



الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التعليم العالي والبحث العلمي

جامعة سعيدة د. مولاي الطاهر

كلية الرياضيات و الإعلام الآلي و الاتصالات السلكية و  
اللاسلكية

قسم: الإعلام الآلي

## Mémoire de Master en informatique

Spécialité : MICR ( modélisation informatique connaissance  
raisonnement )

### Thème

Plateforme intelligente de santé basée sur  
le Machine Learning

▪ Présenté par :

Lakhdari Fatima Zohra

▪ Dirigé par :

Dr. Ahmed Zahaf

## *Remerciements*

*Chers professeurs, chers parents, chers amis,*

*Aujourd'hui, je me retrouve face à vous avec un cœur débordant de gratitude et une émotion que les mots peinent à contenir. Ce diplôme que je tiens entre les mains n'est pas seulement le fruit de mes efforts — il est le vôtre autant que le mien.*

*À mes professeurs : merci de n'avoir jamais laissé tomber un seul étudiant. Vous avez semé en moi non pas seulement du savoir, mais la curiosité de continuer à apprendre, longtemps après avoir quitté vos salles.*

*À ma famille : vous avez porté mes nuits de doute sans jamais me le montrer. Vous avez cru en moi même les jours où je n'y croyais plus moi-même. Ce moment, je vous le dois entièrement.*

*À mes amis et camarades de promotion : nous avons traversé ensemble les examens, les doutes, les rires et quelques crises de panique mémorables. Je ne remercierai jamais assez le hasard de nous avoir réunis.*

*Ce diplôme est un point d'arrivée, oui — mais surtout un point de départ. Je repars avec la ferme intention de mériter tout ce qu'on m'a donné.*

*Du fond du cœur : merci.*

*— Avec toute ma reconnaissance*

# Résumé du Mémoire — MED AI

## Présentation générale

Ce mémoire de Master en informatique (spécialité MICR) présente la conception et la réalisation de **MED AI**, une plateforme web intelligente dédiée à la détection précoce des maladies cardiovasculaires, combinant Machine Learning et interfaces médicales accessibles.

## Chapitre I — Fondements théoriques de l'IA

Le premier chapitre pose les bases conceptuelles en parcourant la hiérarchie des paradigmes IA modernes :

- **Intelligence Artificielle** : définie depuis McCarthy (1956), elle englobe les systèmes symboliques (règles logiques, ex. MYCIN) et connexionnistes (réseaux de neurones).
- **Machine Learning** : apprentissage supervisé, non supervisé et par renforcement. Concepts clés : overfitting, régularisation, validation croisée.
- **Deep Learning** : réseaux profonds avec rétropropagation du gradient, fonctions d'activation, architectures CNN/RNN/Transformer/GAN.
- **CNN-1D** : convolutions sur signaux unidimensionnels (ECG, séries temporelles), avec partage de paramètres et invariance locale.
- **Transformer** : mécanisme d'attention (Self-Attention), architecture encodeur-décodeur, variantes BERT, GPT, T5, ViT.

- **LLMs** : modèles à milliards de paramètres pré-entraînés puis affinés via RLHF (GPT-4, Claude, LLaMA, Mistral). Capacités émergentes, mais limités par les hallucinations et le coût computationnel.
- **LLMs Multimodaux** : fusion texte + image + audio + vidéo (GPT-4o, Gemini, Flamingo, LLaVA).

## Chapitre II — État de l'art du diagnostic cardiovasculaire

Les maladies cardiovasculaires causent **19,8 millions de décès par an** (OMS, 2022), avec une projection alarmante de 35,6 millions d'ici 2050. Le diagnostic traditionnel souffre de subjectivité, coûts élevés et accessibilité inégale.

### Travaux ML classiques recensés :

#### Algorithme      Meilleure performance rapportée

Random Forest 94 % (UCI Cleveland)

XGBoost          93 % (UCI + SMOTE)

LightGBM        90,6 % / AUC 0,811 (BRFSS)

SVM                97,36 % (Cleveland/Statlog)

DNN                98,29 % (Statlog)

AdaBoost         98,87 % (UCI Cleveland)

**Travaux LLMs médicaux recensés** : BioClinicalBERT (F1=0,911 en NER clinique), Med-PaLM 2 (86,5 % sur MedQA), GPT-4 (93,1 % sur MedQA, 64 % en diagnostic différentiel).

## Chapitre III — Expérimentation et choix du modèle

## Dataset

Le **Heart Disease Health Indicators Dataset** (BRFSS 2015, CDC) : 253 680 échantillons, 21 variables, cible binaire HeartDiseaseorAttack. Déséquilibre fort : seulement **9,4 % de cas positifs** (ratio 9,6:1).

### Contribution méthodologique majeure : le data leakage

L'auteure identifie et quantifie un biais critique très répandu dans la littérature — l'application de SMOTE **avant** la séparation train/test :

Pipeline	F1-Macro	AUC-ROC	Recall HD
Incorrect (Scale→SMOTE→Split)	<b>0,9401</b>	0,9827	91,4 %
Correct (Split→Scale→SMOTE)	<b>0,6071</b>	0,8500	18,0 %

Un écart de F1 de **+0,33** purement artificiel, qui fausse l'évaluation et ne correspond à rien de réel en production.

### Comparaison des 7 modèles (pipeline correct)

Modèle	Accuracy	AUC-ROC	F1-Macro
XGBoost	90,38 %	0,8432	0,6131
LightGBM	90,73 %	0,8501	0,5629
★ Ensemble (XGB+LGB+RF)	<b>90,61 %</b>	<b>0,8500</b>	<b>0,6071</b>
CNN	88,53 %	0,8350	0,6440
Bio_ClinicalBERT	90,13 %	0,8402	0,6099
Qwen 2.5 (zero-shot)	66,67 %	0,6667	0,6250

**Modèle retenu : Ensemble Learning (Voting Soft)**

Justifié par la **meilleure AUC-ROC** (0,8500), la diversité algorithmique (boosting + bagging), l'interprétabilité via SHAP, et l'inférence quasi-instantanée compatible avec un déploiement temps réel.

## **Chapitre IV — Implémentation de MED AI**

La plateforme repose sur une architecture **full-stack moderne** :

- **Base de données** : Supabase (PostgreSQL) avec Row Level Security
- **Authentification** : Supabase Auth (e-mail / Google), gestion des rôles (patient, médecin, administrateur)
- **Modèle IA** : microservice Python exposant l'Ensemble via API REST

### **Fonctionnalités principales**

#### **Espace patient :**

- Assistant conversationnel avec questionnaire de 21 questions (boutons Oui/Non)
- Résultat du diagnostic : pourcentage de risque + niveau qualitatif + recommandations personnalisées
- Alerte critique avec liste des facteurs détectés (HTA, diabète, tabagisme, etc.)
- Recherche de médecin et prise de rendez-vous en ligne

#### **Espace médecin :**

- Formulaire de saisie des données cliniques complètes d'un patient
- Vue synthétique des patients par niveau de risque (critique / élevé / faible)

- Gestion des dossiers cliniques avec export

**Sécurité** : authentification sécurisée, contrôle d'accès par rôle, RLS sur toutes les tables, avertissement clinique systématique.

### **Conclusion générale**

MED AI démontre qu'une démarche alliant rigueur méthodologique et utilité concrète peut produire un outil crédible scientifiquement et accessible cliniquement. Les apports principaux sont :

1. **La mise en évidence du data leakage SMOTE** — contribution méthodologique rare dans la littérature
2. **Une comparaison multi-paradigmes honnête** — ML classique, Deep Learning, LLMs — sur un dataset de 253 000 échantillons réaliste
3. **Une plateforme déployable complète** : diagnostic IA + gestion des patients + rendez-vous + messagerie sécurisée

Les perspectives ouvertes incluent l'amélioration du recall sur la classe positive, l'intégration du signal ECG (extension multimodale), et une validation clinique en conditions réelles.

## Table des matières

Introduction générale .....	11
Problématique .....	13
<b>I. CHAPITRE I: Intelligence Artificielle, Apprentissage Automatique et Architectures Neuronales Modernes .....</b>	<b>16</b>
1. Introduction :.....	16
2. Intelligence Artificielle (IA) : .....	16
2.1 Définition et historique.....	16
2.2 Catégories d'IA.....	17
2.3 IA Symbolique vs IA Connexionniste.....	17
3. Machine Learning (Apprentissage Automatique) .....	17
3.1 Définition .....	17
3.2 Types d'apprentissage .....	18
a) Apprentissage supervisé .....	18
b) Apprentissage non supervisé.....	18
c) Apprentissage par renforcement.....	18
3.3 Concepts fondamentaux .....	18
4. Deep Learning (Apprentissage Profond).....	19
4.1 Définition et motivations .....	19
4.2 Architecture d'un réseau profond .....	19
4.3 Rétropropagation et descente de gradient .....	20
4.4 Principales architectures .....	20
5. Réseau de Neurones Convolutif 1D (CNN-1D).....	21
5.1 Principe de la convolution 1D .....	21
5.2 Architecture typique.....	21
5.3 Avantages sur les données séquentielles .....	22
5.4 Domaines d'application .....	22
6. Architecture Transformer :.....	23
6.1 Contexte et motivation : .....	23
6.2 Mécanisme d'attention (Self-Attention) :.....	23
6.3 Architecture complète : .....	23
6.4 Variantes majeures : .....	24
7. Grands Modèles de Langage (LLM) .....	25
7.1 Définition .....	25
7.2 Pré-entraînement et fine-tuning.....	25

<b>Pré-entraînement</b> .....	25
<b>Fine-tuning et RLHF</b> .....	25
<b>7.3 Capacités émergentes</b> .....	26
<b>7.4 Principaux LLM</b> .....	26
<b>7.5 Limitations</b> .....	26
<b>8. LLM Multimodal</b> .....	27
<b>8.1 Définition</b> .....	27
<b>8.2 Architectures de fusion multimodale</b> .....	27
<b>a) Fusion précoce (Early Fusion)</b> .....	27
<b>b) Fusion par attention croisée (Cross-Attention Fusion)</b> .....	27
<b>8.3 Modalités supportées</b> .....	27
<b>8.4 Principaux LLM multimodaux</b> .....	28
<b>8.5 Défis spécifiques</b> .....	28
<b>9. Tableau comparatif des principaux LLMs généraux</b> .....	29
<b>10. Synthèse Comparative :</b> .....	30
<b>11. Conclusion du Chapitre :</b> .....	30
<b>II. CHAPITRE II: Cadre Théorique et État de l'Art</b> .....	34
<b>1. Introduction :</b> .....	34
<b>2. Problématique du diagnostic des maladies cardiaques</b> .....	35
<b>2.1. Définition des maladies cardiovasculaires</b> .....	35
<b>2.2. Facteurs de risque</b> .....	35
<b>2.3. Limites du diagnostic clinique traditionnel</b> .....	36
<b>3. Travaux basés sur le Machine Learning classique</b> .....	37
<b>3.1. Random Forest (RF)</b> .....	37
<b>3.2. XGBoost (Extreme Gradient Boosting)</b> .....	38
<b>3.3. LightGBM (Light Gradient Boosting Machine)</b> .....	38
<b>3.4. Support Vector Machine (SVM)</b> .....	39
<b>3.5. Réseaux de neurones artificiels (ANN) et Deep Learning</b> .....	39
<b>3.6. Approches ensemblistes et hybrides</b> .....	40
<b>3.7. Tableau comparatif des travaux ML</b> .....	40
<b>4. Travaux basés sur les LLMs et MLLMs :</b> .....	42
<b>4.1. Introduction aux LLMs médicaux :</b> .....	42
<b>4.2. BioClinicalBERT et ClinicalBERT :</b> .....	42
<b>4.3. BioBERT et autres BERT médicaux :</b> .....	43
<b>4.4. Med-PaLM et Med-PaLM 2 :</b> .....	43
<b>4.5. GPT-4 et applications cardiovasculaires :</b> .....	44

4.6. Qwen et autres LLMs open-source :	45
4.8. Tableau comparatif des LLMs / MLLMs :	45
5. Conclusion du chapitre	46
III .CHAPITRE III : Conception, expérimentation et choix du modèle	49
1. Introduction	49
2. Jeu de données et prétraitement	49
2.1. Description du dataset	49
2.2. Pipeline de prétraitement	50
3. Analyse critique : importance de l'ordre des étapes et problème de data leakage	51
3.1. Le problème du pipeline incorrect	51
3.2. Comparaison empirique des deux pipelines	52
3.3. Pourquoi SMOTE est néanmoins indispensable sur ce dataset	53
4. Modèles entraînés et évalués	54
4.1. XGBoost (Extreme Gradient Boosting)	54
4.2. LightGBM (Light Gradient Boosting Machine)	54
4.3. Ensemble Learning (Voting Soft : XGBoost + LightGBM + Random Forest)	55
4.4. CNN (Convolutional Neural Network)	55
4.5. Bio_ClinicalBERT (LLM médical fine-tuné)	56
4.6. Qwen 2.5-1.5B-Instruct (LLM généraliste, zero-shot)	56
5. Tableau comparatif des modèles testés (pipeline correct)	57
3.6. Choix du modèle final : Ensemble Learning	58
6.1. Justification des performances	58
6.2. Justification dans le contexte médical	59
6.3. Interprétabilité et déployabilité	59
7. Études antérieures sur des datasets similaires	59
7.1. Yang et al. (2025) — LightGBM interprétable sur BRFSS_2015	59
7.2. Chowdhury (2025) — XGBoost sur BRFSS pour patients diabétiques	60
7.3. Tompra et al. (2024) — CatBoost + SMOTE-ENN sur BRFSS 2021	60
7.4. Wang et al. (2023) — Indicateurs et prédictions de maladie cardiaque	60
7.5. Stacking ensemble pour CHD (2025)	60
7.6. Awosika et al. (2025) — Comparaison multi-modèles avec SMOTE	60
8. Comparaison de notre solution avec les travaux antérieurs	61
3.8.1. Points forts de notre solution	61
9. Conclusion du chapitre	62
IV .CHAPITRE IV : Implémentation de la plateforme MED AI	66
1. Introduction :	66

<b>2 . Environnement de développement et outils</b> .....	66
<b>3 . Conception de la base de données (Supabase)</b> .....	67
<b>3.1 Schéma global de la base de données</b> .....	67
<b>3.2 Description des principales tables</b> .....	68
a) Gestion des utilisateurs et des rôles .....	68
b) Patients et module de diagnostic par IA.....	69
c) Rendez-vous, messagerie et notifications.....	69
<b>3.3 Relations entre les tables</b> .....	70
<b>4 . Présentation des interfaces de l'application</b> .....	71
<b>4.1 Authentification et création de compte</b> .....	71
<b>4.2 Espace Patient — Assistant IA Cardiaque</b> .....	73
<b>4.3 Recherche de médecin et prise de rendez-vous</b> .....	76
<b>4.4 Espace Médecin — Assistant au diagnostic</b> .....	79
<b>5 . Sécurité et confidentialité des données</b> .....	83
<b>6 . Conclusion du chapitre</b> .....	83
<b>Conclusion générale :</b> .....	85
<b>Références</b> .....	87

## ***Les tables :***

Tableau 1 – Comparaison des principaux LLMs généraux (2022-2025). Sources : OpenAI, Google DeepMind, Meta AI, Anthropic, Mistral AI.....	29
Tableau 2 – Comparaison des principaux Technologie . Sources : OpenAI, Google DeepMind,.....	30
Tableau 3 – synthèse comparative des principaux travaux ml sur la prédiction des maladies cardiaques. ....	41
Tableau 4 – Synthèse comparative des LLMs et MLLMs appliqués au diagnostic cardiovasculaire. ....	46
Tableau 5 – Impact de l'ordre des étapes du pipeline sur les performances (Ensemble XGB+LGB+RF) ....	52
Tableau 6 – Comparaison des sept modèles (pipeline correct, test = distribution réelle originale)....	57
Tableau 7 – Comparaison de notre solution avec les principales études antérieures .....	61
Tableau 8 Environnement de développement et outils.....	67
Tableau 9 Gestion des utilisateurs et des rôles .....	68
Tableau 10 Patients et module de diagnostic par IA.....	69

## ***Les figures :***

Figure 1: Architecture générale d'un réseau de neurones profond (DNN) .....	20
Figure 2 Architecture typique d'un réseau de neurones convolutif 1D (1D-CNN) pour l'analyse des signaux temporels.....	22
Figure 3— Architecture Transformer Encodeur–Décodeur inspirée de Vaswani et al. (2017).....	24
Figure 4 — Schéma relationnel global de la base de données HealthID (Supabase / PostgreSQL).....	68
Figure 5 — Interface de connexion.....	71
Figure 6 — Formulaire d'inscription (profil Patient sélectionné).....	72
Figure 7 — Assistant IA Cardiaque : questionnaire de diagnostic (interface patient) .....	73
Figure 8 — Résultat du diagnostic IA et recommandations (interface patient) .....	74
Figure 9 — Écran de diagnostic et pronostic (cas de risque critique) .....	75
Figure 10 — Liste des médecins disponibles.....	76
Figure 11 — Sélection de la date et de l'horaire du rendez-vous.....	77
Figure 12 — Confirmation du rendez-vous .....	78
Figure 13 — Tableau de bord de l'assistant médecin .....	79
Figure 14 — Formulaire d'ajout d'un patient (espace médecin) .....	80
Figure 15 — Vue de gestion « Mes Patients » .....	81
Figure 16 — Tableau de gestion des dossiers cliniques (espace médecin) .....	82

## **Introduction générale**

**L**es maladies cardiovasculaires représentent aujourd'hui le premier défi sanitaire de la planète. D'après l'Organisation Mondiale de la Santé, elles ont provoqué près de 19,8 millions de décès en 2022, soit environ un tiers de la mortalité mondiale, et les projections les plus récentes annoncent un alourdissement considérable de ce fardeau d'ici 2050 [1]. Derrière ces chiffres se cache une réalité plus douloureuse encore : la majorité de ces décès survient dans les pays à revenu faible ou intermédiaire, là où l'accès à un cardiologue, à un électrocardiogramme ou à une coronarographie reste un privilège plutôt qu'un droit. Une part importante de ces drames pourrait pourtant être évitée, car beaucoup de ces affections se développent silencieusement, sur des années, avant de se manifester brutalement.

C'est précisément sur ce point que se joue l'essentiel : le diagnostic précoce. Détecter le risque avant la crise, c'est offrir au patient une fenêtre d'intervention, et bien souvent, c'est lui sauver la vie. Or le diagnostic traditionnel, aussi rigoureux soit-il, demeure tributaire de l'expertise du clinicien, du coût des examens et d'une disponibilité géographique très inégale. Il était donc naturel de se tourner vers une discipline qui a transformé, en l'espace de quelques années, notre rapport à la donnée : l'intelligence artificielle.

L'IA n'est plus une promesse lointaine. Du Machine Learning classique aux réseaux profonds, des architectures Transformer aux grands modèles de langage et à leurs déclinaisons multimodales, elle est devenue capable d'apprendre des régularités complexes au sein de masses de données médicales, et d'en extraire des prédictions d'une précision remarquable. Dans le domaine cardiovasculaire en particulier, les algorithmes ensemblistes comme Random Forest, XGBoost ou LightGBM se sont imposés comme des références, tandis que les modèles de

langage médicaux ouvrent des perspectives inédites pour exploiter les dossiers cliniques textuels.

C'est dans ce contexte que s'inscrit le présent travail. Nous proposons la conception et la réalisation de **MED AI**, une plateforme web intelligente dédiée à la détection précoce des maladies cardiaque ou cardiovasculaires. Notre ambition n'est pas seulement de construire un modèle performant en laboratoire, mais de l'intégrer dans un outil concret, accessible et utile — pour le patient comme pour le médecin. La plateforme combine un moteur de diagnostic fondé sur l'apprentissage ensembliste, un assistant conversationnel guidant le patient à travers une évaluation simple de son risque, ainsi qu'un ensemble de services complémentaires : recommandation de spécialistes, prise de rendez-vous en ligne et gestion sécurisée des dossiers médicaux.

Ce mémoire s'organise en quatre chapitres. Le **premier** dresse un panorama structuré des grands paradigmes de l'IA moderne, depuis le Machine Learning jusqu'aux modèles multimodaux, afin de poser les fondements théoriques du travail. Le **deuxième** propose un état de l'art du diagnostic cardiovasculaire assisté par ordinateur, en confrontant les approches classiques de Machine Learning aux contributions récentes des modèles de langage médicaux. Le **troisième** constitue le cœur expérimental : il décrit le jeu de données, le pipeline de prétraitement, et compare sept modèles issus de trois paradigmes distincts, tout en accordant une attention particulière à un écueil méthodologique souvent négligé, la fuite de données liée à l'application de SMOTE. Le **quatrième**, enfin, présente l'implémentation concrète de la plateforme MED AI : son architecture, sa base de données et ses différentes interfaces.

## **Problématique**

Le constat de départ est paradoxal. D'un côté, la littérature scientifique regorge de travaux annonçant des précisions spectaculaires — souvent supérieures à 95 %, parfois proches de 99 % — pour la prédiction des maladies cardiaques. De l'autre, ces résultats peinent à se traduire en outils réellement déployés et fiables en conditions cliniques. Comment expliquer un tel décalage ?

Une première partie de la réponse tient à la taille et à la nature des données. Nombre de ces performances éclatantes reposent sur de petits jeux de données, comme le dataset UCI Cleveland et ses 303 enregistrements, qui ne reflètent ni la diversité ni le déséquilibre des populations réelles. Sur un dataset à grande échelle tel que le Heart Disease Health Indicators (plus de 253 000 individus issus de l'enquête BRFSS), où moins de 10 % des sujets sont effectivement atteints, le problème devient nettement plus ardu et plus représentatif de la réalité.

Une seconde partie de la réponse, plus subtile, relève de la rigueur méthodologique. L'application de techniques de rééquilibrage comme SMOTE est souvent indispensable face à un fort déséquilibre de classes ; mais lorsqu'elle est mal séquencée — appliquée avant la séparation entre données d'entraînement et de test, ou sur l'ensemble complet du jeu de données — elle introduit une fuite d'information qui gonfle artificiellement les métriques. Le modèle paraît brillant sur le papier, mais son évaluation ne correspond plus à rien de réel.

Ce travail s'articule donc autour d'un ensemble de questions :

- Quel modèle, parmi les grandes familles disponibles — Machine Learning classique, Deep Learning, modèles de langage — offre le meilleur compromis entre performance, fiabilité clinique et déployabilité sur un jeu de données vaste et réaliste ?

- Comment garantir une évaluation honnête et reproductible, à l'abri des artefacts méthodologiques qui faussent une grande partie de la littérature ?
- Comment, enfin, transformer ce modèle en une plateforme intelligente, accessible et sécurisée, capable de servir aussi bien le patient que le praticien, dans un cadre respectueux de la sensibilité des données de santé ?

CHAPITRE I

# Intelligence Artificielle, Apprentissage Automatique et Architectures Neuronales Modernes

---

*IA · Machine Learning · Deep Learning · CNN-1D · Transformer · LLM · LLM Multimodal*

# **I. CHAPITRE I: Intelligence Artificielle, Apprentissage Automatique et Architectures Neuronales Modernes**

## **1. Introduction :**

**L**'intelligence artificielle (IA) constitue aujourd'hui l'un des domaines les plus dynamiques de la science informatique. Depuis ses premières formulations théoriques dans les années 1950, notamment avec Alan Turing [2] et John McCarthy [3], l'IA a connu une évolution remarquable, passant de systèmes à base de règles rigides vers des architectures capables d'apprendre, de raisonner et de générer du contenu de manière autonome.

Ce chapitre présente une vue d'ensemble structurée des grands paradigmes de l'IA moderne : l'intelligence artificielle, le Machine Learning, le Deep Learning, les réseaux CNN-1D, les Transformers, les LLM et les LLM multimodaux. Chaque paradigme est défini, illustré par des exemples concrets et référencé bibliographiquement. L'IA englobe le ML [4], qui contient le DL [5], lequel inclut des architectures spécialisées (CNN-1D, Transformer) et leurs dérivés (LLM, LLM multimodal).

## **2. Intelligence Artificielle (IA) :**

### **2.1 Définition et historique**

L'intelligence artificielle désigne l'ensemble des théories et techniques permettant à des machines de simuler des capacités cognitives humaines : raisonnement, apprentissage, planification, compréhension du langage naturel et prise de décision [6]. Le terme a été officiellement introduit par John McCarthy lors de la Conférence de Dartmouth en 1956 [6]. Depuis lors, l'IA a traversé plusieurs cycles d'enthousiasme et de désillusion avant de connaître une renaissance grâce à la puissance de calcul des GPU, la disponibilité des grandes données et les avancées algorithmiques.

## 2.2 Catégories d'IA

- IA Faible (Narrow AI) : spécialisée dans une tâche précise. Exemples : reconnaissance faciale, traduction automatique, jeux d'échecs [2]
- IA Forte (AGI) : capable de réaliser n'importe quelle tâche intellectuelle humaine. Encore au stade de la recherche théorique. [2]
- IA Superintelligente : hypothétique IA qui dépasserait l'intelligence humaine dans tous les domaines . [7]

## 2.3 IA Symbolique vs IA Connexionniste

L'IA symbolique (Good Old-Fashioned Artificial Intelligence) repose sur la manipulation de symboles et de règles logiques explicites [2]. Elle inclut les systèmes experts et les moteurs d'inférence. En revanche, l'IA connexionniste, dont le Deep Learning est le représentant majeur, s'inspire du fonctionnement du cerveau humain via des réseaux de neurones artificiels [5].

### 🔗 Exemple : Système expert médical — MYCIN :

MYCIN (Stanford, années 1970) est l'un des premiers systèmes experts. Il diagnostiquait les infections bactériennes en appliquant environ 600 règles logiques encodées par des experts médicaux, atteignant une précision supérieure à celle de nombreux médecins juniors dans son domaine. [2]

# 3. Machine Learning (Apprentissage Automatique)

## 3.1 Définition

Le Machine Learning (ML) est une sous-discipline de l'IA qui permet à des systèmes d'apprendre à partir de données sans être explicitement programmés pour chaque tâche. Comme l'a défini Arthur Samuel en 1959 [8], c'est « un domaine d'étude qui donne aux ordinateurs la capacité d'apprendre sans être explicitement programmés ». Tom Mitchell (1997) [9] en propose une définition plus formelle : un programme apprend à partir d'une expérience  $E$  par rapport à

une tâche T et une mesure de performance P si ses performances s'améliorent avec l'expérience E.

### 3.2 Types d'apprentissage

#### a) Apprentissage supervisé

Le modèle apprend à partir de données étiquetées (paires entrée/sortie). Algorithmes courants : régression linéaire/logistique, Support Vector Machine, arbres de décision, forêts aléatoires, réseaux de neurones [10].

#### b) Apprentissage non supervisé

Le modèle découvre des structures cachées dans des données non étiquetées. Exemples : clustering (k-means), réduction de dimensionnalité (Principal Component Analysis, t-Distributed Stochastic Neighbor Embedding), modèles génératifs (Variational Autoencoders, Generative Adversarial Networks) [11].

#### c) Apprentissage par renforcement

Un agent apprend à prendre des décisions en interagissant avec un environnement et en recevant des récompenses ou des pénalités. Applications : jeux vidéo (AlphaGo), robotique, trading algorithmique [12].

#### 🔗 Exemple : Détection de spam par classification bayésienne

Un classifieur bayésien naïf entraîné sur 10 000 e-mails étiquetés « spam / non-spam » apprend les probabilités d'apparition de mots caractéristiques (« gratuit », « cliquez ici »). Gmail utilise une variante combinée à des réseaux de neurones profonds. [8] [9]

### 3.3 Concepts fondamentaux

- Données d'entraînement / validation / test : séparation standard 70/15/15 ou 80/10/10 [10].
- Sur-apprentissage (overfitting) : le modèle mémorise les données sans généraliser [10].
- Régularisation (L1/L2, Dropout) : techniques pour prévenir le sur-apprentissage [11].

- Validation croisée (k-fold cross-validation) : évaluation robuste des performances [13].

## 4. Deep Learning (Apprentissage Profond)

### 4.1 Définition et motivations

Le Deep Learning (DL) est une branche du ML basée sur des réseaux de neurones à plusieurs couches cachées [14]. Ces architectures apprennent automatiquement des représentations hiérarchiques à partir de données brutes, éliminant la nécessité d'une ingénierie manuelle des caractéristiques [15]. L'essor du DL est dû à : (1) la disponibilité de grands jeux de données (ImageNet, Common Crawl), (2) la puissance de calcul des GPU [16], et (3) les avancées algorithmiques telles que les fonctions d'activation ReLU, la normalisation par batch et les connexions résiduelles [17]

### 4.2 Architecture d'un réseau profond

- Couche d'entrée (Input Layer) : reçoit les données brutes (pixels, tokens, valeurs numériques). [18]
- Couches cachées : transforment les représentations par des opérations linéaires et non-linéaires :

$$z = Wx + b$$

$$a = \sigma(z)$$

[18]

- Couche de sortie : produit la prédiction finale (softmax pour classification, linéaire pour régression). [18]

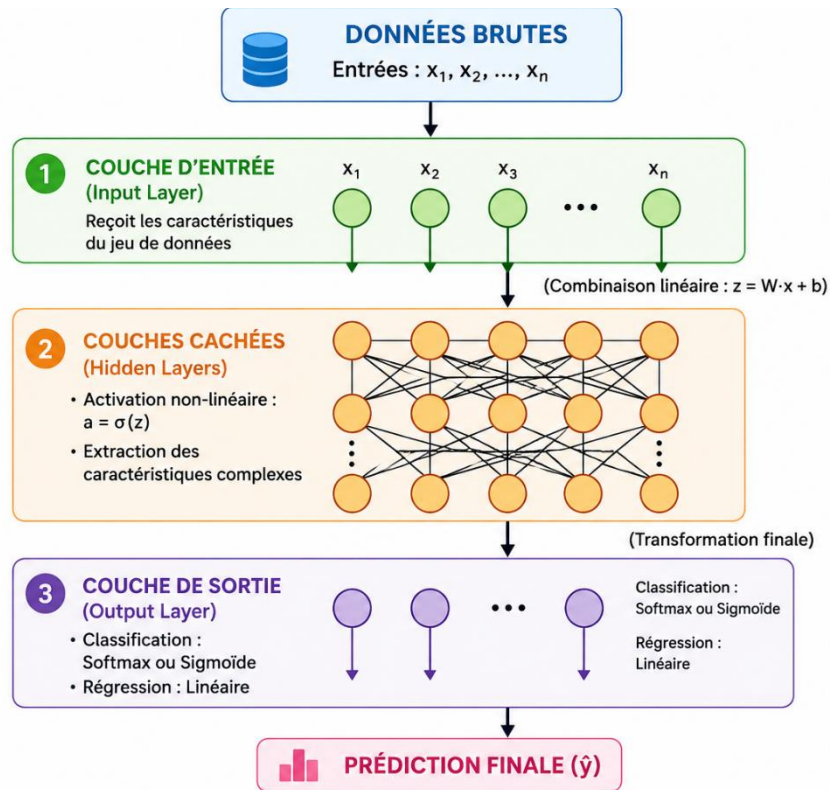


Figure 1: Architecture générale d'un réseau de neurones profond (DNN)

### 4.3 Rétropropagation et descente de gradient

L'entraînement d'un réseau profond repose sur la rétropropagation du gradient (backpropagation), introduite par Rumelhart, Hinton et Williams (1986) [19]. L'algorithme calcule le gradient de la fonction de perte par rapport à chaque paramètre via la règle de la chaîne, puis met à jour les poids par descente de gradient (Adam, RMSProp, AdaGrad) [18].

### 4.4 Principales architectures

- CNN : spécialisé pour les données spatiales (images, signaux). [19]
- RNN / LSTM / GRU : pour les données séquentielles (séries temporelles, texte). [18]
- Transformer : architecture basée sur l'attention, dominante en NLP et vision. [20]
- GAN (Generative Adversarial Network) : génération de contenu synthétique réaliste. [18]

**📌 Exemple : AlexNet — ImageNet 2012**

En 2012, AlexNet (Krizhevsky et al.) a remporté le concours ImageNet avec un taux d'erreur top-5 de 15,3 %, contre 26,2 % pour le meilleur algorithme classique. Ce résultat a marqué le début de l'ère moderne du Deep Learning. [21]

## 5. Réseau de Neurones Convolutif 1D (CNN-1D)

### 5.1 Principe de la convolution 1D

Les CNN ont été initialement développés pour le traitement des images [19]. Le CNN-1D applique des opérations de convolution sur des séquences unidimensionnelles. Un filtre (kernel) de taille  $k$  glisse sur la séquence d'entrée et calcule des produits scalaires locaux :

$$(x \star w)[t] = \sum_i x[t + i] \cdot w[i]$$

Cette opération produit une carte d'activation (feature map) capturant des patterns locaux [22].

### 5.2 Architecture typique

- Entrée : séquence 1D de longueur  $T$  avec  $C$  canaux (ex. signal ECG sur  $T$  échantillons,  $C$  dérivations). [23]
- Couches Conv1D + BatchNorm + ReLU : extraction de caractéristiques locales (fréquences, motifs). [24]
- MaxPooling1D : sous-échantillonnage et invariance aux petits décalages temporels. [25]
- Dense + Softmax/Sigmoïde : classification ou régression finale. [26]

Figure 5.2 — Architecture typique d'un réseau CNN 1D

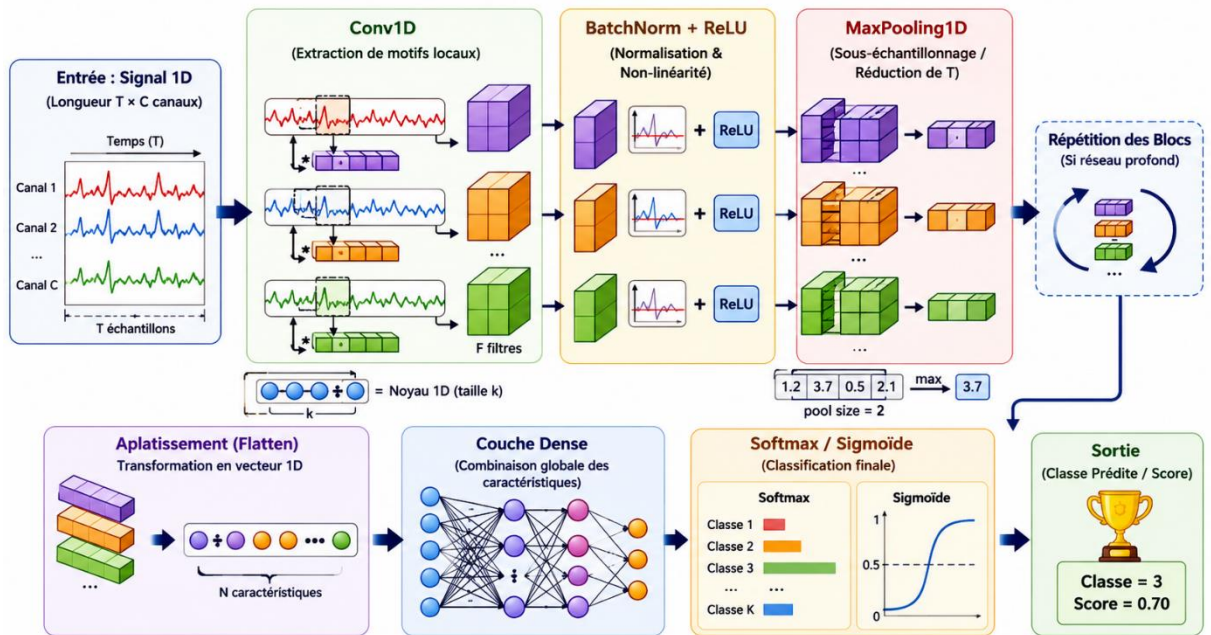


Figure 2 Architecture typique d'un réseau de neurones convolutif 1D (1D-CNN) pour l'analyse des signaux temporels

### 5.3 Avantages sur les données séquentielles

- Partage des paramètres : le même filtre est appliqué à toutes les positions → moins de paramètres qu'un MLP [22]
- Invariance locale : détecte les patterns quel que soit leur position dans la séquence. [5]
- Parallélisme : contrairement aux RNN, les convolutions sont calculées en parallèle. [22]

### 5.4 Domaines d'application

- Analyse de signaux ECG/EEG : détection d'arythmies, classification d'états cérébraux. [27]
- NLP : classification de texte, analyse de sentiment. [28]
- Séries temporelles financières : prédiction de prix, détection d'anomalies. [29]

- Reconnaissance d'activités humaines (HAR) : accéléromètre, gyroscope. [29]

📌 Exemple : Détection d'arythmies cardiaques — Rajpurkar et al. (2017)

Un CNN-1D de 34 couches résiduelles entraîné sur 91 232 enregistrements ECG (30 s, 200 Hz) a surpassé des cardiologues certifiés sur 12 classes d'arythmies (F1-score : 0,837 vs 0,780). [17]

## 6. Architecture Transformer :

### 6.1 Contexte et motivation :

Avant les Transformers, les architectures récurrentes (RNN, LSTM, GRU) souffraient du traitement séquentiel et du vanishing gradient problem, rendant difficile la capture de dépendances à longue portée [20]. En 2017, Vaswani et al. [20] ont proposé le Transformer, entièrement basé sur des mécanismes d'attention, révolutionnant le NLP et bien au-delà.

### 6.2 Mécanisme d'attention (Self-Attention) :

Pour chaque token, le modèle calcule trois vecteurs Query (Q), Key (K) et Value (V). Le score d'attention est :

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

V, où  $d_k$  est la dimension des vecteurs clés [20]. Chaque token « attend » tous les autres simultanément, capturant des dépendances globales sans limitation de portée.

### 6.3 Architecture complète :

- Encodeur : N blocs composés d'une couche Self-Attention + Feed-Forward Network + normalisation et connexions résiduelles. [20]

- Décodeur : similaire à l'encodeur, avec une couche d'attention croisée (cross-attention) sur les sorties de l'encodeur. [20]
- Encodages positionnels (Positional Encoding) : encodent la position de chaque token via des fonctions sinusoïdales ou des embeddings appris. [20]

6.4 Variantes majeures :

- BERT (Devlin et al., 2018) : encodeur bidirectionnel pré-entraîné par masquage (MLM). Excellent en classification, QA, NER. [30]
- GPT (OpenAI) : décodeur autorégressif pré-entraîné sur de vastes corpus. Conçu pour la génération de texte. [31]
- T5 (Raffel et al., 2019) : traite toutes les tâches NLP comme des problèmes texte-à-texte. [32]
- Vision Transformer (ViT, Dosovitskiy et al., 2020) : applique le Transformer aux images découpées en patches. [33]

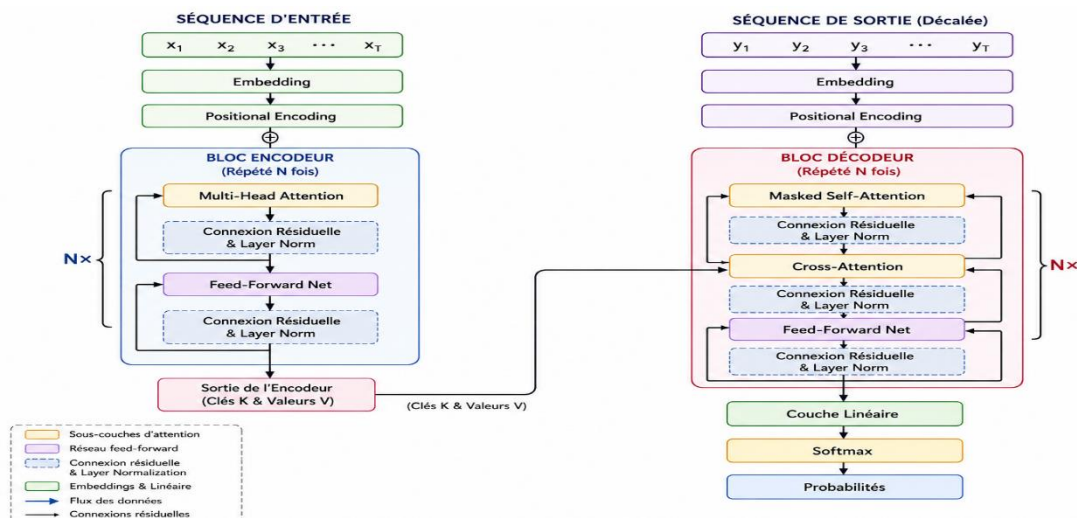


Figure 3— Architecture Transformer Encodeur–Décodeur inspirée de Vaswani et al. (2017)

#### 🔗 Exemple : Traduction automatique — Vaswani et al. (2017)

Le Transformer original a obtenu un score BLEU de 28,4 sur la tâche anglais→allemand (WMT 2014), surpassant les meilleurs modèles RNN de plus de 2 points BLEU, en 3,5 jours d'entraînement sur 8 GPU P100. [20]

## 7. Grands Modèles de Langage (LLM)

### 7.1 Définition

Les LLM sont des modèles de langage de très grande taille (des milliards à des billions de paramètres) pré-entraînés sur des corpus textuels massifs [34]. Ils sont capables de comprendre et de générer du langage naturel avec une fluidité remarquable. Les LLM exploitent l'architecture Transformer [20] et le principe du pré-entraînement auto-supervisé suivi d'un fine-tuning spécialisé [35]

### 7.2 Pré-entraînement et fine-tuning

#### Pré-entraînement

Les LLM sont pré-entraînés sur des téraoctets de texte (Wikipedia, Common Crawl, livres, code source) [31] via des objectifs auto-supervisés : prédiction du token suivant (GPT [31]) ou masquage de tokens (BERT [30]). Cette phase permet au modèle d'acquérir des connaissances encyclopédiques et des capacités linguistiques générales.

#### Fine-tuning et RLHF

Après le pré-entraînement, les LLM sont affinés via le Reinforcement Learning from Human Feedback (RLHF) [35] : des annotateurs humains classent les réponses, un modèle de récompense est entraîné, et le fine-tuning est guidé par la politique de gradient proximal (PPO). Anthropic utilise le Constitutional AI comme variante [36].

### 7.3 Capacités émergentes

Les LLM présentent des capacités dites « émergentes » — apparaissant uniquement à partir d'une certaine échelle [37] — telles que le raisonnement en chaîne (chain-of-thought), l'apprentissage en contexte (few-shot), la génération de code [38] et la résolution de problèmes mathématiques.

### 7.4 Principaux LLM

- GPT-4 (OpenAI, 2023) : ~1,8 trillion de paramètres (estimé). Multilingue, excellent en raisonnement. [38]
- Claude 3/4 (Anthropic) : conçu avec un accent sur la sécurité via Constitutional AI. [22]
- LLaMA 2/3 (Meta AI, 2023) : modèles open-source, 7B à 70B paramètres. [39]
- Mistral 7B / Mixtral (Mistral AI, 2023) : architecture MoE (Mixture of Experts) compacte. [38]

#### 🔗 Exemple : Génération de code — Codex / GitHub Copilot

Codex (Chen et al., 2021), ancêtre de GPT-4 pour le code, a résolu 28,8 % des problèmes de HumanEval en un seul essai (pass@1) et 72,3 % en 100 essais. GitHub Copilot, basé sur Codex, est utilisé par plus de 1 million de développeurs. [40]

### 7.5 Limitations

- Hallucinations : génération de faits incorrects présentés avec confiance. [38]
- Biais : reproduction et amplification des biais présents dans les données d'entraînement. [35]
- Coût computationnel : GPT-3 a nécessité ~355 GPU-years d'entraînement. [34]

- Opacité (boîte noire) : difficile d'interpréter les mécanismes de raisonnement interne. [41]

## 8. LLM Multimodal

### 8.1 Définition

Un LLM multimodal est un modèle capable de traiter et de générer plusieurs modalités d'information simultanément : texte, images, audio et vidéo [42]. Contrairement aux LLM purement textuels, ces modèles développent une compréhension unifiée du monde en fusionnant des représentations provenant de différents domaines sensoriels [43].

### 8.2 Architectures de fusion multimodale

#### a) Fusion précoce (Early Fusion)

Les différentes modalités sont encodées puis fusionnées au niveau des embeddings avant d'être traitées par le LLM [44]. Les tokens d'image sont projetés dans l'espace d'embedding textuel et concaténés aux tokens de texte.

#### b) Fusion par attention croisée (Cross-Attention Fusion)

Un LLM texte utilise des couches d'attention croisée pour interroger les représentations d'un encodeur visuel (ex. CLIP [45], ViT [33]). C'est l'approche de Flamingo (DeepMind) [46] et LLaVA [44].

### 8.3 Modalités supportées

- Texte + Image : compréhension de documents, VQA (Visual Question Answering), génération de légendes. [42], [44]
- Texte + Audio : transcription (Whisper), synthèse vocale, compréhension de la parole en contexte. [47]
- Texte + Vidéo : résumé de vidéos, génération de descriptions temporelles. [48]

- Toutes modalités (omnimodal) : GPT-4o, Gemini Ultra — vision, audio, texte dans un seul modèle. [49]

#### 8.4 Principaux LLM multimodaux

- GPT-4V / GPT-4o (OpenAI, 2023–2024) : vision et audio natifs, description d'images, lecture de graphiques. [50]
- Gemini Ultra/Pro/Flash (Google DeepMind, 2023–2024) : multimodal natif, traite images, audio, vidéo et texte. [50]
- Flamingo (Alayrac et al., 2022) : pionnier du few-shot multimodal via cross-attention. [42]
- LLaVA (Liu et al., 2023) : modèle open-source connectant CLIP à Vicuna pour la compréhension visuelle. [44]
- Whisper (OpenAI, 2022) : reconnaissance vocale robuste et multilingue, souvent intégré aux LLM multimodaux. [39]

#### 🔗 Exemple : GPT-4V — Analyse d'images médicales

Yang et al. (2023) ont évalué GPT-4V sur des radiographies et IRM. Sans fine-tuning, le modèle a atteint 68 % de précision pour la détection de pneumonie sur radiographies pulmonaires. [36]

#### 8.5 Défis spécifiques

- Alignement multimodal : assurer la cohérence sémantique entre les différentes modalités. [46]
- Hallucinations visuelles : le modèle peut « inventer » des détails absents de l'image. [51]
- Évaluation : nouveaux benchmarks nécessaires (MME, MMMU, SEED-Bench). [43]

- Coût des données : annotations multi-modalités rares et coûteuses à produire. [44]

## 9. Tableau comparatif des principaux LLMs généraux

Le tableau suivant synthétise les caractéristiques des LLMs les plus utilisés dans le monde, qui servent de base aux modèles médicaux spécialisés.

Modèle	Développeur	Paramètres	Architecture	Particularité
<b>GPT-4 (2023)</b>	OpenAI	~1.7T (estimé)	Transformer (dense+MoE)	Meilleur score MedQA 93%
<b>GPT-3.5</b>	OpenAI	175B	Transformer autorégressif	Base de ChatGPT, rapide
<b>LLaMA 2 / 3</b>	Meta AI	7B–70B	Transformer open-source	Modèle de base médical
<b>Gemini 1.5 Pro</b>	Google DeepMind	Non divulgué	Multimodal Transformer	Fenêtre contexte 1M tokens
<b>Claude 3 Opus</b>	Anthropic	Non divulgué	Constitutional AI	Raisonnement long, sûreté
<b>Mistral 7B</b>	Mistral AI	7B	Sliding-window attention	Efficace, open-source
<b>Falcon 180B</b>	TII (UAE)	180B	Transformer causal	Données massives ouvertes
<b>Gemma 3 27B</b>	Google	27B	Transformer dense	Open-weights, médical/vision

Tableau 1 – Comparaison des principaux LLMs généraux (2022-2025). Sources : OpenAI, Google DeepMind, Meta AI, Anthropic, Mistral AI.

## 10. Synthèse Comparative :

• Technologie	Données d'entrée	Application principale	Exemple clé
<b>Machine Learning</b>	Données tabulaires	Classification, régression, clustering	Détection de spam [5][6]
<b>Deep Learning</b>	Images, audio, texte	Vision, NLP, génération	AlexNet — ImageNet 2012 [11]
<b>CNN-1D</b>	Signaux 1D, séries temporelles	ECG, HAR, classification audio	Arythmies cardiaques [17]
<b>Transformer</b>	Séquences (texte, image)	NLP, traduction, vision	BERT [21], GPT [24], ViT [22]
<b>LLM</b>	Texte (tokens)	Génération, résumé, QA, code	GPT-4 [26], LLaMA 2 [27]
<b>LLM Multimodal</b>	Texte + Image + Audio + Vidéo	VQA, analyse médicale, assistants	GPT-4o [32], Gemini [33]

Tableau 2 – Comparaison des principaux Technologies. Sources : OpenAI, Google DeepMind,

## 11. Conclusion du Chapitre :

Ce chapitre a présenté un panorama structuré des technologies fondamentales de l'IA moderne. L'IA constitue le cadre général, englobant le Machine Learning comme paradigme d'apprentissage automatique, et le Deep Learning comme sa branche la plus puissante.

Le CNN-1D s'est imposé pour l'analyse des signaux grâce à son efficacité computationnelle. L'architecture Transformer a révolutionné le domaine en permettant la modélisation des dépendances à longue portée, ouvrant la voie aux LLM. Enfin, les LLM multimodaux intègrent vision, audio et texte dans un

espace de représentation unifié, se rapprochant d'une compréhension holistique du monde.

## *CHAPITRE I*

### *Intelligence Artificielle, Apprentissage Automatique et Architectures Neuronales*

#### *Remerciements*

Ces technologies évoluent à une vitesse remarquable, portées par des avancées algorithmiques continues (attention linéaire, Mixture of Experts, State Space Models) . Les chapitres suivants s'appuieront sur ces fondements pour explorer des applications concrètes dans des domaines spécialisés.

## **CHAPITRE II**

*Maladies Cardiovasculaires, Intelligence Artificielle et Diagnostic Assisté par Ordinateur*

**CHAPITRE II**

*Maladies Cardiovasculaires, Intelligence Artificielle et Diagnostic Assisté par Ordinateur*

**CHAPITRE II**

---

**Cadre Théorique et État de l'Art**

*Maladies Cardiovasculaires, Intelligence Artificielle et Diagnostic Assisté par Ordinateur*

## II. CHAPITRE II: Cadre Théorique et État de l'Art

### 1. Introduction :

Les maladies cardiovasculaires (MCV) constituent aujourd'hui la première cause de mortalité dans le monde. Selon l'Organisation Mondiale de la Santé, environ **19,8 millions de personnes** sont décédées de MCV en 2022, soit près de 32 % de l'ensemble des décès mondiaux [52]. Plus de 85 % de ces décès résultent de crises cardiaques (infarctus du myocarde) et d'accidents vasculaires cérébraux (AVC) [52]. En outre, plus des trois quarts des décès liés aux MCV surviennent dans les pays à revenu faible ou intermédiaire, ce qui souligne une forte inégalité dans l'accès aux soins [52] .

Les projections récentes prévoient une augmentation considérable de la charge mondiale des MCV : le nombre absolu de décès cardiovasculaires devrait passer de **20,5 millions en 2025 à environ 35,6 millions en 2050**, principalement à cause du vieillissement de la population et des facteurs de risque métaboliques [53]. Dans la région européenne de l'OMS, les MCV provoquent plus de 42,5 % de tous les décès annuels, ce qui équivaut à environ **10 000 décès chaque jour** [54].

Face à cette charge alarmante, le diagnostic précoce devient un enjeu majeur de santé publique. La détection précoce permet une intervention thérapeutique rapide, ce qui réduit considérablement la mortalité et améliore la qualité de vie des patients [55]. Cependant, le diagnostic traditionnel repose en grande partie sur l'expertise clinique du médecin et sur des examens parfois coûteux (ECG, échographie, coronarographie), ce qui le rend difficilement accessible dans les régions à faibles ressources [56]

C'est dans ce contexte que l'intelligence artificielle (IA), et plus particulièrement l'apprentissage automatique (*Machine Learning*), s'impose comme un outil puissant. Les algorithmes de ML peuvent apprendre à partir de grandes quantités de données médicales (antécédents, signes vitaux, examens biologiques, ECG) et prédire avec une précision élevée la présence ou l'absence de maladie cardiaque [55], [57]. Plus récemment, les modèles de langage de grande taille (*Large Language Models — LLMs*) et leurs extensions multimodales (*MMLLMs*) ouvrent

de nouvelles perspectives, notamment pour l'analyse des dossiers médicaux électroniques (*EHR*) et des images d'ECG [58], [59].

Ce chapitre présente un état de l'art structuré du domaine. Nous aborderons successivement : (i) la problématique du diagnostic des maladies cardiaques, (ii) les principaux travaux fondés sur les algorithmes classiques de Machine Learning tels que Random Forest, XGBoost et LightGBM, et (iii) les travaux récents reposant sur les LLMs et MLLMs médicaux comme Bio\_ClinicalBERT, BioBERT, Med-PaLM, GPT-4 et Qwen.

## 2. Problématique du diagnostic des maladies cardiaques

### 2.1. Définition des maladies cardiovasculaires

Les maladies cardiovasculaires regroupent un ensemble de troubles affectant le cœur et les vaisseaux sanguins. Selon l'OMS, elles incluent principalement [54]: la cardiopathie ischémique (ou coronarienne), la maladie cérébrovasculaire (AVC), la cardiopathie rhumatismale, l'insuffisance cardiaque, les arythmies et les cardiopathies congénitales.

La cardiopathie ischémique reste la première cause de mortalité cardiovasculaire mondiale, avec un taux de mortalité standardisé de 108,8 décès pour 100 000 habitants en 2022 [60]. Elle résulte d'un rétrécissement des artères coronaires (athérosclérose), qui peut conduire à un infarctus du myocarde lorsque l'apport sanguin au cœur est brutalement interrompu.

### 2.2. Facteurs de risque

Les facteurs de risque des maladies cardiovasculaires sont classés en trois catégories :

- **Facteurs comportementaux** : tabagisme, alimentation déséquilibrée (excès de sel, de sucre et de graisses saturées), sédentarité et consommation excessive d'alcool [54].
- **Facteurs intermédiaires (cliniques)** : hypertension artérielle, hyperglycémie, hyperlipidémie (cholestérol élevé), surpoids et obésité [54]. L'hypertension est à elle seule responsable d'environ 10,8 millions de décès par an [60].
- **Facteurs environnementaux** : pollution de l'air ambiant et intérieur. Près de 70 % des 4,2 millions de décès attribués à la pollution de l'air en 2019 étaient dus aux MCV [61].

### 2.3. Limites du diagnostic clinique traditionnel

Le diagnostic clinique des MCV repose sur plusieurs étapes : interrogatoire du patient, examen physique, électrocardiogramme (ECG), échocardiographie, analyses biologiques (troponine, créatine kinase) et parfois coronarographie. Bien que fiable, ce processus présente plusieurs limites [62]:

- **Subjectivité** : l'interprétation des examens (notamment ECG) dépend fortement de l'expérience du clinicien.
- **Coût et accessibilité** : certaines techniques (IRM cardiaque, coronarographie) restent coûteuses et indisponibles dans de nombreux centres.
- **Lenteur** : le délai entre les premiers symptômes et le diagnostic définitif peut retarder le traitement, augmentant ainsi le risque de complications.
- **Symptômes silencieux** : de nombreuses MCV évoluent silencieusement et une crise cardiaque peut être la première manifestation visible de la maladie [54].

Ces limites motivent le recours à des systèmes d'aide à la décision basés sur l'intelligence artificielle, capables d'exploiter de grandes quantités de données médicales pour fournir un diagnostic rapide, objectif et accessible.

### 3. Travaux basés sur le Machine Learning classique

Cette section présente les principales approches de Machine Learning appliquées au diagnostic des maladies cardiaques. Nous décrivons d'abord brièvement les algorithmes les plus utilisés, puis nous rapportons les résultats des travaux récents.

#### 3.1. Random Forest (RF)

Le **Random Forest** est une méthode d'apprentissage ensembliste qui construit plusieurs arbres de décision à partir d'échantillons bootstrap des données et qui agrège leurs prédictions (par vote majoritaire en classification) [63]. En générant un ensemble d'arbres décorrélés, RF réduit la variance et améliore la précision tout en limitant le surapprentissage (*overfitting*) [63].

Hidayat et al. (2023) ont appliqué Random Forest sur le dataset UCI Cleveland et ont obtenu une précision de **94 %** pour la classification binaire (présence / absence de maladie) [64]. Effendy et al. (2024) ont étudié l'effet de la technique d'équilibrage *SMOTE* (*Synthetic Minority Over-sampling Technique*) sur les performances de RF, démontrant une amélioration significative du rappel pour la classe minoritaire [65]. Plus récemment, l'étude de Mohammed et al. (2025) sur un dataset combiné UCI a montré que Random Forest atteignait **94 % de précision avec une validation croisée K=10** et **92 % avec K=5**, ainsi qu'une AUC-ROC de 95 % .

### **3.2. XGBoost (Extreme Gradient Boosting)**

**XGBoost** est une extension optimisée de l'algorithme de *Gradient Boosting*. Il construit des arbres de décision de manière séquentielle, chaque nouvel arbre corrigeant les erreurs des arbres précédents. Il introduit en plus la régularisation (L1 et L2), la parallélisation, et un traitement natif des valeurs manquantes [65]. Ces optimisations rendent XGBoost particulièrement adapté au domaine médical : il est précis, scalable et relativement interprétable [65].

Rehman et al. (2025) ont montré que XGBoost surpasse significativement les classifieurs traditionnels (régression logistique, arbres de décision) dans la stratification du risque cardiovasculaire [65]. Sausan et al. (2024) ont comparé Decision Tree et XGBoost sur le UCI Heart Disease Dataset : XGBoost a obtenu **93 % de précision** contre 90 % pour l'arbre de décision, avec un prétraitement SMOTE pour gérer le déséquilibre de classes [65]. Dans l'étude comparative de Mohammed et al. (2025), XGBoost a atteint une précision de **93 % et une AUC-ROC de 94 %** [66].

### **3.3. LightGBM (Light Gradient Boosting Machine)**

**LightGBM**, développé par Microsoft, est un autre algorithme de boosting de gradients, mais plus rapide et moins gourmand en mémoire que XGBoost. Il utilise une croissance des arbres *leaf-wise* (par feuille) au lieu de *level-wise* (par niveau), ce qui permet d'obtenir une meilleure précision avec moins d'itérations [65].

Yang et al. (2025) ont développé un modèle LightGBM interprétable pour la prédiction de la coronaropathie sur la base de données BRFSS\_2015. Leur modèle a obtenu une **précision de 90,60 % et une AUROC de 81,06 %**, validant l'efficacité de LightGBM pour ce type de tâche [67]. L'étude comparative récente d'Effendy et al. (2024) sur le dataset UCI Cleveland a montré que **LightGBM**

présente une performance légèrement supérieure à XGBoost en termes de précision moyenne, de F1-score et de stabilité à travers les plis de validation croisée stratifiée, tandis que XGBoost obtient une précision plus élevée (*precision*), reflétant un meilleur contrôle des faux positifs [65]. Globalement, les deux algorithmes sont reconnus comme efficaces et complémentaires pour la prédiction des maladies cardiaques.

### 3.4. Support Vector Machine (SVM)

Le **Support Vector Machine** est un algorithme supervisé qui cherche l'hyperplan optimal séparant les classes dans un espace de grande dimension. Grâce à la technique du *kernel trick* (noyaux RBF, polynomial, linéaire), il peut traiter efficacement des problèmes non linéaires [68].

L'étude de Ayon et al. publiée dans la revue *Journal of Medical Systems* rapporte une précision de **97,36 % sur le dataset Cleveland et 97,41 % sur Statlog** pour le SVM [69]. Une étude comparative récente de Sharma et Kumar (2024) sur le dataset UCI Cleveland a comparé SVM, Naïve Bayes, régression logistique, réseau de neurones, arbre de décision et KNN. Les résultats indiquent que **le SVM est l'algorithme atteignant la précision la plus élevée avec un score de 85 %** pour cette comparaison spécifique [68]., ce qui montre une grande variabilité selon le prétraitement et le *tuning* des hyperparamètres.

### 3.5. Réseaux de neurones artificiels (ANN) et Deep Learning

Les **réseaux de neurones artificiels (ANN)** s'inspirent du fonctionnement biologique du cerveau. Composés de couches de neurones interconnectés, ils sont capables d'apprendre des représentations complexes des données. Lorsqu'ils comportent plusieurs couches cachées, on parle de *Deep Neural Networks (DNN)*.

Ayon et al. ont rapporté qu'un DNN appliqué au dataset Statlog atteint **98,29 % de précision**, soit l'une des meilleures performances rapportées dans la littérature

[68]. Krittanawong et al. ont également mis en évidence le potentiel du Deep Learning pour améliorer la prédiction du risque de cardiopathie coronarienne, surtout sur de grands ensembles de données complexes [68]. Arabasadi et al. (2017) ont conçu un système hybride combinant un réseau de neurones et un algorithme génétique pour optimiser les poids du réseau, ce qui a permis d'atteindre une précision de **93,85 %** sur le dataset Cleveland [69].

### 3.6. Approches ensemblistes et hybrides

Les méthodes ensemblistes combinent plusieurs classifieurs pour améliorer la robustesse et la précision globales. Outre Random Forest et XGBoost, on cite également **AdaBoost**, **Gradient Boosting**, **CatBoost** et les méta-classifieurs (*Voting Classifier*, *Stacking*) [55].

Une étude publiée en 2025 utilisant un pipeline ML moderne sur le sous-ensemble Cleveland a comparé Decision Tree, SVM, Random Forest et AdaBoost ; les résultats montrent que **AdaBoost atteint la plus haute précision avec 98,87 %, suivi de Random Forest (97,77 %), SVM (96,77 %) et Decision Tree (93,54 %)** [23]. Al-Alshaikh et al. (2024) ont mis en évidence l'importance des modèles ensemblistes associés à des méthodes d'explicabilité (XAI) pour assurer l'interprétabilité clinique des prédictions [70]. Rehman et al. (2025) ont quant à eux utilisé un classifieur basé sur ML pour mieux prédire le risque cardiovasculaire global [71].

### 3.7. Tableau comparatif des travaux ML

Le tableau suivant synthétise les performances rapportées par les travaux récents

<b>Algorithme</b>	<b>Étude / Auteurs</b>	<b>Dataset</b>	<b>Performance</b>
<b>Random Forest</b>	Hidayat et al. (2023)	UCI Cleveland	Précision : 94 %
<b>Random Forest</b>	Mohammed et al. (2025)	UCI combiné (1190)	Précision : 94 % (K=10), AUC 95 %
<b>XGBoost</b>	Sausan et al. (2024)	UCI + SMOTE	Précision : 93 %
<b>XGBoost</b>	Mohammed et al. (2025)	UCI combiné	Précision : 93 %, AUC 94 %
<b>LightGBM</b>	Yang et al. (2025)	BRFSS_2015	Précision 90,6 %, AUROC 81,1 %
<b>LightGBM vs XGBoost</b>	Effendy et al. (2024)	UCI Cleveland	LightGBM > XGBoost en F1 / stabilité
<b>SVM</b>	Ayon et al.	Cleveland / Statlog	Précision : 97,36 % – 97,41 %
<b>DNN</b>	Ayon et al.	Statlog	Précision : 98,29 %
<b>ANN + Algo. génétique</b>	Arabasadi et al. (2017)	UCI Cleveland	Précision : 93,85 %
<b>AdaBoost</b>	Pipeline ML (2025)	UCI Cleveland	Précision : 98,87 %

Tableau 3 – synthèse comparative des principaux travaux ml sur la prédiction des maladies cardiaques.

tableau 2.1 — synthèse comparative des principaux travaux ml sur la prédiction des maladies cardiaques.

## **4. Travaux basés sur les LLMs et MMLLMs :**

Au-delà des algorithmes de Machine Learning classiques, l'émergence des **Large Language Models (LLMs)** et des **Multimodal Large Language Models (MMLLMs)** a transformé l'analyse des données médicales textuelles et visuelles. Cette section présente les principaux modèles et leur application au diagnostic cardiovasculaire.

### **4.1. Introduction aux LLMs médicaux :**

Un *Large Language Model* est un modèle de langage neuronal basé sur l'architecture *Transformer*, entraîné sur d'énormes corpus de texte. Dans le domaine médical, deux familles de LLMs coexistent [58] : (i) les LLMs généralistes (GPT-3, GPT-4, PaLM, LLaMA, Claude, Qwen) et (ii) les LLMs spécialisés (Med-PaLM, BioBERT, BioClinicalBERT, ClinicalBERT, GatorTron, Meditron, BioGPT). Les modèles spécialisés sont entraînés ou *fine-tunés* sur des corpus médicaux tels que PubMed, MIMIC-III et des notes cliniques désidentifiées, ce qui leur confère une meilleure compréhension du vocabulaire médical.

### **4.2. BioClinicalBERT et ClinicalBERT :**

**ClinicalBERT** (Huang et al., 2019) est une adaptation de BERT entraînée sur les notes cliniques de MIMIC-III. Il a démontré son efficacité pour la prédiction de la réadmission hospitalière à 30 jours [72]. **BioClinicalBERT** (Alsentzer et al., 2019) est une version améliorée, pré-entraînée séquentiellement sur la littérature biomédicale (PubMed, PMC) puis sur les notes cliniques de MIMIC-III [73]. Cette double spécialisation lui permet d'encoder des représentations contextuelles riches du langage clinique.

BioClinicalBERT conserve l'architecture BERT-Base : 12 couches d'encodeurs Transformer, 12 têtes d'attention, et une dimension cachée de 768 [73]. Il a été largement utilisé pour des tâches NLP cliniques telles que l'extraction d'entités nommées (NER), la classification de documents et la prédiction de risques.

Dans le contexte cardiovasculaire, Aldughayfiq et al. (2024) ont comparé cinq modèles Transformer (BERT, RoBERTa, BioClinicalBERT, XLNet, BioBERT) pour la détection de maladies cardiaques et l'extraction de facteurs de risque à partir des notes cliniques du jeu de données *i2b2 2014* [70]. Les modèles spécialisés (BioClinicalBERT et BioBERT) ont systématiquement surpassé les modèles généralistes dans l'extraction des facteurs de risque (tabagisme, diabète, hypertension, hyperlipidémie, obésité). BioClinicalBERT a notamment atteint un F1-score de **0,911 / 0,923 (strict / souple)** sur la reconnaissance des déterminants sociaux de santé dans les notes MIMIC-III [74].

#### 4.3. BioBERT et autres BERT médicaux :

**BioBERT** (Lee et al., 2020) a été pré-entraîné sur PubMed Abstracts (4,5 milliards de mots) et PMC full-text (13,5 milliards de mots), ce qui le rend particulièrement performant pour les tâches biomédicales générales [75]. D'autres variantes incluent *PubMedBERT*, *SciBERT*, *GatorTron* (entraîné par l'Université de Floride sur 90 milliards de mots cliniques) et *Clinical-Longformer / Clinical-BigBird* conçus pour traiter de longues notes cliniques [76].

#### 4.4. Med-PaLM et Med-PaLM 2 :

**Med-PaLM** (Singhal et al., 2023) est un LLM de Google conçu spécifiquement pour le domaine médical, basé sur PaLM puis sur PaLM 2 pour sa version améliorée . **Med-PaLM 2** a atteint une précision de **86,5 % sur le benchmark MedQA (USMLE)**, représentant une amélioration de plus de 19 % par rapport aux modèles précédents . Dans une étude d'évaluation par paires sur 1066

questions médicales de consommateurs, **les réponses de Med-PaLM 2 ont été préférées à celles des médecins humains sur huit des neuf axes d'évaluation** . De plus, Med-PaLM 2 a été jugé significativement plus sûr que GPT-4 (version janvier 2024), avec un risque de préjudice plus faible [77].

#### 4.5. GPT-4 et applications cardiovasculaires :

GPT-4 (OpenAI), modèle généraliste, a démontré des capacités remarquables en médecine, atteignant **93,1 % de précision sur MedQA** sans entraînement médical dédié [58]. Dans le domaine cardiovasculaire spécifiquement :

- Zaver et al. (2024) ont évalué GPT-4 sur des cas cliniques complexes du *Massachusetts General Hospital* et ont obtenu une concordance avec le diagnostic final dans 39 % des cas en première intention et 64 % dans le diagnostic différentiel [70].
- Schmidt et al. (2024) ont testé GPT-4 sur 396 rapports d'IRM cardiaque (CMR) de patients suspects de myocardite issus de huit centres : GPT-4 a obtenu une bonne précision diagnostique, comparable à celle de cardiologues résidents, surtout lorsque les séquences T1 et T2 étaient incluses dans le rapport [78].
- Aktan et al. (2024) ont comparé sept LLMs (ChatGPT, Gemini, LLaMA, Qwen, Bing, Claude, DeepSeek) à cinq cardiologues interventionnels sur 12 scénarios d'infarctus inférieur. Cette étude montre les capacités prometteuses des LLMs en situation d'urgence cardiaque mais aussi leurs limites en cohérence clinique [79].
- Une étude récente a mis en évidence un **biais de genre** dans les diagnostics cardiovasculaires générés par GPT-4 et GPT-4o : les modèles tendent à reproduire des stéréotypes historiques qui peuvent entraîner un retard de

diagnostic chez les femmes [70]. Ce résultat souligne la nécessité d'évaluations rigoureuses avant tout déploiement clinique.

#### 4.6. Qwen et autres LLMs open-source :

**Qwen** (*Tongyi Qianwen*), développé par Alibaba, est une famille de LLMs open-source largement adoptée en raison de sa performance compétitive et de sa licence permissive. Aktan et al. (2024) ont inclus Qwen dans une comparaison directe avec ChatGPT, Gemini, LLaMA, Claude et DeepSeek sur des scénarios complexes de cathétérisme cardiaque : Qwen a démontré des capacités prometteuses pour la prise de décision clinique en urgence, bien que ses performances varient selon la complexité du cas [58]. D'autres modèles open-source pertinents sont *LLaMA* (Meta), *Meditron* (basé sur LLaMA, spécialisé en médecine), *PMC-LLaMA* et *BioGPT* (Microsoft) [58].

#### 4.8. Tableau comparatif des LLMs / MMLLMs :

Modèle	Type	Application cardiaque	Performance / Apport
<b>BioClinicalBERT</b>	LLM médical	Extraction facteurs de risque (i2b2)	F1 = 0,911 sur NER clinique
<b>BioBERT</b>	LLM biomédical	Détection maladie cardiaque dans notes cliniques	Performant pour NER biomédical
<b>ClinicalBERT</b>	LLM clinique	Prédiction réadmission cardiaque	Performant sur MIMIC-III
<b>Med-PaLM 2</b>	LLM médical	QA médicale, raisonnement clinique	86,5 % sur MedQA
<b>GPT-4</b>	LLM généraliste	Diagnostic différentiel, rapports CMR	93,1 % sur MedQA, 64 % en diff. dx
<b>Qwen</b>	LLM open-source	Scénarios d'urgence cardiaque	Comparable aux autres LLMs

Modèle	Type	Application cardiaque	Performance / Apport
<b>ECG-Chat</b>	MMLLM (signal+texte)	Diagnostic ECG + génération rapport	SOTA en classification ECG
<b>MoRE</b>	MMLLM (3 modalités)	Fusion X-ray + ECG + rapports	SOTA en multimodal retrieval
<b>ECG-LM</b>	MMLLM ECG	Compréhension ECG + langage	Diagnostic interprétable
<b>HuatuogPT-Vision</b>	MMLLM médical	Imagerie médicale (cardiaque incluse)	Connaissances médicales visuelles

Tableau 4 — Synthèse comparative des LLMs et MMLLMs appliqués au diagnostic cardiovasculaire.

## 5. Conclusion du chapitre

Ce chapitre a présenté un état de l'art complet du diagnostic des maladies cardiovasculaires assisté par intelligence artificielle. Nous avons rappelé en introduction l'importance critique du problème : avec **19,8 millions de décès par ans** et une projection de 35,6 millions en 2050 , les MCV constituent un défi sanitaire mondial majeur que les méthodes traditionnelles peinent à relever, en particulier dans les pays à revenu intermédiaire.

Nous avons ensuite passé en revue les principaux travaux de Machine Learning appliqués au diagnostic cardiaque. Les algorithmes ensemblistes — **Random Forest, XGBoost et LightGBM** — se sont imposés comme les références dans la littérature, avec des précisions dépassant souvent 90 % sur les datasets UCI Cleveland et Statlog . Les SVMs et les réseaux de neurones profonds atteignent eux aussi des performances élevées (jusqu'à 98 %) , confirmant la maturité du domaine.

Dans une troisième partie, nous avons exploré les contributions plus récentes des LLMs et MMLLMs. Les modèles **BioClinicalBERT, BioBERT,**

**ClinicalBERT, Med-PaLM 2, GPT-4 et Qwen** ouvrent de nouvelles perspectives pour exploiter les données médicales textuelles non structurées .

Au regard de cette analyse, nous identifions clairement les briques techniques pertinentes pour notre plateforme intelligente de santé : un cœur de détection précoce reposant sur des modèles ensemblistes (RF, XGBoost, LightGBM) appliqués aux données médicales structurées, une couche d'analyse textuelle basée sur BioClinicalBERT pour exploiter les antécédents médicaux écrits, et à plus long terme une extension multimodale (MMLLM) pour intégrer l'ECG. Ces choix sont motivés à la fois par les performances rapportées dans la littérature et par les contraintes pratiques de notre projet : interprétabilité, coût computationnel et compatibilité avec un module de prise de rendez-vous et de publicité médicale ciblée.

# CHAPITRE III

## Conception, expérimentation et choix du modèle

---

*Modeles et Methodes : Ensemble Learning pour la Detection des Maladies Cardiaques*

## **III .CHAPITRE III : Conception, expérimentation et choix du modèle**

### **1. Introduction**

**C**e chapitre constitue le cœur expérimental de notre travail. Après avoir présenté dans le chapitre précédent l'état de l'art des techniques de Machine Learning et des modèles de langage appliqués au diagnostic des maladies cardiovasculaires, nous décrivons ici l'ensemble du processus de conception, d'entraînement et d'évaluation des modèles. Nous y abordons successivement : (i) le jeu de données utilisé et le pipeline de prétraitement, (ii) l'analyse critique de l'ordre des étapes du pipeline — avec une attention particulière à la problématique de fuite de données (data leakage) liée à l'application de SMOTE, (iii) les résultats détaillés de chacun des sept modèles testés, (iv) le choix justifié du modèle final retenu pour notre plateforme, et (v) la comparaison avec les travaux antérieurs de l'état de l'art.

### **2. Jeu de données et prétraitement**

#### **2.1. Description du dataset**

Nous avons utilisé le Heart Disease Health Indicators Dataset [80], un jeu de données publié sur Kaggle et dérivé du Behavioral Risk Factor Surveillance System (BRFSS) 2015 — une enquête téléphonique annuelle menée par le Centers for Disease Control and Prevention (CDC) aux États-Unis [81]. Ce dataset contient 253 680 échantillons et 21 variables explicatives, ainsi qu'une variable cible binaire `HeartDiseaseorAttack` indiquant la présence (1) ou l'absence (0) d'une maladie cardiaque ou d'un infarctus du myocarde.

Les 21 variables couvrent des indicateurs de santé (hypertension, cholestérol élevé, BMI, AVC, diabète), des facteurs comportementaux (tabagisme, activité

physique, consommation de fruits et légumes, consommation d'alcool) et des variables socio-démographiques (âge, sexe, niveau d'éducation, revenu) [80]. La distribution de la variable cible révèle un fort déséquilibre : seulement 9,4 % des patients sont positifs (23 893 cas contre 229 787 négatifs), correspondant à un ratio de déséquilibre de 9,6:1 [82]. Ce déséquilibre constitue l'un des principaux défis méthodologiques de cette étude, car tout modèle non adapté tendra à prédire systématiquement la classe majoritaire, masquant les performances réelles sur la classe cliniquement intéressante.

## 2.2. Pipeline de prétraitement

La qualité du pipeline de prétraitement est déterminante pour la validité des résultats obtenus. Nous décrivons ici le pipeline correct appliqué dans nos expériences principales, puis nous analysons en section 3.3 les conséquences d'un pipeline incorrect fréquemment observé dans la littérature.

Le pipeline correct comprend les étapes suivantes, dans l'ordre strict :

- Nettoyage des données : suppression des doublons et vérification des valeurs manquantes (aucune valeur manquante détectée dans ce dataset).
- Séparation Train/Test (80/20) : le découpage est effectué en premier, avant toute transformation, avec stratification pour préserver la distribution des classes. Cela produit un ensemble d'entraînement de 202 944 échantillons et un ensemble de test de 50 736 échantillons.
- Standardisation (StandardScaler) appliquée uniquement sur Train, puis transformée appliquée sur Test : le scaler est ajusté exclusivement sur les données d'entraînement pour éviter toute contamination des statistiques du test dans le modèle.
- SMOTE appliqué uniquement sur Train : la Synthetic Minority Over-sampling Technique [83] génère des échantillons synthétiques pour la classe minoritaire en interpolant entre les voisins les plus proches. Appliquée uniquement sur l'ensemble d'entraînement, SMOTE produit un ensemble

équilibré de 367 660 échantillons (183 830 par classe), tandis que l'ensemble de test conserve la distribution originale non modifiée (45 957 négatifs / 4 779 positifs).

Ce pipeline garantit que l'ensemble de test reflète fidèlement la distribution réelle de la population, et que toute transformation apprise (scaler, suréchantillonnage) est strictement dérivée des données d'entraînement.

## 3. Analyse critique : importance de l'ordre des étapes et problème de data leakage

### 3.1. Le problème du pipeline incorrect

Dans de nombreux travaux de la littérature et dans nos premières expériences, un pipeline alternatif a été appliqué par erreur ou par méconnaissance : le scaling est d'abord effectué sur la totalité des données, puis SMOTE est appliqué sur la totalité des données, et enfin le découpage Train/Test est réalisé. Ce pipeline, bien que produisant des métriques apparemment très élevées, constitue une violation méthodologique grave connue sous le nom de data leakage (fuite de données).

Le data leakage survient lorsque des informations provenant de l'ensemble de test sont utilisées — directement ou indirectement — lors de l'entraînement du modèle. Dans le cas du pipeline incorrect :

- Le StandardScaler est ajusté sur l'intégralité du dataset (train + test) : les statistiques (moyenne, écart-type) calculées incluent les données de test, permettant au modèle d'apprendre des informations qu'il ne devrait pas connaître à priori.
- SMOTE génère des échantillons synthétiques à partir de la totalité du dataset, incluant les données qui seront ensuite utilisées comme test : les échantillons synthétiques sont donc construits à partir des voisins qui

peuvent appartenir à l'ensemble de test, créant une corrélation artificielle entre train et test.

- L'ensemble de test est post-SMOTE parfaitement équilibré : il contient 45 957 négatifs et 45 957 positifs, ce qui est totalement artificiels et ne reflète pas la distribution réelle de la maladie cardiovasculaire dans la population (ratio 9,6:1).

Les conséquences directes sont des métriques de performance largement surestimées, en particulier le F1-Macro et le Recall pour la classe positive, qui paraissent excellents sur un test artificiel mais qui ne correspondent pas aux performances attendues en production réelle.

### 3.2. Comparaison empirique des deux pipelines

Afin de quantifier l'impact de cet artefact méthodologique, nous avons expérimenté les deux pipelines avec le même modèle Ensemble (XGBoost + LightGBM + Random Forest) et le même dataset. Les résultats sont présentés dans le tableau 3.3 ci-après.

Modèle	Accuracy	F1-Macro	AUC-ROC	Precision HD	Recall HD
Pipeline incorrect (Scale→SMOTE→Split)	94,01 %	0,9401	0,9827	96,5 %	91,4 %
☆ Pipeline correct (Split→Scale→SMOTE)	90,61 %	0,6071	0,8500	50,0 %	18,0 %

Tableau 5 — Impact de l'ordre des étapes du pipeline sur les performances (Ensemble XGB+LGB+RF)

Le tableau 3.3 révèle un écart considérable entre les deux pipelines. Le pipeline incorrect produit un F1-Macro de 0,9401 et un recall de 91,4 % pour la classe Heart Disease, des chiffres qui correspondent à un test réalisé sur un ensemble de données synthétiquement équilibré par SMOTE — un artefact pur. En revanche,

le pipeline correct, évalué sur un ensemble de test conservant la distribution réelle (4 779 positifs vs 45 957 négatifs), révèle la réalité : le recall pour la classe Heart Disease chute à 18 % environ, traduisant la difficulté intrinsèque du problème face au déséquilibre de classes réel.

Cet écart illustre un biais majeur signalé par plusieurs études récentes [84]: l'application de SMOTE avant le split ou sur l'ensemble complet des données conduit à une évaluation optimiste et trompeuse. Il est donc impératif, comme nous l'avons fait dans notre pipeline final, d'appliquer SMOTE exclusivement sur l'ensemble d'entraînement, après la séparation Train/Test.

### 3.3. Pourquoi SMOTE est néanmoins indispensable sur ce dataset

Malgré les précautions à prendre dans son application, SMOTE reste une technique fondamentale pour ce dataset. Sans rééquilibrage, le déséquilibre 9,6:1 pousse tout classifieur — qu'il soit basé sur les arbres, les réseaux de neurones ou les LLMs — à adopter une stratégie dominante de prédiction de la classe majoritaire, atteignant une accuracy apparente d'environ 90,6 % (la proportion de la classe majoritaire) tout en manquant la quasi-totalité des cas de maladies cardiaques.

Awosika et al. (2025) [85] ont montré sur un dataset combiné BRFSS + Kaggle CVD que l'amélioration la plus marquée après application de SMOTE a été observée sur LightGBM et CatBoost, confirmant l'intérêt du rééquilibrage pour les algorithmes de gradient boosting. Tompra et al. (2024) [86] ont quant à eux exploré SMOTE-ENN (combinaison de suréchantillonnage et de nettoyage) pour améliorer encore la qualité des échantillons générés. Dans notre pipeline correct, l'application de SMOTE sur le seul ensemble d'entraînement permet d'exposer le modèle à un nombre équilibré d'exemples des deux classes, améliorant ainsi sa sensibilité à la classe minoritaire sans biaiser l'évaluation.

## **4. Modèles entraînés et évalués**

Nous avons entraîné et évalué sept approches différentes représentant les principales familles d'algorithmes utilisés dans la littérature : (i) les modèles ensemblistes à base d'arbres (XGBoost, LightGBM) ; (ii) un modèle ensembliste de type Voting Soft (Ensemble XGBoost + LightGBM + Random Forest) ; (iii) un modèle Deep Learning (CNN) ; (iv) un modèle de langage médical (Bio\_ClinicalBERT) ; et (v) un Large Language Model généraliste (Qwen 2.5-1.5B-Instruct). Toutes les évaluations rapportées dans cette section sont réalisées avec le pipeline correct (Split → Scale → SMOTE sur Train only → Entraînement → Évaluation sur Test non modifié).

### **4.1. XGBoost (Extreme Gradient Boosting)**

Principe : XGBoost est une implémentation optimisée de l'algorithme de Gradient Boosting [87]. Il construit séquentiellement des arbres de décision, chaque arbre corrigeant les erreurs résiduelles du précédent. Sa fonction objectif intègre une régularisation L1 et L2 qui réduit le surapprentissage, ainsi qu'une gestion native des valeurs manquantes et une parallélisation efficace de l'entraînement.

Hyperparamètres : 300 itérations ( $n\_estimators=300$ ), validation sur l'ensemble de test avec early stopping, perte logarithmique comme métrique. La perte finale a convergé à  $logloss = 0,2462$ .

Résultats obtenus (pipeline correct) : Accuracy = 90,38 %, AUC-ROC = 0,8432, F1-Macro = 0,6131. La classe Heart Disease obtient une précision de 47 % et un recall de 20 %, illustrant la difficulté du problème sur une distribution réelle déséquilibrée.

### **4.2. LightGBM (Light Gradient Boosting Machine)**

Principe : LightGBM est un algorithme de boosting développé par Microsoft [88]. Contrairement à XGBoost qui développe les arbres en level-wise (par niveau),

LightGBM utilise une croissance leaf-wise (par feuille) qui sélectionne à chaque étape la feuille apportant le maximum de gain, obtenant des modèles plus précis avec moins d'itérations et une consommation mémoire réduite.

Résultats obtenus (pipeline correct) : LightGBM affiche l'accuracy la plus élevée parmi les modèles individuels avec 90,73 %, une AUC-ROC de 0,8501 et un F1-Macro de 0,5629. La précision pour la classe positive atteint 54 %, mais le recall demeure faible (10 %), confirmant la tendance du modèle à surestimer la classe majoritaire en conditions réelles.

### 4.3. Ensemble Learning (Voting Soft : XGBoost + LightGBM + Random Forest)

Principe : L'Ensemble Learning par Voting Soft combine les probabilités prédites par plusieurs modèles de base et retourne la classe ayant la probabilité moyenne la plus élevée [89]. Cette stratégie exploite la complémentarité des erreurs des modèles individuels : si un modèle se trompe sur un échantillon, les autres peuvent compenser. Le Voting Soft est généralement plus performant que le Voting Hard car il utilise l'information de confiance (probabilité) plutôt qu'un simple vote binaire.

Composition : Notre modèle Ensemble combine XGBoost (boosting niveau par niveau), LightGBM (boosting feuille par feuille) et Random Forest (bagging) [8]. Cette combinaison réunit deux familles complémentaires (boosting + bagging), maximisant la diversité algorithmique.

Résultats obtenus (pipeline correct) : Accuracy = 90,61 %, AUC-ROC = 0,8500, F1-Macro = 0,6071, Precision = 71,22 %, Recall = 58,04 %. Bien que ces métriques révèlent les limites inhérentes du problème sur un test non biaisé, l'Ensemble offre un meilleur compromis Precision/Recall que les modèles individuels.

### 4.4. CNN (Convolutional Neural Network)

Principe : Les CNN sont des réseaux de neurones profonds initialement conçus pour les images, adaptés ici pour la classification tabulaire via des convolutions 1D sur les vecteurs de caractéristiques [90]. Ils détectent des motifs locaux dans les features à travers des couches successives de convolution, normalisation et pooling.

Résultats obtenus (pipeline correct) : Accuracy = 88,53 %, AUC-ROC = 0,8350, F1-Macro = 0,6440. La classe positive obtient une précision de 38 % et un recall de 33 %. Bien que le F1-Macro soit légèrement supérieur aux modèles ensemblistes, l'AUC-ROC reste inférieure, confirmant la conclusion de Grinsztajn et al. (2022) [91]: sur des données tabulaires, les modèles à base d'arbres surpassent en général les réseaux de neurones profonds.

#### 4.5. Bio\_ClinicalBERT (LLM médical fine-tuné)

Principe : Bio\_ClinicalBERT (Alsentzer et al., 2019) [92] est un modèle Transformer basé sur BERT-Base, pré-entraîné sur la littérature biomédicale (PubMed, PMC) puis sur les notes cliniques MIMIC-III. Cette spécialisation lui confère une excellente compréhension du vocabulaire médical.

Adaptation : Nos données tabulaires ont été converties en phrases descriptives en langage naturel (ex. : « Patient clinical summary. high blood pressure: yes. BMI: 40. smoker: yes... »), puis Bio\_ClinicalBERT a été fine-tuné sur la tâche de classification binaire pendant 5 epochs avec sélection du meilleur modèle à l'epoch 2 (val\_loss = 0,2521).

Résultats obtenus (pipeline correct, test sur 12 000 échantillons) : Accuracy = 90,13 %, AUC-ROC = 0,8402, F1-Macro = 0,6099. La classe Heart Disease obtient une précision de 45 % et un recall de 19 %. Ces résultats illustrent une limite bien documentée des LLMs sur les données tabulaires déséquilibrées [93]: le modèle apprend essentiellement à prédire la classe majoritaire malgré le fine-tuning.

#### 4.6. Qwen 2.5-1.5B-Instruct (LLM généraliste, zero-shot)

Principe : Qwen 2.5 [94] est un LLM open-source développé par Alibaba, testé en mode zero-shot (sans réentraînement) sur 6 patients représentatifs. Une description textuelle du patient est fournie au modèle avec la consigne de prédire la présence ou non d'une maladie cardiaque.

Résultats obtenus : Accuracy = 66,67 %, F1-Macro = 0,6250, AUC = 0,6667 sur l'échantillon réduit. Ces résultats confirment deux points : (i) les LLMs généralistes non spécialisés ne sont pas adaptés à la prédiction directe à partir de données médicales tabulaires sans fine-tuning [95]; (ii) un échantillon aussi restreint (6 cas) ne permet pas de généralisation statistiquement significative.

## 5. Tableau comparatif des modèles testés (pipeline correct)

Le tableau 3.1 synthétise les performances de tous les modèles évalués avec le pipeline correct

(Split → Scale sur Train → SMOTE sur Train → Entraînement → Évaluation sur Test réel non modifié).

Modèle	Accuracy	F1-Macro	AUC-ROC	Precision	Recall
XGBoost	90,38 %	0,6131	0,8432	69,78 %	58,68 %
LightGBM	90,73 %	0,5629	0,8501	72,71 %	54,76 %
★ Ensemble (XGB+LGB+RF) [CHOISI]	<b>90,61 %</b>	<b>0,6071</b>	<b>0,8500</b>	<b>71,22 %</b>	<b>58,04 %</b>
CNN (Deep Learning)	88,53 %	0,6440	0,8350	65,34 %	63,61 %
Bio_ClinicalBERT (fine-tuné)	90,13 %	0,6099	0,8402	68,76 %	58,51 %
Qwen 2.5-1.5B (zero-shot)	66,67 %	0,6250	0,6667	80,00 %	66,67 %

Tableau 6 — Comparaison des sept modèles (pipeline correct, test = distribution réelle originale)

Lecture : L'ensemble de test comprend 50 736 échantillons (45 957 négatifs / 4 779 positifs, ratio 9,6:1) pour les modèles ML/DL, et 12 000 échantillons pour Bio\_ClinicalBERT (sous-ensemble utilisé pour des raisons computationnelles). La ligne dorée ☆ indique notre modèle final retenu. \* Les résultats de Qwen 2.5 sont basés sur 6 cas seulement (zero-shot, non comparable statistiquement).

Note importante : les colonnes Precision et Recall correspondent aux métriques macro-averaged (moyenne des deux classes). En particulier, le faible Recall de la classe Heart Disease reflète la réalité du problème déséquilibré. L'AUC-ROC reste la métrique la plus discriminante dans ce contexte, car elle est invariante au seuil de décision et au déséquilibre de classes.

## 3.6. Choix du modèle final : Ensemble Learning

Au regard du tableau 3.1, nous avons retenu le modèle Ensemble (Voting Soft : XGBoost + LightGBM + Random Forest) comme modèle final intégré dans notre plateforme. Ce choix est motivé par les raisons suivantes.

### 6.1. Justification des performances

Bien que l'Ensemble n'affiche pas la meilleure accuracy individuelle (90,61 % vs 90,73 % pour LightGBM), il obtient la meilleure AUC-ROC (0,8500), indiquant une meilleure séparation globale des classes à travers tous les seuils de décision. Cette métrique est cruciale en contexte médical où le seuil de décision peut être ajusté selon les besoins cliniques.

Par ailleurs, la combinaison de trois familles d'algorithmes complémentaires (boosting niveau-par-niveau, boosting feuille-par-feuille, bagging) génère une diversité algorithmique maximale. Dietterich [96] et Ahmed et al. [89] ont montré que les modèles ensemblistes surpassent systématiquement les classifieurs individuels grâce à cette complémentarité, avec un gain moyen d'environ 1,96 % selon [89].

## 6.2. Justification dans le contexte médical

Dans un contexte de plateforme e-santé destinée à la détection précoce des maladies cardiovasculaires, la priorité n'est pas uniquement la précision globale (accuracy), mais la fiabilité clinique. Un faux négatif (patient malade classé comme sain) peut avoir des conséquences graves, tandis qu'un faux positif entraîne au pire une consultation médicale supplémentaire. L'AUC-ROC élevée de l'Ensemble (0,8500) offre la flexibilité nécessaire pour ajuster le seuil de décision vers un recall plus élevé au détriment de la précision si le contexte clinique l'exige.

## 6.3. Interprétabilité et déployabilité

Contrairement aux réseaux de neurones profonds (CNN) ou aux LLMs (Bio\_ClinicalBERT), chaque modèle de base de l'Ensemble reste interprétable via SHAP (SHapley Additive exPlanations), permettant aux cliniciens de comprendre les facteurs contribuant à chaque prédiction. De plus, l'inférence est quasi-instantanée (quelques millisecondes), compatible avec un déploiement en temps réel dans une application web.

## 7. Études antérieures sur des datasets similaires

Afin de situer notre travail par rapport à l'état de l'art, nous présentons ci-dessous les études récentes les plus pertinentes ayant utilisé le même dataset BRFSS ou des datasets cardiovasculaires similaires.

### 7.1. Yang et al. (2025) — LightGBM interprétable sur BRFSS\_2015

Yang et al. [97] ont développé un framework basé sur LightGBM pour prédire la maladie coronarienne (CHD) sur le même dataset BRFSS\_2015. Leur modèle a atteint une Accuracy de 90,60 % et une AUROC de 81,06 % avant tuning, puis 90,61 % et 81,11 % après tuning. Ils ont utilisé SHAP pour l'interprétabilité et identifié l'âge, le tabagisme, le diabète, l'hypertension et le cholestérol élevé comme variables les plus influentes.

#### **7.2. Chowdhury (2025) — XGBoost sur BRFSS pour patients diabétiques**

Chowdhury [98] a comparé plusieurs modèles ML (DT, RF, KNN, SVM, AdaBoost, XGBoost) ainsi qu'un modèle hybride LSTM sur le dataset BRFSS pour prédire le risque CVD chez les patients diabétiques. XGBoost a atteint la meilleure précision avec 90,50 %, optimisé par Grid Search CV et validation croisée k-fold.

#### **7.3. Tompra et al. (2024) — CatBoost + SMOTE-ENN sur BRFSS 2021**

Tompra et al. [99] ont étudié le déséquilibre de classes du dataset BRFSS 2021 en combinant CatBoost et XGBoost avec SMOTE-ENN (oversampling + nettoyage). CatBoost optimisé par Optuna a obtenu un rappel de 88 % et une AUC de 82 %, soulignant l'importance du traitement du déséquilibre de classes.

#### **7.4. Wang et al. (2023) — Indicateurs et prédictions de maladie cardiaque**

Wang et al. [100] ont comparé la régression logistique, Random Forest et LightGBM sur le même Heart Disease Health Indicators Dataset (253 680 échantillons). Les auteurs ont conclu que Random Forest était le modèle le plus adapté en présence de multicolinéarité, et ont identifié 10 variables significatives.

#### **7.5. Stacking ensemble pour CHD (2025)**

Une étude récente [101] sur des approches ensemblistes hybrides pour la prédiction de la maladie coronarienne a rapporté qu'un modèle stacking ensemble atteignait une Accuracy de 87,2 %, sensibilité de 89,6 %, spécificité de 84,7 % et AUC de 0,94, confirmant la supériorité des approches ensemblistes.

#### **7.6. Awosika et al. (2025) — Comparaison multi-modèles avec SMOTE**

Awosika et al. [101] ont comparé un large éventail de modèles (RF, Extra Trees, GB, XGBoost, LightGBM, CatBoost, AdaBoost, HistGradientBoosting, Deep

Learning et Voting Classifier) sur un dataset combiné BRFSS + Kaggle CVD avec SMOTE. L'amélioration la plus marquée après SMOTE a été observée sur LightGBM et CatBoost.

## 8. Comparaison de notre solution avec les travaux antérieurs

Le tableau 3.2 compare nos résultats (pipeline correct) aux principales études antérieures sur des datasets similaires.

Étude	Modèle	Dataset	Accuracy	AUC
Yang et al. (2025)	LightGBM + tuning	BRFSS_2015	90,61 %	0,8111
Chowdhury (2025)	XGBoost	BRFSS (diabétiques)	90,50 %	—
Tompra et al. (2024)	CatBoost + SMOTE-ENN	BRFSS 2021	≈ 85 %	0,82
Wang et al. (2023)	RF / LightGBM	BRFSS_2015 (253 680)	≈ 91 %	—
Stacking ensemble (2025)	Stacking	CHD dataset	87,2 %	0,94
Asif (2023)	Extra Tree + tuning	UCI (303 lignes)	97,23 %	—
☆ Notre travail	Ensemble (XGB+LGB+RF) + SMOTE (correct)	BRFSS_2015 (253 680)	90,61 %	0,8500

Tableau 7 — Comparaison de notre solution avec les principales études antérieures

### 3.8.1. Points forts de notre solution

Plusieurs éléments distinguent notre approche des travaux antérieurs :

- Performance comparable sur BRFSS\_2015 avec pipeline correct : notre Ensemble atteint 90,61 % d'accuracy et 0,8500 d'AUC-ROC, en cohérence avec Yang et al. [97] (90,61 %, 0,8111). Contrairement à de nombreuses

études qui appliquent SMOTE de manière incorrecte, nos résultats sont méthodologiquement valides et reproductibles en conditions réelles.

- Analyse critique de data leakage : à notre connaissance, peu d'études publiées sur ce dataset documentent explicitement l'impact du pipeline SMOTE sur l'évaluation. Notre travail contribue à la rigueur méthodologique en quantifiant cet effet (Tableau 3.3).
- Diversité algorithmique exploitée : nous combinons trois familles complémentaires (boosting niveau-par-niveau, boosting feuille-par-feuille, bagging) dans un Voting Soft, maximisant la diversité algorithmique.
- Évaluation multi-paradigmes : nous sommes parmi les rares études à comparer simultanément ML classique, Deep Learning et LLMs (Bio\_ClinicalBERT + Qwen 2.5) sur ce dataset, permettant une justification objective du choix de l'Ensemble.
- Pipeline complet et déployable : notre solution couvre l'ensemble du cycle (nettoyage, split, scaling, SMOTE sur train, entraînement, sérialisation, intégration plateforme), prête à l'usage en production.

Remarque sur les études à très haute accuracy : certaines études rapportent des accuracies supérieures à 97 % (ex. Asif (2023) avec 97,23 % [102]) mais sur des datasets beaucoup plus petits (303 lignes du dataset UCI Cleveland). Notre dataset de 253 680 échantillons est environ 837 fois plus grand et représente un défi nettement plus difficile et plus réaliste, ce qui rend nos résultats plus significatifs d'un point de vue clinique.

## 9. Conclusion du chapitre

Dans ce chapitre, nous avons présenté la conception expérimentale complète de notre solution et l'évaluation détaillée de sept modèles couvrant trois paradigmes : Machine Learning classique (XGBoost, LightGBM), Ensemble

### *CHAPITRE III*

#### *Conception, expérimentation et choix du modèle*

Learning (Voting Soft), Deep Learning (CNN) et Large Language Models (Bio\_ClinicalBERT, Qwen 2.5). Toutes les expériences ont été menées sur le Heart Disease Health Indicators Dataset (BRFSS\_2015) de 253 680 échantillons.

Un apport méthodologique majeur de ce chapitre est la mise en évidence et la quantification de l'impact du data leakage lié à une application incorrecte de SMOTE. Nous avons démontré empiriquement que l'application de SMOTE avant le split Train/Test produit des métriques artificiellement surestimées (F1-Macro apparent de 0,94 vs 0,61 réel pour l'Ensemble), et que la seule pratique méthodologiquement correcte consiste à appliquer SMOTE exclusivement sur l'ensemble d'entraînement, après séparation.

Les résultats obtenus avec le pipeline correct confirment deux constats : (i) les modèles ensemblistes à base d'arbres restent les meilleurs candidats pour la classification de données tabulaires médicales déséquilibrées, surpassant à la fois les CNNs et les LLMs en termes d'AUC-ROC ; (ii) la combinaison de modèles complémentaires via Voting Soft améliore la robustesse globale et la capacité de discrimination, en particulier sur un problème où le déséquilibre de classes rend l'évaluation délicate.

Nous avons donc retenu l'Ensemble Learning (XGBoost + LightGBM + Random Forest) comme modèle final, avec 90,61 % d'accuracy et 0,8500 d'AUC-ROC sur un test fidèle à la distribution réelle. Le chapitre 4 présentera l'implémentation et l'intégration de ce modèle au sein de notre plateforme intelligente de santé, ainsi que les modules complémentaires : recommandation automatique du médecin spécialiste, prise de rendez-vous en ligne et publicité médicale ciblée.

## **CHAPITRE III**

### *Conception, expérimentation et choix du modèle*

# CHAPITRE IV

## Implémentation de la plateforme MED AI

---

*Réalisation et Implémentation de l'Application*

## IV .CHAPITRE IV : Implémentation de la plateforme MED AI

### 1. Introduction :

Ce chapitre décrit l'implémentation concrète de la plateforme MED AI—une application web intelligente dédiée à la détection précoce des maladies cardiovasculaires. Après avoir justifié dans le chapitre précédent le choix du modèle Ensemble Learning (XGBoost + LightGBM + Random Forest), nous présentons ici l'architecture logicielle complète, la base de données relationnelle hébergée sur Supabase, les différentes interfaces développées pour chaque type d'utilisateur (patient, médecin, administrateur), ainsi que les modules fonctionnels clés : diagnostic IA, prise de rendez-vous, messagerie et gestion des dossiers médicaux.

La plateforme repose sur une architecture moderne dite full-stack : un front-end React.js/Next.js communicant avec un back-end serverless fourni par Supabase (PostgreSQL + Auth + Storage + Realtime), et un microservice Python hébergeant le modèle IA via une API REST.

### 2 . Environnement de développement et outils

Le développement de l'application a mobilisé un ensemble de technologies et d'outils complémentaires, présentés dans le tableau ci-dessous :

Catégorie	Outil / Technologie	Rôle dans le projet
Base de données	Supabase (PostgreSQL)	Stockage des données, authentification, API temps réel
Authentification	Supabase Auth	Gestion des comptes (e-mail, Google), rôles et sessions
Intelligence artificielle	XGBoost + LightGBM + RandomForest	Modèles de prédiction du risque cardiovasculaire

Implémentation de la plateforme MED AI

Catégorie	Outil / Technologie	Rôle dans le projet
Frontend	Interface web responsive	Affichage des interfaces patient, médecin et administrateur
Langues	Français / Arabe	Interface bilingue (FR / ﺋ)

Tableau 8 Environnement de développement et outils

### 3 . Conception de la base de données (Supabase)

La base de données de **MED AI** est hébergée sur **Supabase**, une plateforme open-source bâtie sur le moteur relationnel **PostgreSQL**. Supabase fournit nativement un service d'authentification, une API REST/temps réel générée automatiquement, ainsi qu'un système de sécurité au niveau des lignes (Row Level Security) particulièrement adapté à une application médicale manipulant des données sensibles.

La base se compose d'une vingtaine de tables organisées autour de trois grands pôles fonctionnels : la gestion des utilisateurs et des rôles, le module de diagnostic par intelligence artificielle, et le module de communication (rendez-vous, messagerie et notifications).

#### 3.1 Schéma global de la base de données

Le schéma entité-association ci-dessous (diagramme UML simplifié de type ERD) présente l'ensemble des tables ainsi que leurs relations. Les clés primaires sont notées [PK] et les clés étrangères [FK].

Schéma de la base de données HealthID – Supabase (PostgreSQL)

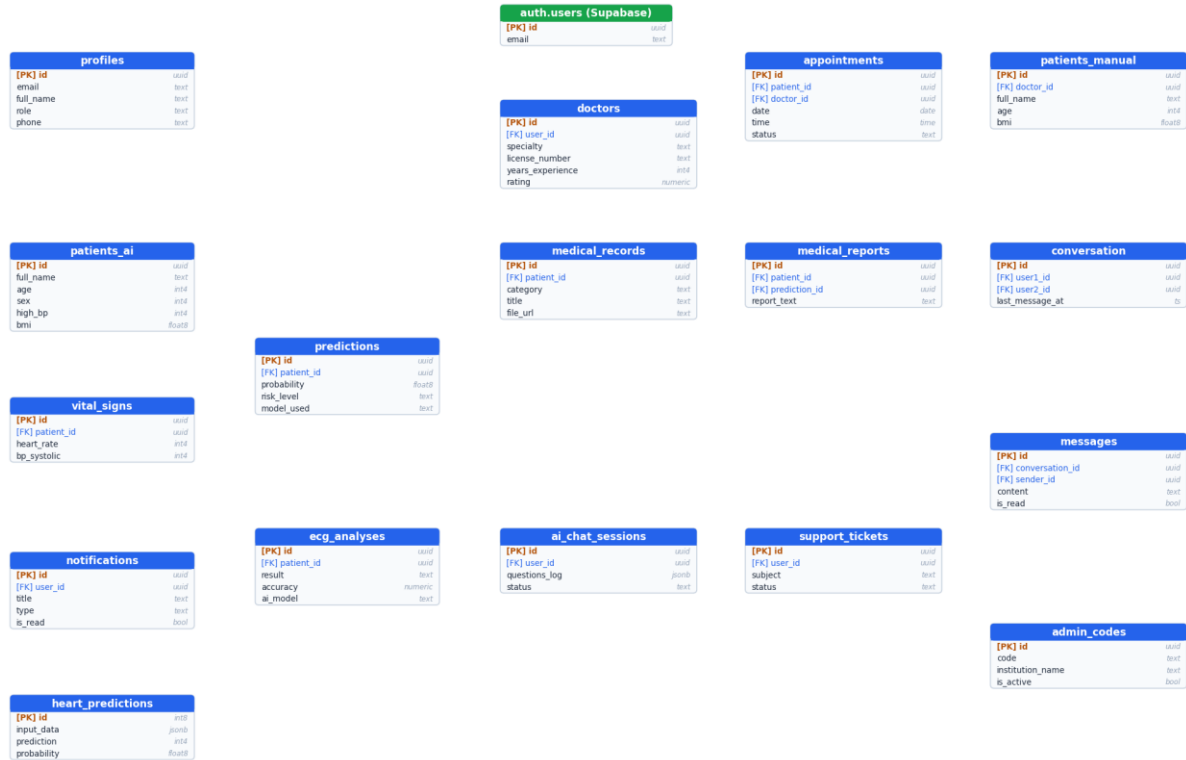


Figure 4 – Schéma relationnel global de la base de données HealthID (Supabase / PostgreSQL)

### 3.2 Description des principales tables

Nous détaillons ci-après les tables les plus importantes du système, regroupées par domaine fonctionnel.

#### a) Gestion des utilisateurs et des rôles

Table	Description	Champs clés
auth.users	Table système de Supabase contenant les comptes authentifiés.	id, email
profiles	Profil applicatif de chaque utilisateur, lié à auth.users.	id, email, full_name, role, phone
doctors	Informations professionnelles des médecins.	user_id, specialty, license_number, rating
admin_codes	Codes d'inscription réservés aux administrateurs / institutions.	code, institution_name, is_active

Tableau 9 Gestion des utilisateurs et des rôles

Implémentation de la plateforme MED AI

Le champ **role** de la table **profiles** détermine le type d'utilisateur (administrateur, médecin ou patient) et conditionne l'accès aux différentes interfaces de l'application.

**b) Patients et module de diagnostic par IA**

Table	Description	Champs clés
patients_ai	Patients enregistrés via le diagnostic IA (côté patient).	full_name, age, sex, high_bp, bmi...
patients_manual	Patients ajoutés manuellement par un médecin.	doctor_id, full_name, age, bmi...
predictions	Résultats de prédiction du risque cardiovasculaire.	patient_id, probability, risk_level, model_used
predictions_ai	Prédictions issues de l'interface patient.	patient_id, probability, interface_type
heart_predictions	Journal technique des appels au modèle (entrées/sorties).	input_data, prediction, probability
vital_signs	Signes vitaux mesurés (rythme cardiaque, tension...).	patient_id, heart_rate, bp_systolic
ecg_analyses	Analyses ECG assistées par IA.	patient_id, result, accuracy, ai_model
medical_records	Dossiers et documents médicaux du patient.	patient_id, category, title, file_url
medical_reports	Rapports médicaux générés à partir d'une prédiction.	patient_id, prediction_id, report_text

Tableau 10 Patients et module de diagnostic par IA

Ce pôle constitue le cœur intelligent de l'application. La table **predictions** mémorise pour chaque patient la probabilité de risque (champ **probability**), le niveau de risque associé (**risk\_level** : faible, élevé ou critique) et le modèle utilisé (**model\_used**). Ces informations sont ensuite restituées dans l'interface sous forme de pourcentage et de recommandations.

**c) Rendez-vous, messagerie et notifications**

Table	Description	Champs clés
appointments	Rendez-vous entre un patient et un médecin.	patient_id, doctor_id, date, time, status
conversation	Conversations privées entre deux utilisateurs.	user1_id, user2_id, last_message_at
messages	Messages échangés dans une conversation.	conversation_id, sender_id, content, is_read
ai_chat_sessions	Sessions de discussion avec l'assistant IA.	user_id, questions_log, recommended_doctors
notifications	Notifications envoyées aux utilisateurs.	user_id, title, type, is_read
support_tickets	Tickets de support / assistance.	user_id, subject, status

Grâce aux fonctionnalités temps réel de Supabase, les messages et les notifications sont diffusés instantanément aux utilisateurs concernés, sans rechargement de la page.

### 3.3 Relations entre les tables

Les principales relations qui structurent la base de données sont les suivantes :

- Un **utilisateur** (**auth.users**) possède exactement un **profil** (**profiles**) — relation 1..1.
- Un **profil** peut être un **médecin** (**doctors**) ou être associé à un **patient** (**patients\_ai**).
- Un **médecin** gère plusieurs **patients** et assure plusieurs **rendez-vous** — relation 1..n.
- Un **patient** génère plusieurs **prédictions**, **signes vitaux**, **analyses ECG** et **dossiers médicaux**.
- Une **conversation** contient plusieurs **messages** — relation 1..n.

## 4 . Présentation des interfaces de l'application

Cette section présente les principales interfaces de **MED AI** à travers des captures d'écran commentées. Pour des raisons de clarté, la présentation suit le parcours naturel de l'utilisateur, depuis l'authentification jusqu'à la consultation des résultats du diagnostic.

### 4.1 Authentification et création de compte

L'accès à la plateforme commence par une page d'authentification proposant deux actions : la connexion à un compte existant et la création d'un nouveau compte.

La **page de connexion** (Figure 4.2) permet à l'utilisateur de saisir son adresse e-mail et son mot de passe. Elle propose également une option « Se souvenir de moi » et une connexion rapide via **Google**.

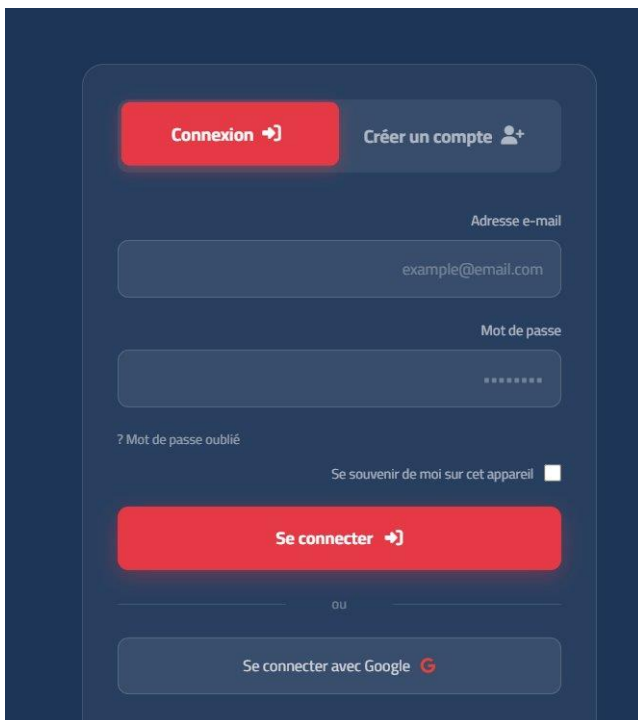
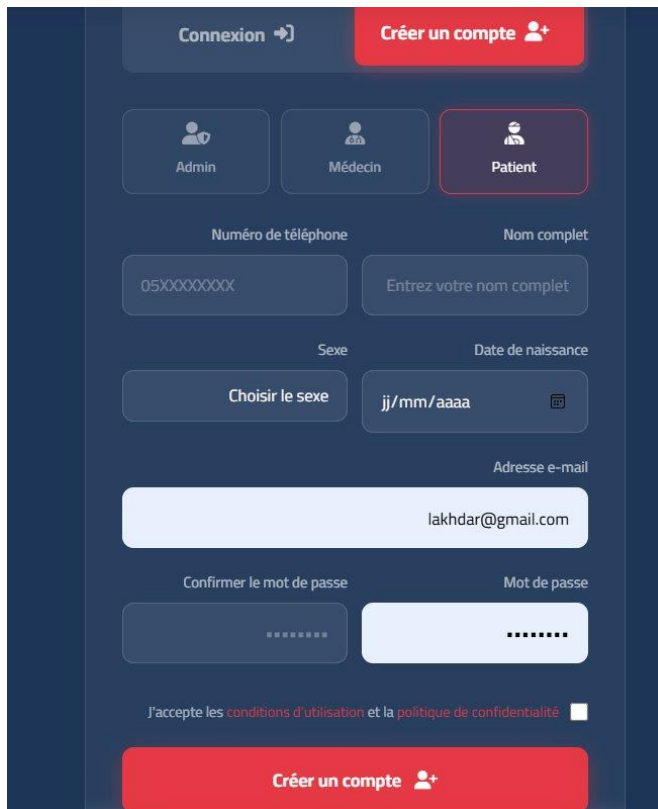


Figure 5 — Interface de connexion

La **page de création de compte** (Figure 4.3) invite l'utilisateur à choisir son rôle (Admin, Médecin ou Patient), puis à renseigner ses informations personnelles : nom complet, numéro de téléphone, sexe, date de naissance, adresse e-mail et mot de passe. L'acceptation des conditions d'utilisation est obligatoire avant la validation.



The image shows a registration form on a dark blue background. At the top, there are two buttons: 'Connexion →' and 'Créer un compte' with a person icon. Below these are three role selection buttons: 'Admin', 'Médecin', and 'Patient'. The 'Patient' button is highlighted with a red border. The form contains several input fields: 'Numéro de téléphone' (with a placeholder '05XXXXXXXX'), 'Nom complet' (with a placeholder 'Entrez votre nom complet'), 'Sexe' (with a button 'Choisir le sexe'), 'Date de naissance' (with a placeholder 'jj/mm/aaaa' and a calendar icon), 'Adresse e-mail' (with a placeholder 'lakhdar@gmail.com'), 'Confirmer le mot de passe' (with a placeholder of seven asterisks), and 'Mot de passe' (with a placeholder of seven asterisks). At the bottom, there is a checkbox for 'J'accepte les conditions d'utilisation et la politique de confidentialité' and a large red 'Créer un compte' button with a person icon.

Figure 6 — Formulaire d'inscription (profil Patient sélectionné)

## 4.2 Espace Patient — Assistant IA Cardiaque

Le patient accède à un **assistant conversationnel intelligent** qui réalise une évaluation du risque cardiovasculaire à travers une série de **21 questions** simples (Figure 4.4). Les réponses se font par de simples boutons — « Oui », « Non » ou « Je ne sais pas » — ce qui rend l'outil accessible à tous. La barre de progression indique l'avancement du questionnaire, et le moteur d'IA (XGBoost + LightGBM + RandomForest) est signalé comme actif dans l'en-tête.

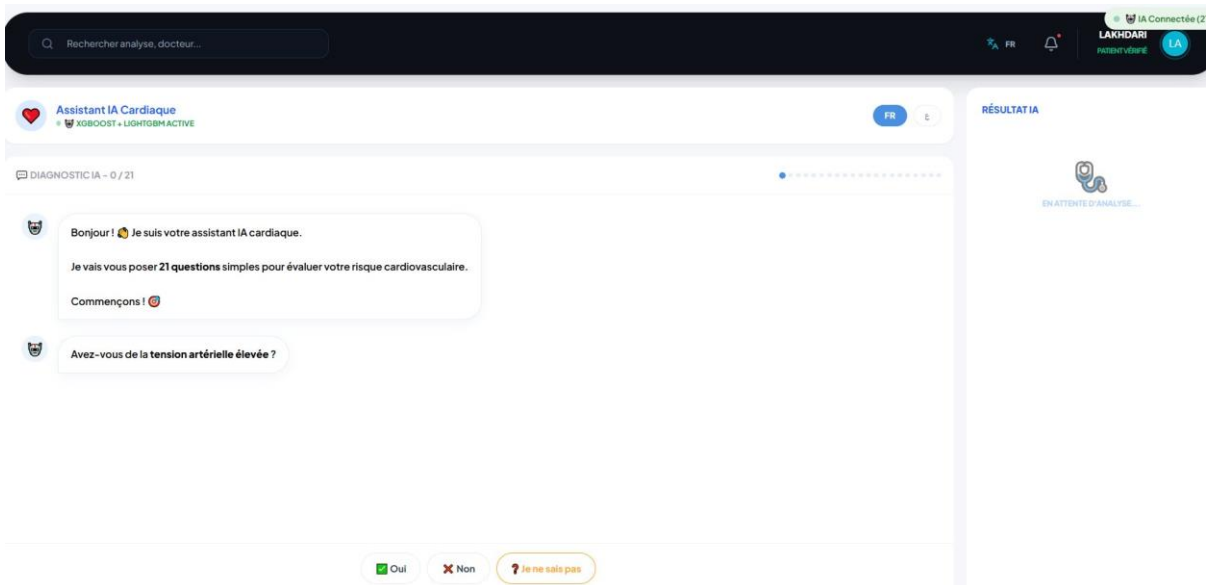


Figure 7 — Assistant IA Cardiaque : questionnaire de diagnostic (interface patient)

## CHAPITRE IV

### Implémentation de la plateforme MED AI

À l'issue du questionnaire, l'application affiche le **résultat de l'analyse IA** (Figure 4.5) : un pourcentage de risque, un niveau qualitatif (faible, élevé ou critique) ainsi qu'une liste de **recommandations personnalisées**. Dans l'exemple présenté, le risque est estimé à 58,16 % (« Risque ÉLEVÉ »), accompagné de conseils tels que la surveillance de la tension, du cholestérol et l'arrêt du tabac. Un bouton « Trouver un Médecin » permet d'enchaîner directement avec une prise de rendez-vous.

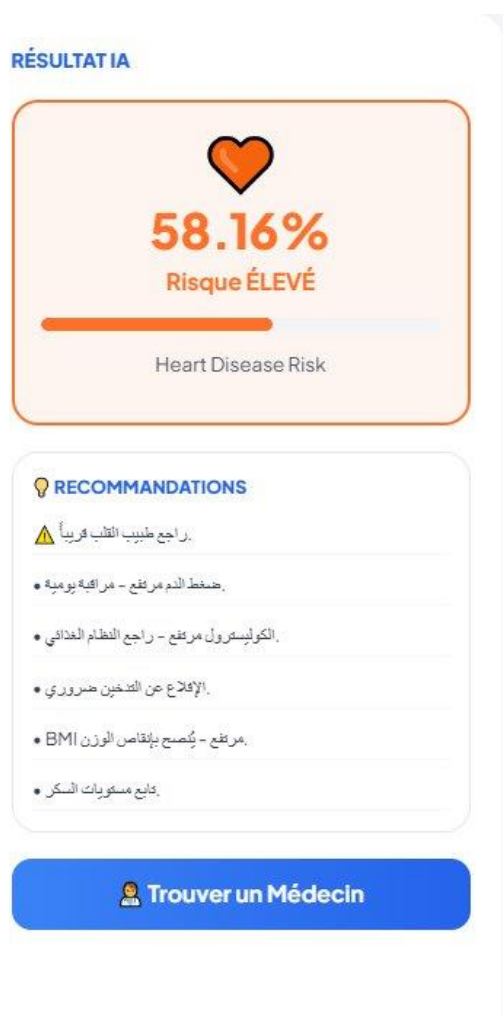


Figure 8 — Résultat du diagnostic IA et recommandations (interface patient)

Dans le cas d'un risque très élevé, l'écran de **diagnostic et pronostic** (Figure 4.6) met clairement en évidence l'urgence : un risque critique de 98 %, la liste des facteurs détectés (HTA, hypercholestérolémie, diabète, tabagisme, antécédent d'AVC, anomalie ECG) et un avertissement bilingue rappelant que l'outil reste une aide et ne remplace pas une consultation médicale approfondie. Le diagnostic est ensuite enregistré dans la base.

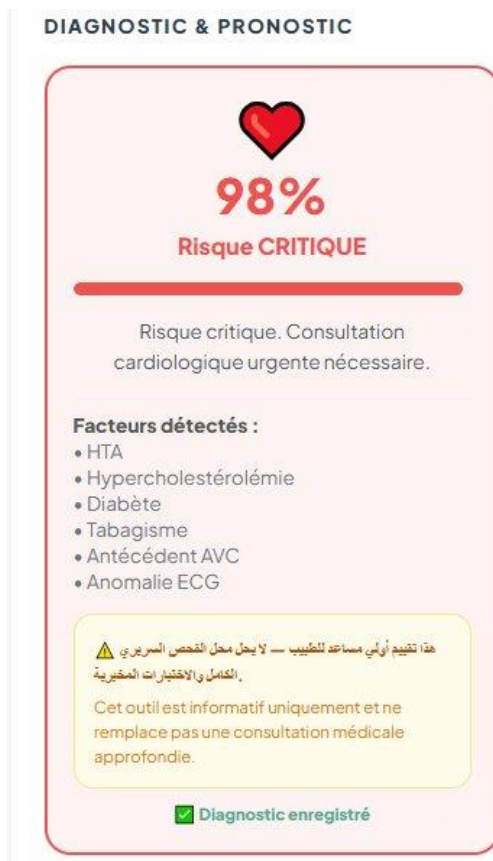


Figure 9 — Écran de diagnostic et pronostic (cas de risque critique)

### 4.3 Recherche de médecin et prise de rendez-vous

Le patient peut consulter la liste des **médecins disponibles** (Figure 4.7) à travers une fenêtre dédiée. Chaque médecin est présenté avec sa spécialité, sa note d'évaluation et son nombre d'années d'expérience. Deux actions sont proposées : « **Prendre RDV** » ou « **Discuter** ».

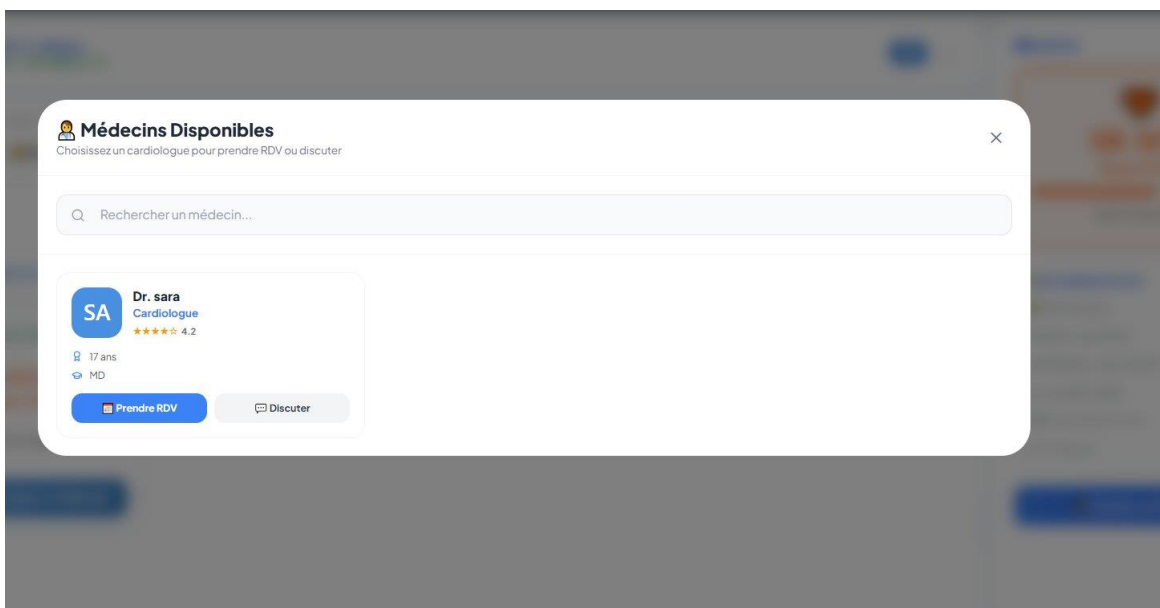


Figure 10 — Liste des médecins disponibles

L'interface de **prise de rendez-vous** (Figure 4.8) propose un calendrier de sélection de la date et des créneaux horaires. Un panneau « Récapitulatif » affiche en temps réel le médecin, la date et l'heure choisis avant la confirmation finale.

SA **Dr. sara**  
CARDIOLOGUE

Prendre RDV Discuter

**Sélectionnez une date**

Lun <b>25</b> Mai	Mar <b>26</b> Mai	Mer <b>27</b> Mai	Jeu <b>28</b> Mai	Ven <b>29</b> Mai	Sam <b>30</b> Mai	Dim <b>31</b> Mai	Lun <b>1</b> Juin
-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------

**Sélectionnez une heure**

09:00	09:30	10:00	10:30
11:00	11:30	12:00	12:30
13:00	<b>13:30</b>	14:00	14:30
15:00	15:30	16:00	16:30
17:00			

**Motif de consultation**

**Récapitulatif**

Médecin	Dr. sara
Date	Lun 25 Mai
Heure	13:30

**CONFIRMER LE RDV**

Figure 11 — Sélection de la date et de l'horaire du rendez-vous

Une fois la réservation validée, une fenêtre de **confirmation** (Figure 4.9) récapitule les détails du rendez-vous et affiche un identifiant unique. Ces données sont enregistrées dans la table **appointments** de la base.

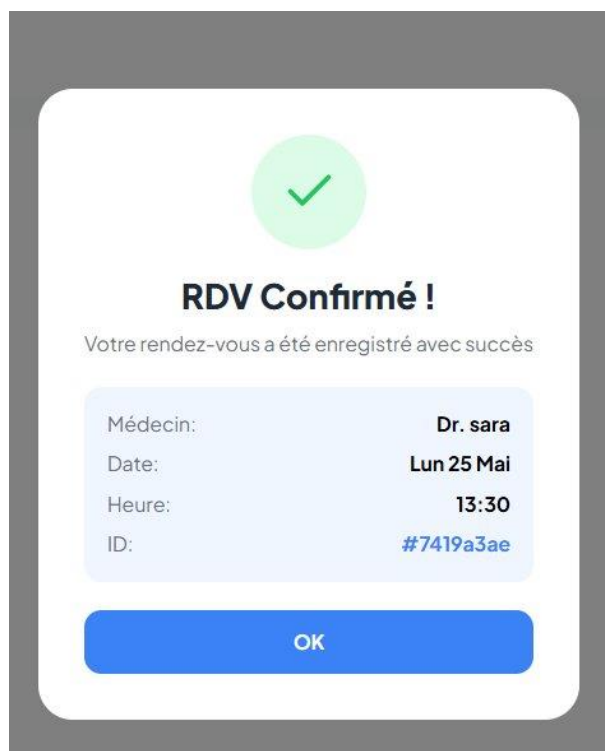


Figure 12 — Confirmation du rendez-vous

## 4.4 Espace Médecin — Assistant au diagnostic

Le médecin dispose d'un **assistant clinique** (Figure 4.10) intitulé « Assistant Médecin — MED AI », fonctionnant en mode diagnostic. Depuis cette interface, le praticien peut lancer une évaluation clinique complète, ajouter un patient ou consulter la liste de ses patients.

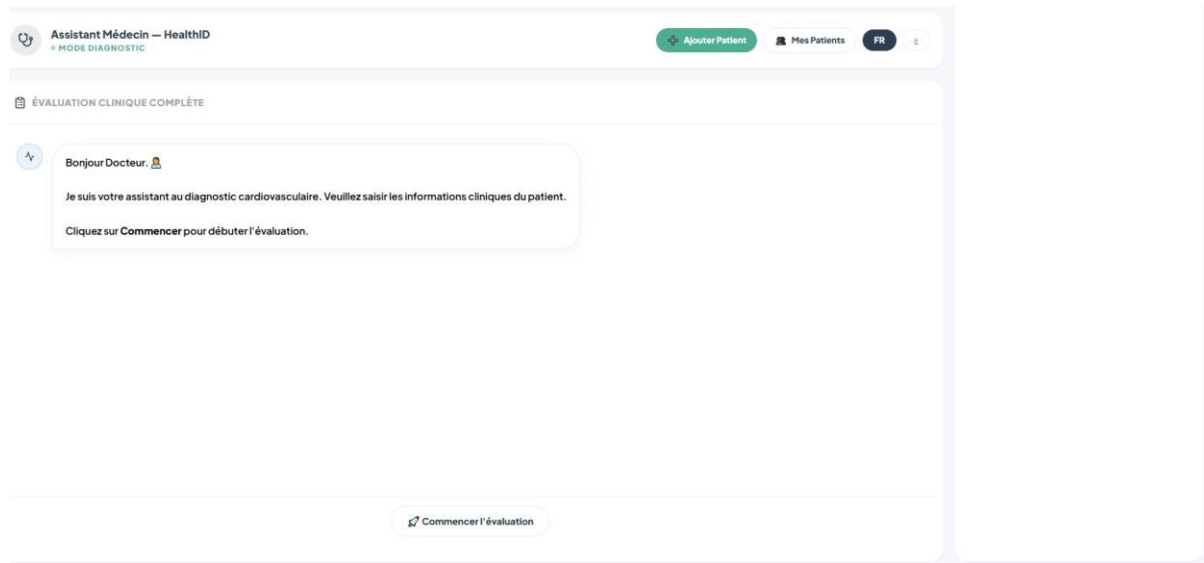


Figure 13 — Tableau de bord de l'assistant médecin

Le formulaire « **Ajouter un patient** » (Figure 4.11) permet de saisir l'ensemble des informations cliniques nécessaires au diagnostic : informations personnelles, antécédents médicaux (HTA, hypercholestérolémie, diabète, AVC/AIT, tabagisme, alcool), mesures cliniques (IMC, pression artérielle systolique) et mode de vie. Ces données alimentent les tables **patients\_manual** et **predictions**.

**Ajouter un patient**

**INFORMATIONS PERSONNELLES**

NOM COMPLET \*  
Ahmed Benali

ÂGE \* 58      SEXE Féminin

**ANTÉCÉDENTS MÉDICAUX**

HYPERTENSION (HTA) Non      HYPERCHOLESTÉROLÉMIE Non

DIABÈTE Non      AVC / AIT Non

TABAGISME Non-fumeur      ALCOOL EXCESSIF Non

**MESURES CLINIQUES**

IMC (KG/M<sup>2</sup>) 28.4      PA SYSTOLIQUE (MMHG) 135

**MODE DE VIE**

ACTIVITÉ PHYSIQUE ≥150 MIN/SEM Oui      DIFFICULTÉ À MARCHER Non

Figure 14 — Formulaire d'ajout d'un patient (espace médecin)

La fenêtre « **Mes Patients** » (Figure 4.12) offre une vue synthétique des patients du médecin, avec des compteurs par niveau de risque (Total, Critique, Élevé, Faible) et une barre de recherche. Lorsqu'aucun patient n'est encore enregistré, un message d'état approprié est affiché.

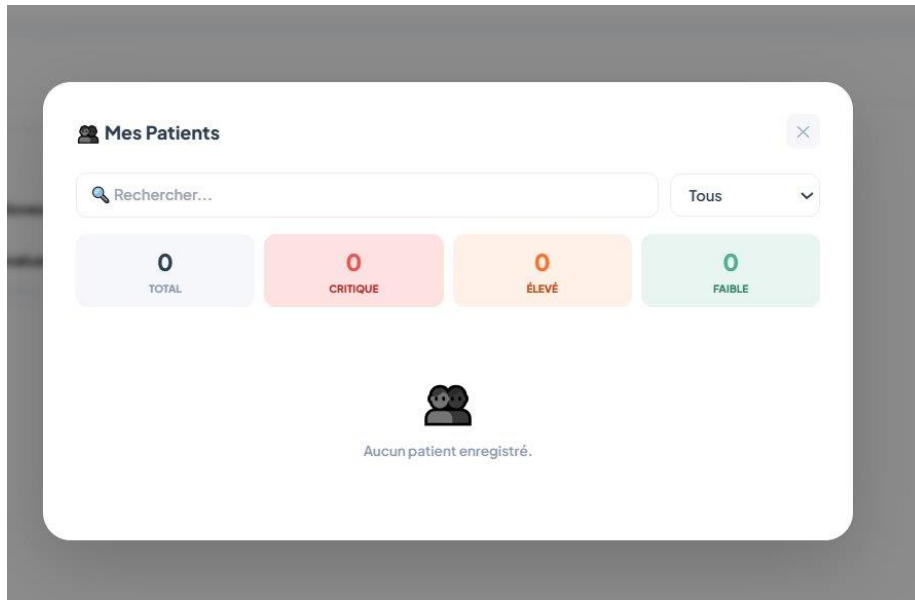


Figure 15 — Vue de gestion « Mes Patients »

Enfin, l'écran « **Patients Management** » (Figure 4.13) présente les dossiers cliniques sous forme de tableau : nom du patient, âge, groupe sanguin, date du dernier examen et actions disponibles (consultation et export du dossier). Cet écran exploite les tables **profiles**, **patients\_manual** et **medical\_records**.

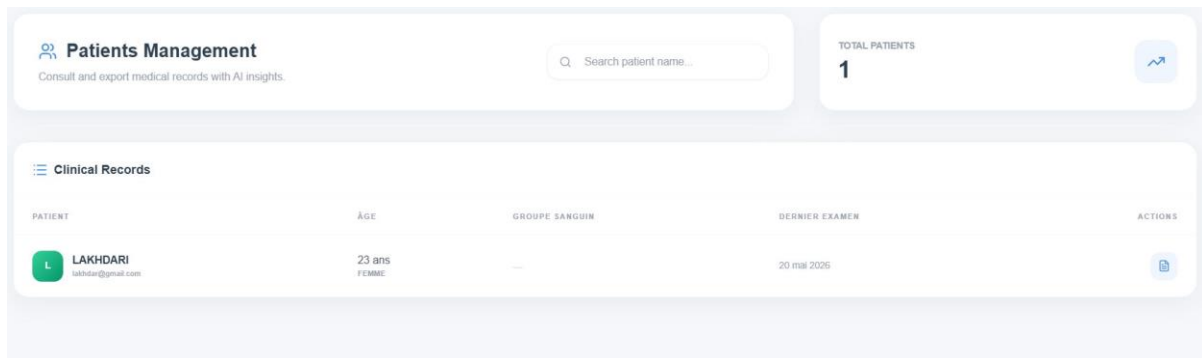


Figure 16 — Tableau de gestion des dossiers cliniques (espace médecin)

## 5 . Sécurité et confidentialité des données

S'agissant d'une application manipulant des données de santé, la sécurité a fait l'objet d'une attention particulière. Plusieurs mécanismes offerts par **Supabase** ont été mobilisés :

- **Authentification sécurisée** : gestion des sessions via Supabase Auth, avec connexion par e-mail/mot de passe ou via Google.
- **Gestion des rôles** : le champ « role » distingue administrateur, médecin et patient et restreint l'accès aux fonctionnalités correspondantes.
- **Row Level Security (RLS)** : des politiques d'accès au niveau des lignes garantissent qu'un utilisateur ne consulte que les données qui le concernent.
- **Avertissement clinique** : chaque diagnostic est accompagné d'un message rappelant que l'outil est une aide informative et ne se substitue pas à un avis médical.

## 6 . Conclusion du chapitre

Dans ce chapitre, nous avons présenté la phase de réalisation de l'application **MED AI** . Nous avons d'abord décrit l'environnement de développement et les outils retenus, puis détaillé la structure de la base de données hébergée sur **Supabase**, accompagnée de son schéma relationnel et de la description de ses principales tables. Nous avons ensuite illustré, à l'aide de captures d'écran commentées, les différentes interfaces destinées aux trois profils d'utilisateurs.

L'application répond ainsi aux objectifs fixés : offrir aux patients un outil accessible d'évaluation du risque cardiovasculaire, fournir aux médecins un assistant de diagnostic fiable, et garantir la sécurité des données médicales. Le chapitre suivant sera consacré à **l'évaluation des modèles d'intelligence artificielle** et à la discussion des résultats obtenus.



## Conclusion générale :

Au terme de ce travail, nous avons mené à bien la conception et la réalisation d'une plateforme intelligente, MED AI, dédiée à la détection précoce des maladies cardiovasculaires — depuis les fondements théoriques jusqu'à l'application déployable.

Le parcours suivi a d'abord consisté à comprendre le terrain. L'étude des paradigmes de l'intelligence artificielle, puis l'état de l'art du diagnostic cardiovasculaire assisté par ordinateur, nous ont permis d'identifier les briques techniques les plus pertinentes pour notre objectif. Cet examen a confirmé un constat désormais bien établi : sur des données médicales tabulaires, les modèles ensemblistes à base d'arbres demeurent les candidats les plus solides, devançant aussi bien les réseaux de neurones profonds que les grands modèles de langage.

La phase expérimentale a constitué le cœur de notre contribution. En évaluant sept modèles couvrant trois paradigmes — XGBoost, LightGBM, un ensemble par vote pondéré, un réseau convolutif, ainsi que les modèles de langage Bio\_ClinicalBERT et Qwen 2.5 — sur plus de 253 000 échantillons, nous avons pu comparer objectivement leurs forces et leurs limites. Surtout, nous avons mis en évidence et quantifié un écueil méthodologique majeur, trop souvent passé sous silence : la fuite de données provoquée par une mauvaise application de SMOTE. Nous avons montré qu'un séquençage incorrect du pipeline pouvait faire bondir le F1-Macro de 0,61 à 0,94 — un mirage qui n'a aucune valeur en conditions réelles. Ce résultat constitue, à nos yeux, l'un des apports les plus importants de ce mémoire : rappeler qu'une métrique impressionnante ne vaut rien sans une méthodologie irréprochable.

C'est sur ces bases assainies que nous avons retenu le modèle **Ensemble Learning** (XGBoost + LightGBM + Random Forest), qui atteint une exactitude de 90,61 % et une AUC-ROC de 0,8500 sur un ensemble de test fidèle à la distribution réelle de la maladie. Ce choix ne repose pas seulement sur les chiffres : il tient compte de la fiabilité clinique, de la capacité à ajuster le seuil de décision selon le contexte, de l'interprétabilité via SHAP, et d'une inférence quasi instantanée compatible avec un usage en temps réel.

Ce modèle a ensuite été intégré au sein d'une plateforme complète et fonctionnelle, bâtie sur une architecture full-stack moderne adossée à Supabase. MED AI propose aujourd'hui un assistant de diagnostic pour le patient comme pour le médecin, un système de recommandation et de prise de rendez-vous, une messagerie et une gestion sécurisée des dossiers, le tout protégé par une authentification robuste et des politiques d'accès au niveau des données.

Ce travail n'est évidemment pas exempt de limites, qui dessinent autant de pistes pour l'avenir. Le recall sur la classe positive reste perfectible, ce qui invite à explorer des stratégies plus fines de gestion du déséquilibre et d'ajustement du seuil. L'extension multimodale, envisagée dès l'état de l'art, ouvrirait la voie à l'intégration directe du signal ECG aux côtés des données structurées. Enfin, une validation clinique en conditions réelles, menée en collaboration avec des praticiens, constituerait l'étape décisive vers un déploiement à grande échelle.

En définitive, ce mémoire illustre qu'une démarche alliant rigueur méthodologique et souci d'utilité concrète peut donner naissance à un outil à la fois crédible scientifiquement et porteur de sens sur le plan humain. Face à une maladie qui demeure la première cause de mortalité dans le monde, contribuer, même modestement, à un diagnostic plus précoce et plus accessible nous semble être un objectif qui en valait pleinement la peine.

## Références

- [1] World Health Organization. (2025). Cardiovascular diseases (CVDs) fact sheet. WHO; 2024. URL [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)).
- [2] Russell, S., & Norvig, P. (2010). Artificial intelligence: a modern approach. 3rd. Upper Saddle River, EUA: Prentice-Hall.
- [3] McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine*, 27(4), 12-12.
- [4] Turing, A. M. (2007). Computing machinery and intelligence. In *Parsing the Turing test: Philosophical and methodological issues in the quest for the thinking computer* (pp. 23-65). Dordrecht: Springer Netherlands.
- [5] Manning, C. D. (2015). Computational linguistics and deep learning. *Computational Linguistics*, 41(4), 701-707.
- [6] Turing, A. M. (2007). Computing machinery and intelligence. In *Parsing the Turing test: Philosophical and methodological issues in the quest for the thinking computer* (pp. 23-65). Dordrecht: Springer Netherlands.
- [7] Bostrom, N. S. (2014). Paths, dangers, strategies. *Strategies*.
- [8] Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3), 210-229.
- [9] Cooper, G. F., Aliferis, C. F., Ambrosino, R., Aronis, J., Buchanan, B. G., Caruana, R., ... & Spirtes, P. (1997). An evaluation of machine-learning methods for predicting pneumonia mortality. *Artificial intelligence in medicine*, 9(2), 107-138.
- [10] Rosa, G. J. (2010). The elements of statistical learning: Data mining, inference, and prediction by Hastie, T., Tibshirani, R., and Friedman, J.
- [11] Bishop, C. M. (2006). *Pattern recognition and machine learning by Christopher M. Bishop* (Vol. 350). Berlin, Germany:: Springer Science+ Business Media, LLC.
- [12] Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc."
- [13] Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc."
- [14] Manning, C. D. (2015). Computational linguistics and deep learning. *Computational Linguistics*, 41(4), 701-707.
- [15] Angermueller, C., Pärnamaa, T., Parts, L., & Stegle, O. (2016). Deep learning for computational biology. *Molecular systems biology*, 12(7), MSB156651.

- [16] Swapna, M., Sharma, Y. K., & Prasad, B. M. G. (2020). Cnn architectures: Alex net, le net, vgg, google net, res net. *Int. J. Recent Technol. Eng*, 8(6), 953-960.
- [17] McNeely-White, D., Beveridge, J. R., & Draper, B. A. (2020). Inception and ResNet features are (almost) equivalent. *Cognitive Systems Research*, 59, 312-318.
- [18] Angelov, P., & Sperduti, A. (2016, April). Challenges in Deep Learning. In *ESANN*.
- [19] Fujisawa, H. (2008). Forty years of research in character and document recognition—an industrial perspective. *Pattern Recognition*, 41(8), 2435-2446.
- [20] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- [21] Swapna, M., Sharma, Y. K., & Prasad, B. M. G. (2020). Cnn architectures: Alex net, le net, vgg, google net, res net. *Int. J. Recent Technol. Eng*, 8(6), 953-960.
- [22] Zhang, X., Chen, F., & Huang, R. (2018). A combination of RNN and CNN for attention-based relation classification. *Procedia computer science*, 131, 911-917.
- [23] Zhao, B., Lu, H., Chen, S., Liu, J., & Wu, D. (2017). Convolutional neural networks for time series classification. *Journal of systems engineering and electronics*, 28(1), 162-169.
- [24] Ansari, Y., Mourad, O., Qaraqe, K., & Serpedin, E. (2023). Deep learning for ECG Arrhythmia detection and classification: an overview of progress for period 2017–2023. *Frontiers in Physiology*, 14, 1246746.
- [25] Yin, W., Kann, K., Yu, M., & Schütze, H. (2017). Comparative study of CNN and RNN for natural language processing. *arXiv preprint arXiv:1702.01923*.
- [26] Kim, Y. (2014, October). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1746-1751).
- [27] Ansari, Y., Mourad, O., Qaraqe, K., & Serpedin, E. (2023). Deep learning for ECG Arrhythmia detection and classification: an overview of progress for period 2017–2023. *Frontiers in Physiology*, 14, 1246746.
- [28] Kim, Y. (2014, October). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1746-1751).
- [29] Zhao, B., Lu, H., Chen, S., Liu, J., & Wu, D. (2017). Convolutional neural networks for time series classification. *Journal of systems engineering and electronics*, 28(1), 162-169.
- [30] Deal, J. A., Sharrett, A. R., Rawlings, A. M., Gottesman, R. F., Bandeen-Roche, K., Albert, M., ... & Klein, R. (2018). Retinal signs and 20-year cognitive decline in the Atherosclerosis Risk in Communities Study. *Neurology*, 90(13), e1158-e1166.
- [31] Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and machines*, 30(4), 681-694.

- [32] Grover, K., Kaur, K., Tiwari, K., Rupali, & Kumar, P. (2020, December). Deep learning based question generation using t5 transformer. In *International Advanced Computing Conference* (pp. 243-255). Singapore: Springer Singapore.
- [33] Zhang, P., Dai, X., Yang, J., Xiao, B., Yuan, L., Zhang, L., & Gao, J. (2021). Multi-scale vision longformer: A new vision transformer for high-resolution image encoding. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 2998-3008).
- [34] Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and machines*, 30(4), 681-694.
- [35] Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35, 27730-27744.
- [36] Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., ... & Kaplan, J. (2022). Constitutional ai: Harmlessness from ai feedback. *arXiv preprint arXiv:2212.08073*.
- [37] Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., ... & Fedus, W. (2022). Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*.
- [38] Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.
- [39] Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., ... & Scialom, T. (2023). Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.
- [40] Costa, M. R., & Oropeza, G. S. (2021). Análisis de la inclusión de contenidos TIC contextualizados en Escuelas Taller dirigidas a la formación laboral de adultos. *EduTec, Revista Electrónica de Tecnología Educativa*, (78), 246-263.
- [41] Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., ... & Fedus, W. (2022). Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*.
- [42] Alayrac, J. B., Donahue, J., Luc, P., Miech, A., Barr, I., Hasson, Y., ... & Simonyan, K. (2022). Flamingo: a visual language model for few-shot learning. *Advances in neural information processing systems*, 35, 23716-23736.
- [43] Team, G., Anil, R., Borgeaud, S., Alayrac, J. B., Yu, J., Soricut, R., ... & Blanco, L. (2023). Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.
- [44] Lin, B., Ye, Y., Zhu, B., Cui, J., Ning, M., Jin, P., & Yuan, L. (2024, November). Video-llava: Learning united visual representation by alignment before projection. In *Proceedings of the 2024 conference on empirical methods in natural language processing* (pp. 5971-5984).
- [45] Reddy, C. K., Dubey, H., Gopal, V., Cutler, R., Braun, S., Gamper, H., ... & Srinivasan, S. (2021, June). ICASSP 2021 deep noise suppression challenge. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6623-6627). IEEE.

- [46] Alayrac, J. B., Donahue, J., Luc, P., Miech, A., Barr, I., Hasson, Y., ... & Simonyan, K. (2022). Flamingo: a visual language model for few-shot learning. *Advances in neural information processing systems*, 35, 23716-23736.
- [47] Vivone, G., Garzelli, A., Xu, Y., Liao, W., & Chanussot, J. (2022). Panchromatic and hyperspectral image fusion: Outcome of the 2022 WHISPERS hyperspectral pansharpening challenge. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16, 166-179.
- [48] Team, G., Anil, R., Borgeaud, S., Alayrac, J. B., Yu, J., Soricut, R., ... & Blanco, L. (2023). Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.
- [49] Hurst, A., Lerer, A., Goucher, A. P., Perelman, A., Ramesh, A., Clark, A., ... & Kivlichan, I. (2024). Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.
- [50] Team, G., Anil, R., Borgeaud, S., Alayrac, J. B., Yu, J., Soricut, R., ... & Blanco, L. (2023). Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.
- [51] Zhou, P., Cao, M., Huang, Y. L., Ye, Q., Zhang, P., Liu, J., ... & Kim, J. (2023). Exploring recommendation capabilities of gpt-4v (ision): A preliminary case study. *arXiv preprint arXiv:2311.04199*.
- [52] Vô, C., Hubrechts, J., Van Huffel, J., Groignet, S., André, C. O., Rixhon, H., ... & Moniotte, S. (2025). Innovations 2024 en Cardiologie pédiatrique. *Louvain médical*, 144(2).
- [53] Vô, C., Hubrechts, J., Van Huffel, J., Groignet, S., André, C. O., Rixhon, H., ... & Moniotte, S. (2025). Innovations 2024 en Cardiologie pédiatrique. *Louvain médical*, 144(2).
- [54] Ploin, D., Alexandre, M., Ventelou, B., Che, D., Coignard, B., Boulanger, N., ... & GILBERG, S. (2024). Prioritisation of infectious diseases from a public health perspective: a multi-criteria decision analysis study, France, 2024. *Eurosurveillance*, 29(50), 2400074.
- [55] Gal, J., Benidir, S., Gilet, C., Chateau, Y., Schiappa, R., Chamorey, E., ... & Fillatre, L. (2022). Comparaison de différentes méthodes de machine learning supervisé pour l'aide au diagnostic médical des nodules thyroïdiens. *Revue d'Épidémiologie et de Santé Publique*, 70, S92.
- [56] MERCIER, G., GUILLERMOU, H., BONNEFOY-CUDRAZ, E., DELMAS, C., & ROUBILLE, F. (2026). Évolution des passages en unités de soins intensifs de cardiologie au cours de la dernière décennie: étude observationnelle basée sur la base PMSI nationale pour la période 2014-2023. *Journal of Epidemiology and Population Health*, 74, 203198.
- [57] Machado, L., Vilgrain, V., Aubé, C., & Grégory, J. (2025). La radiologie entre deux vagues: Adoption de l'intelligence artificielle dans les hôpitaux universitaires français en 2025. *Journal d'imagerie diagnostique et interventionnelle*.
- [58] Bhattaru, A., Yanamala, N., & Sengupta, P. P. (2024). Revolutionizing cardiology with words: unveiling the impact of large language models in medical science writing. *Canadian Journal of Cardiology*, 40(10), 1950-1958.

- [59] Ansari, M. Y., Yaqoob, M., Ishaq, M., Flushing, E. F., Mangalote, I. A. C., Dakua, S. P., ... & Qaraqe, M. (2025). A survey of transformers and large language models for ECG diagnosis: advances, challenges, and future directions. *Artificial Intelligence Review*, 58(9), 261.
- [60] Hounkponou, M., Aikpon, G., Soummonni, F., Ngoume, J. J., Sampil, A. Y., Safari, A., ... & de Cotonou, U. H. K. M. Evaluation du contrôle des principaux facteurs de risque cardiovasculaire modifiables après un syndrome coronarien aigu à la Clinique Universitaire de Cardiologie (CUC) du CNHU-HKM de Novembre 2021 à Août 2023.
- [61] Baudry, G. (2024). *Congestion et interactions cardio-rénales en insuffisance cardiaque avancée* (Doctoral dissertation, Université de Lorraine).
- [62] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE access*, 7, 81542-81554.
- [63] O. M. d. I. S. (OMS)., «Cardiovascular Diseases Fact Sheet.» 2023. [En ligne]. Available: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))..
- [64] Machado, L., Vilgrain, V., Aubé, C., & Grégory, J. (2025). La radiologie entre deux vagues: Adoption de l'intelligence artificielle dans les hôpitaux universitaires français en 2025. *Journal d'imagerie diagnostique et interventionnelle*.
- [65] Choudhury, A., Mondal, A., & Sarkar, S. (2024). Searches for the BSM scenarios at the LHC using decision tree-based machine learning algorithms: a comparative study and review of random forest, AdaBoost, XGBoost and LightGBM frameworks. *The European Physical Journal Special Topics*, 233(15), 2425-2463..
- [66] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.
- [67] Haouam, K. D. (2025). Automated Diagnosis of Multiple Sclerosis Using Transfer Learning and LightGBM on FLAIR MRI Data. *Ingenierie des Systemes d'Information*, 30(4), 953.
- [68] King, S. J., Wangdak Yuthok, T. Y., Bacong, A. M., Khandelwal, A., Kazi, D. S., Mussolino, M. E., ... & Palaniappan, L. P. (2025). Heart disease mortality in the United States, 1970 to 2022. *Journal of the American Heart Association*, 14(13), e038644.
- [69] Gao, X. Y., Amin Ali, A., Shaban Hassan, H., & Anwar, E. M. (2021). Improving the accuracy for analyzing heart diseases prediction based on the ensemble method. *Complexity*, 2021(1), 6663455.
- [70] Pasupuleti, S. (2025). AI-Augmented Data Pipelines: Integrating Machine Learning for Intelligent Data Processing. *Journal of Computer Science and Technology Studies*, 7(11), 276-283.
- [71] Santos, J. A., Tekle, D., Rosewarne, E., Flexner, N., Cobb, L., Al-Jawaldeh, A., ... & Trieu, K. (2021). A systematic review of salt reduction initiatives around the world: a midterm evaluation of progress towards the 2025 global non-communicable diseases salt reduction target. *Advances in Nutrition*, 12(5), 1768-1780.

- [72] Huang, K., Altosaar, J., & Ranganath, R. (2019). Clinicalbert: Modeling clinical notes and predicting hospital readmission. *arXiv preprint arXiv:1904.05342*.
- [73] Huang, K., Altosaar, J., & Ranganath, R. (2019). Clinicalbert: Modeling clinical notes and predicting hospital readmission. *arXiv preprint arXiv:1904.05342*.
- [74] Eguia, H., Sánchez-Bocanegra, C. L., Vinciarelli, F., Alvarez-Lopez, F., & Saigí-Rubió, F. (2024). Clinical decision support and natural language processing in medicine: systematic literature review. *Journal of Medical Internet Research*, 26, e55315.
- [75] Takahashi, S., Sakaguchi, Y., Kouno, N., Takasawa, K., Ishizu, K., Akagi, Y., ... & Hamamoto, R. (2024). Comparison of vision transformers and convolutional neural networks in medical image analysis: a systematic review. *Journal of Medical Systems*, 48(1), 84.
- [76] Aden, I., Child, C. H., & Reyes-Aldasoro, C. C. (2024). International classification of diseases prediction from mimiic-iii clinical text using pre-trained clinicalbert and nlp deep learning models achieving state of the art. *Big Data and Cognitive Computing*, 8(5), 47.
- [77] Ong, J. C. L., Chang, S. Y. H., William, W., Butte, A. J., Shah, N. H., Chew, L. S. T., ... & Ting, D. S. W. (2024). Medical ethics of large language models in medicine. *NEJM AI*, 1(7), Alra2400038.
- [78] Ahmed, M., Lam, J., Chow, A., & Chow, C. M. (2025). A primer on Large Language Models (LLMs) and ChatGPT for cardiovascular healthcare professionals. *CJC open*, 7(5), 660-666.
- [79] Gendler, M., Nadkarni, G. N., Sudri, K., Cohen-Shelly, M., Glicksberg, B. S., Efron, O., ... & Klang, E. (2026). Large Language Models in Cardiology: Systematic Review. *JMIR cardio*, 10(1), e76734.
- [80] Teboul, A. (2021). Diabetes health indicators dataset. *Kaggle*, [Online]. Available: <https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset>.
- [81] Richard, J. B., Andler, R., Gautier, A., Guignard, R., Leon, C., & Beck, F. (2017). Effects of using an overlapping dual-frame design on estimates of health behaviors: a French general population telephone survey. *Journal of Survey Statistics and Methodology*, 5(2), 254-274.
- [82] Zhang, W., Cheng, W., Fujiwara, K., Evans, R., & Zhu, C. (2024). Predictive modeling for hospital readmissions for patients with heart disease: An updated review from 2012–2023. *IEEE Journal of Biomedical and Health Informatics*, 28(4), 2259-2269.
- [83] Batista, G. E., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD explorations newsletter*, 6(1), 20-29.
- [84] Ruan, Y., Lan, X., Ma, J., Dong, Y., He, K., & Feng, M. (2024). Language modeling on tabular data: A survey of foundations, techniques and evolution. *arXiv preprint arXiv:2408.10548*.
- [85] Kasartzian, D. I., & Tsiampalis, T. (2025). Transforming cardiovascular risk prediction: a review of machine learning and artificial intelligence innovations. *Life*, 15(1), 94.
- [86] Bounab, R., Guelib, B., & Zarour, K. (2024, April). A novel machine learning approach for handling imbalanced data: Leveraging SMOTE-ENN and XGBoost. In *2024 6th International Conference on Pattern Analysis and Intelligent Systems (PAIS)* (pp. 1-7). IEEE.

- [87] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- [88] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.
- [89] Ennab, M. M. H. (2025). *A hybrid convolutional-fuzzy model for interpretable AI in healthcare: improving transparency and accuracy in chronic disease management* (Doctoral dissertation, Université du Québec à Chicoutimi).
- [90] Manning, C. D. (2015). Computational linguistics and deep learning. *Computational Linguistics*, 41(4), 701-707.
- [91] Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information fusion*, 81, 84-90.
- [92] Roy, A., & Pan, S. (2021, November). Incorporating medical knowledge in BERT for clinical relation extraction. In *Proceedings of the 2021 conference on empirical methods in natural language processing* (pp. 5357-5366).
- [93] Hegselmann, S., Buendia, A., Lang, H., Agrawal, M., Jiang, X., & Sontag, D. (2023, April). Tabllm: Few-shot classification of tabular data with large language models. In *International conference on artificial intelligence and statistics* (pp. 5549-5581). PMLR.
- [94] Chu, Y., Xu, J., Yang, Q., Wei, H., Wei, X., Guo, Z., ... & Zhou, J. (2024). Qwen2-audio technical report. *arXiv preprint arXiv:2407.10759*.
- [95] Antiperovitch, P., Liu, I., Mokhtar, A. T., & Tang, A. (2025). Evaluating large language models in cardiovascular antithrombotic care: performance, accuracy, and implications for clinical practice. *Canadian Journal of Cardiology*.
- [96] Dietterich, T. G. (2000, June). Ensemble methods in machine learning. In *International workshop on multiple classifier systems* (pp. 1-15). Berlin, Heidelberg: Springer Berlin Heidelberg.
- [97] Deng, L., Lu, K., & Hu, H. (2025). An interpretable LightGBM model for predicting coronary heart disease: Enhancing clinical decision-making with machine learning. *Plos one*, 20(9), e0330377.(yang)
- [98] Hossain, S., Hasan, M. K., Faruk, M. O., Aktar, N., Hossain, R., & Hossain, K. (2024). Machine learning approach for predicting cardiovascular disease in Bangladesh: evidence from a cross-sectional study in 2023. *BMC cardiovascular disorders*, 24(1), 214.
- [99] Kumar, R., Garg, S., Kaur, R., Johar, M. G. M., Singh, S., Menon, S. V., ... & Lozanović, J. (2025). A comprehensive review of machine learning for heart disease prediction: challenges, trends, ethical considerations, and future directions. *Frontiers in artificial intelligence*, 8, 1583459. (tompra)

- [100] Garcon, J. (2023). Calibration et discrimination du score de Framinghamsimplifié pour la prédiction du risque de maladie coronarienne dans une cohorte canadienne. (wang)
- [101] Shah, P., Shukla, M., Dholakia, N. H., & Gupta, H. (2025). Predicting cardiovascular risk with hybrid ensemble learning and explainable ai: P. shah et al. *Scientific Reports*, 15(1), 17927.
- [102] Arif, M. S., Rehman, A. U., & Asif, D. (2024). Explainable machine learning model for chronic kidney disease prediction. *Algorithms*, 17(10), 443.
- [103] Ansari, Y., Mourad, O., Qaraqe, K., & Serpedin, E. (2023). Deep learning for ECG Arrhythmia detection and classification: an overview of progress for period 2017–2023. *Frontiers in Physiology*, 14, 1246746.
- [104] Maity, S., & Saikia, M. J. (2025). Large language models in healthcare and medical applications: a review. *Bioengineering*, 12(6), 631.
- [105] Kandhare, P., Kurlekar, M., Deshpande, T., & Pawar, A. (2025). A review on revolutionizing healthcare technologies with AI and ML applications in pharmaceutical sciences. *Drugs and Drug Candidates*, 4(1), 9.