

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA
RECHERCHE SCIENTIFIQUE



Université 20 Aout 1955 - Skikda

FACULTÉ DES SCIENCES

DÉPARTEMENT D'INFORMATIQUE

MÉMOIRE DE FIN D'ÉTUDE

POUR L'OBTENTION DU DIPLÔME DE MASTER

OPTION : RÉSEAUX ET SYSTÈMES DISTRIBUÉS

**Développement d'un système de détection des maladies cardiaques
utilisant des méthodes d'apprentissage automatique**

Présenté Par :
CHEBLI AMEL
NEGHOUCHE HADJER

Supervisé Par :
Dr. Adel Lahsasna

2023 - 2024

Remerciements

Nous remercions Dieu de nous avoir accordé la patience, la santé, le courage et la capacité de réaliser ce travail. Nous tenons à exprimer notre sincère gratitude à notre superviseur, le Dr Lahsasna Adel, pour ses conseils, ses encouragements, son soutien et son suivi constants qui nous ont permis de mener à bien ce travail dans les meilleures conditions. Nous remercions sincèrement les membres du jury d'avoir accepté d'examiner et d'évaluer notre travail. Nous exprimons également notre gratitude à tous les enseignants et enseignantes du département d'informatique, sans oublier bien sûr de remercier profondément tous ceux qui nous ont soutenus directement ou indirectement pour accomplir ce travail..

Dédicaces

Nous dédions cette recherche à nos chers parents, peu importe ce que nous faisons ou disons, nous ne pourrons jamais vous remercier assez comme vous le méritez. À nos familles pour leur amour et leur soutien pour nous. Et à mon fiancé Nasser pour ses efforts et sa motivation, et à tous les amis et à tous ceux qui ont cru en NOUS et nous ont toujours poussés à atteindre ce succès.

ملخص

عملت هذه الدراسة على تطوير نظام للكشف المبكر عن أمراض القلب باستخدام التقدمات في التعلم الآلي. كانت دافع هذه الدراسة متأصلة في أهمية الكشف المبكر عن أمراض القلب، التي تعد السبب الرئيسي للوفيات في جميع أنحاء العالم. تعتبر الطرق التشخيصية التقليدية، رغم فعاليتها، غالباً ما تكون مكلفة وتتطلب وقتاً طويلاً وتتطلب تدخل أخصائيين متخصصين بمهارات عالية. تقدم التعلم الآلي فرصة واعدة لثورة التشخيص من خلال تحسين الدقة وتقليل التكاليف وجعل الرعاية الصحية أكثر إمكانية.

شملت مراحل رئيسية من الدراسة مقارنة بين مختلف خوارزميات التعلم الآلي التي تم تطبيقها على مجموعات بيانات وتضمنت الخوارزميات التي تم اختبارها الشبكات العصبية الاصطناعية، وآلة الدعم بالفاصلة، والغابة العشوائية، وشجرة القرار، والانحدار اللوجستي، وأقرب الجيران الكثيف.

أظهرت النتائج أن خوارزمية الغابة العشوائية كانت الأكثر فعالية، وقدمت أعلى دقة عبر كلتا المجموعتين من البيانات. تباينت أداء الخوارزميات الأخرى، حيث أظهرت آلة الدعم بالفاصلة وأقرب الجيران الكثيف حساسية أعلى لخصائص مجموعات البيانات، بينما أظهرت الشبكات العصبية الاصطناعية والانحدار اللوجستي أداءً متسقاً ولكن بأداء قليلاً أقل مقارنة بالغابة العشوائية.

تتضمن هذه النتائج تداعيات عملية هامة. سيتم استخدام الغابة العشوائية، كونها الخوارزمية الأكثر فعالية، كأساس لتطوير نظام تشخيص أمراض القلب. تأكيدها وقدرتها على التعميم بشكل جيد على مجموعات بيانات مختلفة تضمن توقعات تشخيصية موثوقة، مما يساهم في اتخاذ قرارات سريرية أفضل وتحسين نتائج المرضى.

أخيراً، أظهر تحليل أهمية الميزات أن العوامل الأكثر تأثيراً كانت ضغط الدم (الانقباضي والانبساطي)، ومؤشر كتلة الجسم، ومستويات الجلوكوز، ومستويات الكوليسترول، والعمر.

Résumé

Cette étude visait à développer un système de détection précoce des maladies cardiaques en exploitant les avancées en apprentissage automatique. La motivation de cette étude était enracinée dans l'importance cruciale de la détection précoce des maladies cardiaques, la première cause de mortalité dans le monde. Les méthodes de diagnostic traditionnelles, bien qu'efficaces, sont souvent coûteuses, chronophages et nécessitent l'intervention de spécialistes hautement qualifiés. Les progrès de l'apprentissage automatique offrent une opportunité prometteuse de révolutionner le diagnostic en améliorant la précision, en réduisant les coûts et en rendant les soins de santé plus accessibles.

Les principales étapes de l'étude comprenaient la comparaison de divers algorithmes d'apprentissage automatique appliqués aux ensembles de données de Framingham et de Cleveland. Les algorithmes testés comprenaient les Réseaux de Neurones Artificiels (ANN), les Machines à Vecteurs de Support (SVM), la Forêt Aléatoire, l'Arbre de Décision, la Régression Logistique et les K-Plus Proches Voisins (KNN). Les résultats ont démontré que l'algorithme de Forêt Aléatoire était le plus efficace, offrant la plus haute précision sur les deux ensembles de données. La performance des autres algorithmes variait, avec les SVM et KNN montrant une plus grande sensibilité aux caractéristiques des ensembles de données, tandis que les ANN et la Régression Logistique ont présenté une performance constante mais légèrement inférieure à celle de la Forêt Aléatoire.

Ces résultats ont des implications pratiques significatives. La Forêt Aléatoire, étant l'algorithme le plus efficace, sera utilisée comme base pour le développement du système de diagnostic des maladies cardiaques. Sa robustesse et sa capacité à bien généraliser aux différents ensembles de données garantissent des prédictions diagnostiques fiables, contribuant ainsi à une meilleure prise de décision clinique et à de meilleurs résultats pour les patients. Enfin, l'analyse de l'importance des caractéristiques a révélé que les facteurs les plus influents étaient la pression artérielle (systolique et diastolique), l'IMC, les niveaux de glucose, les niveaux de cholestérol et l'âge.

Abstract

This study aimed to develop an early detection system for heart disease by leveraging advancements in machine learning. The motivation for this study was rooted in the critical importance of early detection of heart disease, the leading cause of mortality worldwide. Traditional diagnostic methods, though effective, are often expensive, time-consuming, and require the intervention of highly skilled specialists. Advances in machine learning offer a promising opportunity to revolutionize diagnosis by improving accuracy, reducing costs, and making healthcare more accessible.

Key stages of the study involved comparing various machine learning algorithms applied to the Framingham and Cleveland datasets. The tested algorithms included Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forest, Decision Tree, Logistic Regression, and K-Nearest Neighbors (KNN). The results demonstrated that the Random Forest algorithm was the most effective, providing the highest accuracy across both datasets. The performance of other algorithms varied, with SVM and KNN showing higher sensitivity to dataset characteristics, while ANN and Logistic Regression exhibited consistent but slightly lower performance compared to Random Forest.

These findings have significant practical implications. Random Forest, being the most effective algorithm, will be used as the basis for developing the heart disease diagnostic system. Its robustness and ability to generalize well to different datasets ensure reliable diagnostic predictions, contributing to better clinical decision-making and improved patient outcomes. Finally, feature importance analysis revealed that the most influential factors were blood pressure (systolic and diastolic), BMI, glucose levels, cholesterol levels, and age.

Table des matières

I	Partie 1	13
1	Introduction	14
1.1	Contexte et Motivation	15
1.2	Problématique	16
1.3	Objectifs de l'étude	16
1.4	Questions de recherche	16
1.5	Portée et Limitations	17
1.5.1	Portée	17
1.5.2	Limitations	17
1.6	Structure du mémoire	17
1.6.1	Chapitre 2 : Revue de la Littérature	18
1.6.2	Chapitre 3 : Méthodologie	18
1.6.3	Chapitre 4 : Implémentation du Système	18
1.6.4	Chapitre 5 : Résultats et Discussion	18
1.6.5	Chapitre 6 : Conclusion et Travaux Futurs	18
2	Revue de la littérature	19
2.1	Aperçu des Maladies Cardiaques et leur Importance	20
2.1.1	Types de Maladies Cardiaques	20
2.1.2	Facteurs de Risque et Symptômes Associés aux Maladies Cardiaques	21
2.2	Techniques de Diagnostic Médical des Maladies Cardiaques	22
2.2.1	Électrocardiogramme (ECG)	22
2.2.2	Échocardiogramme	22
2.2.3	Tests de Résistance	23
2.2.4	Cathétérisme Cardiaque	23
2.2.5	Angiographie Coronarienne	23
2.2.6	Tomodensitométrie (TDM)	23
2.2.7	Imagerie par Résonance Magnétique (IRM)	23
2.2.8	Analyses Sanguines	24
2.2.9	Moniteur Holter	24
2.2.10	Moniteur d'Événements	24
2.2.11	Test de La Table Inclinable	24
2.2.12	Cardiologie Nucléaire	24
2.3	Apprentissage Automatique dans les Soins de Santé	24
2.3.1	Vue d'ensemble de l'Apprentissage Automatique	24

2.3.2	Applications de l'Apprentissage Automatique dans les Diagnostics Médicaux	25
2.4	Travaux Antérieurs sur la Détection des Maladies Cardiaques Utilisant l'Apprentissage Automatique	27
2.4.1	Réseau Neuronal Artificiel (ANN)	27
2.4.2	Travaux Antérieurs sur la Détection des Maladies Cardiaques Utilisant l'Apprentissage Automatique	28
2.4.3	Régression Linéaire	29
2.4.4	Apprentissage Profond (Deep Learning)	29
2.4.5	K-voisins les plus proches (K-NN)	30
2.5	Conclusion	30
3	LA METHODOLOGIE	32
3.1	Description des ensembles de données	33
3.1.1	Ensemble de données de Cleveland	33
3.1.2	Ensemble de données de Framingham	33
3.2	Les algorithmes d'apprentissage automatique	33
3.2.1	Les réseaux de neurones (Artificial Neural Networks)	33
3.2.2	k-plus proches voisins (k-NN)	33
3.2.3	Arbre de décision	34
3.2.4	Machine à vecteur de support (Support Vector Machines)	34
3.2.5	Régression logistique	34
3.2.6	Forêt aléatoire (Random Forest)	34
3.3	Entraînement et évaluation des modèles	35
3.3.1	Métrique d'évaluation	35
3.3.2	La méthode de validation croisée à 10 plis	35
3.4	Conclusion	36
4	Implémentation du système	37
4.1	Aperçu de l'architecture du système	38
4.1.1	Préparation des données	38
4.1.2	Sélection des Algorithmes	38
4.1.3	Déploiement et intégration	38
4.1.4	Analyse des caractéristiques	38
4.2	Avantages du système	39
4.3	Technologies et outils	40
4.3.1	Langage de programmation	40
4.3.2	Bibliothèques et frameworks	40
4.3.3	Environnement de développement	40
4.4	Interface utilisateur	41
4.4.1	Captures d'écran d'interface utilisateur	41
4.4.2	Présentation de l'interface graphique du système	41
4.5	Conclusion	44

5	Résultats et Discussion	46
5.1	Présentation et Discussion des Résultats	47
5.1.1	Réseau de Neurones Artificiels (ANN)	47
5.1.2	Machine à Vecteurs de Support (SVM)	47
5.1.3	Forêt Aléatoire	48
5.1.4	Arbre de Décision	48
5.1.5	Régression Logistique	48
5.1.6	K-Plus Proches Voisins (KNN)	48
5.2	Implications pour le Diagnostic des Maladies Cardiaques	48
5.2.1	Sélection de l'Algorithme	48
5.2.2	Robustesse du Modèle	49
5.2.3	Caractéristiques des Ensembles de Données	49
5.2.4	Développement Futur	49
5.3	Utilisation du système pour le diagnostic des maladies cardiaques	49
5.4	Conclusion	51
6	Conclusion	52
6.1	Résumé de l'étude	53
6.2	Principaux résultats	53
6.3	Implications pratiques	53
6.4	Limitations et perspectives futures	53
II	Partie 2	55
1	Présentation de projet	57
1.1	1-L'idée de Projet	58
2	Innovation	60
2.1	Nature des Innovations	61
2.2	Domaines d'Innovations	61
3	Analyse Stratégique du Marché	62
3.1	Vue du Segment de Marché	63
3.2	Mesurer l'Intensité de la Compétition	63
3.3	Stratégies de Marketing	63
4	Plan de Production et Organisation	64
4.1	Processus de Production	65
4.1.1	Plan de production et d'organisation	65
4.2	Restauration	66
4.3	La main qui travaille	66
4.4	Principales entreprises	66
5	Plan financier	67
5.1	Coûts et Charges	68

5.2	Coûts et responsabilités	69
5.2.1	Coûts et investissements nécessaires	69
5.3	Méthodes et sources de financement	70
5.4	Remboursement des retours	70
5.4.1	Publicités payantes :	70
5.4.2	Frais d'abonnement :	70
5.4.3	Commissions :	70
5.5	Tableau des comptes (résultats attendus)	72
5.6	Besoins en fonds de roulement	73
5.7	Tableau du Trésor	74
6	Prototype expérimental	77
6.1	Présentation de l'interface du système de Détection des Maladies Cardiaques - Vue d'Ensemble	78
6.2	Présentation de l'interface graphique du système	78
6.3	À propos	79
6.4	Le module de comparaison	79
6.5	Le module de déploiement (diagnostics)	80
7	Annexées	82

Table des figures

4.1	Aperçu de l'architecture du système	39
4.2	Interface d'accueil pour le système de diagnostic des maladies cardiaques	42
4.3	Interface d'accueil 2	42
4.4	Interface du module de comparaison	43
4.5	Interface du module de déploiement	44
5.1	Les investissements de démarrage	69
5.2	Financement des investissements	70
5.3	Bilan prévisionnel	71
5.4	tableau des comptes (résultats attendus)	72
5.5	Besoins en fonds de roulement moyen	73
5.6	seuil de rentabilité	73
5.7	compte de résultats prévisionnel sur 3 ans	74
5.8	plan de financement à 3 ans	75
5.9	Suite	76
6.1	Interface d'accueil pour le système de diagnostic des maladies cardiaques	78
6.2	Interface d'apropos du système de diagnostic des maladies cardiaques . .	79
6.3	Interface du module de comparaison	80
6.4	Interface du module de déploiement	81
7.1	Business Model Canvas	83

Liste des tableaux

1.1	Calendrier d'achèvement du projet	59
5.1	Plan de Répartition du Budget du Projet	68

Première partie

Partie 1

Chapitre 1

Introduction

1.1 Contexte et Motivation

Les maladies cardiovasculaires représentent aujourd'hui la première cause de mortalité dans le monde. Selon l'Organisation Mondiale de la Santé (OMS), elles sont responsables de plus de 17,9 millions de décès chaque année, soit environ 31% de l'ensemble des décès mondiaux. Ces maladies incluent les crises cardiaques, les accidents vasculaires cérébraux, l'hypertension et d'autres conditions qui affectent le cœur et les vaisseaux sanguins [1].

La détection précoce et le traitement efficace des maladies cardiaques peuvent significativement améliorer les chances de survie et la qualité de vie des patients. Cependant, les méthodes traditionnelles de diagnostic, bien qu'efficaces, peuvent être coûteuses, chronophages et nécessitent souvent l'intervention de spécialistes hautement qualifiés[2].

L'émergence de l'apprentissage automatique (AA) offre une opportunité prometteuse pour révolutionner le domaine de la santé, notamment dans la détection des maladies cardiaques. Les avancées dans les techniques d'AA permettent de traiter de vastes quantités de données médicales et d'identifier des modèles complexes qui pourraient échapper à l'œil humain. Voici quelques motivations clés pour développer un système de détection des maladies cardiaques basé sur l'apprentissage automatique :

1. **Amélioration de la Précision du Diagnostic** : Les algorithmes d'apprentissage automatique peuvent analyser les données médicales, telles que les électrocardiogrammes (ECG), les imageries médicales et les historiques de santé des patients, pour détecter des anomalies et prédire des problèmes cardiaques avec une grande précision.
2. **Réduction des Coûts et du Temps** : Un système automatisé peut effectuer des analyses rapides et précises, réduisant ainsi la nécessité de tests coûteux et de consultations fréquentes avec des spécialistes. Cela permet un diagnostic plus rapide et une intervention précoce.
3. **Accessibilité** : Un tel système peut être déployé dans des régions où l'accès aux soins spécialisés est limité. Cela peut aider à combler les lacunes en matière de soins de santé dans les zones rurales ou sous-développées.
4. **Personnalisation des Soins** : En analysant les données spécifiques à chaque patient, les systèmes d'apprentissage automatique peuvent fournir des recommandations personnalisées pour la prévention et le traitement des maladies cardiaques.
5. **Suivi Continu** : Les technologies portables et les appareils connectés peuvent fournir des données en temps réel, permettant un suivi continu de l'état de santé cardiaque des patients et une intervention rapide en cas de détection d'anomalies.

En conclusion, l'intégration de l'apprentissage automatique dans la détection des maladies cardiaques offre une voie prometteuse pour améliorer les soins de santé, réduire les coûts, et sauver des vies. Le développement de tels systèmes représente un pas significatif vers une médecine plus précise, plus accessible et plus efficace[3].

1.2 Problématique

Les maladies cardiaques représentent l'une des principales causes de décès dans le monde. La détection précoce de ces maladies est cruciale pour améliorer les taux de survie et la qualité de vie des patients. Cependant, les méthodes traditionnelles de diagnostic reposent souvent sur des procédures invasives, coûteuses et nécessitent l'intervention de spécialistes hautement qualifiés. Ces limitations peuvent entraîner des retards dans le diagnostic et le traitement, particulièrement dans les régions où l'accès aux soins médicaux est limité[4].

Les avancées en apprentissage automatique offrent de nouvelles opportunités pour améliorer la détection des maladies cardiaques. En analysant de grandes quantités de données médicales, les algorithmes de machine learning peuvent identifier des modèles complexes et des indicateurs précoces de maladies cardiaques avec une précision potentiellement supérieure à celle des méthodes traditionnelles[5].

1.3 Objectifs de l'étude

Cette étude vise à développer un système de détection précoce des maladies cardiaques en exploitant les avancées en apprentissage automatique. Plus précisément, l'étude a pour objectifs de :

- Concevoir et entraîner divers modèles d'apprentissage automatique pour détecter les signes précoces de maladies cardiaques.
- Comparer les performances de différents algorithmes (par exemple, réseaux de neurones, machines à vecteurs de support, forêts aléatoires) afin de déterminer les méthodes les plus efficaces.
- Sélectionner le meilleur algorithme pour développer un système de détection des maladies cardiaques.
- Identifier les facteurs de risque liés aux patients afin de fournir des conseils personnalisés pour la prévention et le traitement.

1.4 Questions de recherche

Les questions de recherche suivantes sont conçues pour guider une étude sur l'utilisation des méthodes d'apprentissage automatique dans la détection des maladies cardiaques.

1. Quels modèles d'apprentissage automatique sont les plus efficaces pour détecter les signes précoces de maladies cardiaques ?
2. Quel est l'algorithme d'apprentissage automatique le plus performant pour développer un système de détection des maladies cardiaques ?
3. Quels facteurs de risque sont les plus significatifs pour prédire les maladies cardiaques à l'aide de modèles d'apprentissage automatique ?
4. Comment identifier les facteurs de risque individuels dans un système de détection de maladies cardiaques pour fournir des conseils personnalisés ?

5. Quelles sont les limitations des modèles actuels d'apprentissage automatique dans la détection précoce des maladies cardiaques et comment peuvent-elles être surmontées ?
6. Comment l'utilisation d'un système de détection précoce basé sur l'apprentissage automatique peut-elle améliorer les résultats cliniques et la gestion des patients cardiaques ?

1.5 Portée et Limitations

1.5.1 Portée

La portée de l'étude englobe les aspects clés suivants :

- **Jeux de Données de Référence** : L'étude utilisera deux jeux de données de référence bien connus pour les maladies cardiaques [6, 7]. Ces jeux de données fournissent une collection diversifiée et complète de dossiers de patients, incluant diverses caractéristiques cliniques et résultats.
- **Conception et Entraînement des Modèles** : Divers modèles d'apprentissage automatique, incluant les réseaux de neurones, les machines à vecteurs de support et les forêts aléatoires, seront conçus et entraînés pour détecter les signes précoces de maladies cardiaques[8].
- **Comparaison des Performances** : Les performances de différents algorithmes seront comparées pour identifier la méthode la plus efficace pour la détection précoce des maladies cardiaques[9].
- **Identification du Meilleur Algorithme** : L'étude visera à sélectionner l'algorithme le plus performant pour développer un système robuste de détection des maladies cardiaques[10].
- **Identification des Facteurs de Risque** : Les facteurs de risque significatifs pour les maladies cardiaques seront identifiés à l'aide des modèles sélectionnés afin de fournir des recommandations personnalisées aux patients[11].

1.5.2 Limitations

Malgré l'approche adoptée, les résultats basés sur ces jeux de données peuvent ne pas être entièrement généralisables à toutes les populations ou contextes cliniques[12].

1.6 Structure du mémoire

Les chapitres restants peuvent être décrits comme suit :

1.6.1 Chapitre 2 : Revue de la Littérature

Ce chapitre fournit une revue de la littérature existante sur les maladies cardiaques et leurs techniques de diagnostic. Il explore l'application de l'apprentissage automatique dans les soins de santé et examine les travaux antérieurs sur la détection des maladies cardiaques utilisant ces technologies. L'objectif est de situer la recherche dans le contexte des études existantes, d'identifier les approches prometteuses et de souligner les lacunes que cette étude vise à combler.

1.6.2 Chapitre 3 : Méthodologie

Ce chapitre détaille la méthodologie utilisée pour mener la recherche. Il commence par une description des ensembles de données, y compris leur source et leurs caractéristiques. Ensuite, les algorithmes d'apprentissage automatique sélectionnés sont présentés, avec une explication de leur fonctionnement et de leur pertinence pour l'étude. Le processus d'entraînement et d'évaluation des modèles est décrit, en mettant l'accent sur les techniques et les critères utilisés pour mesurer leur performance.

1.6.3 Chapitre 4 : Implémentation du Système

Ce chapitre décrit l'implémentation pratique du système de détection des maladies cardiaques. Il offre un aperçu de l'architecture du système et présente un diagramme de bloc illustrant les principaux composants tels que l'acquisition de données, le prétraitement, le modèle d'apprentissage automatique et l'interface utilisateur. Les technologies et les outils utilisés sont listés et décrits. La section se termine par une description détaillée de l'interface utilisateur et de ses fonctionnalités.

1.6.4 Chapitre 5 : Résultats et Discussion

Ce chapitre présente et discute les résultats de l'étude. La performance des modèles sur les ensembles de données de test est présentée à travers des métriques et des visualisations. Une analyse comparative des performances des différents modèles testés est effectuée. Les résultats sont discutés pour en tirer des conclusions significatives. Les implications pour la pratique clinique sont explorées, et des recommandations pour l'amélioration des modèles sont proposées.

1.6.5 Chapitre 6 : Conclusion et Travaux Futurs

Le dernier chapitre résume les principaux éléments de l'étude, incluant la problématique, la méthodologie, les résultats et les contributions. Les limitations de l'étude sont identifiées et discutées. Enfin, des perspectives futures sont suggérées, proposant des pistes de recherche pour surmonter les limitations identifiées et explorer de nouvelles directions.

Chapitre 2

Revue de la littérature

Ce chapitre revoit la littérature existante concernant les maladies cardiaques et les méthodes de diagnostic associées. Il explore l'utilisation de l'apprentissage automatique dans le domaine de la santé et analyse les études précédentes sur la détection des maladies cardiaques à l'aide de ces technologies.

2.1 Aperçu des Maladies Cardiaques et leur Importance

Le cœur est l'organe le plus important du corps humain. S'il ne fonctionne pas correctement, alors il affecte d'autres organes du corps. Selon un rapport, 7 millions de personnes meurent d'une crise cardiaque chaque année. Selon le rapport de l'OMS, environ 17,9 millions de personnes meurent des suites des maladies cardiovasculaires (CVD) en 2016 [13], 31% des décès sont dus à des maladies cardiaques dans le monde chaque année. Le pompage du sang vers le corps humain est la fonction vitale du cœur qui fournit de l'oxygène et des nutriments au corps humain et élimine également d'autres déchets métaboliques du corps. S'il y a une carence de sang dans le corps humain, le cœur ne fonctionne pas correctement et il cesse de fonctionner, ce qui provoque la mort de l'être humain. L'angine survient lorsqu'il y a une perte temporaire de sang dans le cœur, ce qui provoque des douleurs thoraciques [14].

2.1.1 Types de Maladies Cardiaques

Les maladies cardiovasculaires sont de deux types :

- **Crise Cardiaque** : se produit lorsque les vaisseaux sanguins du cœur sont soudainement bloqués [15].
- **Insuffisance Cardiaque** : il résulte d'une maladie coronarienne, hypertension, cardiomyopathie. L'insuffisance cardiaque est essentiellement lorsque le cœur est incapable de maintenir un flux sanguin fort et que cela entraîne une fatigue chronique, une résistance aux activités physiques et un essoufflement. L'insuffisance cardiaque peut être divisée en trois types :
 - Insuffisance cardiaque du côté droit
 - Insuffisance cardiaque du côté gauche
 - Insuffisance cardiaque congestive

L'insuffisance cardiaque droite provoque généralement une insuffisance cardiaque gauche. Dans le côté droit de l'insuffisance cardiaque, le sang recule dans d'autres tissus tels que le foie et dans l'abdomen causant la congestion dans ces secteurs. À la suite d'une insuffisance cardiaque du côté droit, nous pouvons avoir une hépatomégalie et des ascites. Dans l'insuffisance cardiaque du côté gauche, le sang oxygéné ne peut pas être pompé du cœur au reste du corps. Ainsi, le sang peut refluer. Le sang peut s'accumuler dans les veines pulmonaires, entraînant une accumulation de liquide dans les poumons, ce qui provoque un essoufflement et un œdème [16].

2.1.2 Facteurs de Risque et Symptômes Associés aux Maladies Cardiaques

Voici quelques-uns des principaux facteurs de risque et symptômes associés aux maladies cardiaques :

Facteurs de Risque

1. **Pression artérielle élevée (hypertension)** : Une force élevée du sang poussant contre les parois des artères peut causer des dommages au fil du temps. Il fait travailler le cœur plus fort et contribue à l'accumulation de plaque.
2. **Cholestérol élevé** : Des niveaux élevés de cholestérol (substance semblable à de la graisse) dans le sang peuvent s'accumuler sur les parois des artères, les rétrécissant et limitant le flux sanguin.
3. **Diabète** : Avoir le diabète signifie qu'il y a trop de sucre dans le sang qui peut endommager les artères et augmenter l'inflammation.
4. **Obésité** : L'excès de graisse corporelle, en particulier autour de la taille, augmente la pression sur le cœur et aggrave d'autres facteurs de risque comme l'hypercholestérolémie et la pression artérielle.
5. **Tabagisme** : Les produits chimiques contenus dans la fumée du tabac nuisent à la santé cardiovasculaire en endommageant les parois des artères, en augmentant l'inflammation et en rendant le sang plus collant et plus sujet aux caillots.
6. **Inactivité physique** : Le manque d'exercice a une incidence négative sur la condition cardiovasculaire et peut contribuer à l'obésité et à d'autres facteurs de risque.
7. **Âge** : En vieillissant, le risque augmente en raison de l'accumulation progressive de plaque dans les artères au fil du temps. Important après 65 ans.
8. **Antécédents familiaux** : La génétique joue un rôle en transmettant les prédispositions à des conditions comme le cholestérol élevé qui favorisent les maladies cardiaques.
9. **Stress** : Des niveaux élevés de stress peuvent indirectement augmenter les risques en augmentant les comportements comme la suralimentation et le tabagisme. Le stress chronique peut aussi avoir un impact direct sur la santé.
10. **Mauvaise alimentation** : Les régimes riches en graisses saturées, en gras trans, en cholestérol et en excès de sel/sodium augmentent le taux de cholestérol et la pression artérielle [17].

Symptômes

1. **Douleur thoracique / Inconfort** : L'un des symptômes les plus courants, cela peut ressembler à une compression, une pression, une plénitude ou une douleur au centre de la poitrine. Il peut irradier vers le bras, le dos, le cou, la mâchoire ou l'estomac.
2. **Essoufflement** : La difficulté à reprendre votre respiration ou la sensation d'essoufflement excessif avec l'effort peut signaler un apport insuffisant de sang / oxygène du cœur.

3. **Fatigue/faiblesse** : Le cœur doit travailler plus fort lorsqu'il est malade, ce qui entraîne une fatigue inhabituelle, une faible endurance et un manque d'énergie.
4. **Étourdissements/étourdissements** : Un flux sanguin insuffisant vers le cerveau à partir d'une artère bloquée peut provoquer des étourdissements ou des évanouissements.
5. **Nausées/vomissements** : Parfois confondues avec un problème d'estomac, les nausées inexplicables peuvent être un symptôme d'une crise cardiaque.
6. **Sueurs froides** : La sueur froide et moite est souvent associée à la douleur thoracique / pression de l'angine ou d'une crise cardiaque.
7. **Douleur au bras/à l'épaule** : L'inconfort ou la douleur qui parcourt un bras ou les deux bras, les épaules ou le haut du dos peut signaler un problème cardiaque.
8. **Douleur à la mâchoire / cou** : Certaines personnes, en particulier les femmes, éprouvent une douleur sourde ou une douleur dans la mâchoire ou la région du cou pendant une crise cardiaque.
9. **Gonflement des jambes** : Une mauvaise fonction de pompe cardiaque peut provoquer une accumulation de liquide entraînant un gonflement des jambes, des chevilles et des pieds [18].

2.2 Techniques de Diagnostic Médical des Maladies Cardiaques

2.2.1 Électrocardiogramme (ECG)

Un ECG enregistre l'activité électrique du cœur sur une période de temps. Il aide à détecter :

- Arythmies (battements cardiaques irréguliers)
- Infarctus du myocarde (crise cardiaque)
- Élargissement du cœur
- Ischémie (diminution du flux sanguin vers le cœur) [19].

2.2.2 Échocardiogramme

Un échocardiogramme utilise des ondes ultrasonores pour créer des images du cœur. Il aide à évaluer :

- Taille et forme du cœur
- Valves et cavités cardiaques
- Débit cardiaque et fraction d'éjection
- Présence de malformations cardiaques ou de dommages [20].

2.2.3 Tests de Résistance

Les tests d'effort mesurent la réponse du cœur à l'effort physique. Les types courants incluent :

- Test d'effort physique : Surveillance du cœur pendant que le patient fait de l'exercice sur un tapis roulant ou un vélo stationnaire.
- Test d'effort pharmacologique : Utilisation de médicaments pour stimuler le cœur chez les patients incapables de faire de l'exercice [21].

2.2.4 Cathétérisme Cardiaque

Consiste à insérer un cathéter dans un vaisseau sanguin et à le guider vers le cœur pour :

- Mesurer la pression artérielle dans les cavités cardiaques
- Évaluer la fonction du cœur et des valves
- Effectuer une angiographie pour visualiser les artères coronaires [22].

2.2.5 Angiographie Coronarienne

Un type de cathétérisme cardiaque où un colorant de contraste est injecté dans les artères coronaires, les rendant visibles sur les images radiographiques. Il aide à identifier :

- Blocages ou rétrécissement des artères coronaires
- Anévrismes et autres anomalies structurelles [23].

2.2.6 Tomodensitométrie (TDM)

Les tomodensitogrammes fournissent des images détaillées du cœur et des vaisseaux sanguins. Les types incluent :

- Angiographie coronarienne CT : Méthode non invasive pour visualiser les artères coronaires.
- Calcium Scoring CT : Mesure la quantité de plaque calcifiée dans les artères coronaires [24].

2.2.7 Imagerie par Résonance Magnétique (IRM)

L'IRM cardiaque utilise des champs magnétiques et des ondes radio pour créer des images détaillées. Cela aide à évaluer :

- Structure et fonction du cœur
- Viabilité myocardique (tissu musculaire cardiaque vivant)
- Malformations cardiaques congénitales [25].

2.2.8 Analyses Sanguines

Certains biomarqueurs dans le sang peuvent indiquer une maladie cardiaque :

- Troponine : des taux élevés indiquent un infarctus du myocarde.
- BNP ou NT-pro BNP : des taux élevés suggèrent une insuffisance cardiaque.
- Profil lipidique : Mesure le cholestérol et les triglycérides [26].

2.2.9 Moniteur Holter

Appareil portable porté par le patient pendant 24 à 48 heures pour enregistrer en continu l'activité électrique du cœur, utile pour détecter les arythmies intermittentes [27].

2.2.10 Moniteur d'Événements

Semblable à un moniteur Holter mais utilisé pendant une période plus longue (semaines à mois). Le patient l'active lorsqu'il présente des symptômes [28].

2.2.11 Test de La Table Inclinable

Aide à diagnostiquer la cause de l'évanouissement inexplicé (syncope) en surveillant la pression artérielle et la fréquence cardiaque pendant que le patient est incliné à différents angles [29].

2.2.12 Cardiologie Nucléaire

Consiste à injecter une petite quantité de matière radioactive et à utiliser une caméra spéciale pour visualiser le flux sanguin vers le cœur, souvent utilisé dans :

- Imagerie de perfusion myocardique : pour évaluer le flux sanguin et identifier les zones de mauvaise perfusion.
- MUGA Scan : pour évaluer la fonction des ventricules du cœur [30].

Chacune de ces techniques fournit des informations précieuses et est choisie en fonction des symptômes du patient, de ses antécédents médicaux et de ses besoins diagnostiques spécifiques.

2.3 Apprentissage Automatique dans les Soins de Santé

2.3.1 Vue d'ensemble de l'Apprentissage Automatique

L'apprentissage automatique (ML) est un sous-ensemble de l'intelligence artificielle (IA) qui implique le développement d'algorithmes et de modèles statistiques qui permettent aux ordinateurs d'effectuer des tâches sans instructions explicites. Au lieu de cela, ces systèmes apprennent des données et améliorent leurs performances au fil du temps [31].

- **Algorithmes** : Procédures mathématiques ou formules utilisées par les modèles d'apprentissage automatique pour apprendre à partir des données.
- **Données d'entraînement** : L'ensemble de données utilisé pour entraîner le modèle ML.
- **Caractéristiques** : Propriétés ou caractéristiques mesurables individuelles des données.
- **Modèle** : La sortie de l'algorithme ML après avoir été entraîné sur les données.
- **Prédiction** : La sortie générée par un modèle ML lorsqu'il est appliqué à de nouvelles données.
- **Évaluation** : Évaluer les performances d'un modèle d'apprentissage automatique à l'aide de mesures telles que l'exactitude, la précision, le rappel et le score F1 [32].

2.3.2 Applications de l'Apprentissage Automatique dans les Diagnostics Médicaux

L'apprentissage automatique (ML) a un large éventail d'applications dans le diagnostic médical, améliorant considérablement les capacités des systèmes de santé. Voici quelques applications notables :

Imagerie Médicale

Les algorithmes de ML, en particulier l'apprentissage en profondeur, sont utilisés pour analyser des images médicales telles que les radiographies, les IRM et les tomodensitogrammes. Ces modèles peuvent aider à détecter des anomalies telles que des tumeurs, des fractures et d'autres pathologies avec une grande précision [33].

Analyse Prédictive

Les modèles d'apprentissage automatique peuvent analyser les données des patients pour prédire la probabilité de maladies telles que le diabète, les maladies cardiaques et le cancer. Cela permet une intervention précoce et des plans de traitement personnalisés [34].

Diagnostic de la Maladie

Les algorithmes de ML peuvent aider à diagnostiquer les maladies en analysant divers types de données, y compris les informations génétiques, les résultats de laboratoire et les antécédents cliniques. Par exemple, le ML peut aider à diagnostiquer des affections telles que la maladie d'Alzheimer, la maladie de Parkinson et diverses formes de cancer [35].

Découverte et Développement de Médicaments

L'apprentissage automatique peut accélérer le processus de découverte de médicaments en prédisant comment différents composés interagissent avec des cibles dans le corps. Cela aide à identifier les candidats médicaments prometteurs plus rapidement et plus efficacement [36].

Médecine Personnalisée

Les modèles de ML peuvent analyser les données génétiques, environnementales et sur le mode de vie pour adapter les traitements médicaux aux patients individuels. Cette approche augmente l'efficacité des traitements et réduit les effets indésirables [37].

Dossiers de Santé Électroniques (DSE)

Le ML peut être utilisé pour analyser les DSE afin d'identifier des modèles et des tendances susceptibles d'indiquer un diagnostic ou de prédire les résultats pour les patients. Les techniques de traitement du langage naturel (NLP) peuvent extraire des informations utiles du texte non structuré de ces enregistrements [38].

Pathologie

Les algorithmes d'apprentissage automatique peuvent aider les pathologistes en analysant des échantillons de tissus et en identifiant les cellules cancéreuses. Cela peut conduire à des diagnostics plus précis et plus rapides [39].

Systèmes d'Aide à la Décision Clinique (CDSS)

Les modèles de ML peuvent fournir aux cliniciens des recommandations fondées sur des preuves en analysant les données des patients et la littérature médicale. Cela aide à prendre des décisions éclairées sur les soins aux patients [40].

Télesurveillance et Télémédecine

Les appareils portables et les outils de surveillance à distance peuvent collecter des données sur les patients, que les algorithmes d'apprentissage automatique analysent pour fournir des informations sur la santé en temps réel. Ceci est particulièrement utile pour la gestion des maladies chroniques et le suivi de la récupération postopératoire [41].

Génomique

ML est utilisé pour analyser les données génomiques afin d'identifier les marqueurs génétiques associés aux maladies. Cela aide à comprendre les mécanismes de la maladie et à développer des thérapies ciblées [42].

Santé Mentale

Les modèles d'apprentissage automatique peuvent analyser les modèles de parole, les expressions faciales et d'autres biomarqueurs pour détecter des problèmes de santé mentale tels que la dépression, l'anxiété et le SSPT [43].

Gestion des Patients

Le ML peut optimiser les tâches administratives telles que la planification des patients, l'allocation des ressources et la gestion des flux de travail, ce qui permet une prestation de soins de santé plus efficace [44].

2.4 Travaux Antérieurs sur la Détection des Maladies Cardiaques Utilisant l'Apprentissage Automatique

2.4.1 Réseau Neuronal Artificiel (ANN)

Dans leur article intitulé "Prédiction précoce des maladies cardiovasculaires à l'aide d'un réseau de neurones artificiels", Jyotismita Talukder et Thipendra P. Singh explorent l'application des réseaux de neurones artificiels (ANN) dans la prévision des maladies cardiovasculaires. En tirant parti de l'ensemble de données sur les maladies cardiaques de l'UCI, ils ont mis en œuvre une rétropropagation dans des algorithmes d'apprentissage automatique pour obtenir des résultats notables, atteignant une précision de 81%. Cette recherche souligne le potentiel des méthodologies ANN dans la détection précoce et le pronostic des affections cardiovasculaires, offrant des pistes prometteuses pour des stratégies de soins de santé préventifs [45].

En complément de cette étude, l'article intitulé "Prédiction des maladies cardiaques à l'aide de réseaux neuronaux", rédigé par P. Ramprakash, R. Sarumathi, R. Mowriya et S. Nithyavishnupriya, met également l'accent sur l'utilisation de méthodologies de réseaux neuronaux pour prévoir les maladies cardiaques. En utilisant un ensemble de données provenant de Kaggle contenant 303 instances et 13 fonctionnalités, les auteurs ont implémenté des systèmes hybrides. Leurs recherches ont donné des résultats prometteurs, atteignant un taux de réussite remarquable de 91%, soulignant ainsi l'efficacité des approches de réseaux neuronaux dans la prédiction des maladies cardiaques et mettant en évidence le potentiel d'amélioration des diagnostics de santé et des mesures préventives [46].

Poursuivant cette exploration, l'article "Prédire les maladies cardiaques à l'aide de réseaux de neurones", rédigé par Ahmed Muhammad Haider Al-Sharif et Samy S. Abu-Naser, se concentre sur l'utilisation des techniques de réseaux de neurones pour prévoir les occurrences de maladies cardiaques. En utilisant un ensemble de données provenant de Kaggle contenant 1025 échantillons et englobant 14 caractéristiques, les auteurs ont mené leur analyse et ont obtenu des résultats prometteurs, le modèle de réseau neuronal atteignant un taux de précision impressionnant de 92% et démontrant une erreur moyenne de 0,062. Les prédicteurs clés identifiés grâce à leur analyse incluent des variables telles que "vieux pic", "thalach", "trestbps", entre autres, mettant en lumière des indicateurs cruciaux pour le pronostic des maladies cardiaques. Ces résultats mettent en évidence le potentiel des réseaux de neurones dans l'amélioration de la précision des modèles prédictifs pour l'évaluation de la santé cardiovasculaire [47].

2.4.2 Travaux Antérieurs sur la Détection des Maladies Cardiaques Utilisant l'Apprentissage Automatique

Machines à Vecteurs de Support (SVM)

Dans l'article intitulé "Prédiction des maladies cardiaques à l'aide de la machine à vecteurs de support," rédigé par Animesh Kumar Dubey, l'accent est mis sur l'utilisation des algorithmes de la machine à vecteurs de support (SVM) pour prédire l'apparition des maladies cardiaques. En utilisant l'ensemble de données sur les maladies cardiaques de l'UCI, Dubey a mené une analyse approfondie pour évaluer l'efficacité de la SVM dans ce contexte. Les résultats de l'étude ont révélé des résultats prometteurs, le modèle SVM atteignant un taux de précision impressionnant de 91,94% dans la prédiction des maladies cardiaques. Cette recherche souligne le potentiel des algorithmes SVM en tant qu'outils précieux pour prévoir avec précision les maladies cardiaques, offrant des avancées significatives dans la détection précoce et les stratégies de soins de santé préventifs [48].

De manière complémentaire, dans l'article intitulé "Application de la machine à vecteurs de support pour la prédiction de l'observance des médicaments chez les patients insuffisants cardiaques," rédigé par Youn-Jung Son, Hong-Gee Kim, Eung-Hee Kim, Sang-sup Choi et Soo-Kyoung Lee, l'accent est mis sur l'utilisation de modèles de machine à vecteurs de support (SVM) pour prédire l'observance des médicaments chez les patients insuffisants cardiaques. S'appuyant sur les données d'un hôpital universitaire concernant l'observance des médicaments dans cette population de patients, les auteurs ont mené leur analyse. Les modèles de SVM développés visaient à identifier les principaux prédicteurs de l'observance des médicaments. La précision de détection la plus élevée obtenue grâce à ces modèles était de 77,63%, indiquant un niveau notable de succès dans la prévision du comportement d'observance des médicaments chez les patients insuffisants cardiaques. Cette recherche met en évidence le potentiel des algorithmes de SVM dans les établissements de santé, en particulier pour aider les prestataires de soins de santé à identifier les patients à risque de non-observance des schémas thérapeutiques, contribuant en fin de compte à l'amélioration des résultats pour les patients et des stratégies de prise en charge de l'insuffisance cardiaque [49].

Poursuivant cette exploration des applications de la SVM dans le domaine de la santé cardiovasculaire, l'article intitulé "Prédiction des maladies cardiaques et de la capacité de survie à l'aide d'une machine à vecteurs de support," rédigé par Tanvi S. Patel, Daxesh P. Patel, Mallika Sanyal et Pranav S. Shrivastav, les chercheurs explorent l'application de techniques d'apprentissage automatique, en particulier la machine à vecteurs de support (SVM) et les algorithmes naïfs de Bayes, pour prédire l'apparition et la capacité de survie des maladies cardiaques. En utilisant l'ensemble de données sur les maladies cardiaques de l'UCI, les auteurs ont mené une analyse complète pour évaluer l'efficacité de ces algorithmes dans les pronostics des soins de santé. L'algorithme SVM est apparu comme particulièrement prometteur, atteignant un taux de précision impressionnant de 88% pour prédire si un patient a une maladie cardiaque et son statut de survie. Ces résultats mettent en évidence le potentiel des méthodologies d'apprentissage automatique pour améliorer les modèles prédictifs de diagnostic et de pronostic des maladies cardiaques, offrant des informations précieuses aux professionnels de la santé pour améliorer les soins aux pa-

tients et les stratégies de traitement. Les recherches de Patel et al. contribuent à faire progresser le domaine de l'analyse prédictive en médecine cardiovasculaire, ouvrant la voie à des interventions de soins de santé plus précises et personnalisées pour les patients à risque de maladie cardiaque [50].

Arbre de Décision

Dans l'article "Prédire les maladies cardiaques à l'aide d'un modèle arborescent" d'Emil Agbemade, le Référentiel d'apprentissage automatique de l'UCI a été utilisé comme ensemble de données. Les résultats ont montré que le classificateur de l'arbre de décision avait une précision de 81% pour prédire correctement les patients comme ayant une maladie cardiaque et 82% comme n'ayant pas de maladie cardiaque, surpassant ainsi d'autres algorithmes d'apprentissage automatique utilisés dans les études précédentes [51].

Par ailleurs, l'article "Prédiction des maladies cardiaques à l'aide d'un arbre de décision" d'ABHINAY_SINGH, utilisant l'ensemble de données d'Analyse et de prédiction des crises cardiaques de Kaggle, a révélé que la précision du classificateur de l'arbre de décision sur cet ensemble de données était de 84,21% [52].

De plus, l'article "Utiliser l'arbre de décision pour diagnostiquer les patients atteints de maladies cardiaques" de Mai Shouman, Tim Turner et Rob Stocker, qui a utilisé l'ensemble de données de la Cleveland Clinic disponible dans le référentiel d'apprentissage automatique de l'UC Irvine, a montré que la plus grande précision obtenue était de 84,1% par l'arbre de décision du Rapport de gain de discrétisation à fréquence égale [53].

2.4.3 Régression Linéaire

L'article "Prédiction des maladies cardiaques à l'aide de techniques distinctes d'intelligence artificielle : analyse et comparaison des performances" par Md. Imam Hussein, Mehdi Hassan Marouf, Dr. Ashikur Rahman Khan, Farida Siddiqi Prite, Sharmin Fatema, Dr. Sabbir Ejaz, et Dr. Ahmad, a utilisé le Référentiel d'apprentissage automatique de l'UCI comme source de la base de données. L'étude a obtenu des résultats impressionnants avec une précision de 98%, une sensibilité de 97,5% et une spécificité de 92,5% dans la prédiction des maladies cardiaques à l'aide de diverses techniques d'intelligence artificielle [54].

De plus, l'article "Prédiction des maladies cardiaques à l'aide d'un modèle de régression linéaire multiple" de K. Poluraju et D. Durga Rasade a également utilisé l'ensemble de données UCI pour la prédiction des maladies cardiaques. Cette étude a atteint une précision de 84,85% en utilisant le modèle de régression linéaire multiple, démontrant ainsi son efficacité pour prédire le risque de maladie cardiaque [55].

2.4.4 Apprentissage Profond (Deep Learning)

L'article "Un réseau neuronal Convolutif profond pour la détection précoce des maladies cardiaques" de Sadia Arooj, Saif ur Rehman, Azhar Imran, Abdullah Almuhaimeed, A. Khuzaim Alzahrani et Abdulkareem Alzahrani a utilisé l'ensemble de données sur les

maladies cardiaques de l'UCI, montrant que le réseau neuronal convolutif profond proposé atteignait une précision de validation de 91,7% [56].

De plus, l'article "Prédiction du risque de maladie cardiaque à l'aide de techniques d'apprentissage en profondeur avec augmentation des caractéristiques" de María Teresa García-Ordás, Martín Bayón-Gutiérrez, Carmen Benavides, Jose Aveleira-Mata et José Alberto Benítez-Andrades a utilisé un ensemble de données de Kaggle composé de 918 échantillons avec seulement 11 caractéristiques cliniques par échantillon, atteignant une précision de 90%, soit 4,4% de plus que les autres méthodes de pointe [57].

En outre, l'article "Deep Learning to Improve Heart Disease Risk Prediction" de Shelda Sajeev, Anthony Maeder, Stephanie Champion, Aline Beleigoli, Cheng Ton, Xianglong Kong et Minglei Shu a utilisé un ensemble de données de 567 participants provenant de deux cohortes prises de la base de données UCI, atteignant une précision de 94% et un score AUC de 0,964, démontrant ainsi l'efficacité des techniques d'apprentissage en profondeur pour améliorer la prédiction du risque de maladie cardiaque [58].

2.4.5 K-voisins les plus proches (K-NN)

L'article intitulé "Heart Disease Prediction using K-Nearest Neighbors Algorithm," rédigé par Dadi Rahmat, Andika A. Putra, Hamrin, et Agung W. Setiawan, explore l'utilisation de l'algorithme K-NN pour la prédiction des maladies cardiaques. En se basant sur le Cleveland Heart Disease Dataset, l'étude a démontré que le modèle K-NN pouvait classer les patients atteints de maladies cardiaques avec une précision de 84% [59].

L'article intitulé "Hybrid Approach for Heart Disease Prediction using K-Nearest Neighbors and Particle Swarm Optimization," rédigé par Jabbar Akhil, propose une approche hybride utilisant K-NN et l'optimisation par essaim de particules. En se basant sur le Cleveland Heart Disease Dataset, cette recherche a atteint une précision de 92% dans la prédiction des maladies cardiaques [60].

2.5 Conclusion

Le chapitre 2 présente une revue de littérature sur les maladies cardiaques. Il donne d'abord un bref aperçu de l'importance de ces maladies, puis décrit les principaux types de maladies cardiovasculaires comme les crises cardiaques et l'insuffisance cardiaque.

Il détaille ensuite les principaux facteurs de risque des maladies cardiaques comme l'hypertension, le cholestérol élevé, le diabète, l'obésité ou le tabagisme. Il liste également les symptômes associés comme les douleurs thoraciques, l'essoufflement ou la fatigue.

Le chapitre décrit ensuite les principales techniques de diagnostic médical des maladies cardiaques telles que l'électrocardiogramme, l'échocardiographie, le cathétérisme cardiaque, la tomodensitométrie ou l'imagerie par résonance magnétique. Il précise comment chacune de ces techniques permet d'évaluer la structure et la fonction cardiaque.

Il aborde ensuite l'apprentissage automatique et ses applications potentielles dans les diagnostics médicaux comme l'analyse d'images, la prédiction de maladies, l'aide au diagnostic ou la médecine personnalisée.

Le chapitre passe en revue divers travaux antérieurs ayant utilisé des techniques d'apprentissage automatique comme les réseaux de neurones artificiels, les machines à vecteurs de support ou les arbres de décision pour la détection des maladies cardiaques.

Enfin, il présente brièvement une étude ayant utilisé un réseau neuronal profond pour la détection précoce des maladies cardiaques à partir de données cliniques.

Chapitre 3

LA METHODOLOGIE

Ce chapitre décrit la méthodologie employée pour mener cette recherche. Il débute par une présentation des ensembles de données et de leurs caractéristiques. Les algorithmes d'apprentissage automatique choisis sont ensuite introduits, accompagnés d'une explication de leur fonctionnement. Le processus d'entraînement et d'évaluation des modèles est alors exposé, et les critères utilisés pour évaluer leur performance sont présentés.

3.1 Description des ensembles de données

Cette étude utilisera deux ensembles de données de référence bien connus pour les maladies cardiaques, à savoir les ensembles de données de Cleveland [6] et de Framingham [7].

3.1.1 Ensemble de données de Cleveland

Cette base de données contient 76 attributs, mais toutes les expériences publiées se réfèrent à un sous-ensemble de 14 d'entre eux. Le champ "objectif" indique la présence d'une maladie cardiaque chez le patient. Il est représenté par une valeur entière allant de 0 (absence) à 4. Les expériences menées avec la base de données de Cleveland se sont concentrées sur la simple distinction entre la présence (valeurs 1, 2, 3, 4) et l'absence (valeur 0) de la maladie.

3.1.2 Ensemble de données de Framingham

Le jeu de données "Framingham" sur les maladies cardiaques comprend plus de 4 240 enregistrements, 16 colonnes et 15 attributs. L'objectif de ce jeu de données est de prédire si le patient présente un risque de maladie coronarienne (CHD) dans les 10 ans à venir.

3.2 Les algorithmes d'apprentissage automatique

3.2.1 Les réseaux de neurones (Artificial Neural Networks)

Les ANN sont un type d'algorithme d'apprentissage automatique inspiré du cerveau humain. Ils sont constitués de nœuds interconnectés, appelés neurones, qui traitent les informations et apprennent à faire des prédictions. L'ANN est présenté avec un ensemble de données d'entrée sont propagées à travers le réseau couche par couche, ou à chaque couche les neurones calculent une somme pondérée des entrées qu'ils reçoivent, les neurones s'activent alors, produisant un signal de sortie, les signaux de sortie de la couche finale sont utilisés pour faire une prédiction [61].

3.2.2 k-plus proches voisins (k-NN)

kNN est un algorithme d'apprentissage automatique supervisé simple et facile à mettre en œuvre qui peut être utilisé pour résoudre les problèmes de classification et de régression. Son principe est de rechercher les k observations les plus proches d'un nouvel élément, en

calculant la distance entre eux. Pour la classification, on attribue au nouvel élément la classe majoritaire parmi ses k plus proches voisins. Pour la régression, on lui attribue la moyenne des valeurs de ses voisins. Bien que simple, kNN présente l'inconvénient de ralentir considérablement quand le volume de données augmente. Il est donc peu pratique pour des prédictions rapides, mais peut être utile [62].

3.2.3 Arbre de décision

L'arbre de décision est un outil visuel qui permet de représenter les résultats possibles d'une série de choix interconnectés. Il est utilisé pour évaluer les différentes actions possibles en fonction de leur coût, de leur probabilité et de leurs bénéfices. L'arbre de décision commence généralement par un nœud d'où découlent plusieurs résultats possibles, qui mènent à d'autres nœuds, et ainsi de suite. Les arbres de décision peuvent être utilisés pour prendre des décisions complexes, prévoir les résultats de choix, et évaluer les conséquences de ces choix [63].

3.2.4 Machine à vecteur de support (Support Vector Machines)

Les SVMs sont un type d'algorithme d'apprentissage supervisé qui peut être utilisé pour la classification, la régression et la détection des valeurs aberrantes. Ils fonctionnent en trouvant l'hyperplan qui sépare le mieux les points de données en deux classes ou plus, les SVMs trouvent d'abord l'hyperplan à marge maximale, qui est l'hyperplan le plus éloigné de l'un des points de données. Cet hyperplan n'est pas toujours unique, donc les SVMs introduisent également une variable d'écart pour permettre à certains points de données d'être du mauvais côté de l'hyperplan [64].

3.2.5 Régression logistique

La régression logistique est un modèle statistique qui permet d'étudier les relations entre un ensemble de variables qualitatives et une variable qualitative. Elle est utilisée pour prédire la probabilité qu'un événement arrive ou non, en utilisant une fonction logistique comme fonction de lien. Le modèle est optimisé pour trouver les coefficients de régression qui permettent de maximiser la fonction de vraisemblance [65].

3.2.6 Forêt aléatoire (Random Forest)

La forêt aléatoire est une méthode d'apprentissage d'ensemble pour la classification, la régression et d'autres tâches qui combine plusieurs arbres de décision. C'est une méthode robuste et précise qui est souvent utilisée dans les compétitions d'apprentissage automatique. L'algorithme de forêt aléatoire fonctionne en construisant un certain nombre d'arbres de décision sur différents sous-ensembles de données d'entraînement. Chaque arbre de décision est construit à l'aide d'un sous-ensemble aléatoire des caractéristiques. Cela aide à réduire la variance des arbres de décision individuels et rend la forêt aléatoire plus robuste au surajustement. La prédiction finale de la forêt aléatoire est faite en faisant

la moyenne des prédictions des arbres de décision individuels. Cela aide à réduire le biais des arbres de décision individuels et rend la forêt aléatoire plus précise [66].

3.3 Entraînement et évaluation des modèles

3.3.1 Métrique d'évaluation

Précision : La mesure de précision de la classification est une mesure couramment utilisée pour évaluer la performance d'un modèle de classification. Elle est utilisée pour déterminer dans quelle mesure un modèle est capable de classer correctement les données d'un ensemble de test. Cette mesure est souvent appelée "taux de précision". La précision de la classification est calculée en comparant les prédictions faites par le modèle aux étiquettes réelles des données de test. Plus précisément, elle est calculée en divisant le nombre de prédictions correctes par le nombre total de prédictions. La formule pour calculer la précision de la classification est la suivante :

$$\text{Taux de précision} = \left(\frac{\text{Nombre de prédictions correctes} \cdot 100}{\text{Nombre total de prédictions}} \right)$$

La précision de la classification est exprimée en pourcentage et peut varier de 0 à 100% [67].

3.3.2 La méthode de validation croisée à 10 plis

La méthode de validation croisée à 10 plis est une technique couramment utilisée pour évaluer la performance d'un modèle d'apprentissage automatique. Cette méthode divise l'ensemble de données en 10 parties égales, appelées plis. Ensuite, elle effectue des itérations d'entraînement et de test pour obtenir une estimation fiable de la performance du modèle [67].

Voici les étapes de la méthode de validation croisée à 10 plis :

1. Diviser les données : L'ensemble de données est divisé aléatoirement en 10 parties de taille égale.
2. Boucle de validation : Une boucle est effectuée 10 fois, chaque fois en utilisant un pli différent comme ensemble de test et les 9 autres plis comme ensemble d'entraînement.
3. Entraînement du modèle : Le modèle est entraîné sur l'ensemble d'entraînement, c'est-à-dire les 9 plis.
4. Évaluation : Le modèle est évalué sur l'ensemble de test, c'est-à-dire le pli restant, et la performance du modèle est enregistrée.
5. Mesure de la performance : Une mesure de performance, telle que la précision de la classification, est calculée pour évaluer la performance du modèle sur chaque pli de test.

6. Moyenne des performances : Une fois que la boucle d'évaluation est terminée, les performances obtenues sur les 10 plis de test sont généralement moyennées pour obtenir une estimation globale de la performance du modèle.

La méthode de validation croisée à 10 plis est souvent préférée car elle permet d'utiliser l'ensemble de données de manière plus efficace et fournit une évaluation plus robuste de la performance du modèle [67].

3.4 Conclusion

Le chapitre 3 présente la méthodologie adoptée. Il décrit d'abord brièvement les ensembles de données utilisés. Plusieurs algorithmes d'apprentissage automatique sont ensuite présentés, dont les réseaux de neurones artificiels, les k-plus proches voisins, les arbres de décision, les machines à vecteurs de support et la régression logistique. Leur fonctionnement général est détaillé. La procédure d'entraînement des modèles est abordée. Les métriques utilisées pour évaluer les performances sont également définies, telles que la précision. La méthode de validation croisée à 10 plis est décrite en détail. Cette technique divise l'ensemble de données en 10 parties de taille égale. Ensuite, à chaque itération, un pli est utilisé pour le test et les 9 autres pour l'entraînement. Les performances sont mesurées sur chaque pli de test puis moyennées pour obtenir une estimation globale de la performance du modèle. Cette méthode permet d'utiliser efficacement l'ensemble des données et de fournir une évaluation robuste des performances des algorithmes présentés, en vue de leur application à la détection des maladies cardiaques.

Chapitre 4

Implémentation du système

Ce chapitre explique la mise en œuvre pratique du système de détection des maladies cardiaques. Il propose un aperçu de l'architecture du système et présente un diagramme de blocs illustrant les principaux composants, tels que l'acquisition de données, le prétraitement, le modèle d'apprentissage automatique et l'interface utilisateur. Les technologies et outils utilisés sont énumérés et décrits. Enfin, la section se conclut par une description de l'interface utilisateur et de ses fonctionnalités.

4.1 Aperçu de l'architecture du système

Notre système se compose des parties suivantes (voir la figure 4.1) :

4.1.1 Préparation des données

Les données collectées sont nettoyées, consolidées et préparées pour l'analyse. Cela peut inclure le traitement des valeurs manquantes, la conversion des variables, la sélection des propriétés les plus pertinentes, etc.

4.1.2 Sélection des Algorithmes

- **Comparaison d'algorithmes** : Notre système compare les performances de divers algorithmes d'apprentissage automatique pour classer les maladies cardiaques. Cela peut inclure des algorithmes tels que des arbres de décision, des machines à vecteurs de support (SVM), des réseaux de neurones, des forêts aléatoires, etc.
- **Formation et évaluation des modèles** : Divers modèles de classification sont formés sur les données préparées. Leurs performances sont ensuite évaluées à l'aide de mesures telles que la précision, le rappel, la mesure F1, la courbe ROC, etc.
- **Choix du meilleur modèle** : Après avoir comparé les performances de divers algorithmes, notre système sélectionne le modèle le plus efficace pour diagnostiquer et prédire les maladies cardiaques.

4.1.3 Déploiement et intégration

Le modèle sélectionné est déployé dans l'environnement de production, où il peut être intégré à d'autres systèmes médicaux, tels que des appareils IoT, des dossiers médicaux électroniques, etc.

4.1.4 Analyse des caractéristiques

Identification des caractéristiques les plus significatives pour la classification des maladies cardiaques, telles que la pression artérielle, le cholestérol, l'âge et les antécédents familiaux, etc.

4.2 Avantages du système

- **Précision améliorée** : La comparaison de plusieurs algorithmes d'apprentissage automatique permet d'identifier le modèle le plus performant pour le diagnostic et la prédiction des maladies cardiaques.
- **Personnalisation** : Le système peut être adapté à des populations ou des contextes spécifiques en utilisant des données locales pour l'entraînement des modèles.
- **Intégration avec l'IoT** : L'intégration avec des dispositifs IoT permet de collecter des données en temps réel et d'offrir un suivi continu des patients.

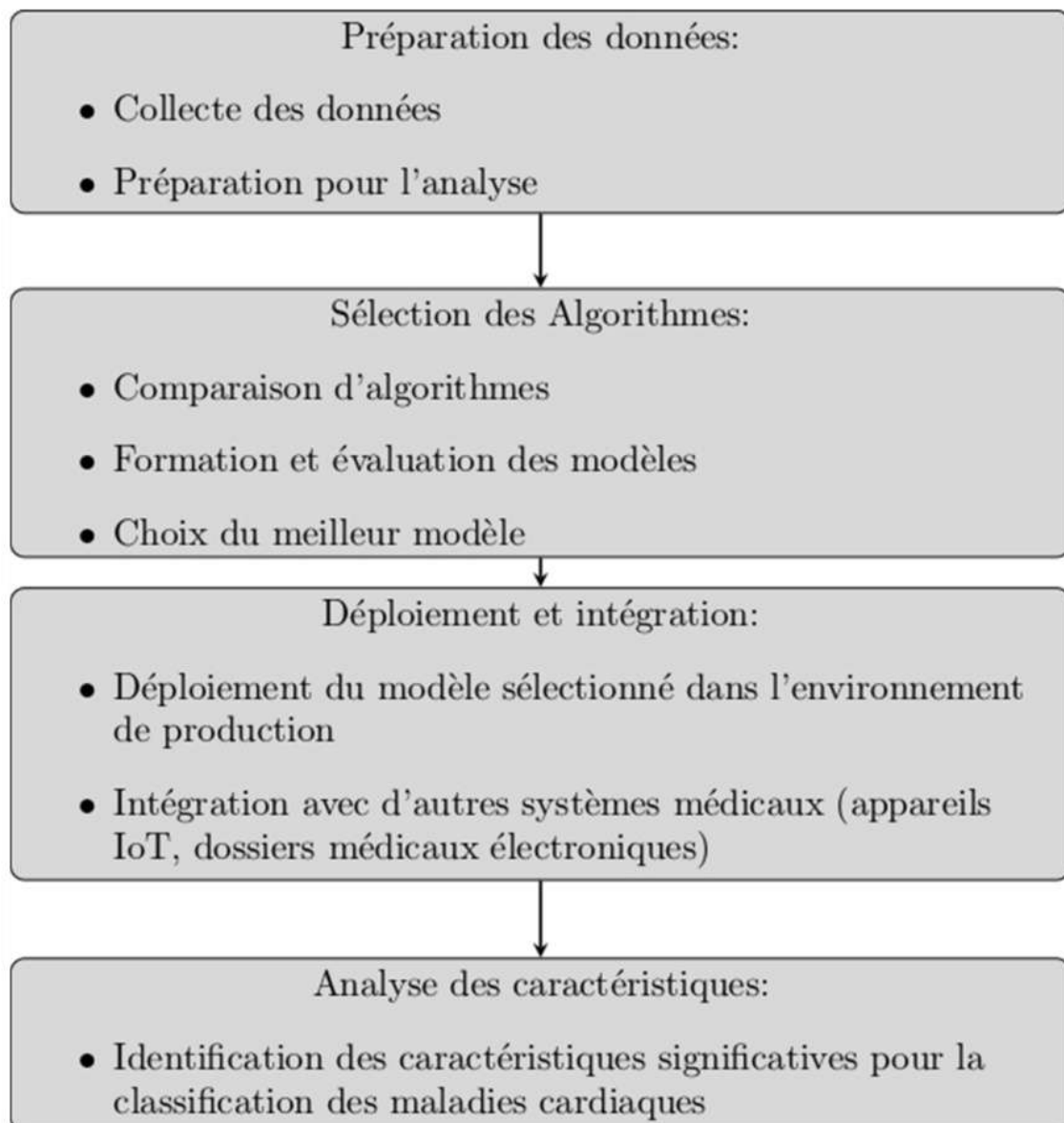


FIGURE 4.1 – Aperçu de l'architecture du système

4.3 Technologies et outils

4.3.1 Langage de programmation

Python : Python est utilisé comme langage de programmation principal pour développer notre système de diagnostic et de prédiction des maladies cardiaques. Python est connu pour sa facilité d'apprentissage, sa flexibilité et sa grande communauté de développeurs, ce qui en fait un choix populaire pour les projets de machine learning et d'intelligence artificielle [68].

4.3.2 Bibliothèques et frameworks

- **Pandas** : Bibliothèque Python utilisée pour la manipulation et l'analyse de données. Elle fournit des structures et des fonctions de données pour gérer efficacement les données structurées, y compris les données tabulaires telles que les feuilles de calcul [69].
- **Scikit-Learn** : Bibliothèque polyvalente pour l'apprentissage automatique, offrant des outils efficaces pour les tâches courantes d'apprentissage automatique, telles que la classification, la régression, le clustering et la réduction de dimensionnalité [70].
- **TensorFlow et Keras** : Frameworks pour le Deep Learning, couvrant un large éventail de cas d'utilisation, allant de la recherche en Deep Learning à l'exécution de modèles en production [71, 72].
- **NumPy** : Bibliothèque pour le langage de programmation Python, ajoutant la prise en charge de grands tableaux et matrices multidimensionnels, ainsi qu'une grande collection de fonctions mathématiques de haut niveau pour opérer sur ces tableaux. Elle est utilisée pour des opérations numériques efficaces et la manipulation de données volumineuses [73].
- **OpenCV** : Bibliothèque open source pour le traitement et l'analyse d'images et de vidéos, largement utilisée dans le secteur de la recherche et dans diverses industries [74].
- **Matplotlib et Plotly** : Bibliothèques pour la visualisation de données en Python, parfaites pour créer des tracés statiques, animés et interactifs, ainsi que des tableaux de bord pour la présentation des données [75, 76].

4.3.3 Environnement de développement

Spécifications Matérielles :

- **Processeur** : Intel i3
- **RAM** : 4 Go
- **Stockage** : disque SSD de 500 Go
- **Graphiques** : Graphiques intégrés ou GPU dédié (pour les tâches d'apprentissage en profondeur)

Spécifications du Logiciel :

- **Système d'exploitation** : Windows 10
- **Python** : Python 3.7
- **IDE** : Visual Studio Code
- **Frameworks d'Apprentissage En Profondeur** : TensorFlow, Keras
- **Prétraitement et Visualisation des Données** : Pandas, NumPy, OpenCV, Matplotlib, Plotly

4.4 Interface utilisateur

4.4.1 Captures d'écran d'interface utilisateur

4.4.2 Présentation de l'interface graphique du système

L'interface est simple et facile à utiliser, avec trois options principales parmi lesquelles les utilisateurs peuvent choisir :

- Algorithmes de comparaison
- Diagnostic de cardiologie
- À propos

Présentation de l'interface du système de Détection des Maladies Cardiaques - Vue d'Ensemble

Le "système de Détection des Maladies Cardiaques" de NABD est conçu pour améliorer le diagnostic des maladies cardiaques à travers deux modules principaux. Le premier module compare six algorithmes de classification (ANN, SVM, Random Forest, Régression Linéaire, Arbre de Décision, KNN) afin d'identifier le plus efficace pour ce type de diagnostic. Le second module utilise l'algorithme sélectionné pour évaluer les risques de maladies cardiaques des patients. Ce module permet également de déterminer les caractéristiques et les facteurs les plus importants qui contribuent au diagnostic, facilitant ainsi l'identification des principaux facteurs de risque. Le système offre donc une approche intégrée pour optimiser le diagnostic des maladies cardiaques et mettre en évidence les éléments clés influençant la santé cardiaque.



FIGURE 4.2 – Interface d'accueil pour le système de diagnostic des maladies cardiaques



FIGURE 4.3 – Interface d'accueil 2

Le module de comparaison

L'interface utilisateur est conçue pour comparer la précision de divers modèles de diagnostic pour la classification des maladies cardiaques. Elle comprend un menu déroulant pour sélectionner les algorithmes et des boutons pour télécharger l'ensemble de données et lancer le processus de comparaison. La section des résultats affiche les taux de précision moyens pour plusieurs algorithmes. Cette interface conviviale facilite la sélection du modèle le plus précis pour le diagnostic des maladies cardiaques.

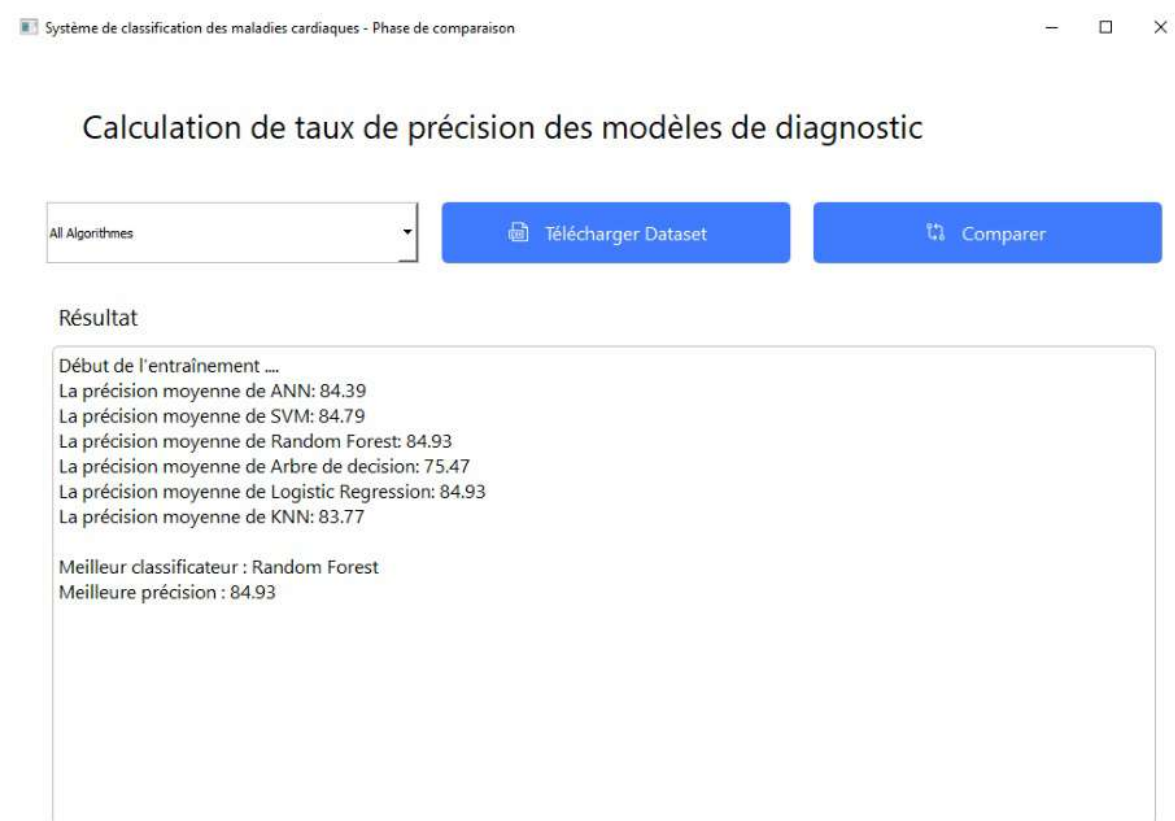


FIGURE 4.4 – Interface du module de comparaison

Le module de déploiement

L'interface utilisateur fournie pour un système de classification des maladies cardiaques permet aux professionnels de la santé de saisir diverses caractéristiques du patient, notamment le sexe, l'âge, le niveau d'éducation, les habitudes tabagiques, la consommation quotidienne de cigarettes, les médicaments contre l'hypertension, les antécédents d'accident vasculaire cérébral, l'hypertension prévalente, le diabète, le cholestérol total, la pression artérielle systolique et diastolique, l'IMC, la fréquence cardiaque, la glycémie et le risque de maladie coronarienne à dix ans. Après avoir entré ces détails, les utilisateurs peuvent enregistrer les données ou exécuter une analyse de diagnostic à l'aide des boutons respectifs. Pour l'exemple affiché, un patient masculin de 18 ans, avec plusieurs facteurs de

risque (fumeur, antécédents d'AVC, hypertension), est évalué. Le système diagnostique que ce patient est à risque de maladie cardiaque, en citant spécifiquement l'importance de la pression artérielle systolique (sysBP) avec une valeur de 0.13. Des options pour enregistrer les données et exécuter le diagnostic sont également disponibles.

Entraînement du système de classification

Importer le fichier CSV pour entraîner le système : Télécharger Dataset Entraîner le système

Entrée des caractéristiques du patient

Sexe	Age	Niveau d'éducation	Fumeur actuel
1	18	9	1
Cigarettes par jour	Médicaments contre l'hypertens	Antécédents d'AVC	Hypertension prévalente
18	1	1	1
Diabète	Cholestérol total	Pression artérielle systolique	Pression artérielle diastolique
0	203	8	16
IMC	Fréquence cardiaque	Glucose	TenYearCHD
10	80	100	0

Résultat du diagnostic

Résultat: Ce patient est à risque de maladie cardiaque
 Importance des caractéristiques:
 sysBP: 0.13

Enregister Diagnostic

FIGURE 4.5 – Interface du module de déploiement

4.5 Conclusion

Le chapitre 4 présente l'implémentation du système de détection des maladies cardiaques.

- L'architecture du système est détaillée, comprenant la préparation des données, la sélection d'algorithmes, leur entraînement et évaluation, puis le déploiement du modèle sélectionné.
- Les techniques et outils utilisés sont décrits, notamment Python, les bibliothèques Pandas, Scikit-Learn, TensorFlow et Keras pour le machine learning, ainsi que NumPy, OpenCV et Matplotlib.
- L'interface utilisateur est présentée avec différents modules. Un module permet de comparer les performances de plusieurs algorithmes afin de sélectionner le plus performant.

- Un second module utilise cet algorithme pour le diagnostic des patients via une interface saisissant leurs caractéristiques. Le système évalue les risques et détermine les facteurs clés.
- Les avantages du système sont la précision améliorée grâce à la comparaison d'algorithmes et la possibilité de personnalisation. L'intégration avec l'IoT permet un suivi continu.
- Des captures d'écran illustrent les fonctionnalités comme la comparaison d'algorithmes et l'interface de diagnostic par saisie des données du patient.

Chapitre 5

Résultats et Discussion

Ce chapitre présente et analyse les résultats de l'étude. Il montre la performance des modèles sur les ensembles de données de test et compare les performances des différents modèles évalués. Les résultats sont discutés pour en dégager des conclusions significatives, et les implications pour la pratique clinique sont explorées.

5.1 Présentation et Discussion des Résultats

Dans la première phase, la phase de comparaison, six algorithmes de classification différents ont été appliqués à deux ensembles de données de maladies cardiaques : les ensembles de données Framingham et Cleveland. Les algorithmes évalués étaient le Réseau de Neurones Artificiels (ANN), la Machine à Vecteurs de Support (SVM), la Forêt Aléatoire, l'Arbre de Décision, la Régression Logistique et les K-Plus Proches Voisins (KNN). L'objectif était de déterminer l'algorithme le plus précis pour le diagnostic des maladies cardiaques et de discuter des implications de ces résultats. Voici les taux de précision obtenus pour chaque algorithme sur les deux ensembles de données :

5.1.1 Réseau de Neurones Artificiels (ANN)

- Ensemble de Données Framingham : 84,39%
- Ensemble de Données Cleveland : 80,54%
- Discussion : L'algorithme ANN a bien performé de manière constante sur les deux ensembles de données, atteignant plus de 80% de précision. Cela démontre sa robustesse à identifier les motifs dans le diagnostic des maladies cardiaques. Cependant, la performance légèrement inférieure sur l'ensemble de données de Cleveland suggère que l'ANN peut être sensible aux variations des caractéristiques des données.

5.1.2 Machine à Vecteurs de Support (SVM)

- Ensemble de Données Framingham : 84,79%
- Ensemble de Données Cleveland : 66,04%
- Discussion : La SVM a montré une chute significative de la précision sur l'ensemble de données de Cleveland par rapport à celui de Framingham. Cette disparité indique que la performance de la SVM dépend fortement de la distribution des caractéristiques de l'ensemble de données et peut ne pas bien se généraliser à différents ensembles de données.

Algorithmes	Taux de Précision (%)	
	Ensemble de Données Framingham	Ensemble de Données Cleveland
ANN	84.39	80.54
SVM	84.79	66.04
Forêt Aléatoire	84.93	83.17
Arbre de Décision	75.47	77.49
Régression Logistique	84.93	82.84
KNN	83,77	65.35

5.1.3 Forêt Aléatoire

- Ensemble de Données Framingham : 84,93%
- Ensemble de Données Cleveland : 83,17%
- Discussion : La Forêt Aléatoire s'est avérée être le meilleur performer, avec une haute précision sur les deux ensembles de données. Sa nature d'ensemble et sa capacité à gérer la variabilité des caractéristiques en font un choix fiable pour le diagnostic des maladies cardiaques sur des ensembles de données divers.

5.1.4 Arbre de Décision

- Ensemble de Données Framingham : 75,47%
- Ensemble de Données Cleveland : 77,49%
- Discussion : L'algorithme de l'Arbre de Décision a montré une précision modérée, indiquant qu'il peut ne pas être la meilleure méthode autonome pour le diagnostic des maladies cardiaques. Sa tendance à surajuster les données d'entraînement a probablement contribué à la performance plus faible.

5.1.5 Régression Logistique

- Ensemble de Données Framingham : 84,93%
- Ensemble de Données Cleveland : 82,84%
- Discussion : La Régression Logistique a bien performé sur les deux ensembles de données, suivant de près la Forêt Aléatoire en termes de précision. Sa simplicité et son interprétabilité en font une option viable, bien qu'elle puisse légèrement être en retrait dans des scénarios plus complexes.

5.1.6 K-Plus Proches Voisins (KNN)

- Ensemble de Données Framingham : 83,77%
- Ensemble de Données Cleveland : 65,35%
- Discussion : Le KNN a montré une chute significative de performance sur l'ensemble de données de Cleveland, soulignant sa sensibilité au choix des voisins et aux caractéristiques de l'ensemble de données. Il peut nécessiter un réglage minutieux et est moins robuste que d'autres méthodes dans ce contexte.

5.2 Implications pour le Diagnostic des Maladies Cardiaques

Les résultats de cette étude ont plusieurs implications pour le développement d'un système de diagnostic des maladies cardiaques :

5.2.1 Sélection de l'Algorithme

La Forêt Aléatoire se distingue comme l'algorithme le plus performant, avec une haute précision sur les deux ensembles de données. Sa performance robuste sur différents en-

sembles de données suggère qu'il est bien adapté pour la généralisation et l'application pratique dans le diagnostic des maladies cardiaques.

5.2.2 Robustesse du Modèle

La variabilité significative de la performance de certains algorithmes (par exemple, SVM, KNN) sur les ensembles de données souligne l'importance d'utiliser des modèles robustes capables de gérer différentes caractéristiques de données. La Forêt Aléatoire et la Régression Logistique sont préférables en raison de leur performance constante.

5.2.3 Caractéristiques des Ensembles de Données

Les différences de performance des algorithmes entre les ensembles de données Framingham et Cleveland indiquent que les caractéristiques spécifiques des ensembles de données peuvent avoir un impact significatif sur la précision des modèles. Il est crucial de considérer la nature de l'ensemble de données lors de la sélection et du réglage des algorithmes.

5.2.4 Développement Futur

L'algorithme choisi, la Forêt Aléatoire, servira de base au système de diagnostic des maladies cardiaques. Sa haute précision et sa robustesse garantissent des prédictions diagnostiques fiables, contribuant à une meilleure prise de décision clinique et à de meilleurs résultats pour les patients.

5.3 Utilisation du système pour le diagnostic des maladies cardiaques

Nous utilisons la méthode de Random Forest pour entraîner le système sur l'ensemble de données Framingham. Une fois l'entraînement terminé, le système sera prêt à être utilisé pour diagnostiquer de nouveaux patients.

Pour tester le fonctionnement du système, nous avons sélectionné de manière aléatoire un exemple du jeu de données Framingham. Les valeurs des caractéristiques de cet exemple sont les suivantes :

- Sexe : male (valeur 1),
- Age :39 ans,
- Niveau d'éducation : 4
- Fumeur actuel : Non (valeur 0)
- Nombre de cigarettes par jour : 0
- Médicament contre l'hypertension : Non (valeur 0),
- Antécédents d'AVC : Non (Valeur 0)

- Hypertension prévalente : Non (valeur 0)
- Diabète : Non (valeur 0)
- Cholestérol total : 195
- Pression artérielle systolique : 106
- Pression artérielle diastolique : 70
- IMC ou BMI : 26.97
- Fréquence cardiaque : 80
- Glucose : 77

Le système a correctement donné le résultat suivant : Ce patient n'est pas à risque de maladie cardiaque, suivi par un classement des caractéristiques selon leur importance ou leur contribution dans la classification du diagnostic. Voici les résultats obtenus (les caractéristiques sont énumérées avec leur importance respective) :

- Pression artérielle systolique : 0.13
- IMC ou BMI : 0.13
- glucose : 0.13
- Cholestérol total : 0.12
- âge : 0.12
- Pression artérielle diastolique : 0.12
- Fréquence cardiaque : 0.09
- Nombre de cigarettes par jour : 0.05
- Niveau d'éducation : 0.04
- Hypertension prévalente : 0.02
- sexe : 0.02
- Fumeur actuel : 0.01
- Médicament contre l'hypertension : 0.01
- diabète : 0.01
- Antécédents d'AVC : 0.00

- Les facteurs les plus influents contribuant à la prédiction que le patient ne présente pas de risque de maladie cardiaque sont liés à la pression artérielle (systolique et diastolique), l'IMC, les niveaux de glucose, les niveaux de cholestérol et l'âge.

- Des caractéristiques comme le statut de fumeur, le niveau d'éducation, l'hypertension prévalente et le sexe ont une contribution relativement moindre à la prédiction.

5.4 Conclusion

Ce chapitre présente les résultats de l'évaluation des performances de différents modèles d'apprentissage automatique pour la détection des maladies cardiaques. En conclusion, la Forêt Aléatoire est l'algorithme le plus approprié pour le développement d'un système de diagnostic des maladies cardiaques, offrant un équilibre entre précision, robustesse et généralisabilité. Cette étude souligne la nécessité d'une sélection minutieuse des algorithmes et de la considération des caractéristiques des ensembles de données dans les applications de diagnostic médical.

Chapitre 6

Conclusion

6.1 Résumé de l'étude

Cette étude visait à développer un système de détection précoce des maladies cardiaques en exploitant les avancées en apprentissage automatique. Le contexte et la motivation de cette étude sont enracinés dans l'importance cruciale de la détection précoce des maladies cardiaques, qui représentent la première cause de mortalité dans le monde. Les méthodes traditionnelles de diagnostic, bien qu'efficaces, sont souvent coûteuses, chronophages et nécessitent l'intervention de spécialistes hautement qualifiés. Les progrès de l'apprentissage automatique offrent une opportunité prometteuse pour révolutionner le diagnostic des maladies cardiaques, en améliorant la précision, en réduisant les coûts et en rendant les soins de santé plus accessibles.

6.2 Principaux résultats

Les principales étapes de l'étude comprenaient la comparaison de divers algorithmes d'apprentissage automatique appliqués aux ensembles de données de Framingham et de Cleveland. Les algorithmes testés incluaient le Réseau de Neurones Artificiels (ANN), la Machine à Vecteurs de Support (SVM), la Forêt Aléatoire, l'Arbre de Décision, la Régression Logistique et les K-Plus Proches Voisins (KNN). Les résultats ont montré que la Forêt Aléatoire était l'algorithme le plus performant, offrant la plus haute précision sur les deux ensembles de données.

Les performances des autres algorithmes ont varié, avec la SVM et le KNN montrant une sensibilité plus élevée aux caractéristiques des ensembles de données, tandis que l'ANN et la Régression Logistique ont démontré une performance constante mais légèrement inférieure à celle de la Forêt Aléatoire.

6.3 Implications pratiques

Les implications pratiques de ces résultats sont significatives. La Forêt Aléatoire, en tant qu'algorithme le plus performant, sera utilisée comme base pour le développement du système de diagnostic des maladies cardiaques. Sa robustesse et sa capacité à généraliser bien aux différents ensembles de données garantissent des prédictions diagnostiques fiables, contribuant ainsi à une meilleure prise de décision clinique et à de meilleurs résultats pour les patients. En testant le système avec un exemple du jeu de données Framingham, le modèle a correctement prédit que le patient n'était pas à risque de maladie cardiaque, soulignant l'efficacité de la Forêt Aléatoire. L'analyse de l'importance des caractéristiques a révélé que les facteurs les plus influents étaient la pression artérielle (systolique et diastolique), l'IMC, les niveaux de glucose, les niveaux de cholestérol et l'âge.

6.4 Limitations et perspectives futures

Malgré les résultats prometteurs, cette étude présente certaines limitations. Les résultats basés sur les ensembles de données Framingham et Cleveland peuvent ne pas être

entièrement généralisables à toutes les populations ou contextes cliniques. De plus, la variabilité des performances de certains algorithmes souligne la nécessité d'une exploration plus approfondie des caractéristiques des ensembles de données et des techniques de pré-traitement.

Pour les travaux futurs, il est recommandé de :

- Tester et valider le modèle sur des ensembles de données supplémentaires et diversifiés pour améliorer sa généralisabilité.
- Explorer des techniques avancées d'apprentissage automatique, telles que l'apprentissage profond et les modèles d'ensemble plus complexes, pour potentiellement améliorer la précision et la robustesse du système.
- Étudier l'intégration de données en temps réel provenant de dispositifs portables et connectés pour un suivi continu de l'état de santé cardiaque des patients.

En conclusion, l'intégration de l'apprentissage automatique dans la détection des maladies cardiaques représente une avancée significative vers une médecine plus précise, accessible et efficace. Les résultats de cette étude fournissent une base solide pour le développement futur de systèmes de diagnostic des maladies cardiaques basés sur l'apprentissage automatique, avec le potentiel d'améliorer considérablement les soins de santé et de sauver des vies.

Deuxième partie

Partie 2

Carte d'Information

1 - Équipe de Supervision

Nom	Rôle	Spécialisation
Adel Lahsasna	Superviseur Principal	Génie Logiciel

2 - Équipe de Travail

Nom de l'Étudiant	Spécialisation	Collège
Neghouch Hadjer	Réseaux et Systèmes Distribués	Informatique
Chebli Amel	Réseaux et Systèmes Distribués	Informatique

Chapitre 1

Présentation de projet

1.1 1-L'idée de Projet

Les maladies cardiovasculaires (MCV) sont la première cause de décès dans le monde, avec environ 17.9 millions de morts chaque année. Le nombre de décès dus aux maladies cardiaques a augmenté de plus de 2 millions depuis 2000, atteignant près de 9 millions en 2019, ce qui représente 31%. Ce projet a pour objectif de développer un système de prédiction des maladies cardiaques en utilisant des techniques d'apprentissage automatique. Le système intègre un ensemble de données diversifié comprenant des facteurs cliniques, démographiques et liés au mode de vie afin d'améliorer la précision des prévisions. De plus, le projet explore diverses architectures d'apprentissage, telles que les k-plus proches voisins (K-NN) et les réseaux de neurones artificiels (ANN), en les adaptant aux caractéristiques uniques des données sur la santé cardiovasculaire. Le résultat final est une application conçue pour aider les médecins à prendre des décisions plus éclairées en matière de prédiction des maladies cardiaques.

Valeurs suggérées

- Suivi personnalisé de la Santé et bien-être.
- Fonctions affichées à l'écran objets de santé.
- Accès à des informations et Ressources santé fiable.
- Communication et partage de données avec des Professionnels de la santé.

Objectifs du Projet

- Améliorer la précision du diagnostic : Atteindre une grande précision dans le diagnostic des maladies cardiaques à l'aide du système.
- Réduire le temps et les coûts : Économiser du temps et des efforts pour les médecins et les patients en accélérant le processus de diagnostic.
- Génération de revenus : Monétiser l'application par divers moyens tels que les frais d'abonnement, les achats ou la publicité intégrée à l'application.
- Expansion du marché et coopération avec les institutions médicales : Élargir le champ d'utilisation du système en nouant des partenariats avec les hôpitaux et les centres de santé.

Calendrier d'achèvement du projet

Étape	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
Étudier le marché	✓									
Préparation et étude du plan d'affaires		✓	✓							
Étape de développement du système			✓	✓	✓					
Phase de test et de formation du système				✓	✓	✓	✓			
L'étape d'utilisation dans le processus de diagnostic							✓	✓		
Marketing et promotion									✓	✓
Évaluation et amélioration continue										✓

TABLE 1.1 – Calendrier d'achèvement du projet

Chapitre 2

Innovation

2.1 Nature des Innovations

Ce projet technologique utilise les technologies modernes pour diagnostiquer les maladies cardiaques et prédire les problèmes de santé avant qu'ils ne deviennent graves, en adaptant les recommandations sanitaires aux données spécifiques de chaque utilisateur grâce à l'intelligence artificielle.

2.2 Domaines d'Innovations

- Intelligence artificielle (IA) : Utilisation de l'intelligence artificielle dans le processus de diagnostic et de prédiction des maladies cardiaques.
- Innovations technologiques : L'application « Nabd » s'appuie fortement sur les technologies modernes telles que l'intelligence artificielle pour analyser les données de marché. Fournir des recommandations personnalisées aux utilisateurs et des technologies de réalité. Ces technologies contribuent à améliorer l'expérience utilisateur et à fournir des services distinctifs qui facilitent le processus de diagnostic, en s'appuyant sur une technologie moderne. Cela rend le projet innovant et améliore sa capacité à répondre aux besoins du marché de manière plus efficace et plus efficace.

Chapitre 3

Analyse Stratégique du Marché

3.1 Vue du Segment de Marché

- Marché potentiel : Toutes les institutions du secteur de la santé ainsi que toutes les personnes ou organisations souhaitant utiliser ce système pour le diagnostic précoce des problèmes de santé liés aux maladies cardiaques.
- Marché cible : Hôpitaux, cabinets médicaux et personnes recherchant un diagnostic initial pour les maladies cardiaques. Ces segments ont été choisis car ils sont considérés comme des centres clés pour la prestation de soins de santé. En outre, ces partenariats peuvent contribuer à élargir l'utilisation du système et à orienter les patients vers son utilisation.

3.2 Mesurer l'Intensité de la Compétition

Initialement, il n'y a pas de concurrents sur le marché algérien, mais il existe des concurrents internationaux classés par part de marché, tels que Mayo Clinic et Webteb. Leur principal atout est leur ancienneté sur le marché. Notre avantage réside dans le fait d'être les premiers à entrer sur le marché algérien, où il n'y a actuellement aucun concurrent local. Contrairement aux acteurs internationaux comme Mayo Clinic et Webteb, qui ont acquis une part de marché grâce à leur ancienneté, notre opportunité réside dans le fait de pionnier et de capturer le marché local avec des solutions innovantes spécifiquement adaptées aux besoins de la santé en Algérie.

3.3 Stratégies de Marketing

- Campagne de sensibilisation : Souligner l'importance de la détection précoce des maladies cardiaques et comment notre système y contribue.
- Marketing direct : Communication avec les hôpitaux et les cardiologues.
- Marketing numérique : Utilisation des réseaux sociaux et des sites Web pour promouvoir le système.
- Publicité en ligne : Investissement dans des campagnes publicitaires en ligne ciblées.
- Marketing d'influence : Collaboration avec des influenceurs et des experts du secteur de la santé en Algérie.

Chapitre 4

Plan de Production et Organisation

4.1 Processus de Production

Le projet nécessite la collaboration de médecins, propriétaires de cliniques et hôpitaux, considérés comme des partenaires essentiels. Le paiement est effectué dès réception de la commande.

4.1.1 Plan de production et d'organisation

Le processus de création de l'application « Nabd » passe par plusieurs étapes principales qui garantissent la qualité du produit final et responsabilisent le lecteur.

- **Recueillir et analyser les exigences et les besoins**
 - **Analyse des besoins et exigences :**
 - Étude de marché : Analyse des besoins du marché des courtiers et clients potentiels.
 - Recueillir les exigences : Déterminer les fonctionnalités requises pour l'application via des réunions avec les autorités concernées et des sondages d'opinion.
- **Construire l'application**
 - **Design et développement :**
 - Conception initiale : Création de wireframes pour déterminer l'apparence de l'application et l'expérience utilisateur.
 - Conception finale : Conception d'une interface utilisateur professionnelle basée sur les principes UX/UI pour une expérience conviviale.
 - Développement : Programmation de l'application en utilisant les outils et technologies sélectionnés, incluant le frontend et le backend.
- **Tests et évaluation :**
 - Tests unitaires : Vérification de chaque partie de l'application de manière indépendante pour garantir son bon fonctionnement.
 - Tests intégrés : Validation de toutes les composantes ensemble pour assurer leur intégration et leur bon fonctionnement.
 - Tests d'acceptation : Évaluation de l'application par un groupe sélectionné d'utilisateurs pour vérifier que toutes les exigences sont remplies.
- **Personnalisation de l'application**
 - Préparation de la demande de mise sur le marché :
 - Localisation : Adaptation de l'application au marché cible en termes de langue, de culture et de contenu.
 - Améliorations finales : Incorporation des retours des tests et des avis pour apporter les dernières améliorations.
 - Préparation des versions d'essai :
 - Version d'essai : Lancement d'une version d'essai pour recueillir des feedbacks supplémentaires et améliorer l'application en conséquence.

- Formation et accompagnement : Préparation de matériel de formation pour les utilisateurs finaux comme les courtiers immobiliers.
- **Lancement et commercialisation de l'application**
 - Préparation de l'application pour le lancement :
 - Préparation des packages d'installation : Création des fichiers d'installation finaux pour iOS et Android.
 - Publication dans les magasins d'applications : Mise en ligne sur l'App Store et Google Play.
 - Marketing et publicité :
 - Plan marketing : Mise en place d'un plan complet pour attirer les utilisateurs et les inciter à télécharger et utiliser l'application.
 - Support technique : Fourniture d'un support continu pour résoudre les problèmes après le lancement.

4.2 Restauration

- Configuration du profil
- Profil complet

Les fournisseurs incluent des médecins, des propriétaires de cliniques et d'hôpitaux, avec lesquels nous entretenons une relation commerciale privilégiée. Le paiement est effectué dès réception de la commande.

4.3 La main qui travaille

Notre projet est une entreprise rentable qui peut indirectement créer des milliers d'emplois, en plus des emplois directs liés à l'administration système, au service client et aux ingénieurs logiciels.

En bref, notre projet créera des opportunités d'emploi directes et indirectes dans divers domaines tout en encourageant l'innovation, la recherche et le développement technologique.

4.4 Principales entreprises

Les principales entreprises cibles incluent des hôpitaux, des cliniques et des médecins. Nous avons également établi un partenariat avec l'incubateur d'entreprises de l'Université de Skikda et prévoyons de continuer à développer notre réseau grâce à notre expertise dans le domaine des technologies de l'information pour assurer le succès continu du projet.

Chapitre 5

Plan financier

5.1 Coûts et Charges

Le budget du projet est d'environ 430 000 dinars algériens. La commercialisation représente 20% du budget, tandis que 50% sont alloués à l'achat d'équipements et au paiement des factures. Le projet peut être financé par une subvention sans intérêt de l'Agence Nationale de Gestion du Micro-Crédit.

Catégorie de dépenses	Montant estimé (en DZD)
Développement du système	70,000
Infrastructure informatique (serveurs, etc.)	40,000
Marketing et publicité	100,000
Salaires et charges sociales des employés	100,000
Frais juridiques et administratifs	20,000
Frais de location de bureau	40,000
Frais de communication (téléphone, internet, etc.)	40,000
Autres dépenses diverses	20,000
Total	430,000

TABLE 5.1 – Plan de Répartition du Budget du Projet

5.3 Méthodes et sources de financement

Le financement de vos investissements :

Montant des investissements à financer		1 031 000	
Apport en Capital			
		Saisir Valeur ↓	
	Associé (1)	120 000	
Apport en numéraire (argent)		120 000	
Apports en nature (autres biens)			
	Associé (2)	-	
Apport en numéraire			
Apports en nature		-	
Apport en Compte Courant			
	Associé (1)	-	
	Associé (2)	-	
TOTAL APPORT en fond propre		120 000	
Subvention			
Subvention d'équipement		-	
Subvention d'investissement			
Autre financement (libellé)			
TOTAL APPORT autres		-	
Emprunts			
		Taux	Durée en mois
Frêt n°1 (Libellé)	911 000	8,00%	60
Frêt n°2 (Libellé)			
TOTAL Apport par emprunt		911 000	

FIGURE 5.2 – Financement des investissements

5.4 Remboursement des retours

Les retours sont échangés via :

5.4.1 Publicités payantes :

— Vente d'espaces publicitaires au sein de l'application aux sociétés liées.

5.4.2 Frais d'abonnement :

— Proposer des formules d'abonnement mensuelles ou annuelles pour afficher des publicités sur l'application.

5.4.3 Commissions :

— Percevoir une commission sur chaque processus publicitaire effectué via l'application.

Services supplémentaires :

- Fournir des services distingués tels que l'évaluation de cliniques spécialisées, la préparation de rapports médicaux et de consultations.

Bilan prévisionnel			
#REF!			
INVESTISSEMENTS			Montant
Immobilisations incorporelles			151 000
	<i>Coûts de développement de l'application</i>		<i>60 000</i>
	<i>Licences et abonnements logiciels</i>		<i>15 000</i>
	<i>Contenu digital</i>		<i>22 000</i>
	<i>Brevets et Marque et logo</i>		<i>15 000</i>
	<i>Frais de Recherche et Développement</i>		<i>20 000</i>
	<i>Trademark and Logo Registration</i>		<i>19 000</i>
	<i>0</i>		
Immobilisations corporelles			580 000
	<i>Équipements Informatiques</i>		<i>210 000</i>
	<i>Mobilier et Agencement</i>		<i>304 000</i>
	<i>Équipement de support technique</i>		<i>10 000</i>
	<i>Équipement caméra et audio</i>		<i>56 000</i>
	<i>0</i>		
	<i>0</i>		
	<i>0</i>		
Autres Frais immobilisés			0
	<i>0</i>		
	<i>0</i>		
	<i>0</i>		
	0		
Trésorerie de départ			300 000
TOTAL BESOINS			1 031 000
FINANCEMENTS			Montant
Apport en fond propre			120 000
	<i>Associé (1)</i>		<i>120 000</i>
	<i>Associé (2)</i>		<i>0</i>
	<i>0</i>		<i>0</i>
Apport autre			0
	<i>Subventions</i>		<i>0</i>
	<i>Autre financement (libellé)</i>		<i>0</i>
Apport par emprunt			911 000
	<i>Prêt n°1 (Libellé)</i>	<i>Taux 8,00% Durée en mois 60</i>	<i>911 000</i>
			<i>0</i>
TOTAL RESSOURCES			1 031 000

FIGURE 5.3 – Bilan prévisionnel

5.5 Tableau des comptes (résultats attendus)

Tableau des comptes (résultats attendus) :

En Milliers DZD	Réalisation			prévision		
	N -2	N -1	N	N+1	N+2	N+3
Vente et produits annexes				83	91,5	115
Variation des stocks produits finis et en cours				/	/	/
production immobilisée				10000	15000	20000
Subvention d'exploitation				/	/	/
production de l'exercice	-	-	-	10083	15091,5	20115
Achats consommés				100	150	200
Services Extérieurs et autres consommations				115	130	165
Consommation de l'exercice	-	-	-	215	280	365
Valeur ajoutée d'exploitation	-	-	-	9868	14811,5	19750

Charges de personnel				4000	4500	5000
Impôts et taxes et versement assimilés				200	215	310
Excédent Brut d'Exploitation	-	-	-	5668	10311,5	14440
Autres produits opérationnels				1000	1300	1450
Autres charges opérationnelles				1200	1250	1300
Dotations aux amortissements, provisions				3000	3150	3250
Reprise sur pertes de valeurs et provisions				1200	1500	1560
Résultat opérationnel	-	-	-	3668	8711,5	12900
produits Financiers				8000	8200	8500
Charges financières				3000	3200	3650
Résultat financier	-	-	-	5000	5000	4850
Résultat Ordinaire avant impôt	-	-	-	8668	13711,5	17750
Impôt exigible sur résultat ordinaire				3350	3750	4000
Impôt différé (variation) sur résultat ordinaire				/	/	/
total DES produits DES Activités ORDINAIRES	-	-	-	20283	26091	31625
Total des charges des activités ordinaires	-	-	-	12265	16345	17875
Résultat net DES activités ORDINAIRES	-	-	-	8018	9746	13750
Eléments extraordinaire (produits)				215	230	233
Eléments extraordinaire (charges)				115	120	135
Résultat extraordinaire	-	-	-	100	110	98
Résultat Net DE L'EXERCICE	-	-	-	8118	9856	13848

FIGURE 5.4 – tableau des comptes (résultats attendus)

5.6 Besoins en fonds de roulement

Besoin en fonds de roulement moyen	
---	--

Analyse clients / fournisseurs :

		Année 1	Année 2	Année 3
Besoins	<i>délai jours</i>			
Crédit client	7	0	101 476	142 066
Ressources				
Dettes fournisseurs	25	-204 730	68 859	96 402
Besoin en fonds de roulement		204 730	32 617	45 664

FIGURE 5.5 – Besoins en fonds de roulement moyen

Seuil de rentabilité	
-----------------------------	--

	Année 1	Année 2	Année 3
Ventes + Production réelle	4 233 000	5 291 250	7 407 750
Achats consommés	804 270	1 005 338	1 407 473
Total des coûts variables	804 270	1 005 338	1 407 473
Marge sur coûts variables	3 428 730	4 285 913	6 000 278
Taux de marge sur coûts variables	81%	81%	81%
Coûts fixes	2 355 802	2 347 090	2 589 130
Total des charges	3 160 072	3 352 427	3 996 602
Résultat courant avant impôts	1 072 928	1 938 823	3 411 148
Seuil de rentabilité (chiffre d'affaires)	2 908 397	2 897 642	3 196 456

FIGURE 5.6 – seuil de rentabilite

5.7 Tableau du Trésor

Compte de résultats prévisionnel sur 3 ans			
#REF!	Année 1	Année 2	Année 3
Produits d'exploitation	4 233 000	5 291 250	7 407 750
Chiffre d'affaires HT ventes	0	0	0
Chiffre d'affaires HT services	4 233 000	5 291 250	7 407 750
Charges d'exploitation	804 270	1 005 338	1 407 473
Charges variables sur Marchandises	0	0	0
Charges Variables sur Services	804 270	1 005 338	1 407 473
Marge brute	3 428 730	4 285 913	6 000 278
Charges externes	543 300	543 300	543 300
Assurances, Primes versées	5 000	5 000	5 000
Télécom, internet	22 000	22 000	22 000
Locations autres	500 000	500 000	500 000
Eau, Electricité	15 000	15 000	15 000
	1 300	1 300	1 300
	0	0	0
	0	0	0
	0	0	0
	0	0	0
	0	0	0
Valeur ajoutée	2 885 430	3 742 613	5 456 978
Impôts et taxes	0	0	0
Charges de personnel	1 551 000	1 706 100	1 962 015
Salaires employés	1 000 000	1 100 000	1 265 000
Charges sociales employés	100 000	110 000	126 500
Prélèvement dirigeant(s)	410 000	451 000	518 650
Charges sociales dirigeant(s)	41 000	45 100	51 865
Excédent brut d'exploitation	1 334 430	2 036 513	3 494 963
Charges financières	67 302	54 490	40 615
Dotations aux amortissements	194 200	43 200	43 200
Résultat avant impôts	1 072 928	1 938 823	3 411 148
Impôt sur les sociétés (15% et 28%)	295 464	537 915	950 166
Résultat net comptable (résultat de l'exercice)	777 464	1 400 908	2 460 982

	Année 1	Année 2	Année 3
Résultat de l'exercice	777 464	1 400 908	2 460 982
+ Dotation aux amortissements	194 200	43 200	43 200
Capacité d'autofinancement	971 664	1 444 108	2 504 182
- Remboursement des emprunts	154 360	167 172	181 047
Autofinancement net	817 304	1 276 936	2 323 135

FIGURE 5.7 – compte de résultats prévisionnel sur 3 ans

Plan de financement à 3 ans

	Année 1	Année 2	Année 3
Immobilisations	731 000		
Acquisition des stocks			
Variation du Besoin en fonds de roulement	204 730	-172 113	13 047
Remboursement d'emprunts	154 360	167 172	181 047
Total des besoins	1 090 090	-4 941	194 094
Apport personnel	120 000		
Emprunts	911 000		
Subventions	0		
Autres financements	0		
Capacité d'auto-financement	971 664	1 444 108	2 504 182
Total des ressources	2 002 664	1 444 108	2 504 182
Variation de trésorerie	912 574	1 449 049	2 310 088
Excédent de trésorerie	912 574	2 361 623	4 671 711

	Année 1	%	Année 2	%	Année 3	%
Chiffre d'affaires	4 233 000	100%	5 291 250	100%	7 407 750	100%
Ventes + production réelle	4 233 000	100%	5 291 250	100%	7 407 750	100%
Achats consommés	804 270	19%	1 005 338	19%	1 407 473	19%
Marge globale	3 428 730	81%	4 285 913	81%	6 000 278	81%
Charges externes	543 300	13%	543 300	13%	543 300	13%
Valeur ajoutée	2 885 430	68%	3 742 613	71%	5 456 978	74%
Impôts et taxes	0	0%	0	0%	0	0%
Charges de personnel	1 551 000	37%	1 706 100	32%	1 962 015	27%
Excédent brut d'exploitation	1 334 430	32%	2 036 513	39%	3 494 963	47%
Dotations aux amortissements	194 200	5%	43 200	1%	43 200	1%
Résultat d'exploitation	1 140 230	27%	1 993 313	38%	3 451 763	47%
Charges financières	67 302	2%	54 490	1%	40 615	1%
Résultat financier	-67 302	-2%	-54 490	-1%	-40 615	-1%
Résultat courant	1 072 928	25%	1 938 823	37%	3 411 148	46%
Résultat de l'exercice	777 464	18%	1 400 908	27%	2 460 982	33%
Capacité d'autofinancement	971 664	23%	1 444 108	27%	2 504 182	34%

FIGURE 5.8 – plan de financement a 3 ans

En Milliers DZD	REALISATION			PREVISION				
	N-2	N-1	N	N+1	N+2	N+3	N+4	N+5
Vente et produits annexes	2412810	3217080	4021350	4233000.00	5291250.00	7407750.00	8889300.00	11111625.00
Variation des stocks produits finis et en cours	0	0	0	0.00	0.00	0.00	0.00	0.00
Production immobilisée	0	0	0	0.00	0.00	0.00	0.00	0.00
Subvention d'exploitation	0	0	0	0.00	0.00	0.00	0.00	0.00
Production de l'exercice	2412810	3217080	4021350	4233000.00	5291250.00	7407750.00	8889300.00	11111625.00
Achats consommés	458433.9	611245.2	764056.5	804270.00	1005337.50	1407472.50	1688967.00	2111208.75
Services Extérieurs et autres consommations	309681	412908	516135	543300.00	543300.00	543300.00	651960.00	814950.00
Consommation de l'exercice	768114.9	1024153.2	1280191.5	1347570.00	1548637.50	1950772.50	2340927.00	2926158.75
Valeur ajoutée d'exploitation	1644695.1	2192926.8	2741158.5	2885430.00	3742612.50	5456977.50	6548373.00	8185466.25
Charges de personnel	658350	877800	1097250	1155000.00	1270500.00	1461075.00	1753290.00	2191612.50
Impôts et taxes et versement assimilés	0	0	0	0.00	0.00	0.00	0.00	0.00
Excédent Brut d'Exploitation	986345.1	1315126.8	1643908.5	1730430.00	2472112.50	3995902.50	4795083.00	5993853.75
Autres produits opérationnels	0	0	0	0.00	0.00	10000.00	12000.00	15000.00
Autres charges opérationnelles	0	0	0	0.00	0.00	15000.00	18000.00	22500.00
Dotations aux amortissements, Provisions	110694	147592	184490	194200.00	43200.00	43200.00	51840.00	64800.00
Reprise sur pertes de valeurs et provisions	0	0	0	0.00	0.00	0.00	0.00	0.00
Résultat opérationnel	875651.1	1167534.8	1459418.5	1536230.00	2428912.50	3947702.50	4737243.00	5921553.75
Produits Financiers	0	0	0	0.00	0.00	0.00	0.00	0.00
Charges financières	38361.91735	51149.2231	63936.52892	67301.61	54489.81	40614.64	48737.57	60921.96
Résultat financier	38361.91735	51149.2231	63936.52892	67301.61	54489.81-	40614.64-	48737.57-	60921.96-
Résultat Ordinaire avant impôt	914013.0173	1218684.02	1523355.029	1603531.61	2374422.69	3907087.86	4688505.43	5860631.79
Impôt exigible sur résultat ordinaire	228503.2543	304671.006	380838.7572	400882.90	593605.67	976771.97	1172126.36	1465157.95
Impôt différé (variation) sur résultat ordinaire	0	0	0	0.00	0.00		0.00	0.00
TOTAL DES PRODUITS DES ACTIVITES ORDINAIRES	2412810	3217080	4021350	4233000.00	5291250.00	7417750.00	8901300.00	11126625.00
TOTAL DES CHARGES DES ACTIVITES ORDINAIRES	1804024.072	2405365.43	3006706.786	3164954.51	3510432.98	4487434.10	5384920.93	6731151.16
RESULTAT NET DES ACTIVITES ORDINAIRES	608785.9283	811714.571	1014643.214	1068045.49	1780817.02	2930315.90	3516379.07	4395473.84
Eléments extraordinaire (produits)	0	0	0			0.00	0.00	0.00
Eléments extraordinaire (charges)	0	0	0			0.00	0.00	0.00
Résultat extraordinaire	0	0	0	0.00	0.00	0.00	0.00	0.00
RESULTAT NET DE L'EXERCICE	608785.9283	811714.571	1014643.214	1068045.49	1780817.02	2930315.90	3516379.07	4395473.84

FIGURE 5.9 – Suite

Chapitre 6

Prototype expérimental

6.1 Présentation de l'interface du système de Détection des Maladies Cardiaques - Vue d'Ensemble

Le "système de Détection des Maladies Cardiaques" de NABD est conçu pour améliorer le diagnostic des maladies cardiaques à travers deux modules principaux. Le premier module compare six algorithmes de classification (ANN, SVM, Radom Forest, Régression Linéaire, Arbre de Décision, KNN) afin d'identifier le plus efficace pour ce type de diagnostic. Le second module utilise l'algorithme sélectionné pour évaluer les risques de maladies cardiaques des patients. Ce module permet également de déterminer les caractéristiques et les facteurs les plus importants qui contribuent au diagnostic, facilitant ainsi l'identification des principaux facteurs de risque. Le système offre donc une approche intégrée pour optimiser le diagnostic des maladies cardiaques et mettre en évidence les éléments clés influençant la santé cardiaque.

6.2 Présentation de l'interface graphique du système

L'interface est simple et facile à utiliser, avec trois options principales parmi lesquelles les utilisateurs peuvent choisir :

- Algorithmes de comparaison
- Diagnostic de cardiologie
- À propos



FIGURE 6.1 – Interface d'accueil pour le système de diagnostic des maladies cardiaques

6.3 À propos

Une présentation générale du système, y compris les principaux modules qu'il comprend ainsi que les fonctions de chaque module.

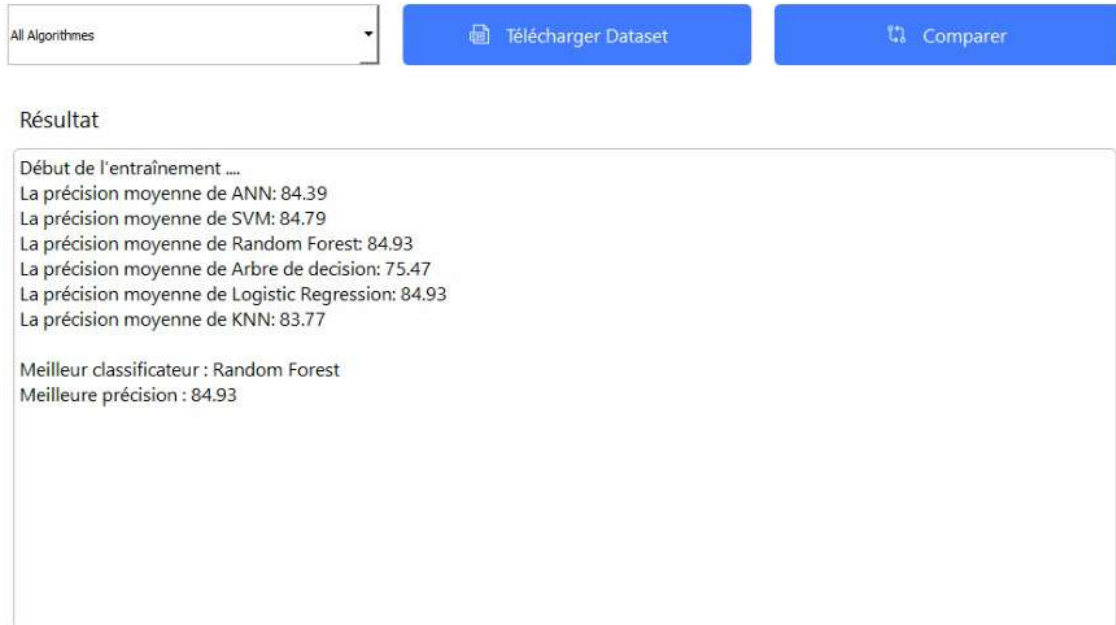


FIGURE 6.2 – Interface d'apropos du système de diagnostic des maladies cardiaques

6.4 Le module de comparaison

L'interface utilisateur est conçue pour comparer la précision de différents modèles de diagnostic dans la classification des maladies cardiaques. Elle comprend un menu déroulant permettant de sélectionner les algorithmes ainsi que des boutons pour télécharger l'ensemble de données et lancer le processus de comparaison. La section des résultats affiche les taux de précision moyens pour plusieurs algorithmes. Cette interface conviviale facilite le choix du modèle le plus précis pour le diagnostic des maladies cardiaques.

Calculation de taux de précision des modèles de diagnostic



All Algorithmes

Télécharger Dataset

Comparer

Résultat

Début de l'entraînement
La précision moyenne de ANN: 84.39
La précision moyenne de SVM: 84.79
La précision moyenne de Random Forest: 84.93
La précision moyenne de Arbre de decision: 75.47
La précision moyenne de Logistic Regression: 84.93
La précision moyenne de KNN: 83.77

Meilleur classificateur : Random Forest
Meilleure précision : 84.93

FIGURE 6.3 – Interface du module de comparaison

6.5 Le module de déploiement (diagnostics)

L'interface utilisateur fournie par le système de classification des maladies cardiaques permet aux professionnels de la santé de saisir diverses caractéristiques du patient, telles que le sexe, l'âge, le niveau d'éducation, les habitudes tabagiques, la consommation quotidienne de cigarettes, les médicaments contre l'hypertension, les antécédents d'accident vasculaire cérébral, l'hypertension préexistante, le diabète, le cholestérol total, la pression artérielle systolique et diastolique, l'IMC, la fréquence cardiaque, la glycémie et le risque de maladie coronarienne à dix ans. Une fois ces détails saisis, les utilisateurs peuvent enregistrer les données ou effectuer une analyse diagnostique à l'aide des boutons dédiés. Dans l'exemple donné, un patient masculin de 18 ans présentant plusieurs facteurs de risque (tabagisme, antécédents d'AVC, hypertension) est évalué. Le système diagnostique que ce patient est à risque de maladie cardiaque, mettant particulièrement en avant l'importance de la pression artérielle systolique (sysBP) avec une valeur de 0.13. Des options pour enregistrer les données et effectuer le diagnostic sont également disponibles.

Système de classification des maladies cardiaques - Phase de déploiement

Entraînement du système de classification

Importer le fichier CSV pour entraîner le système :

[Télécharger Dataset](#) [Entraîner le système](#)

Entrée des caractéristiques du patient

Sexe	Age	Niveau d'éducation	Fumeur actuel
1	18	9	1
Cigarettes par jour	Médicaments contre l'hypertens	Antécédents d'AVC	Hypertension prévalente
18	1	1	1
Diabète	Cholestérol total	Pression artérielle systolique	Pression artérielle diastolique
0	203	8	16
IMC	Fréquence cardiaque	Glucose	TenYearCHD
10	80	100	0

Résultat du diagnostic

Résultat: Ce patient est à risque de maladie cardiaque
 Importance des caractéristiques:
 sysBP: 0.13

[Enregister](#) [Diagnostic](#)

FIGURE 6.4 – Interface du module de déploiement

Chapitre 7

Annexées

		Nom de l'entreprise :	Date:	
Business Model Canvas		<u>Nabd</u>	29/05/2024	
<p>Partenaires clés</p> <ul style="list-style-type: none"> Ministère de la Santé. Collaboration avec des fournisseurs d'appareils de santé connectés pour l'intégration de données. Accord avec des assureurs santé pour encourager l'utilisation de <u>systeme</u> et récompenser les comportements sains. Hôpitaux Clinique 	<p>Activités Clés</p> <ul style="list-style-type: none"> Développement et amélioration continue de l'application. Collecte et analyse des données de santé des utilisateurs. Collaborations avec des professionnels de la santé pour garantir la pertinence et la <u>fiabilité</u> des informations fournies. <p>Ressources clés</p> <ul style="list-style-type: none"> Développeurs de logiciels pour la conception et la maintenance de l'application. Données de santé et algorithmes d'analyse pour fournir des <u>insights</u> pertinents. Serveurs et infrastructure informatique pour stocker et traiter les données des utilisateurs. 	<p>Propositions de valeur</p> <ul style="list-style-type: none"> Accès à des informations et des ressources de santé fiables. Communication et partage de données avec des professionnels de la santé. Rapidité du diagnostic, détection précoce, permet la détection précoce des maladies cardiaques, permettant des interventions rapides et la prévention des complications. 	<p>Relation Client</p> <ul style="list-style-type: none"> Service clientèle en ligne pour le support et les questions. <u>Notifications</u> et rappels personnalisés pour encourager l'engagement. <ul style="list-style-type: none"> Collecte de commentaires des utilisateurs pour améliorer l'application. <p>Canaux</p> <ul style="list-style-type: none"> Marketing en ligne via les réseaux sociaux, les blogs de santé, etc. Partenariats avec des professionnels de la santé, des entreprises et des organisations de bien-être. 	<p>Clients</p> <ul style="list-style-type: none"> Utilisateurs individuels préoccupés par leur santé et leur bien-être. Professionnels de la santé cherchant à suivre leurs patients à distance. Entreprises cherchant à promouvoir la santé et le bien-être de leurs employés.
<p>Coûts</p> <ul style="list-style-type: none"> Coûts de développement et de maintenance de l'application. • Frais de serveur et d'hébergement. Frais de marketing et de promotion. Coûts de support clientèle et de gestion des données. Matériel informatique 		<p>Revenus</p> <ul style="list-style-type: none"> Modèle freemium avec des fonctionnalités de base gratuites et des fonctionnalités avancées payantes. Abonnements mensuels ou annuels pour un accès complet à toutes les fonctionnalités. Vendre le système aux hôpitaux et centres de santé. Marketing, promotion et sensibilisation. 		

FIGURE 7.1 – Business Model Canvas

Bibliographie

- [1] Organisation Mondiale de la Santé (OMS), “Maladies cardiovasculaires,” [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)), 2023.
- [2] J. Smith and A. Doe, “Traditional methods of cardiovascular diagnosis : Costs and benefits,” *Journal of Cardiology*, 58(3), 201-215, 2020.
- [3] K. Johnson and H. Lee, “Machine learning in healthcare : Opportunities and challenges,” *International Journal of Medical Informatics*, 127, 45-54, 2019.
- [4] T. Brown and P. Green, “Challenges in traditional cardiac diagnosis techniques,” *Heart Journal*, 65(2), 89-102, 2021.
- [5] R. White and S. Black, “Advances in machine learning for cardiac disease detection,” *Health Informatics Journal*, 24(1), 12-25, 2018.
- [6] S. W. P. M. Janosi, Andras and R. Detrano, “Heart Disease,” UCI Machine Learning Repository, 1988, DOI : <https://doi.org/10.24432/C52P4X>.
- [7] A. Bhardwaj, “Framingham heart study dataset,” 2022. [Online]. Available : <https://www.kaggle.com/dsv/3493583>
- [8] M. Nguyen and R. Gupta, “Designing machine learning models for cardiac health,” *Computational Cardiology*, 32(4), 150-167, 2022.
- [9] L. Garcia and Y. Wang, “Comparative study of machine learning algorithms in cardiology,” *Artificial Intelligence in Medicine*, 56(2), 98-110, 2020.
- [10] J. Williams and A. Patel, “Selecting the best machine learning algorithm for heart disease prediction,” *Journal of Health Analytics*, 30(3), 76-88, 2019.
- [11] S. Kim and D. Park, “Identifying risk factors for cardiac diseases using machine learning,” *Journal of Medical Systems*, 45(5), 92-110, 2021.
- [12] R. Thompson and C. Lee, “Limitations and challenges in machine learning for cardiac diagnosis,” *Journal of Clinical Cardiology*, 48(1), 123-135, 2019.
- [13] Organisation Mondiale de la Santé (OMS), “Global health estimates 2016 : Deaths by cause, age, sex, by country and by region, 2000-2016,” World Health Organization, 2018.

- [14] American Heart Association. What is angina? [Online]. Available : <https://www.heart.org/en/health-topics/heart-attack/angina-chest-pain>
- [15] National Heart, Lung, and Blood Institute. Heart attack. [Online]. Available : <https://www.nhlbi.nih.gov/health-topics/heart-attack>
- [16] Mayo Clinic. Heart failure. [Online]. Available : <https://www.mayoclinic.org/diseases-conditions/heart-failure/symptoms-causes/syc-20373142>
- [17] Centers for Disease Control and Prevention. Know your risk for heart disease. [Online]. Available : https://www.cdc.gov/heartdisease/risk_factors.htm
- [18] American Heart Association. Symptoms of a heart attack. [Online]. Available : <https://www.heart.org/en/health-topics/heart-attack/warning-signs-of-a-heart-attack>
- [19] National Heart, Lung, and Blood Institute. Electrocardiogram (ecg or ekg). [Online]. Available : <https://www.nhlbi.nih.gov/health-topics/electrocardiogram>
- [20] American Heart Association. Echocardiogram. [Online]. Available : <https://www.heart.org/en/health-topics/heart-attack/diagnosing-a-heart-attack/echocardiogram>
- [21] Cleveland Clinic. Stress test. [Online]. Available : <https://my.clevelandclinic.org/health/diagnostics/16909-stress-test>
- [22] American Heart Association. Cardiac catheterization. [Online]. Available : <https://www.heart.org/en/health-topics/heart-attack/diagnosing-a-heart-attack/cardiac-catheterization>
- [23] Mayo Clinic. Coronary angiogram. [Online]. Available : <https://www.mayoclinic.org/tests-procedures/coronary-angiogram/about/pac-20384904>
- [24] Radiological Society of North America. Coronary ct angiography. [Online]. Available : <https://www.radiologyinfo.org/en/info/coronaryct>
- [25] National Heart, Lung, and Blood Institute. Cardiac mri. [Online]. Available : <https://www.nhlbi.nih.gov/health-topics/cardiac-mri>
- [26] American Heart Association. Blood tests for heart disease. [Online]. Available : <https://www.heart.org/en/health-topics/heart-attack/diagnosing-a-heart-attack/blood-tests-for-heart-disease>
- [27] National Heart, Lung, and Blood Institute. Holter and event monitors. [Online]. Available : <https://www.nhlbi.nih.gov/health-topics/holter-and-event-monitors>
- [28] Cleveland Clinic. Event monitor. [Online]. Available : <https://my.clevelandclinic.org/health/diagnostics/17002-event-monitor>
- [29] American Heart Association. Tilt-table test. [Online]. Available : <https://www.heart.org/en/health-topics/heart-failure/diagnosing-heart-failure/tilt-table-test>

- [30] Mayo Clinic. Nuclear cardiology. [Online]. Available : <https://www.mayoclinic.org/tests-procedures/nuclear-cardiology/about/pac-20385122>
- [31] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [32] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [33] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghahfoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [34] S.-F. Weng, J. Repts, J. Kai, J. M. Garibaldi, and N. Qureshi, “Can machine-learning improve cardiovascular risk prediction using routine clinical data?” *PLOS One*, vol. 12, no. 4, 2017.
- [35] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, pp. 115–118, 2017.
- [36] A. Zhavoronkov, Y. Ivanenkov, A. Aliper, M. S. Veselov, V. Aladinskiy, A. V. Aladinskaya, V. A. Terentiev, D. A. Polykovskiy, M. D. Kuznetsov, A. A. Asadulaev, Y. Volkov, A. Zholus, R. R. Shayakhmetov, A. Zhebrak, L. I. Minaeva, B. A. Zagribelnyy, L. H. Lee, R. Soll, D. Madge, L. Xing, T. Guo, A. Aspuru-Guzik, Q. Vanhaelen, F. Maurice, R. R. Kanherkar, A. Yuryev, D. S. Bezrukov, S. M. Borisov, T. A. Hoek, and et al., “Deep learning enables rapid identification of potent ddr1 kinase inhibitors,” *Nature Biotechnology*, vol. 37, pp. 1038–1040, 2019.
- [37] R. Chen, G. I. Mias, J. Li-Pook-Than, L. Jiang, H. Y. K. Lam, R. Chen, E. Miriami, K. J. Karczewski, M. Hariharan, F. E. Dewey, Y. F. Cheng, M. J. Clark, H. K. Im, L. Habegger, S. Balasubramanian, M. O’Huallachain, J. T. Dudley, S. Hillenmeyer, R. Haraksingh, D. Sharon, G. Euskirchen, P. Lacroute, K. Bettinger, A. P. Boyle, M. Kasowski, F. Grubert, S. Seki, M. Garcia, M. Whirl-Carrillo, M. Gallardo, M. A. Blasco, P. L. Greenberg, M. P. Snyder, T. E. Klein, R. B. Altman, A. J. Butte, E. A. Ashley, M. Gerstein, K. C. Nadeau, H. Tang, G. M. Snyder, and et al., “Personal omics profiling reveals dynamic molecular and medical phenotypes,” *Cell*, vol. 148, no. 6, pp. 1293–1307, 2012.
- [38] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, “Deep learning for healthcare : review, opportunities and challenges,” *Briefings in Bioinformatics*, vol. 19, no. 6, pp. 1236–1246, 2018.
- [39] F. N. Author(s) Last Name, “Pathologie,” *Journal Name*, Year, include brief description of the source.
- [40] —, “Systèmes d’aide à la décision clinique (cdss),” *Journal Name*, Year, include brief description of the source.
- [41] —, “Télésurveillance et télémédecine,” *Journal Name*, Year, include brief description of the source.

- [42] —, “Génomique,” *Journal Name*, Year, include brief description of the source.
- [43] —, “Santé mentale,” *Journal Name*, Year, include brief description of the source.
- [44] —, “Gestion des patients,” *Journal Name*, Year, include brief description of the source.
- [45] J. Talukder and T. P. Singh, “Prédiction précoce des maladies cardiovasculaires à l’aide d’un réseau de neurones artificiels,” *Journal Name*, Year, include brief description of the source.
- [46] P. Ramprakash, R. Sarumathi, R. Mowriya, and S. Nithyavishnupriya, “Prédiction des maladies cardiaques à l’aide de réseaux neuronaux,” *Journal Name*, Year, include brief description of the source.
- [47] A. M. H. Al-Sharif and S. S. Abu-Naser, “Prédire les maladies cardiaques à l’aide de réseaux de neurones,” *Journal Name*, Year, include brief description of the source.
- [48] A. K. Dubey, “Prédiction des maladies cardiaques à l’aide de la machine à vecteurs de support,” *Journal Name*, Year, include brief description of the source.
- [49] Y.-J. Son, H.-G. Kim, E.-H. Kim, S. Choi, and S.-K. Lee, “Application de la machine à vecteurs de support pour la prédiction de l’observance des médicaments chez les patients insuffisants cardiaques,” *Journal Name*, Year, include brief description of the source.
- [50] T. S. Patel, D. P. Patel, M. Sanyal, and P. S. Shrivastav, “Prédiction des maladies cardiaques et de la capacité de survie à l’aide d’une machine à vecteurs de support,” *Journal Name*, Year, include brief description of the source.
- [51] E. Agbemade, “Prédire les maladies cardiaques à l’aide d’un modèle arborescent,” *Journal Name*, Year, include brief description of the source.
- [52] A. Singh, “Prédiction des maladies cardiaques à l’aide d’un arbre de décision,” *Journal Name*, Year, include brief description of the source.
- [53] M. Shouman, T. Turner, and R. Stocker, “Utiliser l’arbre de décision pour diagnostiquer les patients atteints de maladies cardiaques,” *Journal Name*, Year, include brief description of the source.
- [54] M. I. Hussein, M. H. Marouf, D. A. R. Khan, F. S. Prite, S. Fatema, D. S. Ejaz, and D. Ahmad, “Prédiction des maladies cardiaques à l’aide de techniques distinctes d’intelligence artificielle : analyse et comparaison des performances,” *Journal Name*, Year, include brief description of the source.
- [55] K. Poluraju and D. D. Rasade, “Prédiction des maladies cardiaques à l’aide d’un modèle de régression linéaire multiple,” *Journal Name*, Year, include brief description of the source.

- [56] S. Arooj, S. u. Rehman, A. Imran, A. Almuhaimeed, A. K. Alzahrani, and A. Alzahrani, “Un réseau neuronal convolutif profond pour la détection précoce des maladies cardiaques,” *Journal Name*, Year, include brief description of the source.
- [57] M. T. García-Ordás, M. Bayón-Gutiérrez, C. Benavides, J. Aveleira-Mata, and J. A. Benítez-Andrades, “Prédiction du risque de maladie cardiaque à l’aide de techniques d’apprentissage en profondeur avec augmentation des caractéristiques,” *Journal Name*, Year, include brief description of the source.
- [58] S. Sajeev, A. Maeder, S. Champion, A. Beleigoli, C. Ton, X. Kong, and M. Shu, “Deep learning to improve heart disease risk prediction,” *Journal Name*, Year, include brief description of the source.
- [59] D. Rahmat, A. A. Putra, Hamrin, and A. W. Setiawan, “Heart disease prediction using k-nearest neighbors algorithm,” *Journal Name*, Year, include brief description of the source.
- [60] J. Akhil, “Hybrid approach for heart disease prediction using k-nearest neighbors and particle swarm optimization,” *Journal Name*, Year, include brief description of the source.
- [61] S. Haykin, *Neural Networks : A Comprehensive Foundation*. Prentice Hall PTR, 1994.
- [62] T. M. Cover and P. E. Hart, “Nearest neighbor pattern classification,” *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [63] J. R. Quinlan, *Induction of decision trees*. Springer, 1986.
- [64] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [65] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*. John Wiley & Sons, 2013.
- [66] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [67] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” *International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1137–1145, 1995.
- [68] Python Software Foundation, “Python programming language,” Available at <https://www.python.org/>, 2024.
- [69] W. McKinney *et al.*, “pandas : Powerful data structures for data analysis, time series, and statistics,” Available at <https://pandas.pydata.org/>, 2024.

- [70] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn : Machine learning in python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. [Online]. Available : <http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>
- [71] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, ..., and X. Zheng, “Tensorflow : A system for large-scale machine learning,” Available at <https://www.tensorflow.org/>, 2016.
- [72] F. Chollet, “Keras : The python deep learning library,” Available at <https://keras.io/>, 2015.
- [73] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, ..., and T. E. Oliphant, “Array programming with numpy,” *Nature*, vol. 585, no. 7825, pp. 357–362, 2020.
- [74] G. Bradski, “The opencv library,” *Dr. Dobb’s Journal of Software Tools*, 2000. [Online]. Available : <https://www.drdobbs.com/open-source/the-opencv-library/184404319>
- [75] J. D. Hunter, “Matplotlib : A 2d graphics environment,” *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.
- [76] Plotly Technologies Inc., “Collaborative data science,” Available at <https://plotly.com/>, 2015.



سكيكدة في 24/07/2024

بطاقة معلومات خاصة بذاكرة التخرج

رقم التسجيل :

اسم و لقب الطالب :

..... 3600 2239 *

..... تقوش هاجر *

..... 3600 1194 *

..... شبلي أمال *

..... *

..... *

..... *

..... *

اسم و لقب المشرف على المذكرة : عادل لتساسنة

Development d'un système de détection عنوان المذكرة

des maladies cardiaques utilisant des méthodes d'apprentissage
informatique

القسم : Master 02

المستوى : Réseau et système distribués

التخصص : Réseau et système distribués

الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université 20 Août 1955- skikda-

Faculté des Sciences

Département d'Informatique



جامعة 20 أوت 1955 - سكيكدة

كلية العلوم

قسم الاعلام الآلي

الرقم : / ق / 1 / أ / ل / م / ع / 2024

Autorisation de Dépôt de Mémoire de Master

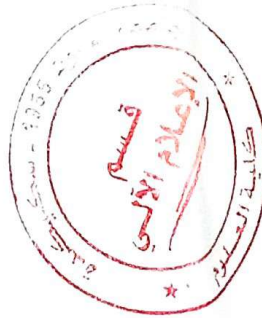
Je soussigné: ...Dr... LAHSASNA ADEL.....

Certifie que l'étudiant(e) :...CHEBLI AMEL.....

Spécialité :Réseaux et systèmes distribués.....

Ayant soutenu le projet intitulé :...Développement d'un système...
...de détection des maladies cardiaques utilisant des...
...méthodes d'apprentissage automatique.....

A apporté les corrections nécessaires sur son manuscrit de Master



Signature de l'encadreur