalence croissante, des stratégies doivent être mises en place pour augmenter le niveau d'activité physique de cette population toujours plus nombreuse. Les arguments en faveur de la mise en œuvre de telles stratégies dans le cadre de la prévention primaire sont également forts [16].

### 1.9.6 Imagerie par résonance magnétique (IRM) du cœur

Grâce à une bonne résolution spatiale et temporelle associée à une approche tridimensionnelle, et à un excellent contraste entre le sang circulant et le myocarde, l'IRM cardiaque constitue une méthode de référence pour l'imagerie des cardiopathies congénitales, des tumeurs cardiaques, des gros vaisseaux et du péricarde. La détermination par ciné-IRM des volumes ventriculaires, de la fraction d'éjection et de la masse myocardique, de la cinétique segmentaire est standardisée, précise et reproductible.

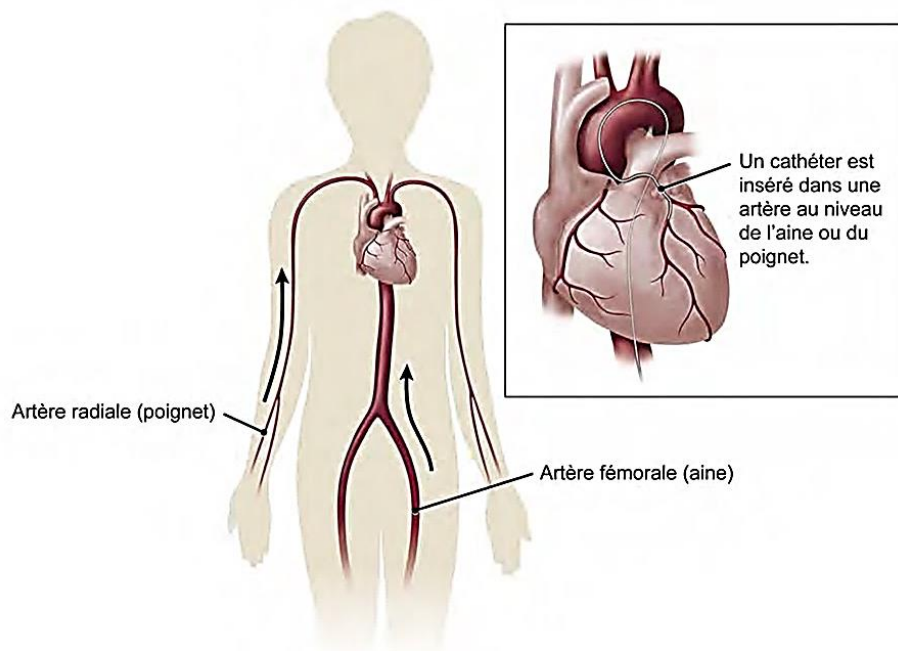
L'IRM de contraste après injection de gadolinium représente un apport diagnostique considérable pour de nombreuses pathologies, qu'elles soient tumorales, inflammatoires, ischémiques, ou pour préciser l'étiologie d'une cardiomyopathie. Dans les cardiopathies

ischémiques, force est de constater que l'IRM est devenue incontestablement la méthode de référence dans l'analyse de la viabilité myocardique.

L'avantage majeur de l'IRM est de pouvoir associer l'étude de la viabilité myocardique à l'analyse de la perfusion myocardique et à une approche fonctionnelle du muscle cardiaque par l'étude des fonctions ventriculaires droite et gauche au repos et lors d'épreuves de provocation d'ischémie myocardique [17].

### 1.9.7 Angiographie coronarienne et cathétérisme cardiaque

Le cathétérisme cardiaque et l'angiographie coronaire sont des méthodes minimalement invasives pour étudier le cœur et les vaisseaux sanguins qui irriguent le cœur (artères coronaires) sans réaliser de chirurgie. Ces tests sont généralement réalisés lorsque les tests non invasifs ne fournissent pas suffisamment d'informations, lorsque les tests non invasifs suggèrent qu'il y a un problème cardiaque ou vasculaire, ou lorsqu'une personne présente des symptômes qui rendent un problème cardiaque ou coronarien très probable. Un avantage de ces tests est que pendant ceux-ci, les médecins peuvent aussi traiter différentes maladies, y compris une maladie des artères coronaires, comme montre la figure 1.7 [2].



**Figure 1.7 :** Cathétérisme cardiaque et coronarographie.

Comme pour toute intervention, la décision de recommander un cathétérisme cardiaque repose sur un rapport risque/bénéfice approprié. En général, le cathétérisme cardiaque diagnostique est recommandé chaque fois qu'il est cliniquement important de définir la présence ou la gravité d'une lésion cardiaque présumée qui ne peut pas être évaluée de manière adéquate par des techniques non invasives.



Le risque de complication majeure lié au cathétérisme cardiaque étant inférieur à 1 % et le taux de mortalité inférieur à 0,08 %, rares sont les patients qui ne peuvent être étudiés en toute sécurité dans un laboratoire actif. Les mesures de la pression intracardiaque et l'artériographie coronarienne sont des procédures qui peuvent être réalisées avec une précision reproductible au mieux par cathétérisme invasif. L'estimation non invasive des pressions intracardiaques peut être avec l'échocardiographie.

L'angiographie coronaire par tomodensitométrie (CT) peut également être utilisée pour évaluer l'anatomie coronaire et fournit des informations complémentaires sur la distribution et la composition de la plaque. Cependant, les limites actuelles de la résolution spatiale, de la variabilité de la fréquence cardiaque, de la coopération du patient, et le dosage des radiations limitent la capacité de la tomodensitométrie à remplacer le cathétérisme cardiaque [18].

### 1.10 Facteurs de risques

Les facteurs de risques d'atteinte des maladies cardiaques sont les suivants [19] :

**L'âge :** Les hommes de 45 ans et plus et les femmes de 55 ans et plus sont plus susceptibles d'avoir une crise cardiaque que les hommes et les femmes plus jeunes.

**Le tabagisme :** Il s'agit du tabagisme et de l'exposition prolongée au tabagisme passif. Si vous fumez, arrêtez.

**L'hypertension artérielle :** Avec le temps, l'hypertension artérielle peut endommager les artères qui mènent au cœur. L'hypertension artérielle associée à d'autres pathologies, telles que l'obésité, l'hypercholestérolémie ou le diabète, augmente encore le risque.

**Cholestérol ou triglycérides élevés :** Un taux élevé de cholestérol à lipoprotéines de faible densité (LDL) (le "mauvais" cholestérol) est le plus susceptible de rétrécir les artères. Un taux élevé de certaines graisses sanguines appelées triglycérides augmente également le risque de crise cardiaque. Le risque de crise cardiaque peut diminuer si le taux de cholestérol à lipoprotéines de haute densité (HDL) - le "bon" cholestérol - se situe dans la fourchette standard.

**L'obésité :** L'obésité est liée à l'hypertension artérielle, au diabète, à des taux élevés de triglycérides et de mauvais cholestérol, et à des taux faibles de bon cholestérol.

**Le diabète :** Le taux de sucre dans le sang augmente lorsque l'organisme ne produit pas une hormone appelée insuline ou ne peut pas l'utiliser correctement. L'hyperglycémie augmente le risque de crise cardiaque.

**Le syndrome métabolique :** Il s'agit d'une combinaison d'au moins trois des éléments suivants : tour de taille (obésité centrale), hypertension artérielle, faible taux de bon cholestérol, taux élevé de triglycérides et taux élevé de sucre dans le sang. Le syndrome métabolique multiplie par deux le risque de développer une maladie cardiaque par rapport à une personne qui n'en est pas atteinte.

**Antécédents familiaux de crises cardiaques :** Si un frère, une sœur, un parent ou un grand-parent a subi une crise cardiaque précoce (avant l'âge de 55 ans pour les hommes et avant l'âge de 65 ans pour les femmes), vous pourriez présenter un risque accru.

**Manque d'exercice :** Le manque d'activité physique (sédentarité) est lié à un risque accru de crise cardiaque. L'exercice physique régulier améliore la santé cardiaque.

**Une alimentation malsaine :** Une alimentation riche en sucres, en graisses animales, en aliments transformés, en graisses trans et en sel augmente le risque de crise cardiaque. Mangez beaucoup de fruits, de légumes, de fibres et d'huiles saines.

**Le stress :** Le stress émotionnel, tel que la colère extrême, peut augmenter le risque de crise cardiaque.

**Consommation de drogues illégales :** La cocaïne et les amphétamines sont des stimulants. Elles peuvent déclencher un spasme de l'artère coronaire susceptible de provoquer une crise cardiaque.

**Antécédents familiaux cardiovasculaires :** Quand un des parents du premier degré a fait un IDM (infarctus du myocarde), un AVC (accident vasculaire cérébral) ou une mort subite de cause cardiaque avant l'âge de 50 ans, ceci constitue un grand facteur de risque que les enfants soient atteints d'une affection cardiovasculaire.

**Une maladie auto-immune :** Une affection telle que la polyarthrite rhumatoïde ou le lupus peut augmenter le risque de crise cardiaque.

## 1.11 Prévention cardiovasculaire

Malgré une diminution régulière depuis une trentaine d'années du nombre de décès liés à un événement cardiovasculaire (CV), ces maladies restent l'une des principales causes de mortalité dans le monde. L'objectif de la prévention CV est de diminuer la morbi-mortalité liée aux événements cardiovasculaires en déterminant des seuils d'intervention thérapeutique.

La prévention primaire s'adresse aux patients jusque-là indemnes de pathologie cardiovasculaire symptomatique. Le dépistage, la prise en charge des FDR CV ont pour but de retarder voire d'éviter la survenue d'événement cardiovasculaire. Pour cela, il faut avoir recours au calcul de risque cardiovasculaire global.

La prévention secondaire s'adresse aux patients ayant déjà une atteinte CV cliniquement exprimée : angor, IDM, AVC, claudication intermittente... Elle cherche à diminuer le risque de récurrences, de complication et de décès [20].

### 1.11.1 Prévenir une crise cardiaque

Il n'est jamais trop tard pour prendre des mesures afin de prévenir une crise cardiaque, même si vous en avez déjà eu une. Quelques moyens de prévention sont [16] :

- Adoptez un mode de vie sain. Ne fumez pas. Maintenez un poids santé en adoptant un régime alimentaire sain pour le cœur. Faites régulièrement de l'exercice et gérez votre stress.
- Gérer les autres problèmes de santé. Certaines maladies, comme l'hypertension artérielle et le diabète, peuvent augmenter le risque de crise cardiaque. Demandez à votre prestataire de soins de santé à quelle fréquence vous devez-vous faire examiner.
- Prenez vos médicaments comme indiqué. Votre médecin peut vous prescrire des médicaments pour protéger et améliorer votre santé cardiaque.
- Il est également conseillé d'apprendre à pratiquer la réanimation cardio-pulmonaire afin de pouvoir aider une personne victime d'une crise cardiaque. Envisagez de suivre une formation accréditée aux premiers secours, comprenant la réanimation cardiopulmonaire (RCP) et l'utilisation d'un défibrillateur externe automatisé (DEA).

### 1.12 Conclusion

Dans ce premier chapitre, nous avons défini et expliqué les maladies cardiaques qui sont progressivement devenues un problème commun de santé publique dans le monde entier, en raison d'une méconnaissance, d'une consommation malsaine et d'un mode de vie médiocre. Aujourd'hui, la prédiction et le diagnostic précis de ces maladies constituent un défi majeur pour les hôpitaux et les praticiens. L'intelligence artificielle dans le domaine de la santé aide les utilisateurs à comprendre l'efficacité des programmes existants et à identifier le traitement qui donne les meilleurs résultats pour les patients en fonction de leur état de santé.

Le développement des technologies informatiques a aidé les établissements de soins de santé à collecter et à stocker des données pour fonder la prise de décision clinique. Le Data Mining transforme les énormes ensembles de données en informations qui sont ensuite utilisées pour faire de meilleures prédictions et prendre de meilleures décisions.

Le chapitre suivant consiste à découvrir le domaine d'intelligence artificielle et expliquer ses deux sous systèmes d'apprentissage, le Machine Learning et le Deep Learning avec leurs différents algorithmes et la différence entre eux.

## **Chapitre 2**

# **Machine Learning & Deep Learning**

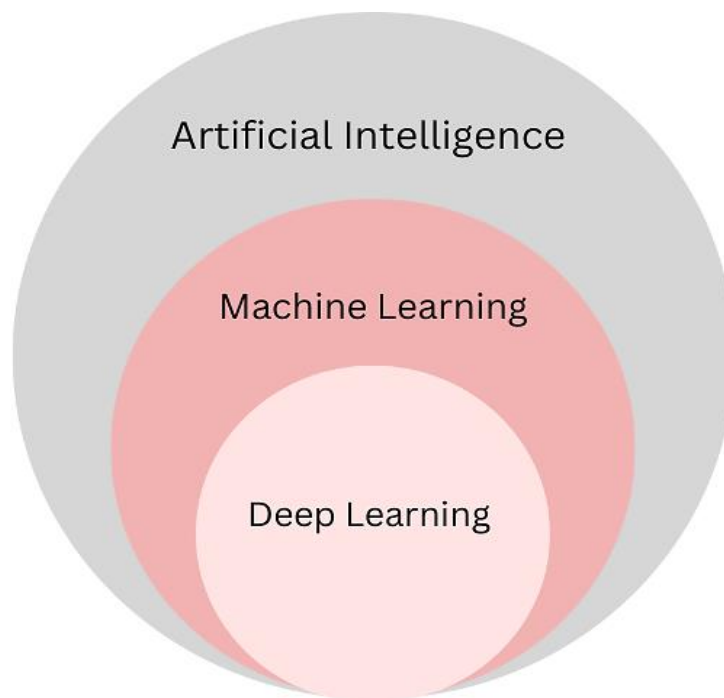
## 2.1 Introduction

L'intelligence artificielle (IA) fait partie de la mise au point de machines intelligentes capables d'effectuer des tâches qui font généralement appel à l'intelligence humaine. Son objectif est de permettre aux automates de comprendre, d'assimiler, de raisonner et de trouver des solutions à divers problèmes de manière automatique, sans nécessiter d'intervention humaine. Elle peut être appliquée dans de vastes domaines d'activité tels que la médecine, la finance, l'agriculture, l'industrie, etc.

L'histoire de l'IA remonte aux années 1950, lorsque des chercheurs ont entamé des travaux sur la modélisation de l'intelligence humaine, et le terme a été inventé en 1955 par le mathématicien américain "John MacCarthy". Depuis lors, les progrès ont été rapides et continus, et l'IA a connu de nombreuses avancées remarquables ces dernières années.

Les technologies de l'IA peuvent être divisées en plusieurs catégories, notamment le Machine Learning (ML) et le Deep Learning (DL).

Le DL est un sous-ensemble du ML, qui est lui-même un sous-ensemble de l'Intelligence Artificielle (IA) comme le montre la figure 2.1 ci-dessous.



**Figure 2.1:** Diagramme de Venn représentant la relation entre AI, ML et DL.

## 2.2 Machine Learning

Le ML est une branche de l'IA qui a été défini par son pionnier Arthur Samuel en 1959 comme le « champ d'étude qui donne aux ordinateurs la capacité d'apprendre sans être explicitement programmés à apprendre [21]. Le principe de base du ML est de fournir à un ordinateur un ensemble de données d'entraînement (ou training set), puis de lui permettre d'analyser ces données pour trouver des modèles et des relations. Une fois que le modèle est établi, il peut être utilisé pour effectuer des prédictions ou des classifications sur de nouvelles données.

Il existe plusieurs types de techniques de ML, notamment :

### 2.2.1 L'apprentissage non supervisé

Il implique également la formation des données, à l'exception du fait que la valeur étiquetée ou la valeur cible n'est pas connue. Dans ce cas, la machine essaie de regrouper les types de données similaires en trouvant le modèle caché. Plutôt que de faire des prédictions, son objectif principal est de découvrir les modèles. La performance du modèle dans l'apprentissage non supervisé ne peut pas être évaluée car la valeur de l'étiquette est absente ou inconnue. Les algorithmes impliqués sont le regroupement K-mean, l'extraction de règles d'association, la modélisation thématique et les techniques de réduction de la dimensionnalité.

### 2.2.2 L'apprentissage par renforcement

Il fonctionne en développant un système qui améliore ses performances en prenant en compte le retour d'information de l'environnement et en adoptant des mesures pour les améliorer. Il s'agit d'un acte d'apprentissage à partir de l'environnement en interagissant avec lui sans aucune aide humaine, c'est un processus itératif.

### 2.2.3 L'apprentissage supervisé

Le modèle est entraîné avec des données étiquetées, c'est-à-dire des données où chaque exemple est associé à une réponse connue. Le but est de prédire la réponse pour de nouvelles données non étiquetées. Il implique diviser les data en deux ensembles, de formation et de test. Le modèle est d'abord entraîné sur l'ensemble d'entraînement, puis ses performances sont testées sur l'ensemble de test. Ils peuvent être évalués à l'aide de mesures de performance.

On se base sur le modèle supervisé, et pour le mieux maîtrisé, il faut comprendre les quatre notions suivantes :

#### 2.2.3.1 Dataset

Dans le ML, il s'agit d'une collection de points de données d'entrée et d'étiquettes de sortie correspondantes ou de valeurs cibles utilisées pour l'entraînement et le test de modèles d'apprentissage automatique. Les ensembles de données peuvent provenir de diverses sources et sont divisés en ensembles de données d'entraînement et de test afin de garantir que les modèles sont capables de bien se généraliser à de nouvelles données non vues [22].

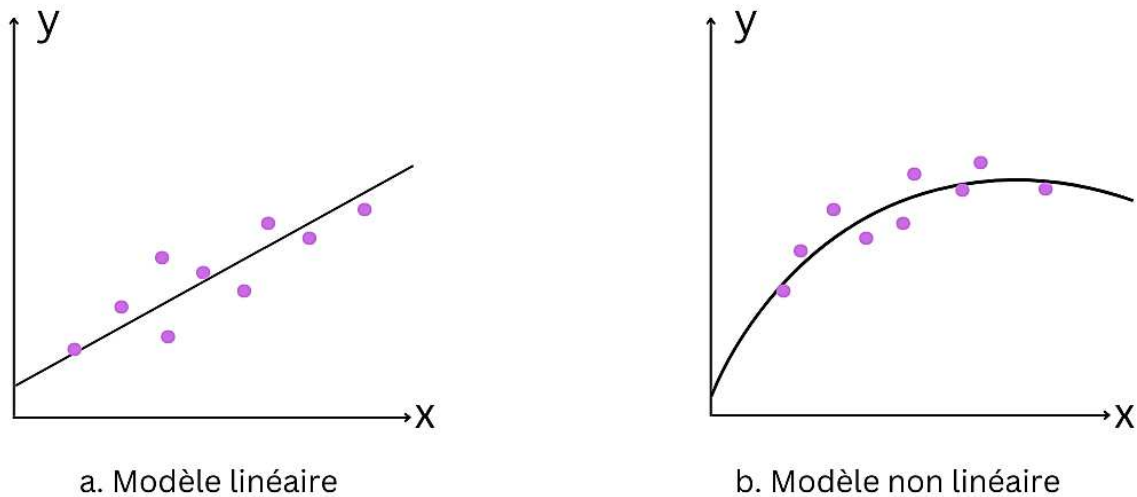
Le regroupement de ces données sera dans un tableau où on a :

- La variable  $y$  porte le nom de Target (la cible). C'est la valeur que l'on cherche à prédire.
- La variable  $x$  porte le nom de feature (facteur ou caractéristique). Un facteur influence la valeur de  $y$ , et on a en général beaucoup de features ( $x_1, x_2, \dots$ ).

### 2.2.3.2 Le modèle et ses paramètres

En ML, un modèle est un algorithme qui est formé sur un jeu de données pour effectuer des prédictions ou prendre des décisions précises sur de nouvelles données inédites. Le modèle est défini par un ensemble de paramètres qui sont appris ou évalués à partir des données de formation et qui définissent le fonctionnement du modèle [22].

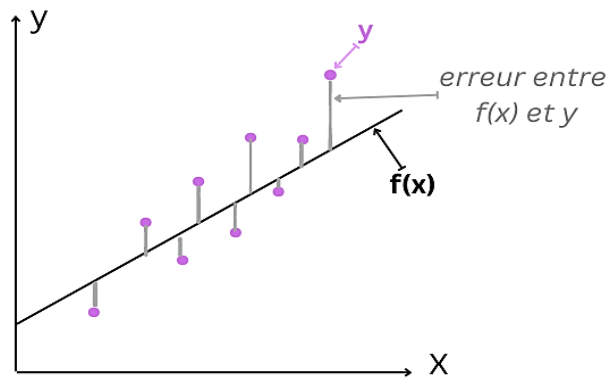
Il peut s'agir d'un modèle linéaire (Figure 2.2.a) ou non linéaire (Figure 2.2.b)



**Figure 2.2 :** Modèle linéaire et modèle non linéaire.

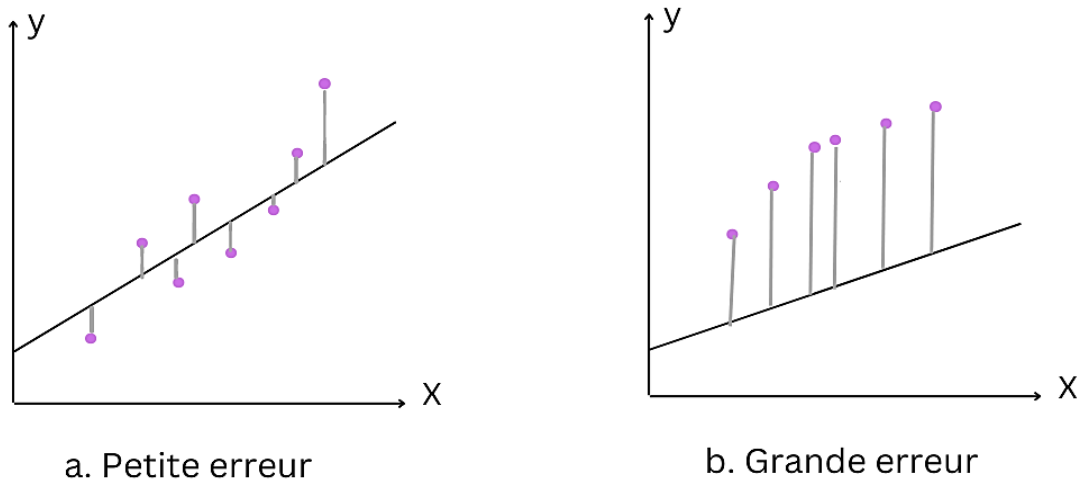
### 2.2.3.3 La Fonction Coût

La fonction de coût est une mesure quantitative de la qualité de prédiction d'un modèle ML. Elle mesure l'erreur ou la différence entre les estimations du modèle et les vraies valeurs de sortie ou étiquettes correspondant aux données d'apprentissage (voir la figure 2.3). Le but de la fonction de coût est de déterminer les valeurs optimales des paramètres du modèle qui minimisent l'erreur de prédiction ou la différence entre les prédictions du modèle et les valeurs réelles de la sortie [22].



**Figure 2.3 :** Fonction coût.

Avoir un bon modèle veut dire avoir un modèle qui donne un minimum d'erreurs c'est-à-dire une petite fonction de coût (voir figure 2.4).



**Figure 2.4 :** Bon modèle (graphe a) vs Mauvais modèle (graphe b).

### 2.2.3.4 L'Algorithme d'apprentissage

Les algorithmes d'apprentissage sont des techniques mathématiques et statistiques permettant de former des modèles ML à base de données. Le principal objectif de l'apprentissage supervisé est de déterminer les paramètres du modèle qui limitent le plus possible la fonction de coût. Les algorithmes d'optimisation permettent d'ajuster les paramètres du modèle afin de réduire au minimum la valeur de la fonction de coût [22].

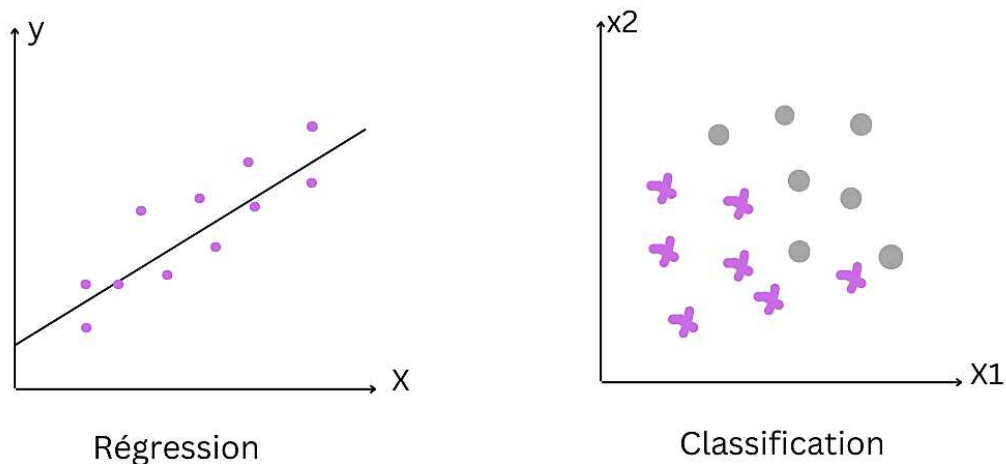


### 2.2.3.5 Les applications d'apprentissage supervisé

Avec l'apprentissage supervisé, nous pouvons élaborer des modèles pour répondre à deux types de problèmes (voir la figure 2.5) [22] :

Les problèmes de régression : Si nous cherchons à prédire la valeur d'une grandeur continue, c'est-à-dire une grandeur qui peut avoir un nombre illimité de possibilités.

Les problèmes de classification : Si nous essayons de classer un objet dans plusieurs classes, c'est-à-dire si nous essayons de prédire la valeur d'une donnée discrète (qui ne peut prendre qu'un nombre fini de valeurs). Dans le cas d'un problème de classification, nous considérons souvent les classes comme des symboles, plutôt que comme des nombres (0, 1, ...).



**Figure 2.5 :** Les deux types de ML.

## 2.3 Types d'algorithmes en Machine Learning

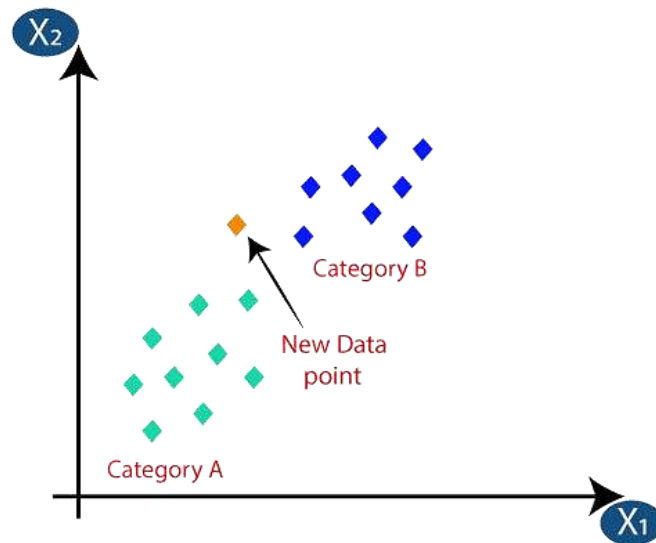
### 2.3.1 K Nearest Neighbour Algorithm (KNN)

Le KNN est un modèle simple et l'un des processus de ML les plus utilisés pour les problèmes de classification, de reconnaissance des formes et de régression. Le KNN détermine les voisins parmi les données en utilisant une distance entre les points de données.

La valeur de  $k$  (où  $k$  est une constante définie par l'utilisateur) permettra de trouver tous les cas de caractéristiques similaires existants avec le nouveau cas.

La valeur de  $k$  est donc importante et doit être choisie avec soin, car le système peut conduire à un sur ajustement (overfitting) si  $k$  est très petit. Il existe quelques limitations, telles que des performances moyenne lorsqu'il existe un vaste ensemble de données d'apprentissage. Le coût de calcul est très élevé car nous devons mesurer la distance entre tous les échantillons d'entraînement et chaque instance de requête [23].

La figure 2.6 représente l'idée de base de l'algorithme KNN pour la classification de nouveau point de données.



**Figure 2.6 :** Classification d'un nouveau point avec KNN.

### 2.3.2 Support Vector Machine (SVM)

L'algorithme machine à vecteur de support (SVM) est un classificateur qui trouve l'hyperplan qui sépare efficacement les observations en leurs étiquettes de classe. Il commence par positionner chaque instance dans un espace de données à  $n$  dimensions, où  $n$  représente le nombre de caractéristiques.

Ensuite, il trace une ligne imaginaire qui sépare clairement les instances appartenant à une étiquette de classe des autres. Un vecteur de support fait référence aux coordonnées d'une instance donnée. En conséquence, le SVM est la frontière qui sépare efficacement les différents vecteurs de support dans un espace de données.

Pour un espace de données bidimensionnel, l'hyperplan est une ligne qui divise l'espace de données en deux sections, chacune représentant une étiquette de classe [23].

- **Le principe de travail**

Le diagramme suivant (voir Figure 2.7) montre un exemple simple de modèle SVM. Les triangles et les points de données circulaires représentent les instances de l'ensemble de données d'entrée, les formes définissant l'étiquette de classe à laquelle chaque instance appartient. La ligne en pointillé indique l'hyperplan qui sépare clairement les points de données, qui est défini en fonction de l'emplacement des points de données dans l'espace de data. Cette ligne est utilisée pour classer les données non vues, représentées par le carré.

Ainsi, les nouvelles instances situées à gauche de la ligne seront classées comme des triangles, tandis que celles situées à droite seront des cercles.

Plus le nombre de features est important, plus l'espace de données aura de dimensions, ce qui rendra impossible la représentation visuelle du modèle [23].

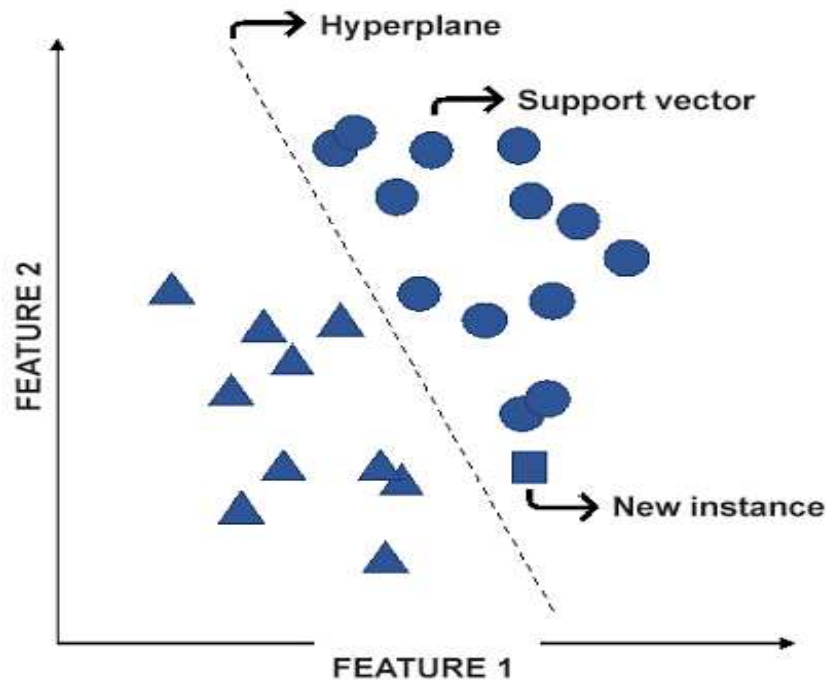


Figure 2.7 : Modèle SVM.

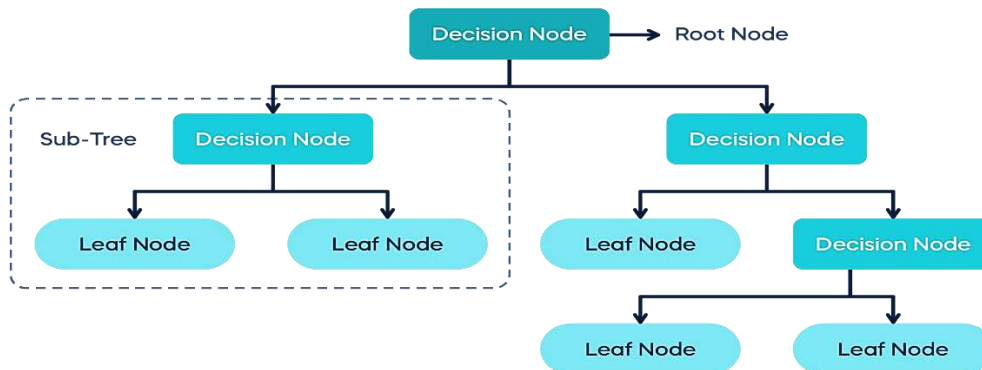
### 2.3.3 Decision Tree Algorithm (DT)

L'algorithme de l'arbre de décision (DT) effectue une classification basée sur une séquence qui ressemble à une structure arborescente. Il fonctionne en divisant l'ensemble de données en petits sous-ensembles qui servent de guides pour développer les nœuds de DT. Les nœuds peuvent être soit des nœuds de décision ou des nœuds de feuilles, les premiers représentent une question ou une décision, et les seconds représentent les décisions prises ou le résultat final.

DT peut traiter des caractéristiques quantitatives et qualitatives, étant donné que les caractéristiques continues sont traitées dans des plages. En outre, les nœuds feuilles peuvent gérer des étiquettes de classe catégoriques ou continues, pour les étiquettes de classe catégoriques, une classification est effectuée, tandis que pour les étiquettes de classe continues, la tâche à accomplir est la régression.

La figure 2.8 représente un exemple d'un DT où les cases en bleu foncé représentent les nœuds de décision, les flèches sont les branches représentant chaque réponse au nœud de

décision, et les cases dont l'arrière-plan est en bleu claire se réfèrent au résultat pour les instances qui suivent la séquence [23].



**Figure 2.8 :** Arbre de décision.

### 2.3.4 Random Forest Classification (RF)

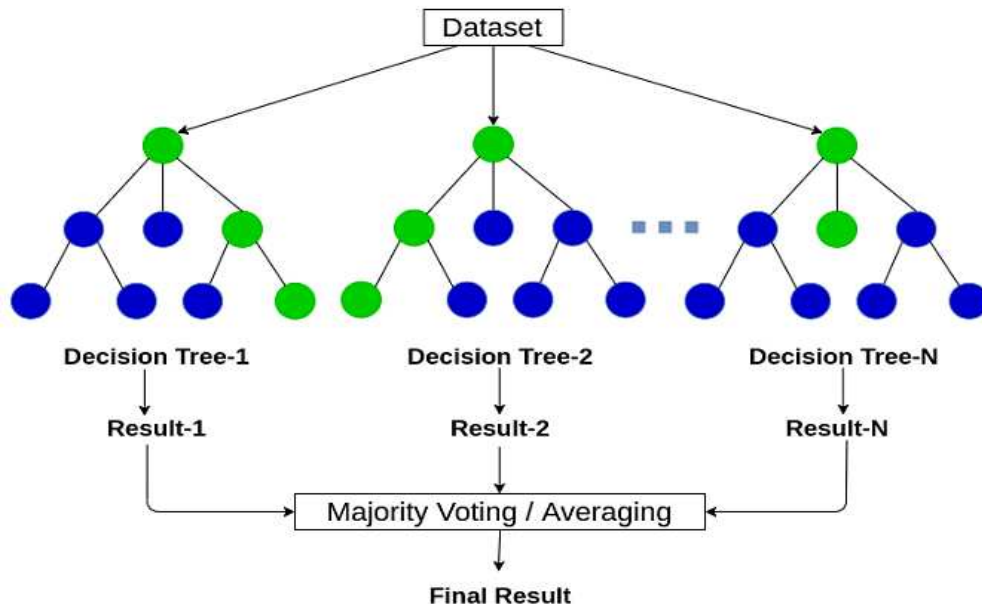
La forêt aléatoire (RF) est un modèle d'ensemble qui peut également être utilisé comme une forme de prédicteur du plus proche voisin. L'idée principale des approches d'ensemble est qu'un groupe de modèles forme un modèle puissant. RF s'accompagne d'une approche ML standard d'arbre de décision en termes d'ensemble.

Dans cet algorithme, une entrée est saisie au sommet, et les données sont réparties en ensembles de plus en plus petits au fur et à mesure qu'elles traversent l'arbre. RF porte cette notion à un niveau supérieur en combinant les arbres avec une notion d'ensemble. L'avantage d'un classificateur RF est que ses temps d'exécution sont concis, non équilibrés et que les données manquantes peuvent être traitées.

Dans la forêt, le nouvel ensemble de données ou les données de test sont distribués à tous les sous-arbres créés [23].

Chaque sous-arbre de décision de la forêt est autorisé à décider de la classe de l'ensemble de données. Le modèle sélectionne ensuite la classe la plus appropriée à partir du vote majoritaire [23].

La figure 2.9 illustre l'idée de base de l'algorithme de forêt aléatoire :



**Figure 2.9 :** Forêt aléatoire.

### 2.3.5 Naïve Bayes Algorithme (NB)

Naïve Bayes (NB) est un algorithme de classification basé sur le théorème de Bayes qui suppose naïvement l'indépendance entre les caractéristiques et attribue le même poids (degré d'importance) à toutes les caractéristiques. Cela signifie que l'algorithme suppose qu'aucune caractéristique n'est en corrélation avec une autre ou ne l'affecte. Par exemple, bien que le poids et la taille soient en quelque sorte corrélés lorsqu'il s'agit de prédire l'âge d'une personne, l'algorithme suppose que chaque caractéristique est indépendante.

En outre, l'algorithme considère que toutes les caractéristiques ont la même importance. Par exemple, même si un diplôme peut influencer les revenus d'une personne dans une plus large mesure que le nombre d'enfants qu'elle a, l'algorithme considère que les deux caractéristiques ont la même importance.

Bien que les ensembles de données de la vie réelle contiennent des caractéristiques qui ne sont pas également importantes ni indépendantes, cet algorithme est populaire parmi les scientifiques car il donne des bons résultats sur de grands ensembles de données. De plus, en raison de l'approche simpliste de l'algorithme, il s'exécute rapidement, ce qui permet de l'appliquer à des problèmes nécessitant des prédictions en temps réel. En outre, il est fréquemment utilisé pour la classification de textes, car il surpasse souvent les algorithmes plus complexes [23].

- **Le principe de fonctionnement**

L'algorithme convertit les données d'entrée en un résumé des occurrences de chaque étiquette de classe par rapport à chaque caractéristique, qui est ensuite utilisé pour calculer la probabilité d'un événement (une étiquette de classe), compte tenu d'une combinaison de caractéristiques. Enfin, cette valeur est normalisée par rapport à la probabilité des autres étiquettes de classe. Le résultat est la probabilité qu'une instance appartienne à chaque label de classe. La somme des chances doit être égale à un, et l'étiquette de classe ayant la probabilité la plus élevée est celle que l'algorithme choisit comme prédiction [23].

## **2.4 Les applications du Machine Learning**

Les applications du ML sont vastes et variées, allant de la reconnaissance d'images à la prédiction de tendances financières.

Nous allons explorer certaines des applications les plus intéressantes du ML, en mettant l'accent sur leur pertinence dans le monde réel.

### **2.4.1 Détection de fraude**

La détection des fraudes est l'une des utilisations les plus importantes de l'apprentissage automatique dans le secteur de la banque et de la finance. Cette technologie est mise en œuvre pour rechercher dans de grands volumes de données relatives aux transactions et repérer des modèles de comportement inhabituel. Chaque achat effectué par un client est évalué en temps réel, et la probabilité que la procédure soit considérée comme frauduleuse est indiquée par un score de fraude. La commande est ensuite bloquée ou gelée en vue d'un contrôle manuel en cas de transaction frauduleuse. L'ensemble de ces processus se déroule en quelques secondes seulement [24].

### **2.4.2 Assistance médicale**

L'assistance médicale est une autre application passionnante du ML. Les systèmes de diagnostic automatisés peuvent aider les médecins à identifier les maladies plus rapidement et avec plus de précision.

Les algorithmes d'apprentissage automatique peuvent également être utilisés pour prédire les résultats du traitement et faciliter la planification des soins de santé.

### **2.4.3 Jeux et éducation**

ML est fréquemment utilisé dans les jeux vidéo et l'éducation. Plusieurs applications éducatives et récréatives utilisent l'IA et aussi le ML. Ces applications incluent Duolingo, qui est une application gratuite d'apprentissage des langues conçue de manière divertissante et interactive. Cette application est bénéfique car elle permet aux utilisateurs de se comprendre comme s'ils jouaient à un jeu sur leur téléphone.

Il collecte des informations à partir de la réponse de l'utilisateur et crée un modèle statique qui peut prédire combien de temps quelqu'un peut se souvenir d'un mot avant de demander une mise à niveau [25].

## 2.5 Deep Learning

### 2.5.1 Définition de l'apprentissage profond

L'apprentissage en profondeur est une forme d'IA issue du concept d'apprentissage automatique. Il permet aux machines d'analyser les données et d'en tirer des conclusions significatives. Cela se fait d'une manière qui restreint le processus de pensée de l'humain, et cette association est finalement inévitable.

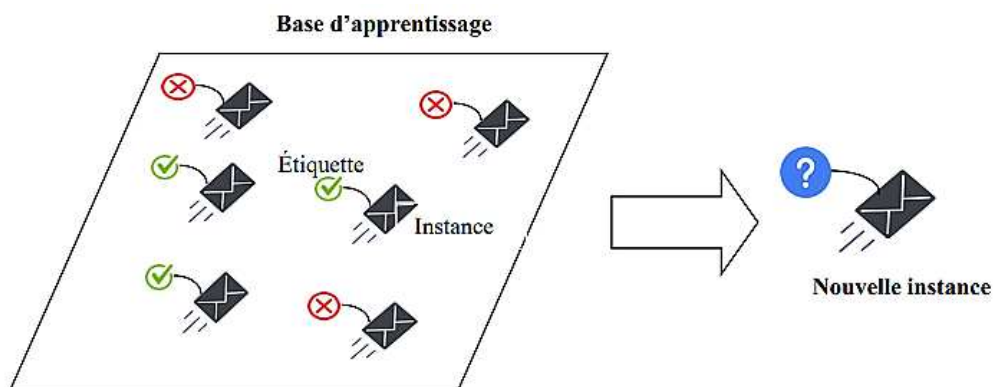
Sa principale particularité est de s'inspirer de l'architecture neuronale du cerveau humain, et de l'analyse de ces données via une méthode spécifique de filtrage et de superposition qui s'appuie plus spécifiquement sur le concept de réseaux de neurones artificiels qui est dépeint comme un réseau d'ensembles d'unités d'exécution d'informations (représentant des neurones) entrelacées en couches et reliées par des connecteurs (synapses).

### 2.5.2 Types d'apprentissage profond

#### 2.5.2.1 Apprentissage supervisé

L'apprentissage supervisé a lieu en présence d'une variable cible. Cette dernière est une variable de l'ensemble de données à prédire, compte tenu des autres variables. L'apprentissage supervisé exige que la variable cible soit connue et que les modèles soient entraînés à prédire correctement cette variable [23].

La figure 2.10 montre un exemple d'apprentissage supervisé :



**Figure 2.10 :** Algorithme supervisé.

### 2.5.2.2 Apprentissage non supervisé

Dans l'apprentissage non supervisé, aucune variable cible n'est fournie dans les données d'apprentissage, mais les modèles visent à assigner une variable cible. Un exemple de technique d'apprentissage non supervisé est le regroupement par k-means. Cet algorithme répartit les données en un nombre défini de classes en fonction de leur proximité par rapport aux points de données voisins. La variable cible assignée peut être soit le numéro de la classe, soit le centre de la classe [23].

La figure 2.11 montre un exemple d'apprentissage non supervisé :

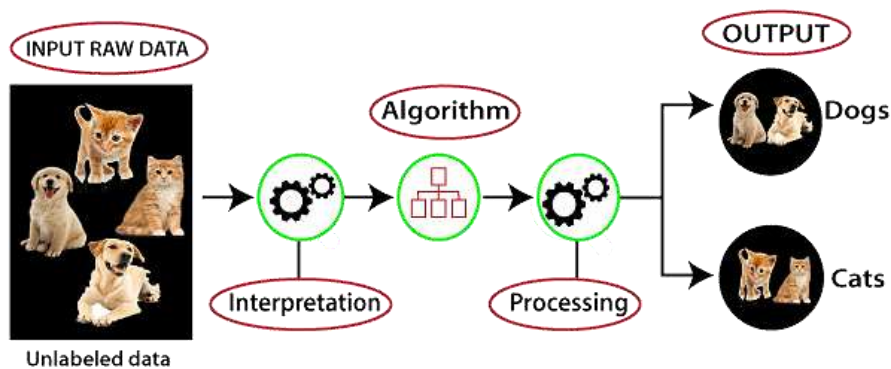


Figure 2.11 : Algorithme non supervisé.

### 2.5.3 Les applications du Deep Learning

Depuis son apparition, le DL a entraîné une évolution significative dans le monde de l'IA et a conduit au développement de nombreuses applications. Grâce à ses améliorations significatives en termes de performance, le DL s'est imposé dans de nombreux domaines qu'on va découvrir maintenant.

#### 2.5.3.1 Détection des retards de développement chez les enfants

Les troubles de la parole, l'autisme et les troubles du développement peuvent priver les enfants souffrant de l'un de ces problèmes d'une bonne qualité de vie. Un diagnostic et un traitement précoces peuvent avoir un effet merveilleux sur la santé physique, mentale et émotionnelle des enfants handicapés. C'est pourquoi l'une des applications les plus nobles de l'apprentissage profond est la détection précoce et la correction de ces problèmes associés aux nourrissons et aux enfants. L'autisme est souvent détecté en le combinant avec des cofacteurs tels que le faible poids après la naissance, l'activité physique, l'indice de masse corporelle, les difficultés d'apprentissage [26].



### 2.5.3.2 Marketing personnalisé

Le marketing personnalisé est un concept qui a connu un grand succès ces dernières années. Les spécialistes du marketing orientent désormais leurs campagnes publicitaires vers les points faibles des consommateurs individuels répondant précisément à leurs besoins. L'apprentissage profond joue un rôle important dans ce domaine.

Aujourd'hui, les consommateurs génèrent d'énormes quantités de données grâce à l'utilisation de plateformes de médias sociaux, d'appareils IoT, de navigateurs web, de dispositifs portables, etc. Pour résoudre ce problème, les entreprises utilisent des modèles de DL personnalisable pour interpréter et distiller des données provenant de différentes sources afin d'extraire des informations précieuses sur les clients. Ils utilisent ensuite ces informations pour prédire le comportement des consommateurs et cibler plus efficacement leurs efforts de marketing.

Vous comprenez donc maintenant comment ces sites de vente en ligne savent quels produits vous recommander [27].

### 2.5.3.3 Détection de fausses nouvelles

Le concept de diffusion de fausses nouvelles pour faire pencher la balance en sa faveur n'est pas nouveau. Cependant, en raison de la popularité explosive d'internet, en particulier des plateformes de médias sociaux, les fausses nouvelles sont devenues omniprésentes.

En plus d'induire les citoyens en erreur, les fausses nouvelles peuvent être utilisées pour modifier campagnes politiques, dénigrer certaines situations et certains individus et commettre d'autres transgressions morales similaires. C'est pourquoi le blocage de toutes les fausses nouvelles est une priorité.

Le DL offre un moyen de lutter contre la menace des fausses nouvelles en utilisant des techniques sophistiquées de détection de langage pour classer les sources d'information frauduleuses.

Cette approche consiste essentiellement à collecter des informations provenant de sources fiables et à les associer avec de nouvelles informations pour vérifier leur validité [27].

## 2.6 Les réseaux de neurones artificiels

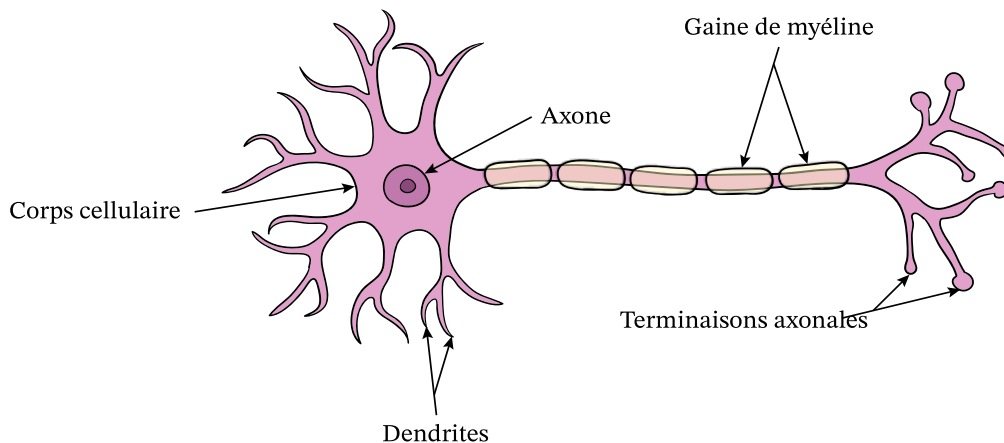
### 2.6.1 Introduction sur les réseaux de neurones

Bien qu'il existe plusieurs algorithmes de ML pour résoudre les problèmes de données, les ANN sont devenues de plus en plus populaires parmi les données scientifiques, en raison de leur capacité à trouver des modèles dans des ensembles de données volumineuses et complexes qui ne peuvent pas être interprétés par des humains.

La partie neuronale du nom fait référence à la ressemblance de l'architecture du modèle avec l'anatomie du cerveau humain. Cette partie est destinée à reproduire la capacité d'un être

humain à apprendre à partir de données historiques en transférant des bits de données d'un neurone à l'autre jusqu'à ce qu'un résultat soit atteint.

Un neurone humain se compose de trois parties principales : le corps cellulaire, les dendrites et l'axone, et fonctionne en générant des potentiels d'action et en formant des connexions synaptiques avec d'autres neurones comme le montre la figure 2.12 [28].



**Figure 2.12 :** Un neurone humain.

### 2.6.2 Fonctionnement des réseaux de neurones artificiels

Les réseaux neuronaux ont une fonction similaire à celle du cerveau humain. Les réseaux sont alimentés par une série de valeurs, dont chacune correspond à un nœud appelé neurone. Les neurones du réseau sont ensuite classés en couches qui constituent le réseau neuronal. Chacun est associé à une valeur numérique qui modifie l'entrée reçue. Les nouvelles valeurs sont absorbées par les neurones et se déplacent ensuite dans le réseau. Cette procédure est illustrée dans la figure 2.13. Une fois l'extrémité du réseau atteinte, on reçoit une sortie qui sera la prédiction calculée par le réseau. Plus il y a de couches et plus elles sont compliquées, plus leurs capacités sont complexes [29].

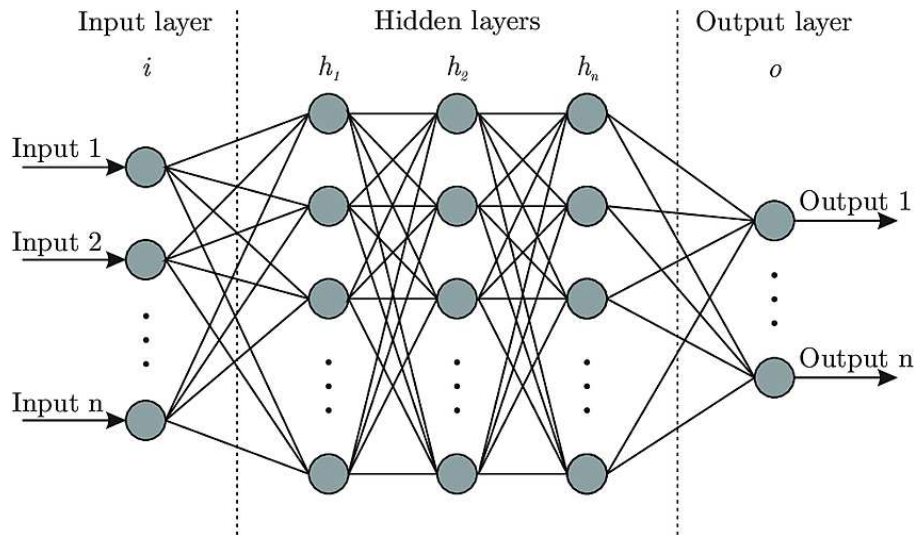


Figure 2.13 : Un réseau artificiel.

## 2.6.3 Types des réseaux de neurones artificiels

### 2.6.3.1 Réseaux neuronaux à propagation directe (FeedForward)

Le FeedForward (FF) est un type des réseaux neuronaux d'architecture ANN le plus simple, dans lequel les informations se diffusent dans une seule et même direction, de l'entrée vers la sortie. Les couches sont entièrement connectées, ce qui signifie que chaque neurone d'une couche est directement relié à toutes les couches suivantes [30] comme le montre la figure 2.14.

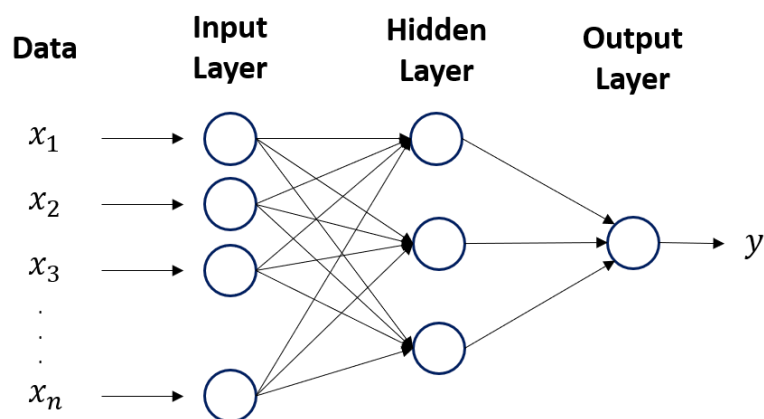


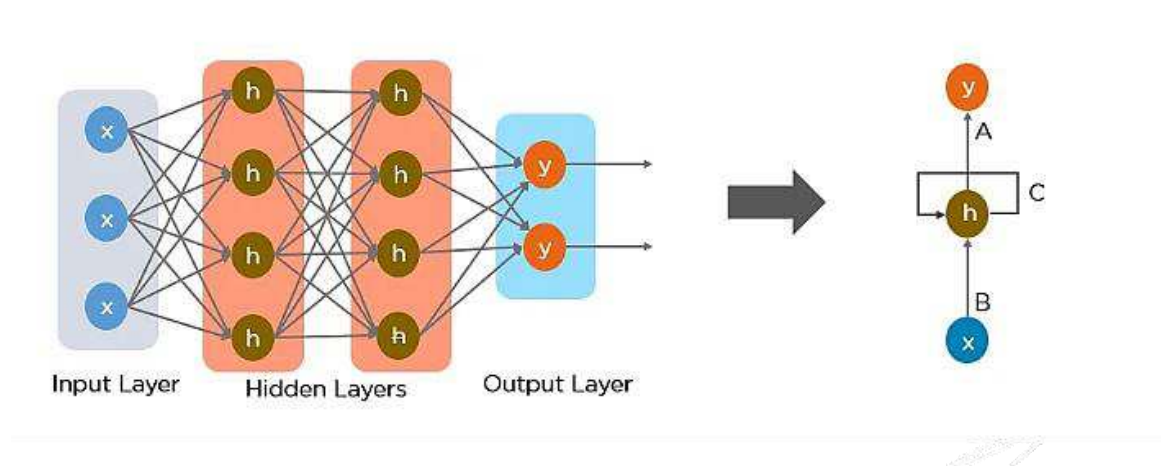
Figure 2.14 : Le réseau FeedForward.

### 2.6.3.2 Réseaux de neurones récurrents RNN

Le réseau neuronal récurrent (RNN) s'agit d'une classe de réseaux neuronaux basés sur le concept de la mémoire séquentielle. Contrairement aux réseaux neuronaux traditionnels, un RNN prédit les résultats dans des données séquentielles.

Le principe de RNN est de renvoyer la sortie d'une couche vers l'entrée, ce principe permet de prédire le résultat de la couche. Dans le processus de calcul, chaque neurone agit comme une cellule de mémoire, le neurone conserve certaines informations lorsqu'il passe au pas de temps suivant, c'est ce qu'on appelle un processus de réseau RNN. Les données à utiliser ultérieurement sont mémorisées et le travail pour l'étape suivante se poursuit dans le processus comme le montre la figure 2.15. La prédiction sera améliorée par la correction d'erreurs, lors de cette opération certains changements sont apportés pour créer la bonne prédiction, le taux d'apprentissage est la vitesse à laquelle le réseau peut faire une prédiction correcte à partir d'une prédiction erronée.

Les réseaux RNN ont de nombreuses applications, et l'une d'entre elles est le modèle de conversion du texte en parole, et il est conçu pour l'apprentissage supervisé [31].



**Figure 2.15 :** Réseaux de neurones récurrents (RNN).

### 2.6.3.3 Réseaux de mémoire à long terme (LSTM)

Dans les RNN, nous pouvons retenir que les informations de l'étape la plus récente. Mais pour un problème tel que la traduction d'une langue, nous avons besoin d'une rétention beaucoup plus importante dont les réseaux LSTM entrent en jeu.

Les LSTM sont un cas particulier de RNN qui a la même structure en chaîne mais avec une structure de module répétitif différent. Cette structure permet au réseau de conserver un nombre beaucoup plus important de valeurs d'étapes antérieures (voir figure 2.16).

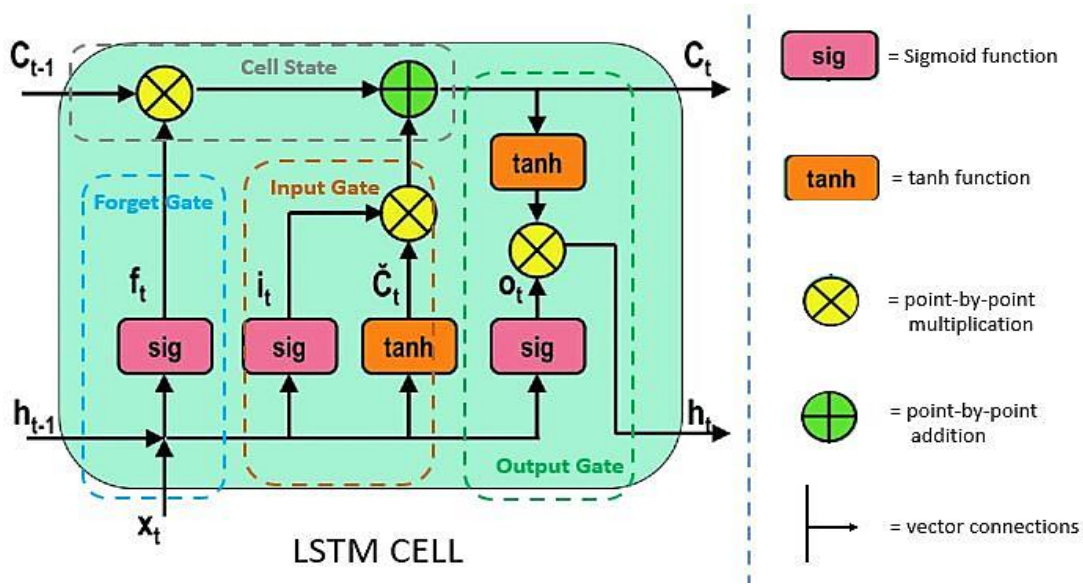


Figure 2.16 : Structure d'un réseau LSTM.

- **Porte d'oubli (forget gate)**

Cette partie décide quelle information peuvent être ignorée et les quelles doivent être conservées pour un usage ultérieur. Les informations provenant de l'entrée actuelle  $X_t$  et de l'état caché  $h_{(t-1)}$  sont transmises à la porte et multiplier par des matrices de poids  $W_f$  puis ajoute un biais  $b_f$  et le résultat sera transmis à une fonction d'activation *Sig* qui génère une sortie binaire. Pour l'état particulier de la cellule, la sortie est 0. Dans ce cas, l'information est oubliée et si la sortie est 1 l'information est conservée pour un usage ultérieur.

$$f_t = \text{Sig}(W_f \cdot [h_{t-1}, X_t] + b_f) \quad (2.1)$$

- **Porte d'entrée (input gate)**

Cette entité est chargée d'augmenter l'état de la cellule avec des informations utiles. Initialement l'information est régulée par la fonction sigmoïde *Sig* et est filtrée afin qu'elle puisse être stockée d'une manière similaire à la porte d'oubli utilisant les entrées  $h_{t-1}$  et  $X_t$ . Après cela, un vecteur est créé  $\check{C}_t$  en utilisant la fonction *tanh* qui génère -1 et +1 qui englobe toutes les valeurs possibles de  $h_{t-1}$  et  $X_t$  les valeurs de sorties produites par les fonctions d'activation qui sont additionnées pour obtenir des informations utiles.

$$I_t = \text{Sig}(W_i \cdot [h_{t-1}, X_t] + b_i) \quad (2.2)$$

$$\check{C}_t = \text{tanh}(W_c \cdot [h_{t-1}, X_t] + b_c) \quad (2.3)$$

- **Porte de sortie (output gate)**

Elle est chargée d'extraire des informations utiles de l'état actuel de la cellule et de les présenter sous la forme d'une sortie. Tout d'abord les valeurs de l'état courant  $X_t$  et de l'état caché précédent  $h_{t-1}$  sont introduites dans la troisième fonction sigmoïde. Le nouvel état de cellule dérivé de l'état de la cellule précédent est ensuite transmis dans la fonction tanh puis ces deux sorties sont additionnées en chaque point en fonction de la valeur finale, puis le réseau décidera quelle information à l'état caché doit contenir et cet état est souvent utilisé pour la prédiction.

$$O_t = \text{Sig}(W_o[h_{t-1}, X_t] + b_o) \quad (2.4)$$

$$H_t = O_t \cdot \tanh(C_t) \quad (2.5)$$

### 2.6.3.4 Réseaux de neurones convolutifs (CNN)

Le CNN est une classe de réseaux neuronaux profonds qui est principalement utilisée dans le domaine de la vision par ordinateur et de l'imagerie. Les CNN sont utilisés pour identifier des images, les regrouper en fonction de leur similarité et mettre en œuvre la reconnaissance d'objets dans des scènes. Les réseaux neuronaux profonds comportent différentes couches, à savoir la couche d'entrée, la couche de sortie et plusieurs couches cachées. Ces couches cachées d'un CNN se composent de couches entièrement connectées, de couches convolutives, d'une couche ReLu comme fonction d'activation, de couches de normalisation et d'une couche de mise en commun [23].

L'architecture d'un réseau CNN [23] :

- **Image d'entrée :** Une image d'entrée constitue le premier composant d'une architecture CNN. Une image peut être de n'importe quel type : un humain, un animal, un paysage, une image radiographique médicale, etc. Chaque image est convertie en une matrice mathématique de zéros et d'uns.
- **Couche de convolution :** La couche de convolution est le point de départ du traitement de l'image.
- **Couche de pooling :** La couche de regroupement nous aide à ignorer les données les moins importantes de l'image et à la réduire encore, tout en préservant ses caractéristiques importantes.
- **Flattening :** L'aplatissement fait partie d'un CNN où l'image est préparée pour servir d'entrée à un ANN. Comme son nom l'indique, l'image regroupée est aplatie et convertie en une seule colonne. Chaque ligne est transformée en colonne et empilée l'une sur l'autre.

La figure 2.17 montre un exemple de réseau convolutif :

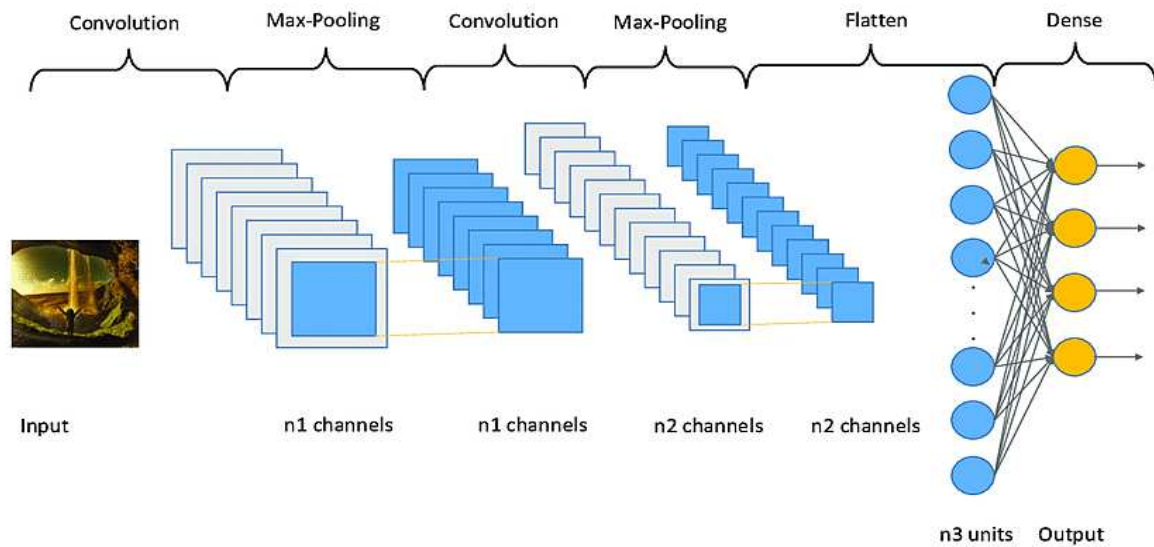


Figure 2.17 : Un réseau de neurones convolutif (CNN).

## 2.6.4 Les fonctions d'activations

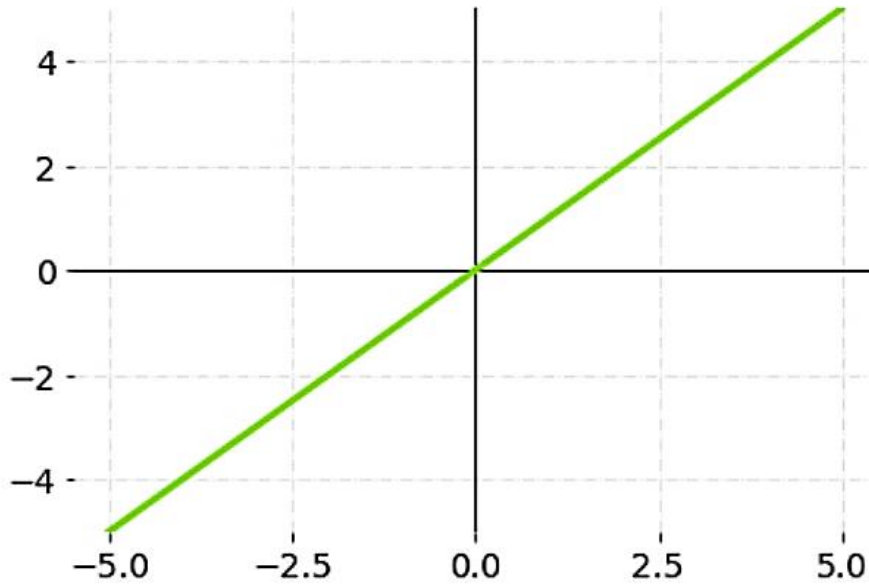
### 2.6.4.1 Introduction sur les fonctions d'activations

Chez l'homme, le cerveau reçoit des informations du monde extérieur, le neurone traite ces informations reçues et active la couche de neurones qui conduit aux décisions nécessaires. Dans les réseaux de neurones, les données sont présentées sous forme d'images, sons, nombre, etc. Le traitement est effectué sur le neurone artificiel, avec un algorithme activant la dernière couche correcte de neurones pour produire des résultats.

La fonction d'activation est utilisée pour déterminer si le neurone est actif ou non (par l'utilisation des opérations mathématiques simples pour analyser si l'entrée du neurone dans le réseau est significative ou non dans le processus de prédiction) et il existe trois types de fonctions d'activation [32].

### 2.6.4.2 Fonction d'activation linéaire

Cette fonction est souvent appelée fonction d'activation d'identité, et est proportionnelle à l'entrée. Elle additionne simplement la somme pondérée des entrées et renvoie le résultat, et la sortie prend n'importe quelle valeur entre  $-\infty$  et  $+\infty$  (voir figure 2.18) [32].



**Figure 2.18 :** Fonction linéaire.

#### 2.6.4.3 Fonction d'activation non linéaire

C'est le type le plus utilisé. Elle permet à un réseau neuronal artificiel de s'adapter facilement à une variété de données et de différencier les sorties, et aussi d'empiler plusieurs couches de neurones car la sortie est une combinaison non linéaire des entrées qui traversent chaque couche. Toute sortie peut être considérée comme une sortie fonctionnelle dans un réseau neuronal qui est calculée par un réseau neuronal. On va parler maintenant des fonctions d'activation non linéaires les plus courantes dans les réseaux de neurones.

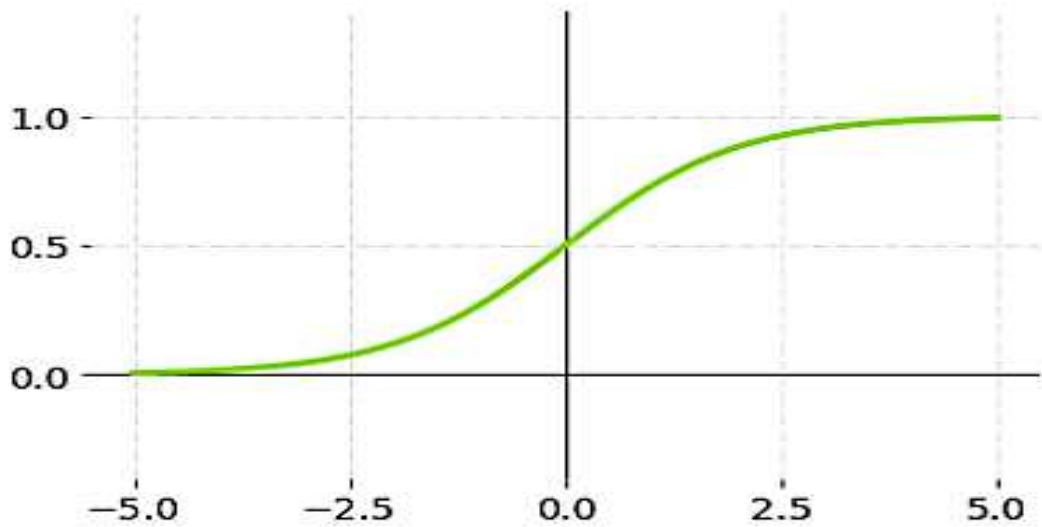
Ci-dessous quelques exemples des fonctions d'activation non linéaires [32] :

- **Fonction Sigmoïde**

Cette fonction est le plus couramment utilisée pour la classification car elle produit une sortie comprise entre 0 et 1 ce qui permet de modéliser la probabilité d'activation d'un neurone (figure 2.19). Elle est définie par [32] :

$$F(x) = \frac{1}{1+e^{-x}} \quad (2.6)$$



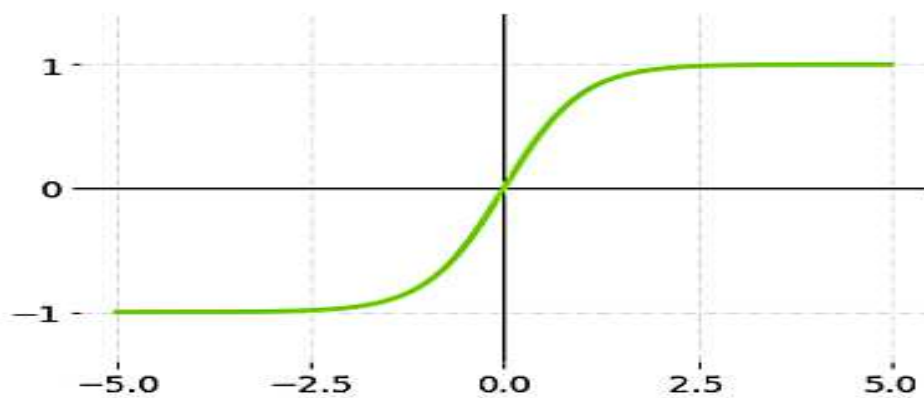


**Figure 2.19 :** Fonction Sigmoide.

- **Fonction Tanh**

La fonction tangente hyperbolique (Tanh) est similaire à la fonction sigmoïde. Elle produit une sortie comprise entre -1 et 1, ce qui permet de modéliser des valeurs négatives (figure 2.20). Son équation mathématique est donnée par [32] :

$$F(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})} \quad (2.7)$$

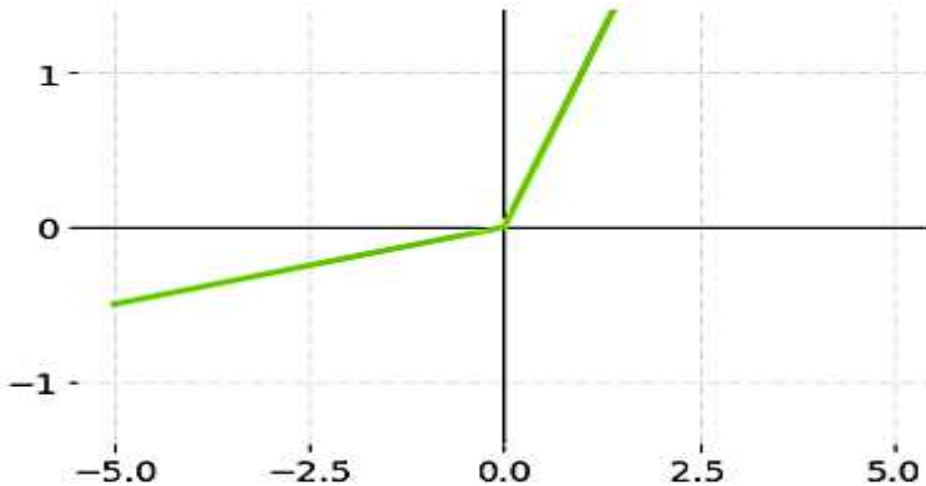


**Figure 2.20 :** Fonction Tanh.

- **Fonction ReLU (Rectified Linear Unit)**

Cette fonction résout le problème de la saturation des neurones. Elle renvoie simplement la valeur d'entrée si celle-ci est positive (figure 2.21), et 0 sinon. Elle est définie par [32] :

$$F(x) = \max(0.1x, x) \quad (2.8)$$

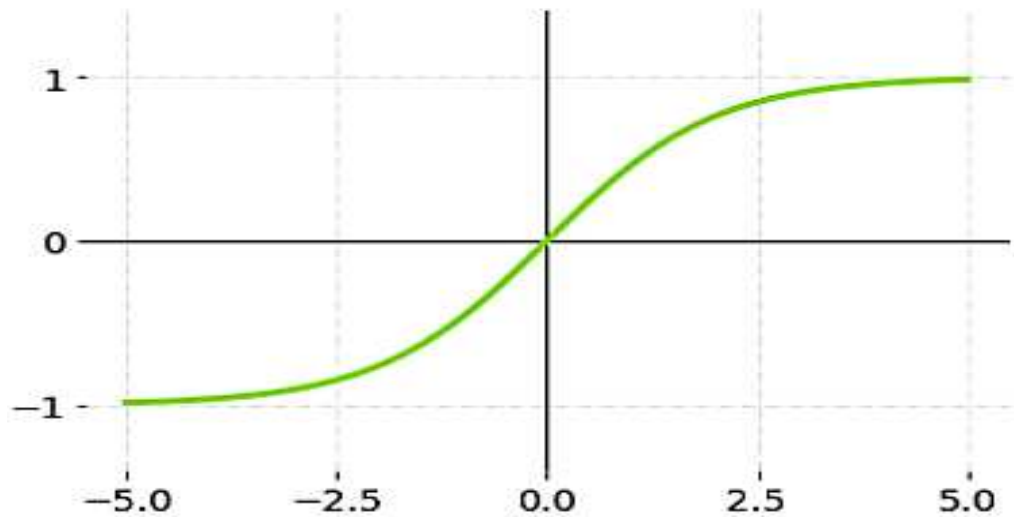


**Figure 2.21** : Fonction ReLU.

- **Fonction Softmax**

Cette fonction est une combinaison de plusieurs sigmoïdes, utilisée dans la classification multi-classes et le plus souvent utilisée dans la dernière couche du réseau neuronal. La fonction Softmax donne la probabilité de la classe actuelle par rapport aux autres (figure 2.22). Elle est définie par [32] :

$$\text{Softmax} = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \quad (2.9)$$



**Figure 2.22 :** Fonction Softmax.

#### 2.6.4.4 Fonction d'activation en escalier

Cette fonction active l'entrée grâce à la comparaison avec un seuil. Si la valeur d'entrée est supérieure à la valeur de seuil, le neurone est activé. Il est désactivé si l'entrée est inférieure au seuil, ce qui signifie que sa sortie n'est pas envoyée à la couche suivante ou masquée.

#### 2.6.5 Fonction de perte (Loss Function)

Lors de l'apprentissage des paramètres optimaux d'un modèle, nous devons définir une fonction pour mesurer l'erreur. Cette fonction est la fonction de perte qui nous fournit une mesure entre les sorties prédites par le réseau et les sorties de l'ensemble de données. Cette fonction de perte peut être définie de différentes manières en fonction du problème et de l'objectif [23].

##### 2.6.5.1 Regression loss

###### ◦ Mean squared error (MSE) / L2 Loss

mean\_squared\_error (MSE) est l'une des fonctions de perte de base utilisée pour résoudre les problèmes de régression. Pour calculer MSE, vous prenez la différence entre vos prédictions et la valeur réelle, vous l'élevez au carré et vous en faites la moyenne sur l'ensemble de vos données [33], comme le montre l'équation suivante :

$$(\text{Résultat réel} - \text{Résultat prédit})^2 \quad (2.10)$$

- **Mean Absolute Error (MAE) / L1 Loss**

mean\_absolute\_error (MAE) est une fonction de perte pour les problèmes de régression. Pour calculer MAE, on prend la différence entre la valeur réelle et la prédiction du modèle et on fait la moyenne sur l'ensemble du jeu de données [34], comme le montre l'équation suivante :

$$\frac{1}{N} \sum_{i=1}^N |valeur\ réelle - valeur\ prédit| \quad (2.11)$$

Avec N= l'ensemble de jeu de données.

### 2.6.5.2 Classifications loss

- **Binary\_crossentropy**

binary\_crossentropy est une fonction de perte pour les problèmes de classification binaire (à deux classes). En général, la perte d'entropie croisée est utilisée pour calculer la perte pour les modèles où le résultat est un nombre de probabilité compris entre 0 et 1 [23].

- **Categorical\_crossentropy**

categorical\_crossentropy est une fonction de perte utilisée pour les modèles de classification multi-classes lorsqu'il y'a plusieurs étiquettes de sortie [35].

## 2.7 Comparaison entre Machine Learning et Deep Learning

Le tableau 2.1 résume la différence entre le ML et le DL [25] :

Paramètre	ML	DL
Fonctionnalité de Data	Permet l'entraînement sur des jeux de données moins vastes, mais requiert davantage d'intervention humaine pour apprendre et corriger ses erreurs.	Requiert de plus larges volumes de données d'entraînement, mais apprend de son propre environnement et de ses erreurs.
Temps d'exécution	Le ML prend moins de temps pour former le modèle que l'apprentissage profond, mais il faut beaucoup de temps pour tester le modèle, mais son niveau de précision est plus faible.	Le DL nécessite un long temps d'exécution pour former le modèle, il peut nécessiter des semaines, mais moins de temps pour le tester.
Côté Hardware	Les modèles ML n'ayant pas besoin d'une grande quantité de jeux de données, ils peuvent fonctionner sur des machines bas de gamme, et être entraînés sur un CPU.	Le modèle DL a besoin d'une énorme quantité de data pour fonctionner efficacement, il a donc besoin de GPU et donc d'une machine haut de gamme.
Feature engineering	Les modèles de ML nécessitent une étape d'extraction des caractéristiques par l'expert, avant d'être poursuivis.	Le DL est la version améliorée de ML ; il n'a donc pas besoin de développer l'extracteur de caractéristiques pour chaque problème ; il tente d'apprendre par lui-même des caractéristiques de haut niveau à partir des données.
Approche de la résolution de problèmes	Pour résoudre un problème donné, le modèle ML traditionnel décompose le problème en sous-parties et, après avoir résolu chaque partie, produit le résultat final.	L'approche de résolution de problèmes d'un modèle DL est différente de celle d'un ML traditionnel, car il prend des données d'entrée pour un problème donné et produit le résultat final. Il suit donc l'approche de bout en bout.
Interprétation des résultats	L'interprétation du résultat pour un problème donné est facile. Comme lorsque nous travaillons avec le ML, nous pouvons interpréter le résultat facilement, c'est-à-dire pourquoi ce résultat s'est produit, quel a été le processus.	L'interprétation du résultat pour un problème donné est très difficile. Comme lorsque nous travaillons avec le modèle DL, nous pouvons obtenir un meilleur résultat pour un problème donné que le modèle ML, mais nous ne pouvons pas trouver pourquoi ce résultat particulier s'est produit, ni le raisonnement.
Data	Les modèles ML nécessitent principalement des jeux de données sous une forme structurée.	Les modèles DL peuvent fonctionner avec des jeux de données structurées et non structurées, car ils s'appuient sur les couches du réseau neuronal artificiel.

**Tableau 2.1:** Différence entre ML et DL.

## **2.8 Conclusion**

Dans ce deuxième chapitre, nous avons brièvement défini l'IA, après nous avons essayé d'expliquer les concepts fondamentaux des deux modèles d'apprentissage ML et DL et leurs types, algorithmes et applications les plus utiles avec une brève comparaison entre eux. Bien que ces deux modèles présentent de nombreux avantages, ils ont également certaines limites. L'IA a été introduite pour la première fois dans les années 1950, ce n'est que très récemment que nous avons été en mesure d'exploiter pleinement ce domaine grâce à des ressources informatiques puissantes, ainsi qu'à l'apparition de grandes bases de données. L'avenir de l'IA est à la fois passionnant et incertain. D'une part, l'IA a le potentiel de résoudre certains des problèmes les plus urgents du monde, tels que le changement climatique et les maladies.

Ce chapitre nous donnera l'occasion de mettre en pratique ce que nous avons appris dans la partie théorique. Le chapitre suivant consiste à créer un modèle capable d'évaluer et de classer l'état de santé cardiaque de chaque individu à l'aide des algorithmes de ML et DL.

## **Chapitre 3**

### **Conception d'un modèle intelligent : résultats et interprétations**

### 3.1 Introduction

Les maladies cardiaques, notamment les crises cardiaques, l'insuffisance cardiaque congestive et les arythmies, restent un problème de santé mondial important. La détection précoce et le diagnostic précis de ces maladies sont essentiels pour une intervention en temps utile et une prise en charge efficace, ce qui améliore en fin de compte les résultats pour les patients. Les techniques du ML et du DL ont montré des résultats prometteurs dans divers domaines, y compris les services de santé, offrant des opportunités de développer des systèmes intelligents qui peuvent aider à la détection et à la prédiction des maladies cardiaques. Ce chapitre se concentre sur le développement d'une variété de modèles du ML ainsi que le DL pour la détection des maladies cardiaques à l'aide du dataset "Personal Key Indicators of Heart Disease (PKIHD)". L'ensemble de données PKIHD est une collection de données médicales individuelles, de mesures cliniques et de facteurs liés au mode de vie qui ont été soigneusement compilées pour faciliter l'identification d'indicateurs clés associées aux maladies cardiaques.

### 3.2 Environnement d'exécution

Pour une bonne réalisation, nous devons choisir des ressources matérielles et des logiciels qui sont capable d'effectuer des évaluations très intenses sur notre travail.

#### 3.2.1 Logiciels

Les logiciels utilisés pour travailler avec cette technique d'apprentissage sont nombreux. Dans cette rubrique, nous avons cité les plus connus et les plus pratiques dans ce domaine.

- **Anaconda**

Anaconda est un logiciel libre qui vous fournit une boîte à outils adaptée à la recherche et à la science. L'installation d'Anaconda vous donne accès à différents environnements qui vous permettent de coder en Python. Ces environnements, également connus sous le nom d'environnements de développement intégré (IDE), sont des plateformes ou des applications qui facilitent grandement le développement du code. Les IDE contiennent de nombreuses fonctionnalités utiles pour écrire, modifier et déboguer du code, visualiser et inspecter des données, stocker des variables, présenter des résultats et collaborer à des projets [36].

- **Jupyter Notebook**

Jupyter Notebook est un IDE basé sur le web qui utilise votre navigateur web par défaut. Chaque bloc de code peut être exécuté séparément, ce qui le rend très flexible et facile à expérimenter. Cela permet d'utiliser différents types de textes dans le même Notebook. Il est ainsi possible de créer et de partager des documents et de présenter votre code et vos résultats de manière organisée et esthétique. Puisqu'il est basé sur le web, il facilite également le partage des blocs-notes avec d'autres personnes et est optimal pour les collaborations [36].



- **Google Colaboratory**

Google Colaboratory (colab) est un service en nuage basé sur les Notebooks Jupyter pour la diffusion de la formation et de la recherche en matière d'apprentissage automatique. Il fournit un moteur d'exécution entièrement configuré pour l'apprentissage profond et un accès gratuit à un GPU robuste [37].

- **Kaggle**

Kaggle est une grande communauté en ligne pour les scientifiques des données qui accueillent et participent à des concours d'apprentissage automatique. Cette plateforme fournit des datasets et des notebooks gratuits avec des différents environnements expérimentaux CPU, GPU et TPU gratuits.

Le tableau 3.1 montre les paramètres de l'environnement utilisé dans le travail :

Les types de notebook	Le score du CPU	Mémoire	Nombre de notebooks pouvant être utilisés en même temps	Quota par semaine
CPU	4	16 GB	10	Illimité
GPU	2	13 GB	2	30 heures
TPU	4	16 GB	2	30 heures

**Tableau 3.1 :** Les paramètres d'environnement kaggle.

Dans le cadre de la réalisation d'un travail performant et précis en termes de résultats, nous avons préféré travailler avec kaggle en raison des avantages qu'il offre et qui nous ont permis d'atteindre nos objectifs.

### 3.2.2 Bibliothèques utilisées

Après avoir choisis l'environnement de travail kaggle, on a besoin des bibliothèques pour appliquer les commandes et les algorithmes de l'apprentissage. Le tableau 3.2 présente les quatre bibliothèques utilisées pour cette réalisation [22].

<b>Bibliothèque</b>	<b>Description</b>
<b>Numpy</b>	Permet de créer et manipuler des matrices simplement et avec efficacité.
<b>Matplotlib</b>	Permet de visualiser nos datasets, nos fonctions, nos résultats sous forme de graphes, courbes et nuages de points.
<b>Scikit-learn</b>	Est la librairie qui contient toutes les fonctions de l'état de l'art du Machine Learning. On y trouve les algorithmes les plus importants ainsi que diverses fonctions de pre-processing.
<b>Pandas</b>	Est une excellente librairie pour importer vos tableaux Excel (et autres formats) dans Python dans le but de tirer des statistiques et de charger votre Dataset dans Sklearn.

**Tableau 3.2 :** Bibliothèques utilisées.

### 3.3 Implémentation

La figure 3.1 illustre les étapes de mise en œuvre de notre approche basée sur des techniques avancées de modélisation et d'analyse de données, notamment le ML et le DL, pour la détection des maladies cardiaques en utilisant le dataset PKIHD. Notre approche repose sur l'utilisation de ces techniques pour entraîner un modèle de détection de maladies cardiaques précis et fiable.

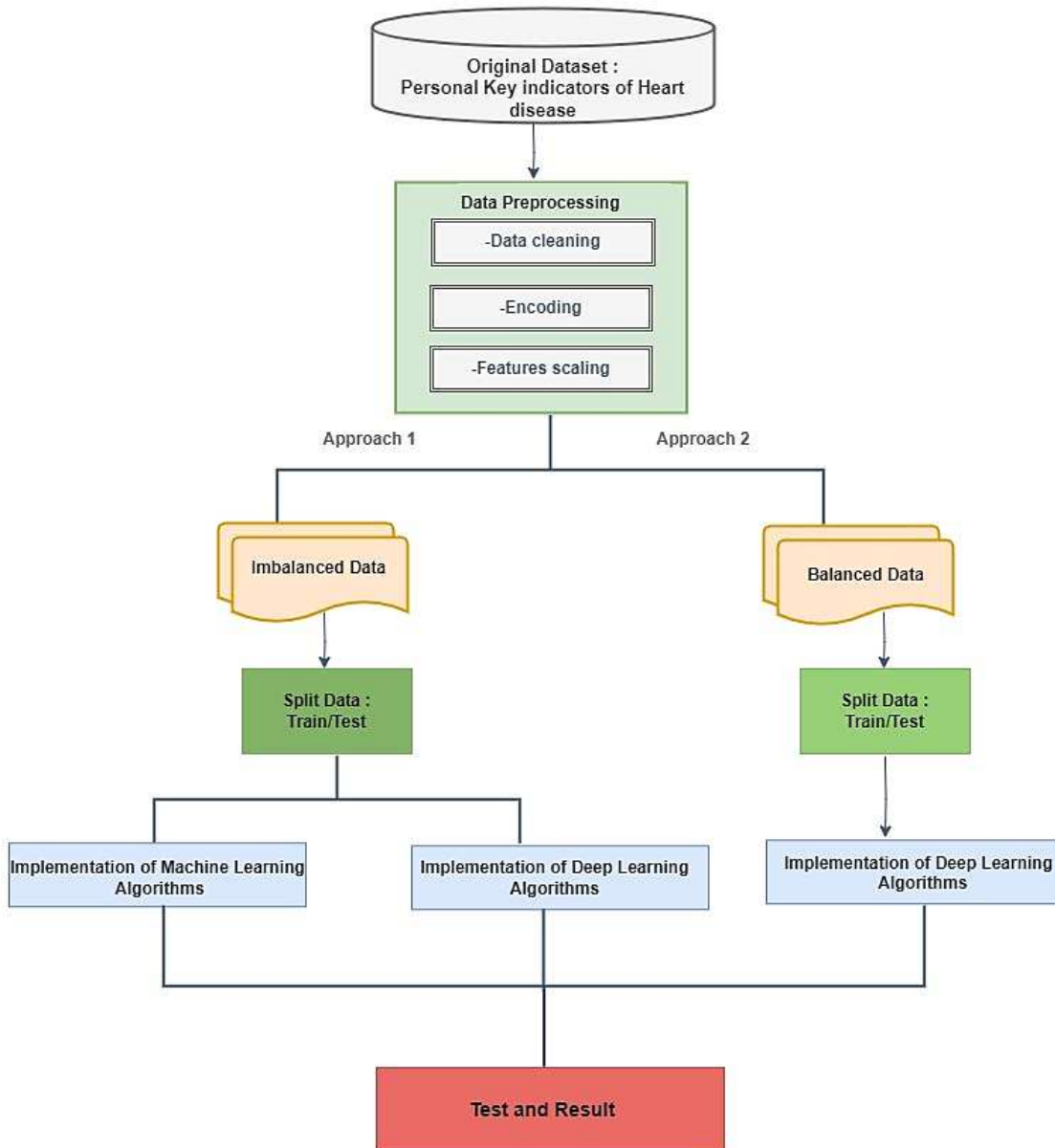


Figure 3.1 : Approche de détection des maladies cardiaques.

### **3.4 Dataset**

Le dataset choisis pour cette étude est "Personal key indicators of heart disease". La construction de cet ensemble de données a été réalisée par l'enquête annuelle du CDC (center for disease control and prevention) en 2020 effectuée sur 400 000 adultes américains basée sur leur état de santé. D'après le CDC, il existe environ la moitié des américains (47%) présentent au moins un des trois principaux facteurs risque des maladies cardiaques : l'hypertension artérielle, l'hypercholestérolémie et le tabagisme. Les chercheurs veulent détecter et prévenir les facteurs qui ont le plus d'impact sur les maladies cardiaques, ce qui est très important dans le domaine des soins de santé, grâce à des modèles qui ont été construits à partir des données à l'aide des méthodes d'apprentissage automatique [38].

### **3.5 Description du dataset**

A l'origine, le dataset contient 300 variables puis ce nombre a été réduit à une vingtaine de variables seulement. Ce dataset peut être utilisé pour appliquer une série de méthodes d'apprentissage automatique. On a traité la variable « Heart Disease » comme une variable binaire. Yes : présente une maladie cardiaque et No : n'a pas de maladie cardiaque [38].

Le tableau 3.3 décrit les attributs du dataset utilisé associés avec leurs types.

	<b>Attributs</b>	<b>Description</b>	<b>Type</b>
1	<b>BMI</b>	Indice de masse corporelle (12 à 94.8 kg).	Numérique
2	<b>Smoking</b>	Fumé au moins 100 cigarettes au cours de votre vie (5 paquets =100 cigarettes).	Catégorique
3	<b>AlcoholDrinking</b>	Gros buveurs (hommes adultes buvant plus de 14 verres par semaine et femme adultes buvant plus de 7 verres par semaine).	Catégorique
4	<b>Stroke</b>	Vérification si vous avez un accident vasculaire cérébral.	Catégorique
5	<b>PhysicalHealth</b>	Le nombre de jours au cours des 30 derniers années ont-ils été consacrés à la santé physique (de 0 à 30).	Numérique
6	<b>MentalHealth</b>	La santé mentale, combien de jours au cours des 30 derniers jours ont été difficiles (de 0 à 30).	Numérique
7	<b>DiffWalking</b>	Est-ce que vous avez de sérieuses difficultés à marcher ou à monter les escaliers.	Catégorique
8	<b>Sex</b>	Le sexe du patient (un homme ou une femme)	Catégorique
9	<b>AgeCategory</b>	Catégorie d'âge de treize niveaux.	Catégorique
10	<b>Race</b>	Valeur attribuée à la race/ethnie (white, Hispanic, black)	Catégorique
11	<b>Diabetic</b>	Si vous avez du diabète.	Catégorique
12	<b>PhysicalActivity</b>	Adulte ayant déclaré avoir fait de l'activité physique ou de l'exercice au cours des 30 derniers jours en dehors de leur travail habituel.	Catégorique
13	<b>GebHealth</b>	L'état de votre santé en général.	Catégorique
14	<b>SleepTime</b>	Nombre d'heures de sommeil avez-vous par période de 24 heures.	Numérique
15	<b>Asthma</b>	Si vous avez de l'asthme.	Catégorique
16	<b>KidneyDisease</b>	Si vous souffriez d'une maladie rénale.	Catégorique
17	<b>SkinCancer</b>	Si vous aviez un cancer de la peau.	Catégorique
18	<b>HeartDisease</b>	Maladie cardiaque (Yes ou No).	Catégorique

**Tableau 3.3 :** Description des attributs de dataset.

Le dataset contient 319795 instances et 18 attributs (1 attribut de décision +17 attributs pour les indicateurs clés du diagnostic). Le tableau 3.4 résume toutes les informations importantes sur le dataset utilisé.

Nom du dataset	Personal key indicators of heart disease
Année de la réalisation	2020
Type du dataset	Classification Binaire
Nombre total des instances	319795
Nombre des attributs	18

**Tableau 3.4 :** Caractéristiques générales du dataset.

### 3.6 Prétraitement de dataset

Les applications des algorithmes de Machine Learning et Deep Learning ont besoin des données. Ces données peuvent être incomplètes, dupliquées ou bien même incorrectes c'est-à-dire de mauvaises qualités, alors l'algorithme qui en résultera sera lui-même assez mauvais puisqu'il est censé reproduire ce qu'il voit dans le data. Donc, il est essentiel de prétraiter ces données avant de les introduire dans le modèle d'apprentissage. Cette phase est connue sous le nom de prétraitement des données.

### 3.7 Nettoyage de dataset

Après avoir analysé l'ensemble de données de notre dataset 'Personal Key Indicators of Heart Disease' qui contient 319795 instances et 18 attribues, et pour garantir la qualité et la fiabilité des données, nous avons éliminé la colonne 'Race' car elle n'a pas une influence sur la détection cardiaque. Puis, nous avons supprimé les lignes dupliquées afin de réduire la taille de notre dataset, à l'aide de la fonction **deduplicated**.

Forme des données avant : (319795, 17)

Nombre des lignes dupliqués : 24345

Forme des données après : (295450, 17)

Après, nous avons numérisé notre dataset en transformant toutes les valeurs présentées dans les features de type catégorique aux valeurs de type Int64 à l'aide de la fonction **replace**.

Pour les features {HeartDisease, Smoking, AlcoholDrinking, Stroke, DiffWalking, PhysicalActivity, Asthma, KidneyDisease, SkinCancer} qui prennent des valeurs de type object (Yes et No), nous avons remplacé 'No' par '0' et 'Yes' par '1'.

Le tableau 3.5 montre la transformation de type du reste des features.

Valeur catégorique	Valeur transformée
AgeCategory	[18-24] :0, [25-29] :1, [30-34] :2, [35-39] :3, [40-44] :4, [45-49] :5, [50-54] :6, [55-59] :7, [60-64] :8, [65-69] :9, [70-74] :10, [75-79] :11, [80 or older] :12
Sex	'Female' :0, 'Male' :1
GenHealth	'Poor' :0, 'Fair' :1, 'Good' :2, 'Very good' :3, 'Excellent' :4
Diabetic	'No' :0, 'No, borderline diabetes' :1, 'Yes (during pregnancy)' :2, 'Yes' :3

**Tableau 3.5 :** Changement de types des features.

### 3.8 Création du modèle de détection de la maladie cardiaque

Dans cette section, nous avons utilisé deux approches pour améliorer les capacités du modèle à identifier automatiquement les facteurs de risques et les indicateurs de la maladie qui influencent sur le côté de diagnostic et du traitement efficace des patients atteints de maladie cardiaque.

#### 3.8.1 Première approche (Imbalanced data)

Dans cette approche, nous avons travaillé sur notre dataset qui était à l'origine déséquilibrée (imbalanced data) en termes de target 'HeartDisease', qui à son tour contenait deux valeurs non balancées comme indiqué dans la figure 3.2 (268202 cas ne présentent pas de maladie alors que 27248 cas présentent une maladie).

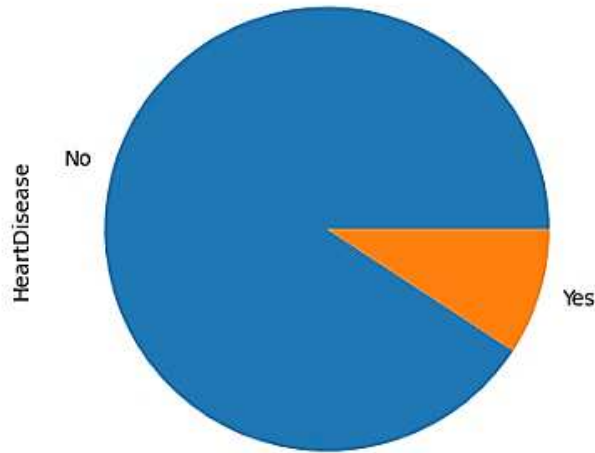


Figure 3.2 : Visualisation de la classe Target.

### 3.8.1.1 Entraînement de modèle

Un modèle est construit sur un ensemble de données d'entraînement. L'évaluation des performances d'un modèle sur le même ensemble est une mauvaise pratique en science des données.

Par conséquent, les data doivent être divisées en deux parties : un ensemble d'entraînement pour la formation du modèle, et un ensemble de test, réservé uniquement à l'évaluation des performances du modèle. La répartition généralement préférée est la suivante : 20 % de données de test, 80 % de données d'entraînement (voir la figure 3.3). Pour ce faire, nous utilisons la fonction **train\_test\_split**. Cette fonction répartit aléatoirement le dataset avant de le diviser en deux parties [23].



Figure 3.3 : Training et test split



### 3.8.1.2 Métriques d'évaluation du modèle

Dans cette section, nous définissons les principaux éléments de base des mesures que nous utiliserons pour évaluer les modèles de classification.

- **Matrices de confusion**

Une matrice de confusion décrit les performances du modèle de classification. Autrement dit, c'est un moyen de résumer les performances d'un classificateur. Le tableau 3.6 montre une représentation de base d'une matrice de confusion et montre comment les résultats prédits par le modèle se comparent aux valeurs réelles [23] :

	<b>Prédit 0</b>	<b>Prédit 1</b>
<b>Réelles 0</b>	TN	FP
<b>Réelles 1</b>	FN	TP

**Tableau 3.6 :** Matrice de confusion.

Chaque colonne de la matrice de confusion représente un facteur d'évaluation. Voici un petit aperçu de ces facteurs :

**True positive (TP) :** Il s'agit du nombre de résultats qui étaient positifs à l'origine et qui ont été prédits comme positifs.

**True négative (TN) :** Il s'agit du nombre de résultats qui étaient négatifs à l'origine et qui ont été prédits négatifs.

**False positive (FP) :** Il s'agit du nombre de résultats qui étaient négatifs à l'origine mais qui ont été prédits positifs. Cette erreur est également appelée erreur de type1.

**False négative (FN) :** Il s'agit du nombre de résultats qui étaient initialement positifs mais qui ont été prédits négatifs. Cette erreur est également appelée erreur de type2.

- **Accuracy**

"Accuracy" est une métrique utilisée pour mesurer la capacité du modèle à classer correctement toutes les instances et elle est considérée comme l'une des plus simples métriques pour mesurer les performances [28]. Elle est définie comme suit :

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3.1)$$

- **Null accuracy**

Null accuracy est la précision que l'on peut obtenir en prédisant la classe la plus fréquente. Si nous n'exécutons pas d'algorithme et que nous nous contentons de prédire la précision sur la base du résultat le plus fréquent, la précision calculée sur la base de cette prédiction est connue sous le nom de précision nulle [23] :

$$\text{Null Accuracy} = \frac{M}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3.2)$$

Avec M= Nombre total d'instances de la classe la plus fréquente.

- **Precision**

Cette métrique mesure la capacité du modèle à classer correctement les étiquettes positives (l'étiquette qui représente l'occurrence de l'événement) en la comparant au nombre total d'instances prédits comme positives comme montre l'équation suivant [28] :

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3.3)$$

- **Sensitivity (Recall)**

Cette métrique mesure le nombre d'étiquettes positives correctement prédites par rapport à l'ensemble des étiquettes positives comme le montre l'équation suivant [28] :

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3.4)$$

- **F1 score**

Le F1 score est une valeur numérique unique comprise entre 0 et 1, utilisée pour évaluer les performances d'un modèle en termes de précision et de rappel [28].

$$\text{F1 score} = 2 \times \left( \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \right) \quad (3.5)$$

### 3.8.1.3 Implémentation des algorithmes de Machine Learning

Dans cette section, nous allons présenter la première approche avec "imbalanced data" en appliquant les algorithmes du ML qui sont mentionnés dans le chapitre précédent.

### 3.8.1.4 Evaluation des résultats

Le tableau 3.7 évalue les résultats obtenus par l'application de 3 algorithmes de ML sur le dataset non équilibré (imbalanced data) en utilisant les métriques d'évaluation. La Figure 3.4 montre ces résultats sous forme d'histogramme.

Algorithmes de ML	Accuracy	Precision	F1-score	Recall
DT	90.72%	90.72%	47.57%	90.72%
RFC	89.56%	89.56%	56.01%	89.56%
KNN	88.99%	88.99%	58.06%	88.99%

**Tableau 3.7 :** Comparaison des performances des algorithmes de ML.

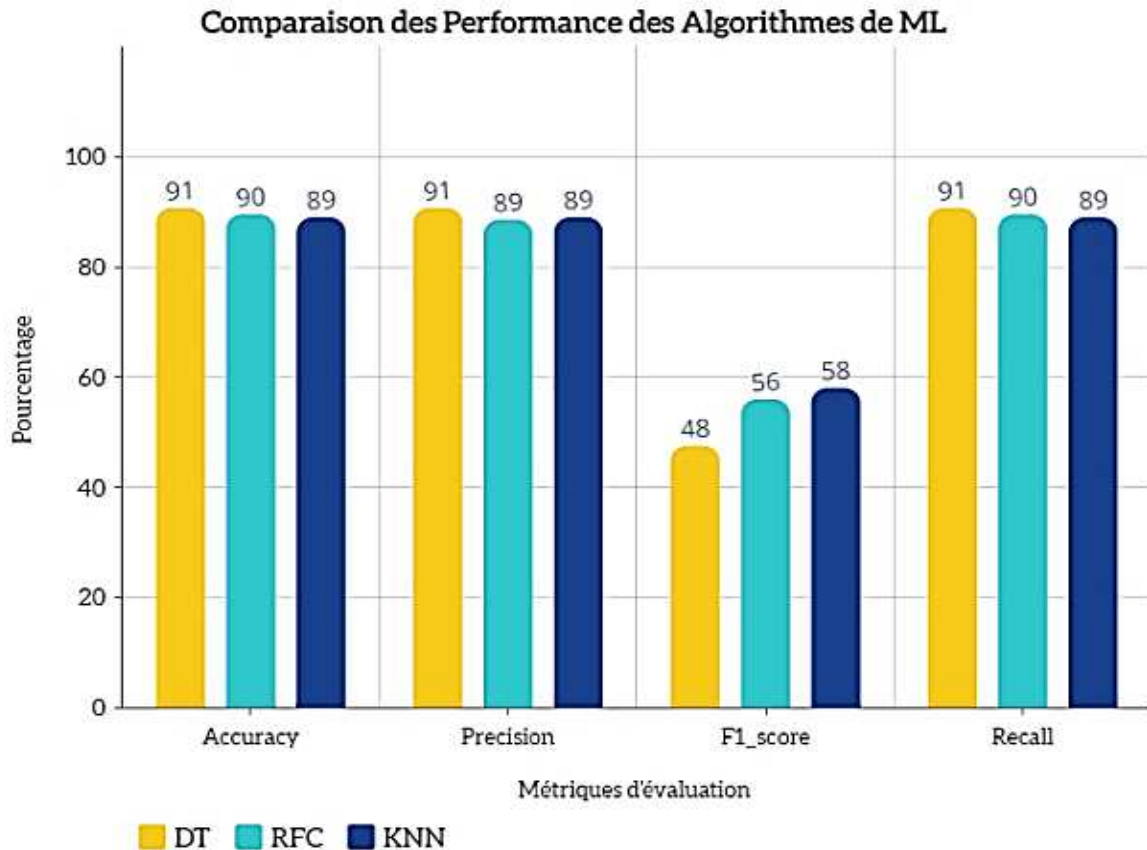


Figure 3.4 : Diagramme de comparaison entre les algorithmes de ML.

### 3.8.1.5 Hyperparamètres de Deep Learning

Pour l'évaluation de notre modèle, nous avons appliqué des paramètres d'entraînements de modèle de DL, dont leur ajustement est important pour obtenir des prédictions précises à partir des données.

- **Les optimiseurs**

Les optimiseurs sont des algorithmes utilisés pour minimiser une fonction d'erreur (fonction de perte) ou pour maximiser l'efficacité de la production et aussi pour ajuster les poids du modèle pendant l'entraînement tels que adam, adamax, rmsprop, etc. L'estimation adaptative du moment (Adam) est un algorithme pour la technique d'optimisation pour entraîner le modèle. Il est très efficace lorsqu'il s'agit de travailler sur des problèmes de grande complexité impliquant un grand nombre de données ou de paramètres [30]. Adamax est une méthode d'optimisation qui est une variante d'Adam basée sur la norme infinie. En raison de sa capacité à ajuster le taux d'apprentissage en fonction des caractéristiques des données, elle est adaptée à l'apprentissage de processus variables dans le temps [39].

- **Le taux d'apprentissage (Learning rate)**

C'est la taille de l'étape que l'algorithme de descente de gradient effectue vers un minimum de la fonction de perte pendant l'apprentissage, ce qui correspond au processus de guidage pour l'optimisation des poids et des biais du modèle [40].

- **Nombre d'itération (epochs)**

C'est un hyperparamètre qui définit le nombre de fois que l'ensemble des données doit être traité par un algorithme d'apprentissage [41].

- **Taille du lot (batch size)**

Il correspond au nombre d'instances à donner au réseau neuronal au cours d'une itération, qui seront utilisées pour effectuer un passage avant et un retour en arrière dans le modèle. Cette procédure permet d'améliorer les capacités du modèle aux données d'apprentissage à cause du changement de ces données à chaque itération [28].

- **Régularisation**

C'est une technique utilisée pour améliorer un modèle qui souffre d'un sur-ajustement (overfitting), ce qui signifie que l'objectif principal de cet hyperparamètre est d'augmenter la capacité de généralisation du modèle. Il existe plusieurs méthodes de régularisation telles que L1 et L2 et dropout qui sont plus connues et fiables [28].

- **Les couches cachées (number of hidden layers)**

Ce sont des couches situées entre les couches d'entrée et les couches de sortie, qui permettent aux neurones artificiels de recevoir un ensemble d'entrées pondérées et de produire une sortie par l'intermédiaire d'une fonction d'activation [42].

- **Nombre d'unités**

Le nombre d'unités est également appelé nœuds, chacun ayant son propre traitement dans chaque couche du réseau neuronal, chaque unité prend en entrée les sorties des unités de la couche précédente et applique une fonction mathématique pour générer sa sortie, puis elle transmet à la couche suivante.

### 3.8.1.6 Evaluation des résultats

Dans cette partie, nous avons utilisé les algorithmes de DL pour une autre phase d'évaluation de la première approche (imbalanced data) en effectuant une hiérarchisation interne, comme indiqué dans la figure 3.5.

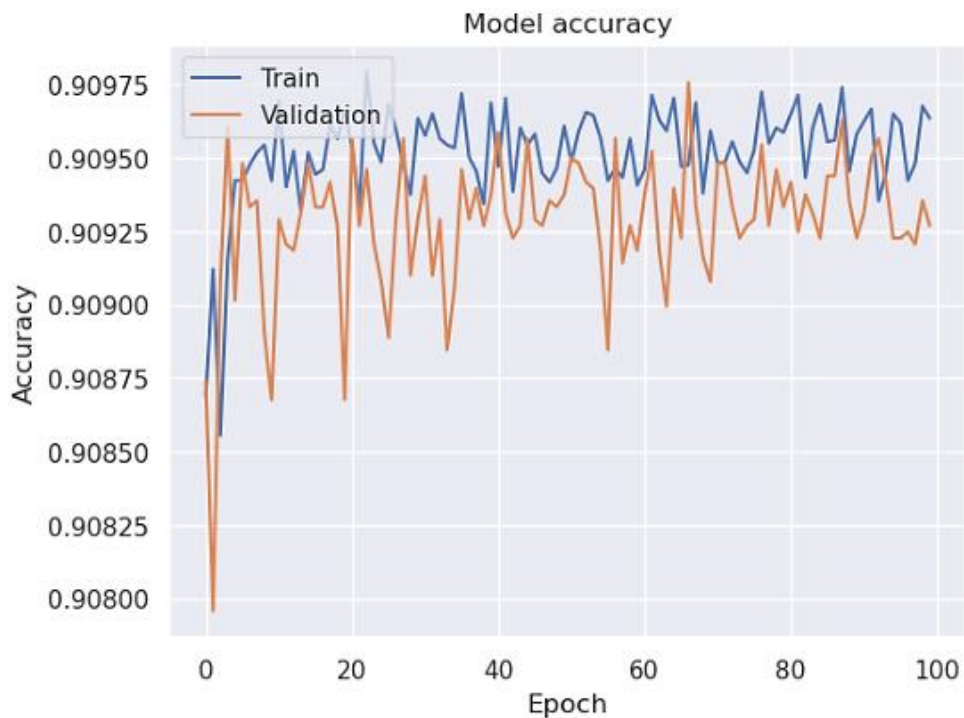
```
Model: "sequential_1"
-----
```

Layer (type)	Output Shape	Param #
dense_5 (Dense)	(None, 128)	2176
dense_6 (Dense)	(None, 64)	8256
dense_7 (Dense)	(None, 32)	2080
dense_8 (Dense)	(None, 32)	1056
dense_9 (Dense)	(None, 1)	33

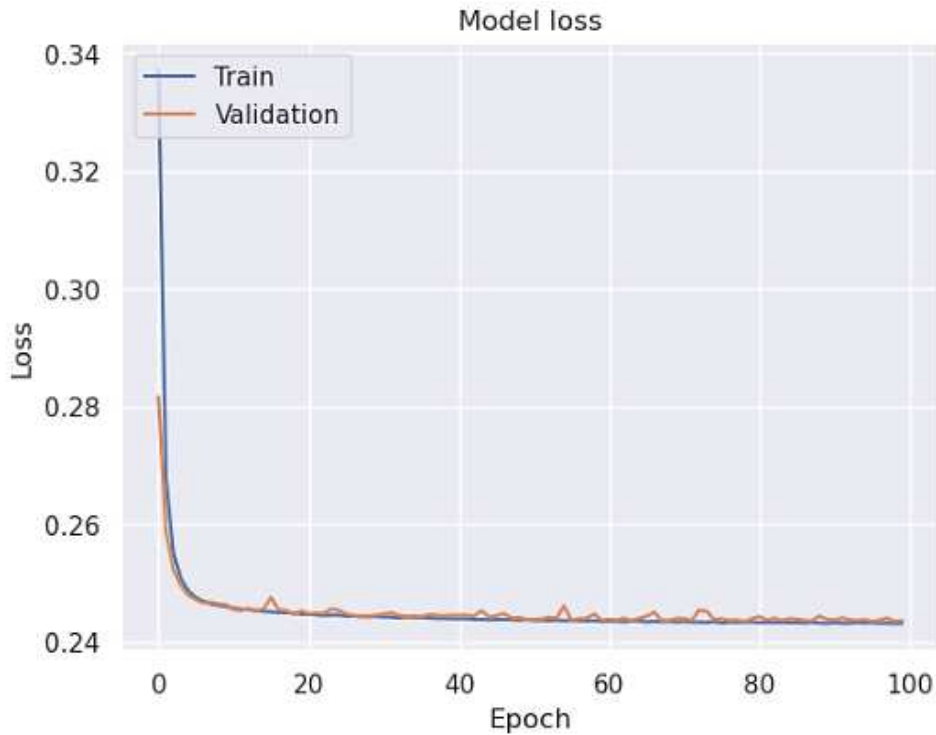
```
-----
Total params: 13,601
Trainable params: 13,601
Non-trainable params: 0
```

**Figure 3.5 :** Résumé du modèle.

Suite à l'application des algorithmes de DL, nous avons visualisé graphiquement les résultats obtenus en termes d'accuracy et loss comme le montre les figures 3.6 et 3.7.



**Figure 3.6 :** Model accuracy.

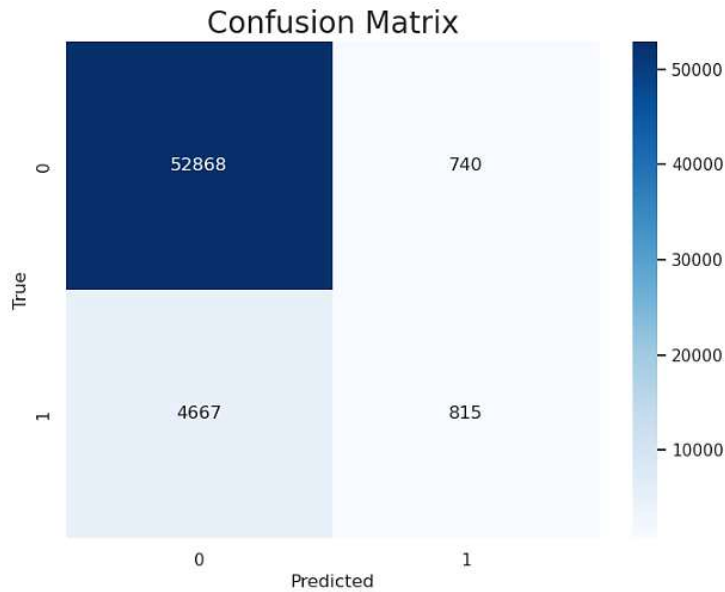


**Figure 3.7 :** Model loss.

Le tableau 3.8 récapitule tous les résultats obtenus en termes de performance du modèle grâce à l'utilisation des métriques d'évaluation. La matrice de confusion correspondante est donnée par la figure 3.8.

Heart Disease	Accuracy	Precision	Recall	F1-score
0	90.95%	92%	99%	95%
1	90.95%	52%	15%	23%

**Tableau 3.8 :** Métriques d'évaluation pour le modèle DL.



**Figure 3.8 :** Matrice de confusion.

Nous constatons que les résultats obtenus avec cette approche sont insuffisants pour établir un modèle précis. Donc, nous avons choisi une autre approche basée sur les données équilibrées (balanced data) de façon à perfectionner notre modèle et améliorer nos résultats.

### 3.8.2 Deuxième approche (balanced dataset)

Les techniques de ML et DL échouent souvent ou donnent des résultats faussement optimistes sur les éléments de data de classification dont la distribution des classes est déséquilibrée. En effet, de nombreux algorithmes sont conçus pour fonctionner sur des données de classification comportant un nombre égal d'observations pour chaque classe. Lorsque ce n'est pas le cas, les algorithmes peuvent apprendre que les quelques exemples de classes majoritaires, les classes minoritaires ne sont pas importants et peuvent être ignorés afin d'obtenir de bonnes performances [43].

Data sampling (l'échantillonnage des données) est un ensemble de techniques qui transforment la partie de données d'apprentissage (training data) afin d'équilibrer ou de mieux balancer la distribution des classes. Une fois équilibrés, les algorithmes de ML et DL standard peuvent être formés directement sur l'ensemble de données transformé, sans aucune modification [43].

#### 3.8.2.1 Méthodes de data Sampling

Parmi les méthodes de data sampling utilisées dans le contexte des problèmes de classification binaire pour adapter le problème de "imbalanced data" et que nous avons choisi pour évaluer notre modèle sont les suivantes [43] :



- **Techniques d'Oversampling**

Ces techniques dupliquent les exemples de la classe minoritaire ou synthétisent de nouveaux exemples à partir des exemples de la classe minoritaire, et parmi ces techniques, on cite :

- **Random Oversampling**

C'est la méthode la plus simple qui consiste à dupliquer aléatoirement des exemples de la classe minoritaire dans l'ensemble de données d'apprentissage.

- **Synthetic Minority Oversampling Technique (SMOTE)**

Cette technique peut-être la plus efficace, elle consiste à sélectionner des exemples proches dans l'espace des features, à tracer une ligne entre les exemples dans l'espace des features et à dessiner un nouvel échantillon (sample) comme un point le long de cette ligne.

- **Techniques Undersampling**

Les méthodes d'Undersampling (sous-échantillonnage) suppriment ou sélectionnent un sous-ensemble d'exemples de la classe majoritaire, et parmi ces techniques nous avons pris :

- **Random Undersampling**

Une technique simple consiste à supprimer au hasard des exemples de la classe majoritaire dans l'ensemble de données d'apprentissage (training data).

### 3.8.2.2 Entraînement du modèle

Dans cette partie, le modèle entraîné est identique à celui utilisé dans la première approche (dédiée pour "imbalanced data") en appliquant les algorithmes de DL.

### 3.8.2.3 Evaluation des résultats

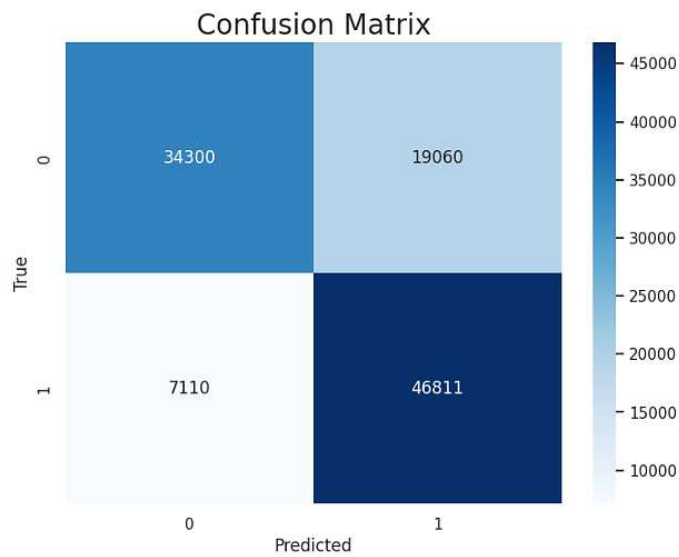
Dans cette section, nous avons utilisé les méthodes de data sampling (Random Oversampling, SMOTE, Random Undersampling) suivi d'une comparaison entre les résultats trouvés des métriques pour chaque technique.

#### **Première méthode (Random Oversampling)**

Les métriques d'évaluation en terme d'accuracy, de précision, de recall et de F1-score ainsi que la matrice de confusion sont données respectivement par le tableau 3.9 et la figure 3.9.

Random Oversampling				
Heart Disease	Accuracy	Precision	Recall	F1-score
0	76%	83%	64%	72%
1	76%	71%	87%	78%

**Tableau 3.9 :** Résultats des métriques de méthode Random Oversampling.



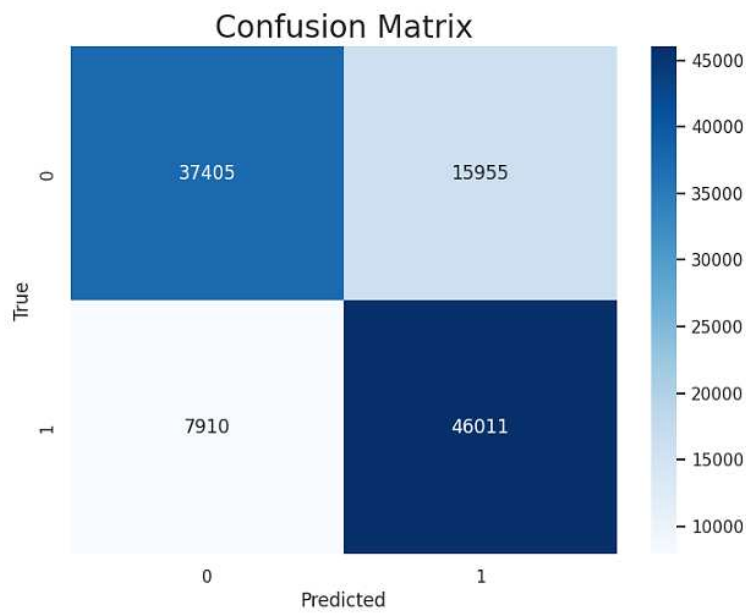
**Figure 3.9 :** Matrice de confusion de méthode Random Oversampling.

### Deuxième méthode (SMOTE)

Les métriques d'évaluation en terme d'accuracy, de précision, de recall et de F1-score ainsi que la matrice de confusion sont données respectivement par le tableau 3.10 et la figure 3.10.

SMOTE				
Heart Disease	Accuracy	Precision	Recall	F1-score
0	78%	83%	70%	76%
1	78%	74%	85%	79%

**Tableau 3.10** : Résultats des métriques de méthode SMOTE.



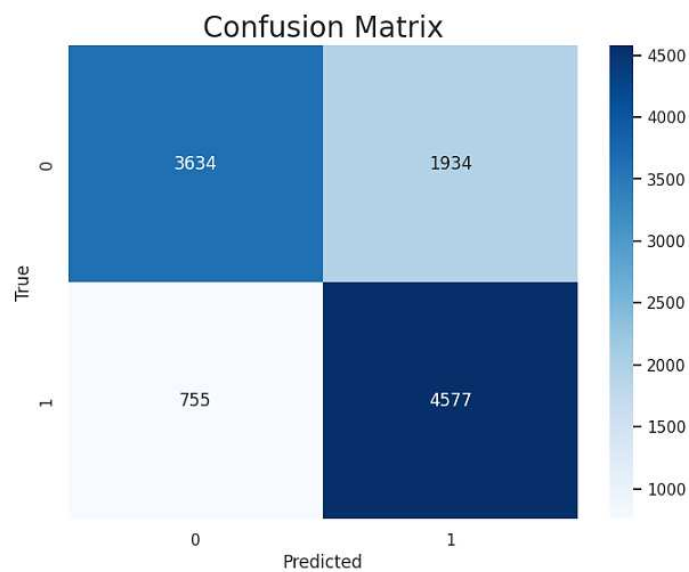
**Figure 3.10** : Matrice de confusion de méthode SMOTE.

### Troisième méthode (Random undersampling)

Les métriques d'évaluation en terme d'accuracy, de précision, de recall et de F1-score ainsi que la matrice de confusion sont données respectivement par le tableau 3.11 et la figure 3.11.

Random Undersampling				
Heart Disease	Accuracy	Precision	Recall	F1-score
0	75%	83%	65%	73%
1	75%	70%	86%	77%

**Tableau 3.11** : Résultats des métriques de méthode Random Undersampling.



**Figure 3.11** : Matrice de confusion de méthode Random Undersampling.

Les résultats trouvés pour cette approche montrent que nous avons pu améliorer notre modèle après la balance du dataset, voyant que le modèle a pris en considération le traitement des données non nul de la classe binaire.

Pour chaque méthode de sampling les métriques d'évaluation (précision, F1-score et recall) ont été améliorés, alors que l'accuracy n'est pas bien améliorée et c'est normal pour le cas d'un dataset équilibré.

### **3.9 Etude comparative**

Notre étude porte sur l'impact de la balance des données sur les résultats d'apprentissage pour une tâche de classification binaire.

Nous avons utilisé le dataset "Personal Key Indicators of Heart Disease (PKIHD)" pour l'étude de deux approches, un ensemble de données non équilibré (imbalanced data) où la classe non nul est sous-représentée par rapport à l'autre, et un ensemble de données équilibré (balanced data), où les deux classes sont représentées de manière égale. Pour chaque approche nous avons utilisé les algorithmes d'apprentissage de Machine Learning et de Deep Learning.

Les critères de comparaison que nous avons pris en considération incluent l'accuracy, la précision, F1-score et la sensibilité (recall).

L'objectif de notre étude était de déterminer comment l'équilibrage des données affecte les résultats d'apprentissage et quel algorithme offre les meilleurs résultats pour une tâche de classification binaire.

Nos résultats ont montré que la balance des données améliorait significativement la précision, le F1-score et le recall pour les algorithmes de DL mais que l'accuracy était mal performé. Cela indique que le DL est plus performant pour discriminer les deux classes, même sur un ensemble de données déséquilibré. Cependant, le ML était plus rapide et plus facile à mettre en œuvre.

### **3.10 Conclusion**

L'objectif de ce dernier chapitre est de réaliser une simulation sur le dataset en deux approches choisies, tout en comparant les performances des métriques d'évaluation du modèle pour chaque application des techniques d'apprentissage de ML et de DL qui répondent à nos besoins pour atteindre un modèle plus amélioré de côté de détection et prédiction efficace et plus adapté au suivi des maladies cardiaques.

# Conclusion générale

L'intérêt croissant pour l'amélioration de la prédiction des maladies cardiaques est un domaine crucial de la médecine préventive qui peut aider à identifier les risques potentiels de cette maladie chez les patients avant qu'ils ne se développent en problèmes plus graves. Les progrès récents dans la technologie de l'apprentissage automatique, notamment l'utilisation de l'apprentissage profond, ont permis de développer des modèles de prédiction de haute précision pour des pathologies cardiaques.

Ces modèles de prédiction peuvent aider les professionnels de la santé à prendre des décisions éclairées pour prévenir les maladies cardiaques et à mettre en place des plans de traitement précoces et appropriés pour les patients à haut risque. Cependant, il est important de souligner que la prédiction des maladies cardiaques doit être utilisée en conjonction avec d'autres méthodes de diagnostic et ne doit jamais remplacer les examens médicaux et les tests diagnostiqués traditionnels.

Le travail mené dans ce mémoire avait pour objectif de détecter les maladies cardiaques à l'aide des techniques du ML et de DL. Notre étude porte sur l'impact de la balance des données sur les résultats d'apprentissage pour une tâche de classification binaire. Nous avons utilisé le dataset "Personal Key Indicators of Heart Disease (PKIHD)" pour l'étude de deux approches, un ensemble de données non équilibré (imbalanced data) où la classe non nul est sous-représentée par rapport à l'autre, et un ensemble de données équilibré (balanced data), où les deux classes sont représentées de manière égale. Pour chaque approche, nous avons utilisé les algorithmes d'apprentissages de Machine Learning et de Deep Learning.

Les critères de comparaison que nous avons pris en considération incluent l'accuracy, la précision, F1-score et la sensibilité (recall).

L'objectif de notre étude était de déterminer comment la balance des données affecte les résultats d'apprentissage et quel algorithme offre les meilleurs résultats pour une tâche de classification binaire.

Nos résultats ont montré que la balance des données améliorait significativement la précision, le F1-score et le recall pour les algorithmes de DL mais que l'accuracy était mal améliorée. Cela indique que le DL est plus performant pour discriminer les deux classes, même sur un ensemble de données déséquilibré. Cependant, le ML était plus rapide et plus facile à mettre en œuvre.

Ce mémoire nous a permis de découvrir de nouveaux aspects, qui nous ont conduits à acquérir des connaissances dans le domaine de l'analyse des données, ce qui nous a ouvert les yeux sur nos besoins et nos manques qui nous ont servi d'améliorer la qualité de notre travail.

Les travaux réalisés durant ce mémoire nous ouvrent différentes perspectives. Une première perspective de recherche vise à utiliser un dataset contenant des données de type images et par conséquent créer un modèle de réseaux de neurones convolutionnels afin de détecter et classifier les maladies cardiaques. Une seconde perspective porte sur l'utilisation d'un dataset contenant des signaux ECG et par suite créer un modèle à base de réseaux de neurones récurrents pour réaliser la tâche de classification.

# Bibliographie

[1] L'anatomie et le fonctionnement du cœur. URL : <https://www.news-medical.net/health/Structure-and-Function-of-the-Heart.aspx> Consulté le 31/03/2023.

[2] L'anatomie et le fonctionnement du cœur, l'arrêt cardiaque et symptômes des maladies cardiaques. URL : <https://www.msmanuals.com/home> Consulté 31/03/2023.

[3] WEINHAUS, Anthony J. et ROBERTS, Kenneth P. Anatomy of the human heart. *Handbook of cardiac anatomy, physiology, and devices*, 2005, p. 51-79.

[4] Circulation sanguine et Insuffisance cardiaque URL : <https://my.clevelandclinic.org/health> Consulté le 31/03/2023.

[5] Le battement du cœur URL : <https://www.hopkinsmedicine.org/health/conditions-and-diseases/anatomy-and-function-of-the-hearts-electrical-system> Consulté le 01/04/2023.

[6] Maladies cardiovasculaires et angine de poitrine URL : <https://www.who.int/health-topics> Consulté le 04/04/2023.

[7] Maladies coronariennes URL : <https://www.webmd.com/heart-disease/heart-disease-types-causes-symptoms> Consulté le 04/04/2023.

[8] Infarctus de myocarde URL : <https://www.nhlbi.nih.gov/health/coronary-heart-disease> Consulté le 06/04/2023.

[9] Asthme cardiaque URL : <https://www.ncbi.nlm.nih.gov/books/NBK537076/#article-25460.r1> Consulté le 06/04/2023.

[10] Arrêt cardiaque URL : <https://www.urmc.rochester.edu/encyclopedia/content.aspx?ContentTypeID=56&ContentID=DM164> Consulté le 10/04/2023.

[11] URL : <https://www.heart.org/en/health-topics/heart-attack/about-heart-attacks/heart-attack-or-sudden-cardiac-arrest-how-are-they-different> Consulté le 10/04/2023.

[12] Crise cardiaque URL : <https://www.coeuretavc.ca/articles/crise-cardiaque-ou-arret-cardiaque> Consulté le 10/04/2023.

[13] Analyse de sang URL : <https://www.heartfailurematters.org/fr/comprendre-linsuffisance-cardiaque/analyses-de-sang/> Consulté le 04/04/2023.

[14] MIRVIS, David M. et GOLDBERGER, Ary L. Electrocardiography. *Heart disease*, 2001, vol. 1, p. 82-128.



- [15] GOTTDIENER, John S., BEDNARZ, James, DEVEREUX, Richard, *et al.* American Society of Echocardiography recommendations for use of echocardiography in clinical trials: a report from the american society of echocardiography's guidelines and standards committee and the task force on echocardiography in clinical trials. *Journal of the American Society of Echocardiography*, 2004, vol. 17, no 10, p. 1086-1119.
- [16] CASILLAS, J. M., GREMEAUX, V., DAMAK, S., *et al.* Exercise training for patients with cardiovascular disease. In : *Annales de réadaptation et de médecine physique*. Elsevier Masson, 2007. p. 403-418.
- [17] FURBER, A., HELFT, G., CHASSAING, S., *et al.* Indications cliniques appropriées de l'IRM en pathologie cardio-vasculaire. *Archives of Cardiovascular Diseases Supplements*, 2009, vol. 1, no 1, p. 34-50.
- [18] DAVIDSON, Charles J. et BONOW, Robert O. Cardiac catheterization. *Libby P*, 1997, vol. 10.
- [19] Facteurs de risques URL : <https://www.mayoclinic.org/diseases-conditions/heart-attack/symptoms-causes/syc-20373106#> Consulté le 05/04/2023.
- [20] ATTIAS, D. BESSE, B et LELLOUCHE, N. KB Cardiologie vasculaire, éd. 2014.
- [21] ML URL : <https://datascientest.com/machine-learning-tout-savoir> Consulté le 16/03/2023.
- [22] SAINT-CIRGUE, Guillaume. Apprendre le Machine Learning en une semaine. *Machine Learnia, Londres*, 2019.
- [23] MOOCARME, Matthew, ABDOLAHNEJAD, Mahla, et BHAGWAT, Ritesh. *The Deep Learning with Keras Workshop : Learn how to define and train neural network models with just a few lines of code*. Packt Publishing Ltd, 2020.
- [24] Application de DL URL : <https://emeritus.org/blog/machine-learning-what-are-machine-learning-applications/> Consulté le 26/03/2023.
- [25] Application de DL URL : <https://www.javatpoint.com/examples-of-machine-learning> Consulté le 26/03/2023.
- [26] Application de DL URL : <https://www.mygreatlearning.com/blog/deep-learning-applications/> Consulté 24/03/2023.
- [27] Application de DL URL : <https://www.interviewbit.com/blog/applications-of-deep-learning/> Consulté le 25/03/2023.
- [28] SALEH, Hyatt. *The The Machine Learning Workshop : Get ready to develop your own high-performance machine learning algorithms with scikit-learn*. Packt Publishing Ltd, 2020.
- [29] Fonctionnement des ANN URL : <https://www.atriainnovation.com/en/what-are-neural-networks-and-their-functions/> Consulté le 09/05/2023.
- [30] Feedforward network URL : <https://www.geeksforgeeks.org/introduction-to-ann-set-4-network-architectures/> Consulté le 15/04/2023.

- [31] réseaux RNN URL : <https://www.upgrad.com/blog/types-artificial-neural-networks-in-machine-language/> Consulté le 12/04/2023.
- [32] Fonction d'activations URL : <https://www.kdnuggets.com/2022/06/activation-functions-work-deep-learning.html> Consulté le 20/03/2023.
- [33] Fonction perte URL : <https://www.datarobot.com/blog/introduction-to-loss-functions/> Consulté le dd/05/2023.
- [34] mean absolute error URL : <https://www.analyticsvidhya.com/blog/2022/06/understanding-loss-function-in-deep-learning/> Consulté le 04/05/2023.
- [35] categorical cross entropy URL : <https://vitalflux.com/keras-categorical-cross-entropy-loss-function/> Consulté le 06/05/2023.
- [36] ROLON-MÉRETTE, Damien, ROSS, Matt, ROLON-MÉRETTE, Thaddé, *et al.* Introduction to Anaconda and Python: Installation and setup. *Quant. Methods Psychol*, 2016, vol. 16, no 5, p. S3-S11.
- [37] CARNEIRO, Tiago, DA NÓBREGA, Raul Victor Medeiros, NEPOMUCENO, Thiago, *et al.* Performance analysis of google colaboratory as a tool for accelerating deep learning applications. *IEEE Access*, 2018, vol. 6, p. 61677-61685.
- [38] Kaggle dataset URL : <https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease> Consulté le 23/04/2023.
- [39] Optimiseurs URL : [https://www.tensorflow.org/api\\_docs/python/tf/keras/optimizers/experimental/Adamax](https://www.tensorflow.org/api_docs/python/tf/keras/optimizers/experimental/Adamax) Consulté le 06/05/2023
- [40] PASSOS, Dário et MISHRA, Puneet. A tutorial on automatic hyperparameter tuning of deep spectral modelling for regression and classification tasks. *Chemometrics and Intelligent Laboratory Systems*, 2022, p. 104520.
- [41] Nombre d'itération URL : <https://www.simplilearn.com/tutorials/machine-learning-tutorial/what-is-epoch-in-machine-learning> Consulté le 07/05/2023.
- [42] Couche caché URL : <https://www.techopedia.com/definition/33264/hidden-layer-neural-networks> Consulté le 07/05/2023.
- [43] BROWNLEE, Jason. *Imbalanced classification with Python : better metrics, balance skewed classes, cost-sensitive learning*. Machine Learning Mastery, 2020.