

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE  
Ministère de L'Enseignement Supérieur et de la Recherche Scientifique  
Université Abderrahmane Mira de Béjaïa  
Faculté des sciences exactes  
Département d'Informatique



**MÉMOIRE DE FIN DE CYCLE  
POUR L'OBTENTION DU DIPLÔME DE MASTER EN  
INFORMATIQUE**

**Domaine** : Mathématiques-informatique

**Filière** : Systèmes informatiques

**Spécialité** : Intelligence artificielle

**Réalisé par :**

Mlle Benali Celine

et

Mlle Mouhoubi Lyna

**Thème**

**L'Internet des objets et le machine learning appliqués  
au suivi des patients diabétiques**

Soutenue, le 19/06/2023, devant le jury composé de :

Mme. AIT ABDELOUHAB Karima	Maître de conférences B	Encadrante
Mme. AZOUI Aicha	Maître assistante A	Co-encadrante
Mme. HOCINI Kenza	Maître de conférences B	Présidente
Mme. LAHLAH Souad	Maître de conférences B	Examinatrice

**Année universitaire 2022-2023**

Nous dédions ce travail à :

Nos chers parents, qui nous ont inculqué les valeurs de l'assiduité et du travail acharné. Nous les remercions pour leur soutien et leur amour inconditionnel.

A nos amis pour leur soutien indéfectible

Enfin, nous souhaitons adresser un remerciement spécial à notre amie Djida pour ses nombreux encouragements à notre égard ainsi que pour son aide précieuse lors de l'élaboration de ce travail.

# Remerciement

Tout d'abord, nous tenons à remercier chaleureusement notre encadrante Mme AIT ABDELOUHAB Karima, MCB à l'université de BEJAIA ainsi que notre co-encadrante AZOUI Aicha, MAA à l'université de BEJAIA, pour leurs précieux conseils, et leur disponibilité sans faille tout au long de la réalisation de ce travail.

Nous remercions également toutes les personnes ayant contribué de près ou de loin à l'élaboration de ce mémoire.

# Table des matières

- Table des figures i
  
- Liste des tableaux iii
  
- Liste des abréviations iv
  
- Introduction Générale 1
  
- 1 Généralités sur l’AIoT 4**
  - 1.1 Introduction . . . . . 4
  - 1.2 l’IoT . . . . . 5
  - 1.3 Qu’est-ce que le cloud . . . . . 6
    - 1.3.1 Modèles de déploiement du cloud . . . . . 6
    - 1.3.2 Types de cloud . . . . . 8
    - 1.3.3 Le cloud et l’IA . . . . . 9
  - 1.4 Définition de l’IA . . . . . 10
    - 1.4.1 Le Machine Learning . . . . . 11
    - 1.4.2 Le Deep Learning . . . . . 14
    - 1.4.3 Différences entre le Machine Learning et le Deep Learning . . . . . 19
  - 1.5 Conclusion . . . . . 20
  
- 2 État de l’art 21**
  - 2.1 Introduction . . . . . 21
  - 2.2 Travaux connexes . . . . . 22
    - 2.2.1 Une architecture intelligente pour la surveillance des patients diabétiques à l’aide d’algorithmes d’apprentissage automatique : . . . 22
    - 2.2.2 Progrès récents dans la détection du cancer à l’aide de l’apprentissage automatique, étude systématique des décennies, comparaisons et défis : . . . . . 24

2.2.3	Une étude du diagnostic des maladies à l'aide de l'apprentissage automatique :	29
2.2.4	Modèles de détection des maladies cardiaques basés sur la technologie d'apprentissage automatique :	31
2.2.5	Une application de l'apprentissage automatique au diagnostic hématologique :	33
2.3	Étude comparative	35
2.4	Conclusion	39
<b>3</b>	<b>Proposition d'une architecture pour le suivi des patients diabétiques</b>	<b>40</b>
3.1	Introduction	40
3.2	Motivations	40
3.3	Statistiques	41
3.4	Définition du diabète	42
3.4.1	Symptômes et diagnostic du diabète	43
3.5	Qu'est-ce que l'hémoglobine glyquée?	45
3.5.1	Pourquoi l'hémoglobine glyquée?	45
3.5.2	L'importance de l'HbA1c dans le suivi du diabète	45
3.5.3	Calcul de l'HbA1c :	46
3.6	Architecture proposée	47
3.7	Les algorithmes utilisés :	49
3.8	Déroulement de l'architecture	54
3.9	Conclusion	56
<b>4</b>	<b>Implémentation d'un système basé sur une architecture pour le suivi des patients diabétiques</b>	<b>57</b>
4.1	Introduction	57
4.2	Environnement de développement	57
4.2.1	Anaconda	58
4.2.2	Spyder	59
4.2.3	kodular	60
4.2.4	Firestore	61
4.2.5	Langage python	62
4.3	Implémentation et résultats	64
4.3.1	Présentation du prototype de l'application mobile	64
4.3.2	Description du dataset	72
4.4	Résultats et évaluation	74

4.4.1	Classification des données . . . . .	74
4.4.2	Interprétation des résultats . . . . .	76
4.5	Conclusion . . . . .	78
	<b>Conclusion générale et perspectives</b>	<b>79</b>
	<b>Bibliographie</b>	<b>81</b>
	<b>Résumé</b>	<b>1</b>

# Table des figures

1.1	Domaines d'application de l'IoT . . . . .	5
1.2	L'architecture du cloud . . . . .	6
1.3	Modèles de déploiement du cloud . . . . .	7
1.4	Les différents services Cloud . . . . .	9
1.5	Rencontre entre l'IA et le Cloud . . . . .	10
1.6	Les domaines résultant de l'Intelligence artificielle . . . . .	11
1.7	Les différents types de Machine Learning . . . . .	14
1.8	Architecture d'un réseau de neurones artificiel . . . . .	15
1.9	Architecture d'un CNN . . . . .	16
1.10	Architecture d'un RNN . . . . .	17
1.11	Processus de fonctionnement d'un GAN . . . . .	18
1.12	la distinction fondamentale entre le machine learning et le deep learning . . . . .	20
2.1	Architecture proposée pour le suivi des patients diabétiques . . . . .	23
2.2	Valeurs d'exactitude, de précision, de rappel et du score F1. . . . .	23
2.3	Système assisté par machine pour la détection du cancer . . . . .	24
2.4	Phases d'entraînement pour la prédiction en utilisant le Deep Learning . . . . .	25
2.5	Architecture MCNN pour la détection du cancer du poumon . . . . .	26
2.6	Cadre d'extraction de fonctionnalités artisanales et CNN pour le cancer de la peau . . . . .	27
2.7	Processus d'extraction avec CNN . . . . .	28
2.8	Architecture Deep Convolutional Framework (DCNN) pour la détection du cancer du sein . . . . .	29
2.9	Liste des algorithmes machine learning utilisés. . . . .	30
2.10	Évaluation des algorithmes de machine learning. . . . .	31
2.11	Évaluation des algorithmes de machine learning. . . . .	32
2.12	Analyse comparative de différentes méthodes de machine learning. . . . .	32

2.13	Représentation schématique du processus de l’algorithme Smart Blood Analytics (SBA).	34
2.14	Comparaison des résultats de prédiction des maladies hématologiques obtenus par les algorithmes de prédiction SBA avec ceux obtenus par des spécialistes en termes de précision.	35
3.1	Statistiques prévisionnelles du nombre de sujets diabétiques sur une durée de 30 ans	41
3.2	Prévalence du diabète selon l’âge et le sexe au niveau de la Wilaya d’Alger (2017)	42
3.3	Carte de contrôle de l’hémoglobine glyquée	47
3.4	Tableau de comparaison du A1C et du eAG	47
3.5	Architecture globale du système	48
3.6	Pseudo-code explicatif de l’algorithme de calcul de l’HbA1c	49
3.7	Fonctionnement de l’algorithme K-Nearest Neighbors	50
3.8	Schéma explicatif du fonctionnement de l’algorithme Decision Tree	51
3.9	Fonctionnement de l’algorithme Random Forest	53
3.10	Schéma décrivant l’ensemble des étapes détaillées du fonctionnement du système	55
4.1	Page d’accueil du Navigateur Anaconda	59
4.2	Interface d’accueil de l’environnement de développement Spyder	60
4.3	Page d’accueil de la plateforme en ligne KODULAR	61
4.4	Page d’accueil de la plateforme Firebase	62
4.5	Écran de démarrage de l’application	65
4.6	Interface de connexion à l’application	66
4.7	Interface d’inscription à l’application	67
4.8	Fiche personnelle	68
4.9	Le menu principal	69
4.10	Le menu latéral	70
4.11	Le menu latéral	71
4.12	Historique des glycémies et des symptômes	72
4.13	Aperçu du dataset	73
4.14	Aperçu du dataset avec les valeurs de l’HbA1c	74
4.15	Résultats de la classification	75
4.16	Aperçu du dataset avec les valeurs de l’HbA1c	76
4.17	Histogrammes des performances des algorithmes utilisés	77

# Liste des tableaux

1.1	Différences entre Machine Learning et Deep Learning [1]	19
2.1	Tableau comparatif des travaux	36
3.1	Avantages et inconvénients de l'algorithme KNN	51
3.2	Avantages et inconvénients de l'algorithme arbre de décision	52
3.3	Avantages et inconvénients de l'algorithme Forêt aléatoire	54
4.1	Performances des algorithmes utilisés selon les quatre critères Accuracy, Precision, Recall et F1 Score	76

# Liste des abréviations

AIOT	l'intelligence artificielle des objets (Artificial Intelligence of Things)
CNN	Réseaux neuronaux convolutifs (Convolutional Neural Network)
CRM	gestion de la relation client (Customer relationship management)
DP	Apprentissage profond (Deep learning)
DSS	Système de support de décision
FaaS	Fonction en tant que service (function as a Service)
GA	algorithme génétique
GAN	Réseaux adversariaux génératifs (Generative adversarial network)
GWT	transformée en ondelettes de Gabor
HbA1c	Hémoglobine glyquée (glycated haemoglobin)
HOG	gradient d'orientation d'histogramme
IaaS	Infrastructure en tant que service (Infrastructure as a Service)
IA	Intelligence artificielle (Artificial intelligence)
IHM	l'interaction homme-machine
IoT	l'internet des objets (Internet of Things)
KNN	K plus proches voisins
LSTM	Réseaux de mémoire à long terme et à court terme (Long short-term memory)
ML	Apprentissage automatique (machine learning)
MLP	Perceptrons multicouches (Multilayer perceptron)
NB	Naive bayes

PaaS	Plateforme en tant que service (Platform as a Service)
PCA	analyse en composantes principales
ReLU	Unité linéaire rectifiée (Rectified Linear Unit)
RNN	Réseaux neuronaux récurrents (recurrent neural networks)
SaaS	Logiciel en tant que service (Software as a Service)
SMO	optimisation minimale séquentielle(Sequential minimal optimization)
SVD	Support vector machine

# Introduction Générale

Aujourd'hui et pour longtemps encore, l'intelligence artificielle (IA) est indispensable à notre quotidien. Si bien que grand nombre d'entre nous en font usage sans même s'en rendre compte. En effet, depuis le développement de cette technologie révolutionnaire, de nombreux domaines ont connu une progression autrefois inespérée, nous épargnant ainsi l'accomplissement de nombreuses tâches qui, en plus d'être éprouvantes, sont terriblement chronophages.

Comme nous le savons, l'intelligence artificielle connaît constamment des changements durant cette dernière décennie. Avec ce changement, naquirent de nouveaux concepts et technologies accessibles à tous issues de l'union de cette dernière avec d'autres techniques dont la plus intéressante reste sans nul doute, l'intelligence artificielle des objets (AIOT). En plus d'être plus rapide et plus efficace que l'internet des objets (IoT) ou l'IA séparés, cette innovation technologique constitue une parfaite combinaison entre l'analyse de données et l'interaction homme-machine (IHM) allant jusqu'à mimer à la perfection le fonctionnement du cerveau humain dans la prise de décisions, et ce, dans de nombreux secteurs facilitant ainsi, grand nombre de métiers et professions.

La compatibilité et flexibilité de l'AIoT avec un grand nombre de domaines est inexorablement un atout de taille pour l'homme. Cependant, si l'on s'intéresse de plus près à ses domaines d'application, nous ne pouvons que constater que le domaine de la santé reste le domaine de prédilection de l'AIoT. Aujourd'hui, il est désormais possible d'obtenir des conseils concernant notre santé personnelle sans avoir à consulter de spécialistes. Pour cela, il suffit d'entrer ses données de santé via un objet connecté (montre, smartphone, tablette) pour obtenir de nombreux services : alimentation à suivre selon sa situation (surpoids, sous-poids, maintien...), hygiène de vie (la pratique d'une activité physique régulière), sans oublier la fonctionnalité la plus intéressante : obtenir un diagnostic à partir des données entrées par un utilisateur grâce à des algorithmes d'intelligence artificielle qui traite les informations de la même façon que l'aurait fait un professionnel de la santé.

Cette nouveauté a totalement chamboulé le monde de la médecine au point où certains médecins s'appuient sur ces mêmes algorithmes afin de confirmer leurs diagnostics et s'assurer de leur fiabilité.

Malgré la réticence qu'elles ont provoquées, ces nouvelles méthodes sont toujours d'actualité et connaissent un succès grandissant. Néanmoins, des questions subsistent : Ces applications sont-elles applicables dans le cadre du suivi de patients atteints de maladies chroniques ? Si oui, contribuent-elles efficacement à l'amélioration des conditions de leurs suivis ?

Afin de répondre à ces interrogations, nous proposons dans le cadre de ce travail, une architecture basée sur la collecte de données, plus précisément de la glycémie capillaire, pour le suivi des patients diabétiques via une application mobile spécialement conçue à cet effet et installée préalablement sur smartphone. Les données collectées propres aux patients seront communiquées par la suite à leurs médecins traitants. La nouveauté réside dans le fait que cette donnée soit habituellement obtenue par le biais d'analyses sanguines effectuées tous les trois mois en laboratoire, pour connaître l'équilibre glycémique du patient sur le long terme. Nous offrons donc aux patients diabétiques par l'intermédiaire de ce projet la possibilité d'obtenir leur score HbA1c (Hémoglobine glyquée) directement sur leurs smartphones sans avoir à faire de prise de sang en laboratoire. L'interprétation des données constitue l'aspect le plus important de notre travail. En effet, grâce aux algorithmes de machine learning (ML) que nous avons utilisé (Random Forest, K-Nearest Neighbors, Decision Tree), nous avons été capables d'interpréter les résultats obtenus lors du calcul de l'hémoglobine glyquée et de les classer en trois catégories distinctes à partir d'intervalles définis préalablement (entrés en paramètres). En réalisant cette classification, nous effectuons nos premiers pas dans la prédiction, qui consiste à prédire la classe à laquelle appartient une donnée.

## Organisation du mémoire

Pour l'organisation de notre mémoire, nous avons suivi la démarche suivante :

**Chapitre 01 :** Ce chapitre comporte les définitions et notions de bases en termes d'IoT, de technologie Cloud ainsi que les différents algorithmes d'intelligence artificielle résultants des sous-domaines de l'IA : le machine learning et le deep learning.

**Chapitre 02 :** Il concerne la présentation de l'état de l'art où sont décrits cinq travaux dont la thématique était l'utilisation de divers algorithmes de machine learning dans la détection et le diagnostic des maladies, suivi d'une étude comparative des résultats obtenus en matière de performances.

**Chapitre 03 :** Cette partie comprend la définition du diabète suivi de ses différents types et symptômes, la notion d'hémoglobine glyquée est aussi définie ainsi que son importance dans le suivi du diabète. Une étude comparative des algorithmes dont nous avons fait usage est également comprise dans ce chapitre. Pour finir, la proposition de notre architecture pour le suivi des patients diabétiques est présentée.

**Chapitre 04 :** Ce quatrième et dernier chapitre a pour contenu la description des différents environnements de développement matériels et logiciels utilisés, mais surtout l'aspect "données" qui représente le corps de notre travail (présentation du dataset, pré-traitement, résultats et performances obtenus).

Pour finir, nous achèverons ce chapitre par une conclusion bilan de notre travail, comprenant les différentes perspectives que notre travail implique.

# Chapitre 1

## Généralités sur l'AIoT

### 1.1 Introduction

De nos jours, l'intelligence artificielle est considérée comme un tournant décisif dans le monde de l'informatique et du numérique. Utilisée dans plusieurs domaines, elle est à l'origine de nombreuses nouvelles technologies révolutionnaires, dont une en particulier qui a grandement impacté l'économie mondiale ainsi que l'expérience utilisateur : l'intelligence artificielle des objets.

L'intelligence artificielle et l'IOT sont omniprésents dans notre quotidien et nous facilitent bien des tâches. En effet, ces technologies ayant été développées durant de nombreuses années offrent déjà une vaste panoplie de gadgets dont la complexité varie, allant de simples montres connectées jusqu'aux maisons intelligentes.

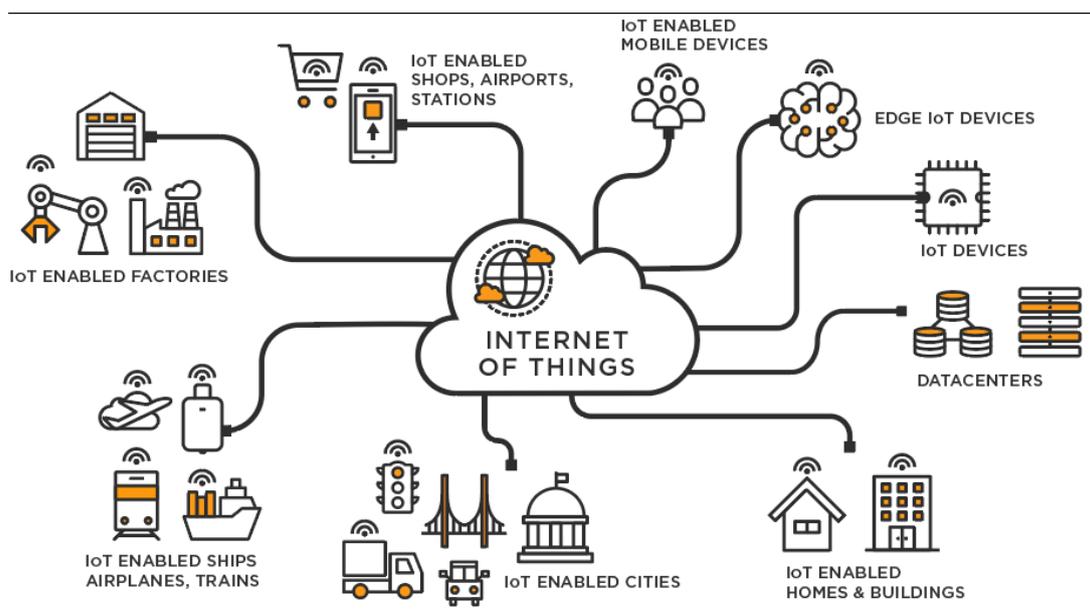
Séparément, l'intelligence artificielle et l'internet des objets sont des technologies qui améliorent considérablement nos vies. Ensemble, elles offrent encore plus de puissance aux applications afin de créer un environnement technologique de pointe. Cette union est alors appelée l'intelligence artificielle des objets (AIoT).

Durant ce chapitre, nous aborderons en détails l'ensemble des éléments nécessaires au développement de l'intelligence artificielle des objets (AIoT) notamment l'intelligence artificielle ainsi que l'internet des objets. La technologie cloud sera également exposée. Pour finir, une section sera consacrée au machine learning et au deep learning (DP ), les domaines les plus connus de l'intelligence artificielle.

## 1.2 l'IoT

L'Internet des objets (ou IoT pour Internet of Things) désigne à la fois le processus de connexion d'objets physiques à Internet et le réseau qui relie ces objets [2].

l'IoT relie tout élément capable de transférer des données sur un réseau. Et, ce sans nécessiter d'interactions entre humains ou entre un humain et un ordinateur. Cependant, l'interaction personne-machine est rendue possible afin de procéder au paramétrage, à la configuration ou simplement pour accéder aux informations. Notons également, qu'un objet connecté doit pouvoir communiquer et transmettre des informations avec d'autres objets connectés. Cette communication de "machine à machine" se fait de façon autonome sans l'intervention d'humains. Les objets récoltent donc de nombreuses données avec l'aide de capteurs, toutes ces informations sont généralement stockées dans le Cloud qui permet ensuite de les manipuler et de les sécuriser [3] comme illustré dans la Figure. 1.1.



**Figure 1.1:** Domaines d'application de l'IoT [4]

L'IoT génère une quantité de données qui exerce une pression énorme sur l'infrastructure Internet. Par conséquent, les entreprises s'efforcent de trouver des moyens d'alléger cette pression et de résoudre le problème des données. Le cloud computing y joue un rôle majeur, notamment en faisant fonctionner tous les appareils connectés. Ajoutons que l'infrastructure en cloud est une solution architecturale adaptée pour l'IoT. En effet, l'IoT bénéficiera des capacités et des ressources illimitées et de l'évolutivité du cloud computing.

## 1.3 Qu'est-ce que le cloud

De plus en plus utilisé par les entreprises de toutes les industries, le cloud computing est la nouvelle forme de stockage de données du 21ème siècle. Le cloud computing est un terme général employé pour désigner la livraison de ressources et de services à la demande par internet. Il désigne le stockage et l'accès aux données par l'intermédiaire d'internet plutôt que via le disque dur d'un ordinateur. Il s'oppose ainsi à la notion de stockage local, consistant à entreposer des données ou à lancer des programmes depuis le disque dur [5]. De manière générale, on parle de cloud computing lorsqu'il est possible d'accéder à des données ou à des programmes depuis internet, ou tout du moins lorsque ces données sont synchronisées avec d'autres informations sur internet. Il suffit donc pour y accéder de bénéficier d'une connexion internet.

Le nom « cloud » ou nuage en français fait référence à l'interconnexion entre de multiples appareils situés à différents emplacements. La forme de ce vaste réseau rappelle celle d'un nuage [6] comme le montre la Figure. 1.2.

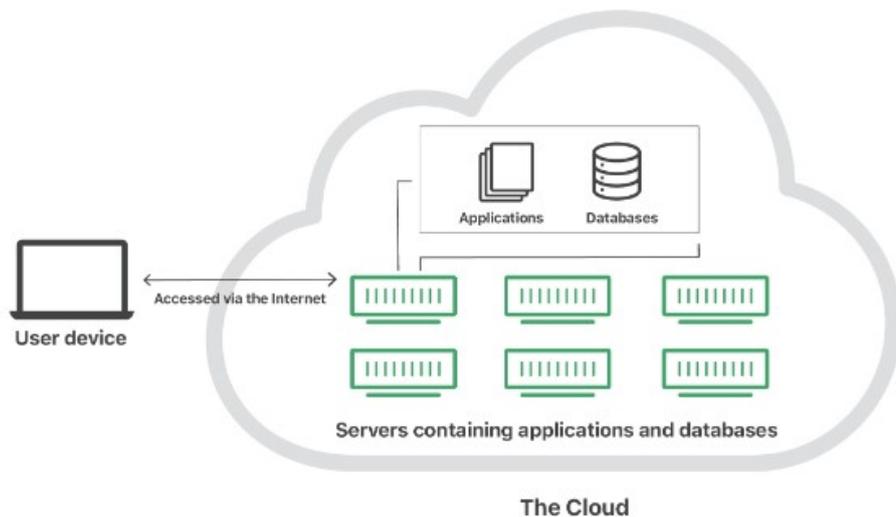


Figure 1.2: L'architecture du cloud [7]

### 1.3.1 Modèles de déploiement du cloud

Les modèles de déploiement dépendent des besoins opérationnels et des exigences techniques. Ils se distinguent en trois catégories : le cloud public, le cloud privé et le cloud hybride [8].

1. **Le cloud public** : il offre un accès prépayé aux ressources informatiques, de stockage et de réseau. Ceux-ci sont fournis par le biais des fournisseurs de Cloud comme Amazon Web Services, Microsoft Azure, Alibaba Cloud, Google Cloud et IBM. Les services de Cloud public nécessitent peu de coûts initiaux et peuvent être déployés rapidement pour une mise sur le marché rapide. Ils sont également bien adaptés aux charges de travail qui peuvent ne durer que peu de temps.
2. **Le cloud privé** : celui-ci est exécuté sur l'infrastructure du centre de données. Cela exige un coût initial plus important et une gestion continue, mais peut permettre de réaliser de potentielles économies sur le long terme. Un cloud privé supporte les charges de travail des missions critiques et les applications héritées difficiles ou impossibles à déplacer vers le cloud public. C'est de plus un modèle idéal permettant d'assurer la conformité aux règles de confidentialité ou de protéger la propriété intellectuelle.
3. **Le cloud hybride** : ce dernier réunit le cloud public et privé afin de pouvoir partager des applications et des données entre eux selon les besoins. Cela donne à l'entreprise la flexibilité d'exécuter des applications de manière à maximiser les économies de coûts potentielles et l'utilisation des ressources tout en répondant aux exigences d'évolutivité et de contrôle.

Il est également possible de cumuler les services de plusieurs fournisseurs de Cloud différents. On parle alors d'une stratégie multi-cloud. La Figure. 1.3 montre ces différents modèles.

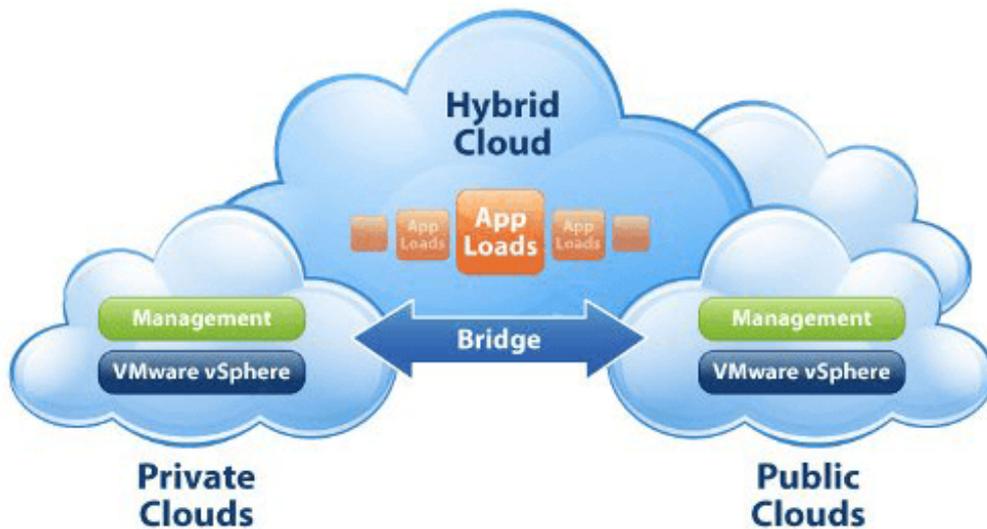


Figure 1.3: Modèles de déploiement du cloud [9]

### 1.3.2 Types de cloud

Il existe trois principaux types de services Cloud : infrastructure en tant que service (IaaS), logiciel en tant que service (SaaS) et plate-forme en tant que service (PaaS) [7].

**1. Infrastructure en tant que service (IaaS : Infrastructure as a Service) :**

L'Infrastructure en tant que service (IaaS) donne accès aux serveurs, au réseau et au stockage. L'IaaS offre plus de contrôle et de flexibilité que tous les modèles de service et peut-être facilement augmenté ou réduit selon les besoins.

Dans ce modèle, une entreprise loue les serveurs et l'espace de stockage dont elle a besoin à un fournisseur de cloud. Elle peut alors utiliser cette infrastructure cloud pour développer ses propres applications. Parmi les fournisseurs de services IaaS, on citera DigitalOcean, Google Compute Engine et OpenStack.

**2. Logiciel en tant que service (SaaS : Software as a service) :**

La plupart des entreprises s'appuient sur de nombreux produits basés sur des logiciels en tant que service (SaaS) pour leurs opérations quotidiennes. Il s'agit d'applications à la demande comme les logiciels de gestion de la relation client (CRM) et les e-mails. Avec le SaaS, les utilisateurs n'ont pas besoin de gérer autre chose que leurs données. Les licences sont achetées sur la base d'un abonnement et les services sont livrés immédiatement. Les services SaaS utilisent une architecture Cloud multi-locataires, ce qui signifie que tous les utilisateurs et toutes les applications partagent une même infrastructure dont la maintenance est assurée de façon centralisée.

**3. Plateforme en tant que service (PaaS : Platform as a service) :**

La Plateforme en tant que service (PaaS) offre les mêmes ressources matérielles que l'IaaS, plus le système d'exploitation et les bases de données. La PaaS permet aux entreprises de développer, d'exécuter et de gérer des applications sans avoir à créer et à maintenir une infrastructure. Elle peut également aider à rationaliser les flux de travail, car plusieurs utilisateurs peuvent accéder simultanément à l'application de développement. Auparavant, le SaaS, le PaaS et l'IaaS étaient les trois principaux modèles de cloud computing, et pratiquement tous les services cloud entraient dans l'une de ces catégories. Toutefois, ces dernières années ont émergé un quatrième modèle qui est FaaS (Function as a Service).

**4. Fonction en tant que service (FaaS : Function as a service) :**

Connu sous le nom de serverless computing ou informatique sans serveur, ce modèle divise les applications cloud en composants plus petits, uniquement exécutés en cas de besoin. Grâce à la Fonction en tant que Service (FaaS), les utilisateurs gèrent purement les fonctions et les données tandis que le fournisseur de cloud gère l'application.

Cela permet aux développeurs d'obtenir les fonctions dont ils ont besoin sans avoir à payer pour des services lorsque le code n'est pas exécuté.

La Figure. 1.4 représente les différents services cloud mentionnés.

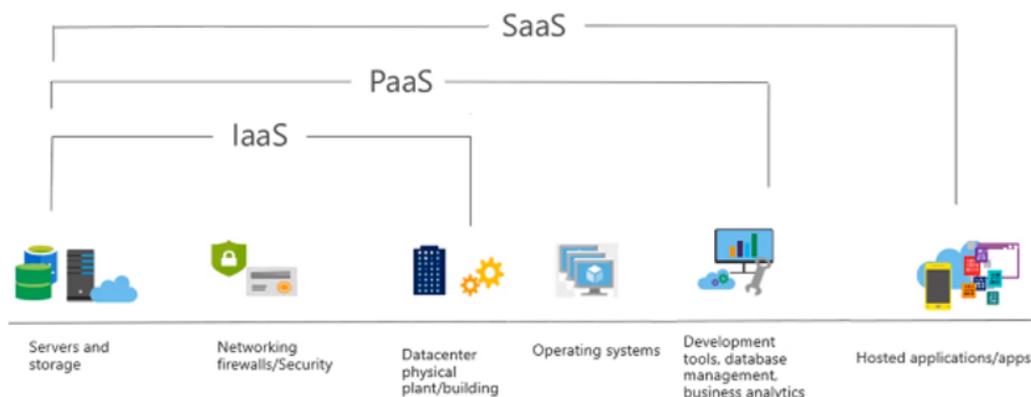


Figure 1.4: Les différents services Cloud [10]

### 1.3.3 Le cloud et l'IA

L'accessibilité, la vitesse et la connectivité offertes par le cloud sont autant d'atouts qui permettent à l'IA de développer toute sa valeur. En intégrant l'IA à une technologie de Cloud hybride, on automatise le contrôle permanent et précis des données. Par exemple, l'IA va se nourrir des données pour optimiser les flux, en les déplaçant dans le cloud, pour les rendre accessibles sans délai à l'usage. Mais quel est le réel intérêt de les coupler ? [11]

L'intérêt se verra dans le contrôle permanent sur les données et l'algorithmique, mais également le désengorgement des outils de stockage, en effet l'Intelligence Artificielle peut également optimiser la capacité de stockage des données. Le machine learning permettra quant à lui d'anticiper le besoin d'accès et la capacité de stockage nécessaire à l'hébergement des nouveaux flux. Enfin, on parlera des coûts réduits au vu du fait qu'on peut opposer au développement de l'IA le fait qu'elle soit, elle-même gourmande en espace. Il est en effet difficile de prédire l'espace de stockage nécessaire lorsque l'on se lance dans un projet d'IA. Mais, le cloud permet d'investir dans un espace de stockage cohérent et évolutif en fonction des besoins et de l'ajuster en quasi-temps réel. Des économies qui peuvent ensuite être allouées au développement de l'IA.

La Figure. 1.5, représente la rencontre en l'IA et le cloud.

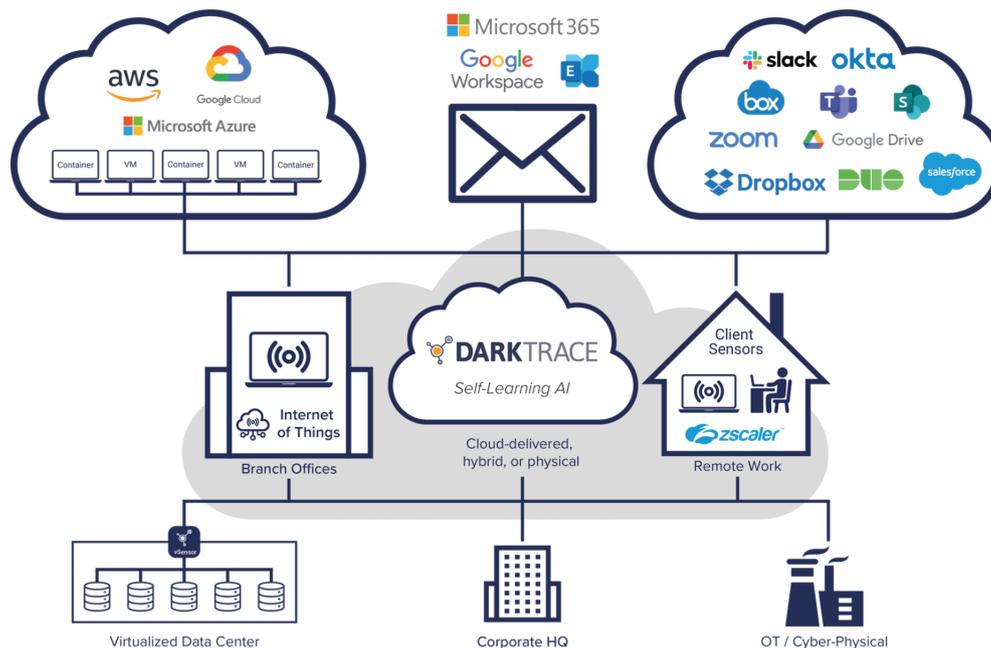


Figure 1.5: Rencontre entre l'IA et le Cloud [12]

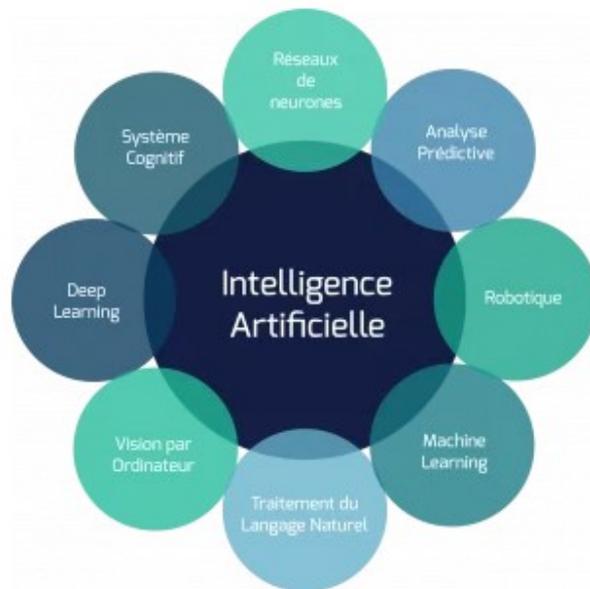
## 1.4 Définition de l'IA

L'intelligence artificielle (IA) est le processus d'imitation de l'intelligence humaine qui repose sur la création et l'application d'algorithmes exécutés dans un environnement informatique dynamique qui peuvent s'améliorer de manière itérative en fonction des informations qu'ils recueillent [13]. Il est à noter que pour imiter le mieux possible le comportement humain, l'intelligence artificielle a besoin d'une quantité de données et d'une capacité de traitement importantes.

### Quelques domaines d'application de l'IA :

- Les chatbots utilisent l'IA pour comprendre les problèmes des clients plus rapidement et répondre plus efficacement.
- Les assistants intelligents utilisent l'IA pour analyser les informations critiques à partir de grands ensembles de données en texte libre afin d'améliorer la planification.
- Les moteurs de recommandation peuvent suggérer automatiquement des émissions télévisées en fonction des habitudes des téléspectateurs.

La Figure. 1.6 illustre les domaines résultants de l'intelligence artificielle.



**Figure 1.6:** Les domaines résultant de l'Intelligence artificielle [14]

Le machine learning (apprentissage automatique) et le deep learning (apprentissage profond) sont deux concepts très importants qui rendent l'intelligence artificielle possible. On confond bien souvent ces deux termes, alors qu'ils désignent deux méthodes bien distinctes employées dans des champs d'application différents [1].

### 1.4.1 Le Machine Learning

Le machine learning est une branche de l'intelligence artificielle et de l'informatique qui se concentre sur l'utilisation de données et d'algorithmes pour imiter la façon dont les humains apprennent sans programmation explicite, améliorant progressivement sa précision. Grâce à l'utilisation de méthodes statistiques, les algorithmes sont formés pour effectuer des classifications ou des prédictions, et pour découvrir des informations clés dans les projets d'exploration de données [15]. Ces informations guident ensuite la prise de décision au sein des applications et des entreprises, impactant idéalement les mesures de croissance clés. Les algorithmes de machine learning sont généralement créés à l'aide de frameworks qui accélèrent le développement de solutions, tels que TensorFlow, PyTorch, Keras...

### 1.4.1.1 Types de machine learning

Le machine learning peut être classé en trois grandes catégories :

#### 1. Machine learning supervisé

Machine learning supervisé ou Supervised Learning (en anglais) est le paradigme d'apprentissage le plus populaire en Machine Learning et en Deep Learning [16]. Les opérateurs présentent à l'ordinateur des exemples d'entrées et les sorties souhaitées, et l'ordinateur recherche des solutions pour obtenir ces sorties en fonction de ces entrées. Le but est que l'ordinateur apprenne la règle générale qui va mapper les entrées et les sorties.

Le machine learning supervisé est utilisé pour faire des prédictions sur des données indisponibles ou futures, on parle alors de "modélisation prédictive". L'algorithme essaie de développer une fonction qui prédit avec précision la sortie à partir des variables d'entrée – par exemple, prédire la valeur d'un bien immobilier (qui représente la sortie) à partir d'entrées telles que le nombre de pièces, l'année de construction, la surface du terrain, l'emplacement, etc [15].

Le machine learning supervisé peut se subdiviser en deux types :

- **Classification** : la variable de sortie est une catégorie.
- **Régression** : la variable de sortie est une valeur spécifique.

Les principaux algorithmes du Machine Learning supervisé sont :

- Forêts aléatoires [17].
- Arbres de décision [18].
- Algorithme K-NN (k-Nearest Neighbors) [19].
- Régression linéaire [20].
- Algorithme de Naïve Bayes [21].
- Machine à vecteurs de support (SVM) [22].
- Régression logistique [23].
- Boosting de gradient [24].

La grande majorité des problèmes de Machine Learning utilisent l'apprentissage supervisé.

## 2. Machine learning non supervisé

Dans le cadre du machine learning non supervisé, l'algorithme détermine lui-même la structure de l'entrée (aucune étiquette n'est appliquée à l'algorithme). Cette approche peut être un but en soi, qui permet de découvrir des structures enfouies dans les données, ou un moyen d'atteindre un certain but. Cette approche est également appelée « apprentissage des caractéristiques » ou feature learning (en anglais).

Un exemple bien connu de machine learning non supervisé est l'algorithme de reconnaissance faciale prédictive de Facebook, qui identifie les personnes sur les photos publiées par les utilisateurs [15].

Il existe deux types de machine learning non supervisé :

- **Clustering** : l'objectif consiste à trouver des regroupements dans les données.
- **Association** : l'objectif consiste à identifier les règles qui permettront de définir de grands groupes de données.

**Les principaux algorithmes du Machine Learning non supervisé :**

- K-means [25].
- Clustering/regroupement hiérarchique [26].
- Réduction de la dimensionnalité [27].

## 3. Machine learning par renforcement

Dans le machine learning par renforcement, un programme informatique interagit avec un environnement dynamique dans lequel il doit atteindre un certain but, par exemple conduire un véhicule ou affronter un adversaire dans un jeu. Le programme-apprenti est conçu sur la base de « récompenses » et de « punitions » pendant qu'il navigue dans l'espace du problème et qu'il apprend à identifier le comportement le plus efficace dans le contexte considéré.

Ce type d'apprentissage se fit connaître en 2013 grâce à un algorithme de machine learning par renforcement (Q-learning) qui s'était rendu célèbre en apprenant comment gagner dans six jeux vidéo Atari sans aucune intervention d'un programmeur [15].

Il existe deux types de machine learning par renforcement :

- **Monte Carlo** : le programme reçoit ses récompenses à la fin de l'état terminal.
- **Machine learning par différence temporelle (TD)** : les récompenses sont évaluées et accordées à chaque étape.

Ces catégories sont également illustrées dans la Figure. 1.7.

**Les principaux algorithmes du machine learning par renforcement sont :**

- Q-learning.
- Deep Q Network (DQN).
- SARSA (State-Action-Reward-State-Action).

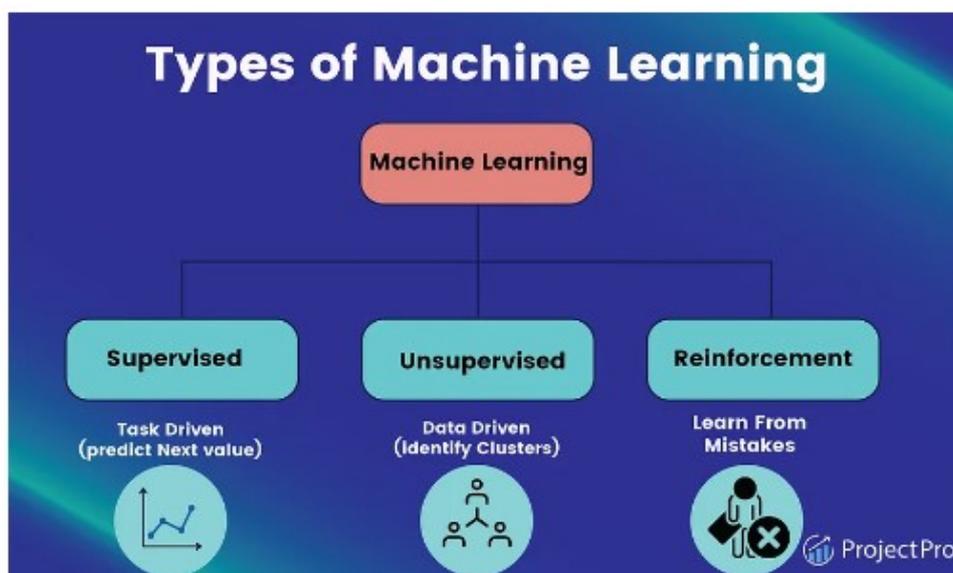


Figure 1.7: Les différents types de Machine Learning [28]

## 1.4.2 Le Deep Learning

Le Deep learning ou apprentissage profond est l'un des sous-ensembles les plus connus du Machine learning. Avec le Deep Learning, nous parlons d'algorithmes capables de mimer les actions du cerveau humain grâce à des réseaux de neurones artificiels. Les réseaux sont composés de dizaines, voire de centaines de «couches» de neurones, chacune recevant et interprétant les informations de la couche précédente [29].

Au fil des années, avec l'émergence du big data et de composants informatiques de plus en plus puissants, les algorithmes de Deep Learning gourmands en puissance et en données ont dépassé la plupart des autres méthodes. Ils semblent être prêts à résoudre bien des problèmes : reconnaître des visages, vaincre des joueurs de poker, permettre la conduite

de voitures autonomes ou encore la recherche de cellules cancéreuses. La Figure. 1.8 nous montre l'architecture de ces derniers.

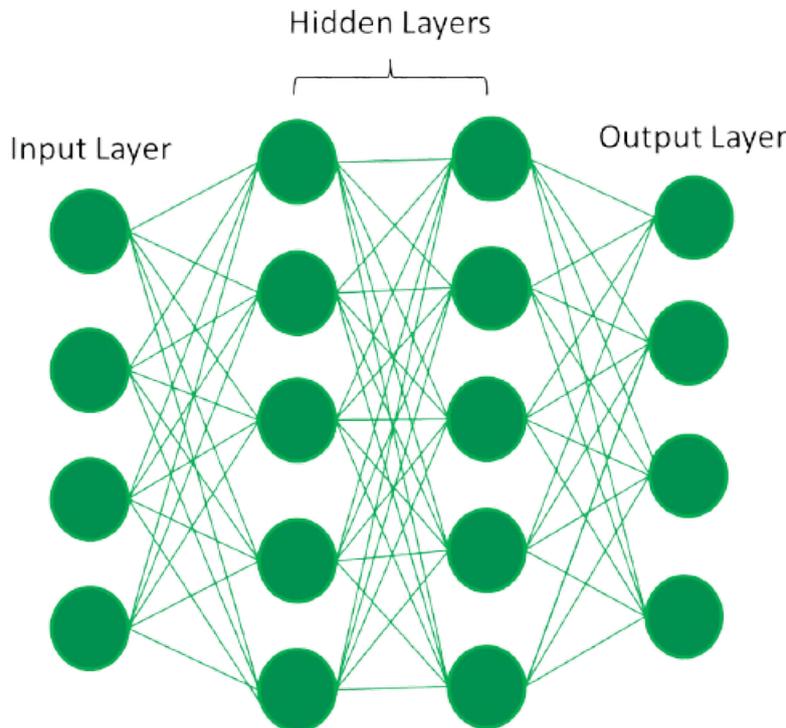


Figure 1.8: Architecture d'un réseau de neurones artificiel [30]

#### 1.4.2.1 Les principaux algorithmes du Deep Learning :

1. **Réseaux neuronaux convolutifs (Convolutional Neural Network, CNN) :** Également appelés ConvNets (CNN), les CNN sont constitués d'une multitude de couches chargées de traiter et d'extraire les caractéristiques des données et sont principalement utilisés pour le traitement d'images et la détection d'objets [31]. Les CNN disposent de plusieurs couches qui traitent et extraient les caractéristiques des données, nous pouvons en citer certaines :
  - **Couche de convolution :** comporte plusieurs filtres pour effectuer l'opération de convolution.
  - **Unité linéaire rectifiée (ReLU) :** la couche ReLU est utilisée afin d'effectuer des opérations sur les éléments. La sortie est une carte de caractéristiques rectifiée.
  - **Couche de mise en commun :** la carte de caractéristiques rectifiée alimente ensuite une couche de mise en commun. La mise en commun est une opération de sous-échantillonnage qui réduit les dimensions de la carte des caractéris-

tiques.

- **La couche de regroupement** : convertit ensuite les tableaux bidimensionnels résultants de la carte d'entités regroupées en un seul vecteur linéaire, long et continu, en les aplatissant.
- **Couche entièrement connectée** : Une couche entièrement connectée se forme lorsque la matrice aplatie de la couche de regroupement est utilisée comme entrée, ce qui permet de classer et d'identifier les images.

la Figure. 1.9 montre un exemple d'implémentation d'un CNN.

## Convolutional Neural Network

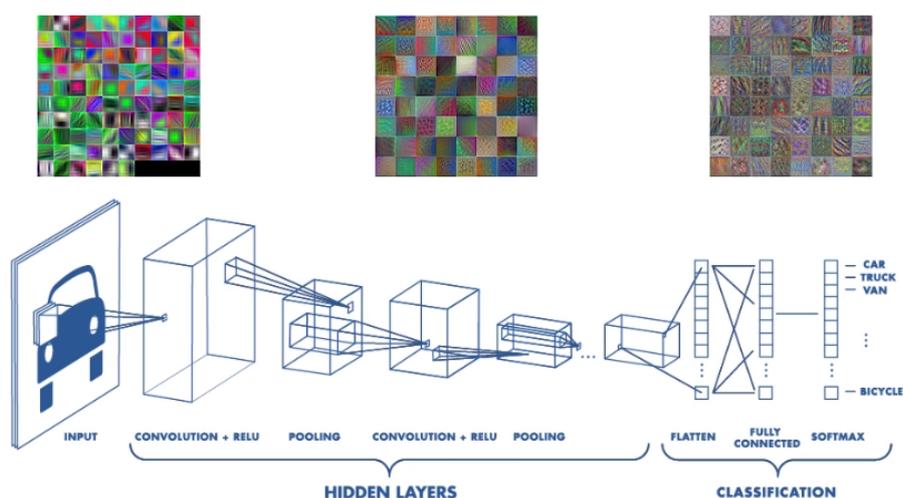


Figure 1.9: Architecture d'un CNN [32]

## 2. Réseaux neuronaux récurrents (RNN) :

Les RNN forment une classe de réseaux qui permet de prédire le futur (jusqu'à un certain point). Ils sont capables d'analyser des séries chronologiques de longueurs quelconques plutôt que sur des données de tailles figées. Ceci les rend très utilisés pour le sous-titrage d'images, l'analyse de séries temporelles, le traitement du langage naturel, la reconnaissance de l'écriture manuscrite et la traduction automatique [31]. La Figure. 1.10 représente le fonctionnement des RNN de manière simplifiée.

# Recurrent Neural Networks

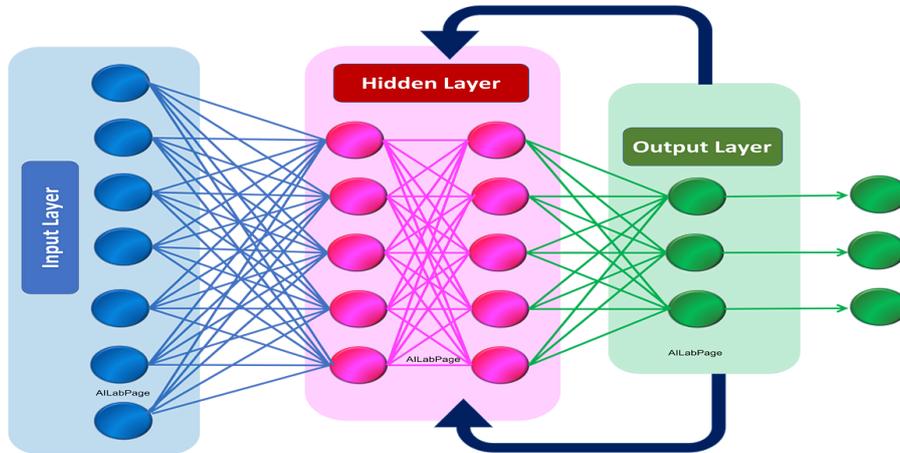


Figure 1.10: Architecture d'un RNN [33]

### 3. Réseaux de mémoire à long terme et à court terme (LSTM) :

Les LSTM sont des dérivés de RNN. Ils peuvent apprendre et mémoriser des dépendances sur une longue durée. Les LSTM conservent ainsi les informations mémorisées sur le long terme. Ils sont particulièrement utiles pour prédire des séries chronologiques, car ils se rappellent des entrées précédentes. Outre ce cas d'utilisation, les LSTM sont également utilisés pour composer des notes de musique et reconnaître des voix [34] ;

### 4. Réseaux adversariaux génératifs (GAN) :

Les GAN créent de nouvelles instances de données qui s'apparentent aux données d'apprentissage profond. Ils possèdent deux principaux composants : un générateur et un discriminateur. Si le générateur apprend à produire des informations erronées, le discriminateur, quant à lui, apprend à exploiter ces fausses informations [34].

L'utilisation des GAN a augmenté au fil du temps. Ils peuvent être utilisés pour améliorer les images astronomiques et simuler les lentilles gravitationnelles pour la recherche sur la matière noire. Les développeurs de jeux vidéo utilisent les GAN pour améliorer les textures 2D à faible résolution des anciens jeux vidéo en les recréant en 4K ou à des résolutions plus élevées via l'apprentissage d'images [31].

Le fonctionnement des GAN est représenté par la Figure. 1.11.

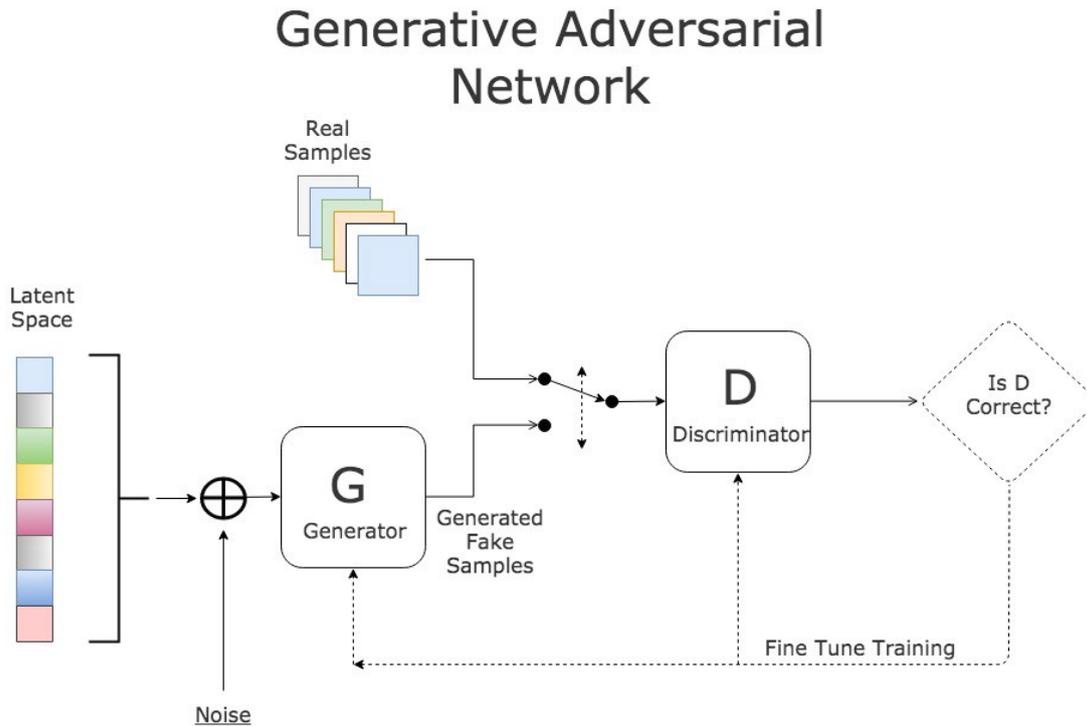


Figure 1.11: Processus de fonctionnement d'un GAN [35]

### 5. Perceptrons multicouches (MLP) :

Les MLP appartiennent à la classe des réseaux neuronaux à anticipation avec plusieurs couches de perceptrons dotés de fonctions d'activation. Les MLP se composent d'une couche d'entrée et d'une couche de sortie qui sont entièrement connectées. Ils ont le même nombre de couches d'entrée et de sortie, mais peuvent avoir plusieurs couches cachées et peuvent être utilisés pour construire des logiciels de reconnaissance vocale, de reconnaissance d'images et de traduction automatique [31].

#### 1.4.2.2 Quelques cas d'utilisation des algorithmes de Deep Learning

De nos jours, le deep learning sert à développer de nombreuses technologies révolutionnaires. Celles-ci sont très utiles dans la vie de tous les jours [34].

- La conduite autonome : Le Deep Learning est l'intelligence artificielle qui a permis de mettre en place la conduite autonome pour les véhicules. Certains algorithmes ont la capacité de pouvoir identifier les panneaux de signalisation tandis que d'autres localisent les piétons. Cette technologie améliore fortement la sécurité routière et l'expérience de conduite des automobilistes.
- Dans le domaine de la médecine : L'IA de Deep Learning permet de distinguer les tumeurs cancéreuses de celles qui ne le sont pas. Elle scanne les photos de radio-

graphie avec une plus grande précision que l'œil humain et permet donc d'anticiper la prise en charge pour allonger les chances de guérison du malade.

- Dans le secteur de l'agriculture : L'agriculture biologique s'appuie sur des drones intelligents capables d'identifier les mauvaises herbes en scannant au survol plusieurs hectares de plantation. Cela permet aux agriculteurs de concentrer uniquement leur énergie sur les zones qui nécessitent un désherbage.

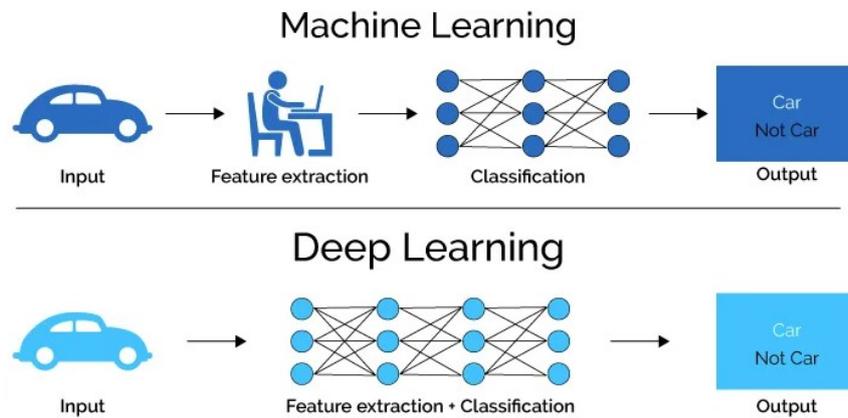
### 1.4.3 Différences entre le Machine Learning et le Deep Learning

Après avoir défini chacune de ces technologies, nous allons maintenant les comparer afin d'identifier puis d'extraire les différences qu'il peut y avoir entre ces dernières dans le Tableau. 1.1 :

Machine learning	Deep learning
Exige des données structurées et bien organisées pour fonctionner	N'a pas besoin de données structurées pour fonctionner
Utilise un algorithme adaptatif à partir des retours faits par l'humain	Utilise des couches de réseaux de neurones qui combinent plusieurs algorithmes en s'inspirant du cerveau humain
Classifie les données qui sont similaires ensembles	Identifie lui-même les caractéristiques discriminantes des données sans catégorisation au préalable
Fonctionne très bien sur une base de données contrôlable	Exige un volume de données considérable pour donner des résultats fiables
Peu coûteuse comparée au Deep Learning	Extrêmement coûteuse en termes de ressources IT
Très intéressante pour une utilisation de masse par les entreprises	Inintéressante pour une utilisation de masse par les entreprises

**Tableau 1.1:** Différences entre Machine Learning et Deep Learning [1]

La Figure. 1.12 illustre le fonctionnement des algorithmes de Machine Learning et Deep Learning



**Figure 1.12:** la distinction fondamentale entre le machine learning et le deep learning [36]

## 1.5 Conclusion

Durant ce chapitre, nous avons évoqué en premier lieu la notion de l'IoT ainsi que le cloud computing qui représentent aujourd'hui les piliers du monde de l'IT, nous avons également mis en évidence leur étroite relation avec l'intelligence artificielle et de ses nombreux domaines d'application.

Dans le chapitre suivant, nous aborderons l'état de l'art où nous allons présenter et analyser divers travaux existants liés aux techniques de l'intelligence artificielle et du machine learning, plus précisément utilisées dans le domaine médical.

# Chapitre 2

## État de l'art

### 2.1 Introduction

L'utilisation de plusieurs technologies telles que l'Internet des objets (IoT), les systèmes embarqués, les technologies de communication, l'intelligence artificielle et les appareils intelligents peut réduire les coûts économiques du système de santé. Différentes technologies de communication ont permis de fournir des services de santé personnalisés et à distance. Afin de répondre aux besoins des futures applications e-santé intelligentes, nous sommes appelés à développer des systèmes de santé intelligents et à multiplier le nombre d'applications connectées au réseau. Pour cela, il a fallu avoir recours à l'une des branches les plus importantes dans le domaine de l'IA qui est le machine learning.

Le machine learning a été appliqué avec succès dans pratiquement tous les domaines comme la robotique, l'éducation, les déplacements vers les soins de santé. Et, dans le domaine de la santé plus particulièrement, les approches de machine learning sont principalement utilisées dans le but de diagnostiquer des maladies au vu du fait que l'approche traditionnelle du diagnostic de la maladie est généralement coûteuse et prend du temps. Et, le potentiel des approches de diagnostic des maladies basées sur le machine learning, rapides et économiques, est prouvé par les chercheurs.

Donc dans ce deuxième chapitre, nous allons présenter un état de l'art où sont décrits cinq travaux abordant le thème du suivi et de la détection des différentes maladies en utilisant les techniques de l'intelligence artificielle et du machine learning plus précisément, par la suite une étude comparative sera établie afin de mettre un accent sur les points forts, mais également les points faibles de chaque approche. Enfin, nous terminons ce chapitre par une conclusion.

## 2.2 Travaux connexes

Dans cette section, nous allons nous intéresser à certains travaux qui ont été réalisés dans le domaine de la santé, plus particulièrement l'utilisation de l'intelligence artificielle et les différents algorithmes du machine learning pour la prédiction et le suivi des patients atteints de maladies chroniques telles que le diabète, le cancer et les maladies cardiovasculaires. En effet, Il existe de nombreux travaux reliés à notre étude dans la littérature, nous avons choisi de présenter les plus significatifs.

### 2.2.1 Une architecture intelligente pour la surveillance des patients diabétiques à l'aide d'algorithmes d'apprentissage automatique :

Une étude réalisée par Amine Rghioui et al [37] dans laquelle, ils ont présenté une architecture intelligente visant à surveiller les patients diabétiques de manière continue, l'objectif principal du travail étant de développer une nouvelle architecture de suivi des patients diabétiques utilisant la technologie 5G. L'architecture permet de surveiller le taux de glycémie du patient en utilisant la technologie de la 5G afin de traiter les informations et de générer des décisions intelligentes.

L'architecture est composée d'un ensemble de capteurs, d'appareils portables, d'une application fonctionnant sur un smartphone et d'un serveur avec une base de données. Le système proposé a été mis en place pour collecter des données sur le taux de glycémie des patients diabétiques, la température et l'activité physique, puis transférer les données du smartphone via la connexion 5G vers une station de base. Par la suite, en utilisant des méthodes d'intelligence artificielle et de machine learning, le système traite intelligemment les données pour aider les utilisateurs à contrôler leur glycémie et à prévoir les changements futurs de leur santé. L'architecture proposée comprend les quatre couches principales suivantes : la partie capteurs, la partie acquisition de données, la partie transmission et la partie base de données comme le montre la Figure. 2.1.

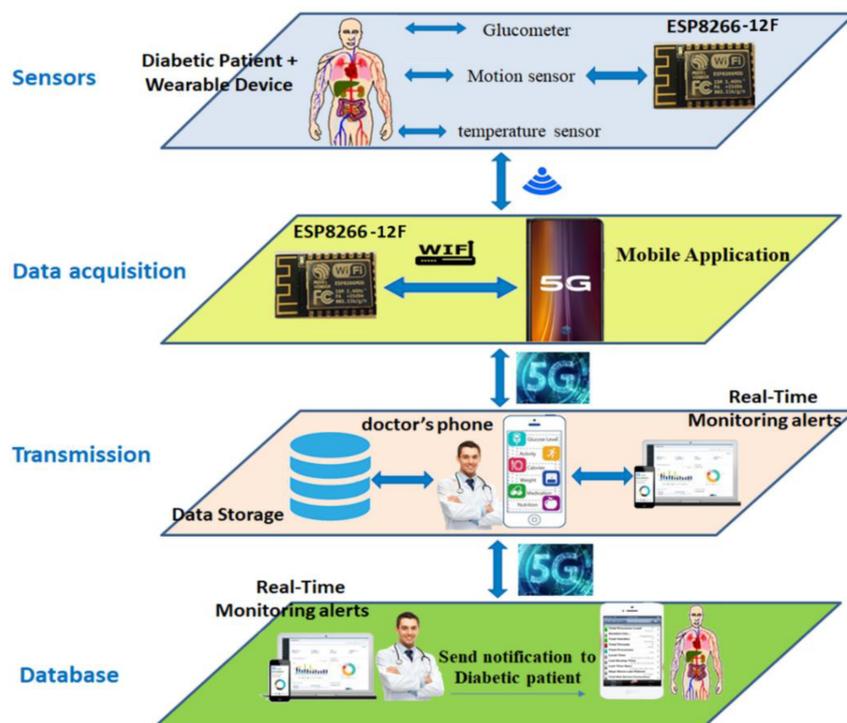


Figure 2.1: Architecture proposée pour le suivi des patients diabétiques

Pour cette étude, la classification des patients diabétiques utilisée est basée sur de l'outil WEKA selon six classificateurs fondés sur des algorithmes de machine learning, à savoir Bayes naïf, J48, ZeroR, SMO (optimisation minimale séquentielle), OneR, forêt aléatoire et logistique simple. Ces algorithmes ont été comparés en termes de précision et d'exactitude. Les résultats de la simulation ont démontré que l'algorithme SMO présentait une excellente classification avec la plus grande précision, sensibilité et exactitude comme le montre la Figure. 2.2.

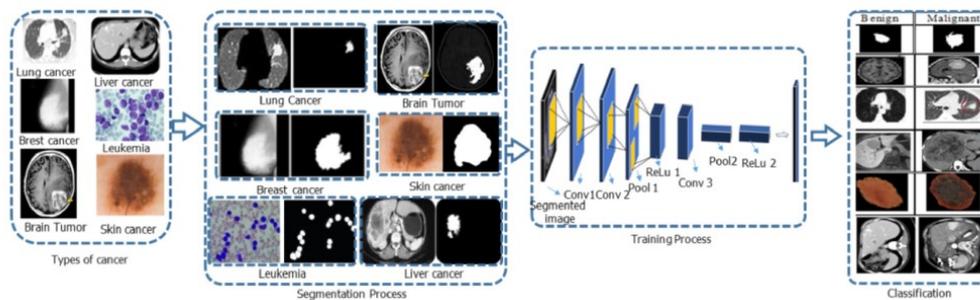
Algorithms	Specificity	Sensitivity	Accuracy	Precision
Naïve Bayes	49.78%	99.08%	84.14%	81.94%
J48	89.49%	99.47%	99.17%	99.32%
SMO	98.92%	99.85%	99.66%	99.66%
ZeroR	0%	100%	69.69%	69.69%
OneR	98.32%	99.47%	99.11%	99.25%
Simple Logistic	92.32%	99.47%	99.11%	99.25%
Random Forest	55.67%	99.06%	85.91%	83.71%

Figure 2.2: Valeurs d'exactitude, de précision, de rappel et du score F1.

## 2.2.2 Progrès récents dans la détection du cancer à l'aide de l'apprentissage automatique, étude systématique des décennies, comparaisons et défis :

Cette étude, réalisée par Tanzila Saba et al [38] aborde les différents types de détection et de classification du cancer à l'aide de machines pour la détection précoce du cancer, permettant ainsi de réduire les déficiences du système manuel. Cet article comporte plusieurs sections traitant les diverses techniques de références dans la détection et la classification des tumeurs cérébrales, du cancer du poumon, de la peau, du sein, de la leucémie ainsi que du cancer du foie.

La représentation picturale de cette étude est montrée par la Figure. 2.3.



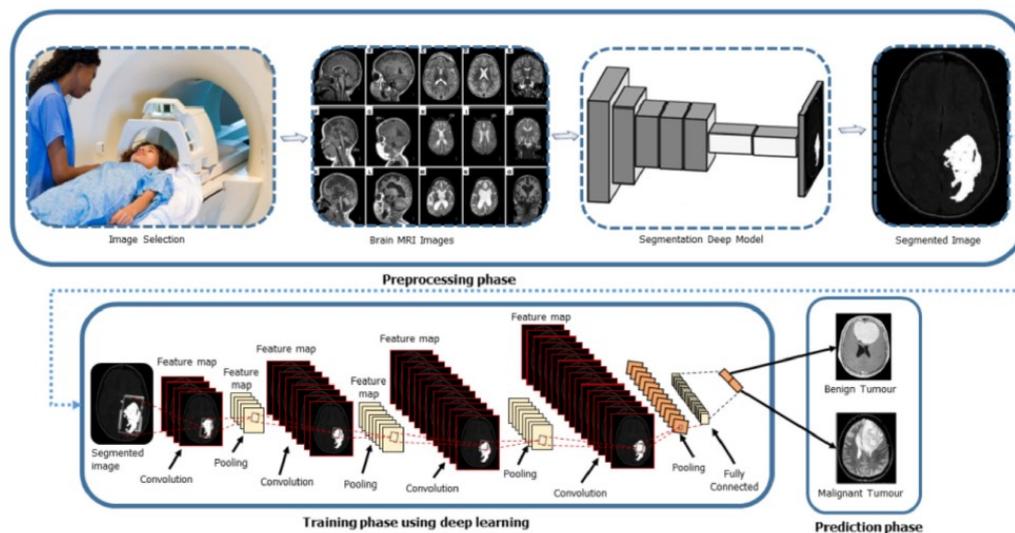
**Figure 2.3:** Système assisté par machine pour la détection du cancer.

Le machine learning est la technique utilisée tout au long de cette étude. Cependant, les algorithmes utilisés diffèrent selon le type de cancer étudié. Dans cette étude, ils ont présenté les algorithmes à appliquer sur six types de cancer différents :

### 1. La tumeur cérébrale :

Parmi les techniques optimales de segmentation des tumeurs cérébrales : Fuzzy Cmeans, K-means clustering, les méthodes de seuil Otsu ou encore l'architecture U-Net qui constitue une des célèbres architectures des réseaux de neurones convolutifs (CNN- Convolutional neural network) utilisées dans la segmentation des tumeurs cérébrales. Pour l'extraction de caractéristiques, le gradient d'orientation d'histogramme (HOG ), la transformée en ondelettes de Gabor (GWT ), les modèles binaires locaux (GWT ) (LBP) et les caractéristiques basées sur la forme. Diverses techniques de sélection et de réduction de caractéristiques sont utilisées comme l'algorithme génétique (GA ) et l'analyse en composantes principales (PCA ) pour une sélection optimale des caractéristiques. L'architecture CNN est considérée comme une des techniques les plus puissantes pour la détection des tumeurs cérébrales. La Figure. 2.4 illustre le processus de prétraitement, de segmentation,

de formation par deep learning et enfin de prédiction des tumeurs cérébrales.



**Figure 2.4:** Phases d'entraînement pour la prédiction en utilisant le Deep Learning.

## 2. Le Cancer du poumon :

Des systèmes appelés systèmes d'aide à la décision sont utilisés. Ces derniers étudient les images par le biais du prétraitement, de la segmentation, de l'extraction de caractéristiques ainsi que le processus de classification. Les réseaux de neurones multi-convolution (MCNN) sont employés pour capturer l'hétérogénéité nodulaire en extrayant des caractéristiques discriminantes à partir de couches empilées en alternance. Dans cette méthode, trois CNN sont utilisés dans le modèle MCNN, dans lequel des patches de nodules parallèles de différentes tailles sont assemblés en tant qu'entrées. La précision de la méthode de segmentation est de 97 La Figure. 2.5 représente l'architecture MCNN utilisée dans la détection du cancer du poumon :

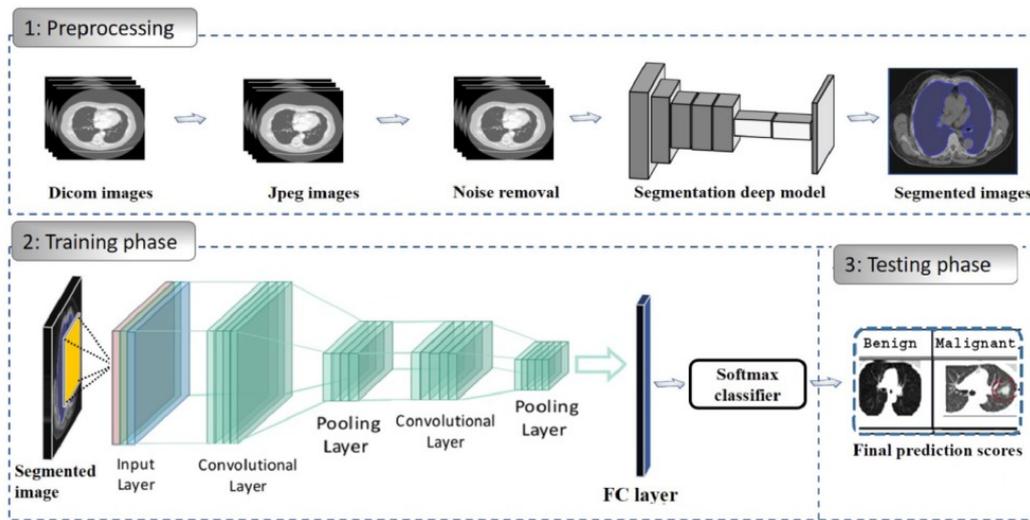


Figure 2.5: Architecture MCNN pour la détection du cancer du poumon.

### 3. le Cancer de la peau :

Dans ce type de cancer, les systèmes sont divisés en deux grandes classes concernant l'extraction de caractéristiques significatives à partir d'images dermoscopiques. Une classe utilise une procédure médicale de diagnostic et extrait automatiquement les mêmes caractéristiques médicales, à savoir la symétrie, plusieurs couleurs, des structures différentielles atypiques.

Quant à l'autre classe, cette dernière est basée sur le machine learning pour reconnaître les modèles statistiques et appliqués aux caractéristiques de l'image, c'est-à-dire les caractéristiques de texture et de couleur. Dans la plupart des travaux, l'accent est mis sur le développement de techniques de machine learning avec extraction de fonctionnalités avancées, comme la règle ABCD, une liste de contrôle en 3 points. La Figure. 2.6 montre le cadre d'extraction de fonctionnalités artisanales et CNN :

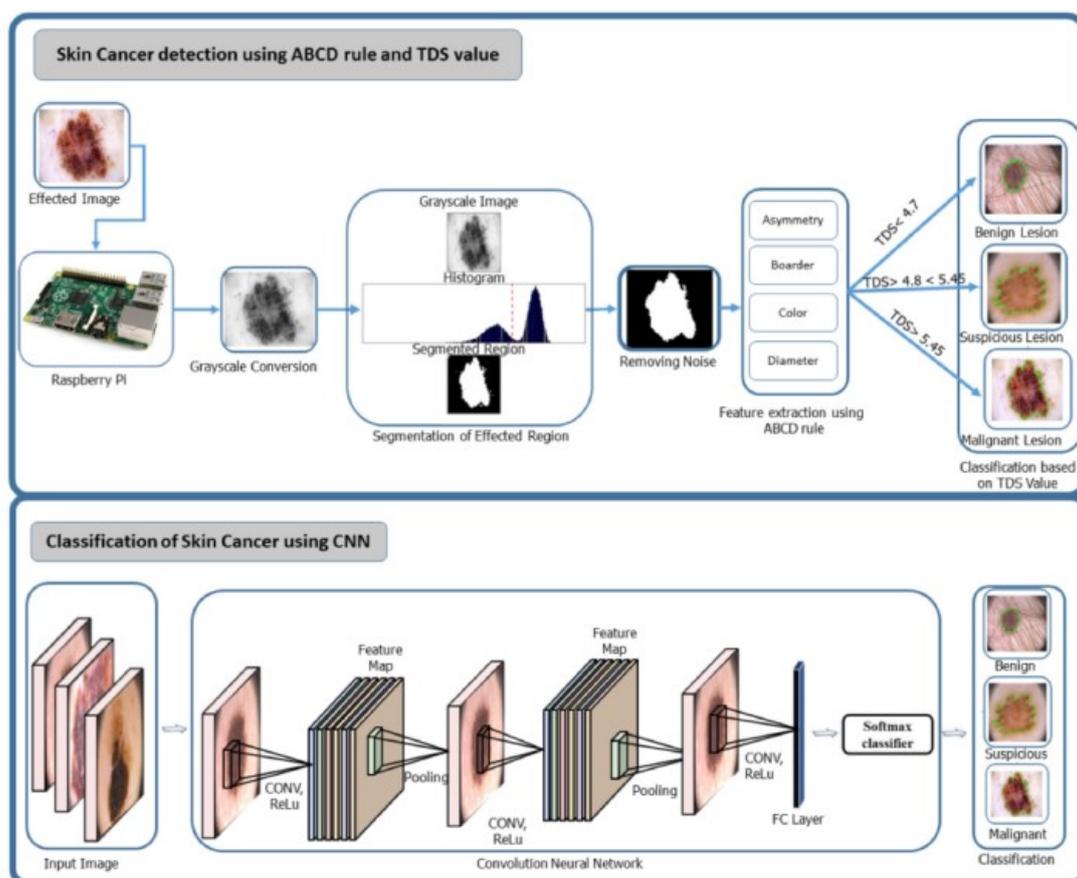


Figure 2.6: Cadre d'extraction de fonctionnalités artisanales et CNN pour le cancer de la peau.

#### 4. La leucémie lymphoblastique aiguë (LAL) :

Pour le diagnostic de la leucémie, l'analyse précise et approfondie des images est indispensable. Dans de nombreux travaux, la segmentation White Blood Cells (WBC) a exploité des observations morphologiques à partir d'images microscopiques en niveaux de gris.

Étant donné que la coloration des globules blancs est plus foncée que les autres composants sanguins, un étirement de contraste a été effectué pour améliorer leurs noyaux. Ensuite, un filtre morphologique a été dérivé en faisant la moyenne des diamètres WBC. L'application de ce filtre morphologique a encore amélioré les noyaux WBC tout en réduisant les composants sanguins plus petits. Ces étapes ont produit des sous-images de dimension fixe contenant des globules blancs situés au centre avec une grande précision.

La détection et la classification de la leucémie lymphoblastique aiguë (LAL) sont présentées par la Figure. 2.7 comme suit :

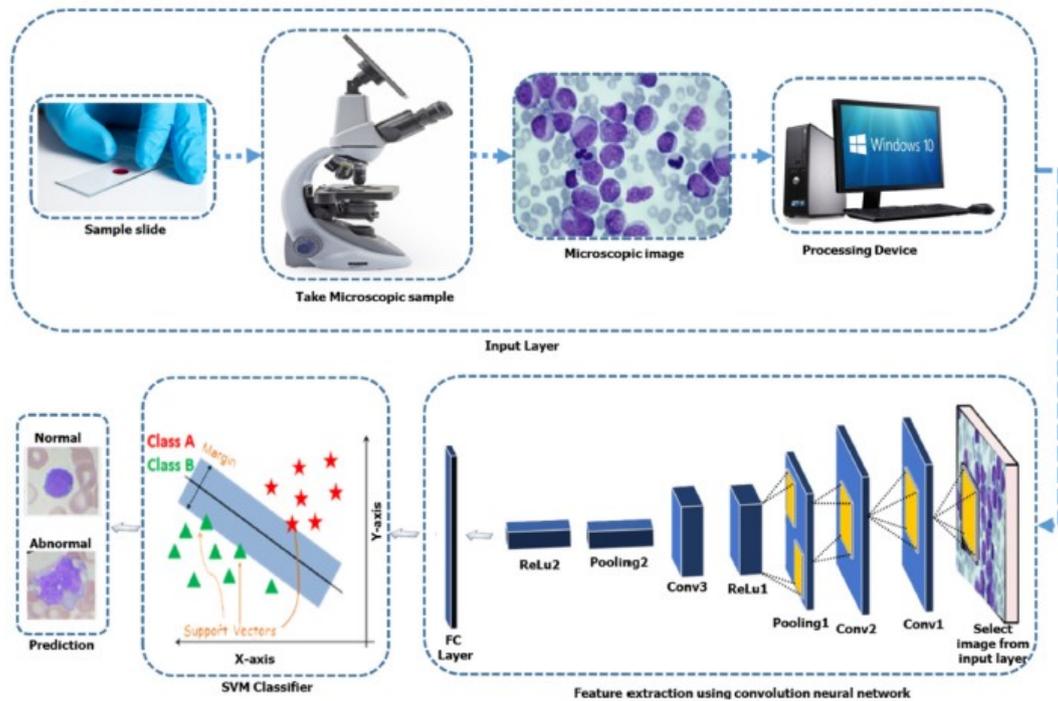
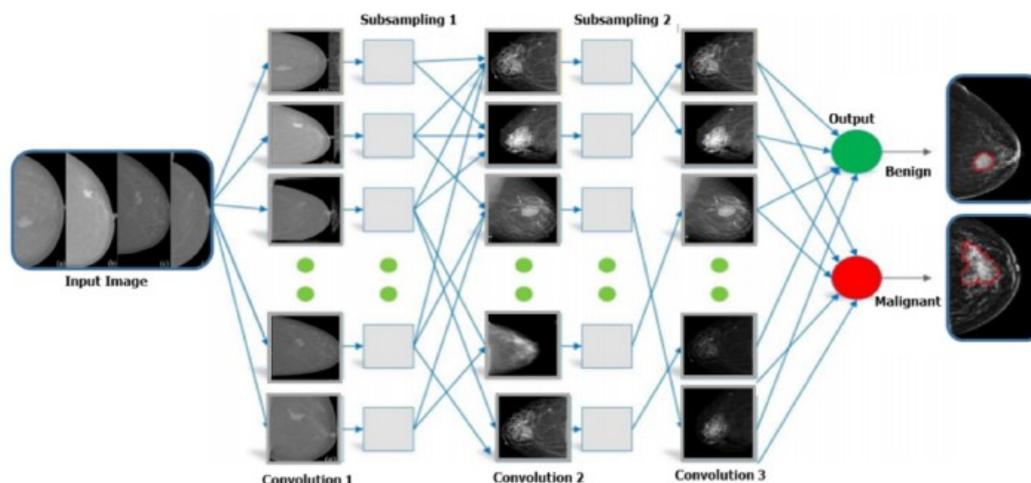


Figure 2.7: Processus d'extraction avec CNN.

## 5. Le cancer du sein :

Le diagnostic du cancer du sein dépend fortement de la performance de la classification des algorithmes utilisés. Plusieurs approches de machine learning telles que les réseaux de neurones, les arbres de décision, les classificateurs KNN, SVM sont appliquées ensemble pour la formation et le test de fonctionnalités permettant de distinguer les objets dans une classe maligne ou bénigne. L'utilisation d'architectures de deep learning a généré des résultats encourageants comme CNN (ainsi que ses sous-ensembles : AlexNet, CiFarNet, GoogLeNet, VGG16 et VGG19) qui sont les plus utilisés dans la détection et la classification des tumeurs mammaires.

Le schéma représenté dans la Figure. 2.8 démontre une architecture de deep learning pour la détection du cancer du sein. La couche d'entrée, la couche de sortie et les autres couches cachées de l'architecture du réseau neuronal convolutif sont des couches complètement liées appelées transformations, puis mises en relation.



**Figure 2.8:** Architecture Deep Convolutional Framework (DCNN) pour la détection du cancer du sein.

## 6. Cancer du foie :

Les systèmes automatiques utilisent trois caractéristiques principales, à savoir l'aspect, la forme et les courbes cinétiques. Les résultats suggèrent que l'approche CNN est plus performante que les autres approches et est prometteuse pour la segmentation des tumeurs du foie. L'efficacité et le contraste d'AdaBoost, RF, SVM ont été expérimentés sur un jeu de données assez limité.

Il est important de retenir que malgré l'immense variété d'algorithmes de machine learning mis au point et utilisés dans le diagnostic et la classification de différents types de cancers, seuls quelques-uns ont prouvé leurs efficacités sur de vastes bases de données.

### 2.2.3 Une étude du diagnostic des maladies à l'aide de l'apprentissage automatique :

Le machine learning (ML) qui est une branche de l'intelligence artificielle, a été appliqué avec succès dans le domaine de la santé pour diagnostiquer des maladies. Les techniques de machine learning ont non seulement été en mesure de diagnostiquer les maladies courantes, mais sont également capables de diagnostiquer les maladies rares. Cette étude réalisée par Samin Poudel [39] vise à trouver l'approche qui permet d'appliquer plusieurs algorithmes de pointe avec des lignes de codes optimales, afin que la recherche de la meilleure méthode ML pour diagnostiquer une maladie particulière puisse être poursuivie efficacement.

La problématique étant que le machine learning offre des algorithmes systématiques et sophistiqués de données cliniques multidimensionnelles, la précision de ce dernier dans le diagnostic des maladies reste une préoccupation. Et l'amélioration de ses performances pour diagnostiquer la maladie est un sujet brûlant dans ce domaine.

L'auteur de la recherche a choisi une approche qui consiste à utiliser des bibliothèques comme AutoGluon qui peuvent aider à trouver l'approche de machine learning la plus performante parmi de nombreuses approches de diagnostic de la maladie pour un ensemble de données donné. Cela réduira la probabilité d'un diagnostic inexact, ce qui est une considération très importante lorsqu'il s'agit de la santé des personnes. Ceci a été fait en testant les performances des approches de 20 algorithmes de machine learning dans le diagnostic du diabète. Et cela, en utilisant des algorithmes de classification de la bibliothèque scikit-learn et de la bibliothèque AutoGluon qui ont été appliqués afin de vérifier la capacité des algorithmes à diagnostiquer la maladie du diabète. Ils sont implémentés avec AWS SageMaker. Certaines de ces classifications utilisées sont : Naïve Bayes, Support Vector Machine (SVM), K Nearest Neighbors (KNN), perceptron et des réseaux de neurones profonds robustes dans AutoGluon comme LightGBM, XGBoost, MXNet et encore plein d'autres. La liste complète est illustrée par la Figure. 2.9.

Library	ML Algorithm	Number of ML approaches
Scikit-Learn	Random Forest Classifier, Decision Tree Classifier, Naïve Bayes Classifier, Perceptron, Multilayer Perceptron, Voting Classifier	6
AutoGluon	WeightedEnsemble_L2, LightGBM_BAG_L1, LightGBM_LARGE_BAG_L1, NeuralNetFastAI_BAG_L1, CATBoost_BAG_L1, ExtraTreesGini_BAG_L1, LightGBMXT_BAG_L1, XGBoost_BAG_L1, RandomForestEntr_BAG_L1, RandomForestGini_BAG_L1, ExtraTreesEntr_BAG_L1, NeuralNetMXNet_BAG_L1, KNeighborsUnif_BAG_L1, KNeighborsDist_BAG_L1	14

**Figure 2.9:** Liste des algorithmes machine learning utilisés.

**La métrique d'évaluation :** Le diagnostic de la maladie est une tâche de classification. De plus, les algorithmes de classification de machine learning sont évalués à l'aide de mesures de précision de classification telles que l'exactitude, la précision, le rappel et le score F1.

L'étude conclut que la plupart des méthodes de machine learning fonctionnent mieux que la référence de précision de base de 65 % pour cet ensemble de données lors du diagnostic du diabète. Environ 77 % de la précision semble être le meilleur cas pour les algorithmes de machine learning de pointe pour l'ensemble de données considéré dans cette étude. Considérant le cas d'avoir des données déséquilibrées, la capacité de la méthode Naïve Bayes à mieux performer a été soulignée, et cela, par rapport aux autres, compte tenu de l'analyse combinée de toutes les mesures d'évaluation. Les résultats ont

été regroupés dans la Figure. 2.10.

S. N	ML Algorithm	Accuracy	F1-score	Precision	Recall
1	Random Forest Classifier (Scikit-learn)	0.74	0.81	0.78	0.84
2	Decision Tree Classifier (Scikit-learn)	0.65	0.73	0.73	0.73
3	Naïve Bayes Classifier (Scikit-learn)	0.77	0.83	0.80	0.86
4	Perceptron (Scikit-learn)	0.49	0.47	0.71	0.35
5	Multilayer Perceptron (Scikit-learn)	0.68	0.76	0.75	0.77
6	Voting Classifier (Scikit-learn)	0.72	0.78	0.79	0.77
7	AutoGluon Best Performer	0.74	0.82	0.76	0.88

Figure 2.10: Évaluation des algorithmes de machine learning.

## 2.2.4 Modèles de détection des maladies cardiaques basés sur la technologie d'apprentissage automatique :

Les maladies cardiovasculaires (MCV) sont un type de maladie cardiaque qui continue d'être une cause majeure de décès dans le monde, représentant plus de 30 % de tous les décès. D'où la détection précoce des maladies cardiaques reste une préoccupation essentielle dans les services de santé (HCS).

Cet article, réalisé par Umarani Nagavelli et al [40], présente les différentes technologies de machine learning basées sur une brève analyse de la détection des maladies cardiaques. L'objectif principal de ce dernier est de donner aux cliniciens un outil pour les aider à diagnostiquer les problèmes cardiaques à un stade précoce. En conséquence, il sera plus facile de traiter efficacement les patients et d'éviter de graves répercussions.

Récemment, un système de gestion hospitalière a été utilisé pour gérer les soins de santé ou les données des patients, ce qui signifie que davantage de données sont produites par ces systèmes. Pour prédire les maladies cardiaques, le DSS (Decision support system) utilisant l'algorithme NB (Naive bayes ) a été développé. Une application Web est construite pour obtenir l'application et l'entrée de l'utilisateur, et elle récupère les principales caractéristiques relatives aux maladies cardiaques à partir d'une base de données historique.

Quatre approches de modèles de machine learning pour la détection des maladies cardiaques sont analysées dans ce travail : ce sont les Naive Bayes avec une prédiction basée sur une approche pondérée, 2 SVM avec une prédiction basée sur XGBoost, une SVM améliorée (ISVM) basée sur une prédiction basée sur la technique d'optimisation de la dualité (DO) et une prédiction basée sur XGBoost.

La technique de machine learning utilisée pour faire la prédiction des maladies car-

diaques est l'approche pondérée bayésienne naive (NBwa). Le système proposé est une application web qui peut être classée en deux modules : côté administrateur et côté utilisateur, respectivement.

L'administrateur soumet l'ensemble de données dans la base de données sous la forme d'une feuille Excel, qui a ensuite sera traduite en poids et stockée. Des clusters sont formés à partir des enregistrements. Selon le poids moyen ou chacun des groupes est déterminé, la classification d'attribut et dans la base de données cette caractéristique peut être stockée. À l'aide des champs fournis, les entrées de l'utilisateur et les données du patient sont converties en poids, et l'algorithme de NB est appliqué aux données. Le résultat peut être affiché à l'utilisateur en fonction de la probabilité du système. Les probabilités conditionnelles et marginales de divers événements sont comparées par l'algorithme de Naive Bayes. Pour les échantillons donnés, cet algorithme est utile pour calculer la valeur la plus proche possible. Le théorème de Bayes est utile pour calculer la probabilité diagnostique lorsque l'état de santé du patient est surveillé sur la base de quelques symptômes. La Figure. 2.11 montre le processus plus amplement.

S. N	ML Algorithm	Accuracy	F1-score	Precision	Recall
1	Random Forest Classifier (Scikit-learn)	0.74	0.81	0.78	0.84
2	Decision Tree Classifier (Scikit-learn)	0.65	0.73	0.73	0.73
3	Naïve Bayes Classifier (Scikit-learn)	0.77	0.83	0.80	0.86
4	Perceptron (Scikit-learn)	0.49	0.47	0.71	0.35
5	Multilayer Perceptron (Scikit-learn)	0.68	0.76	0.75	0.77
6	Voting Classifier (Scikit-learn)	0.72	0.78	0.79	0.77
7	AutoGluon Best Performer	0.74	0.82	0.76	0.88

**Figure 2.11:** Évaluation des algorithmes de ML.

Selon l'analyse des résultats faite dans cet article, il a été trouvé que les paramètres d'exactitude, de précision, de rappel et de mesure F1 sont élevés dans la détection des maladies cardiaques basée sur l'algorithme XGBoost et seule la précision est faible pour le Naive Bayes avec une approche pondérée que les autres, et la précision restante, rappel, et les valeurs de mesure F1 sont faibles dans SVM avec le modèle d'optimisation de dualité (DO). La Figure. 2.12 nous montre les résultats.

Methods	Accuracy	Precision	Recall	F1-measure
Naïve Bayes weighted approach	86.00	82.34	87.25	89.21
2 SVM's and XGBoost	94.03	86.56	94.78	92.79
SVM and DO	89.4	66.1	81.3	82.1
XGBoost	95.9	97.1	94.67	95.35

**Figure 2.12:** Analyse comparative de différentes méthodes de machine learning.

### 2.2.5 Une application de l'apprentissage automatique au diagnostic hématologique :

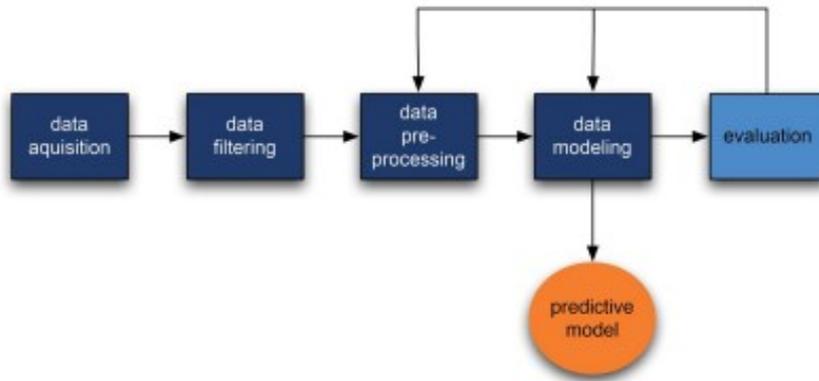
Le machine learning a connu un développement important au cours de la dernière décennie et est utilisé avec succès dans de nombreuses applications intelligentes. Il n'existe cependant aucune application réussie de l'apprentissage automatique qui aborde des domaines plus complexes du diagnostic médical, tels que l'hématologie. Le diagnostic clinique des maladies hématologiques repose sur des tests sanguins de laboratoire.

Cela dit, nous avons constaté que même le spécialiste en hématologie le plus qualifié peut laisser échapper un diagnostic, car il n'a pas pu faire de corrélation entre les divers paramètres sanguins et les diagnostics possibles. Les algorithmes d'apprentissage automatique, eux, peuvent facilement gérer des centaines de paramètres et sont capables de détecter la relation entre ces nombreux attributs, rendant le domaine de la médecine particulièrement intéressant pour les applications du Machine Learning.

Dans cette étude réalisée par Grégor Guncar et al [41] deux modèles hématologiques-prédictifs Smart Blood Analytics (SBA) seront évalués, ceux-ci sont basés sur deux ensembles différents de résultats de tests sanguins de laboratoire clinique, des maladies codées et leurs évaluations. Les deux modèles ont été évalués à l'aide d'une validation croisée stratifiée 10 fois, ainsi que de 20 cas hématologiques supplémentaires sélectionnés au hasard, et leurs performances comparés à une évaluation réalisée par des spécialistes en hématologie.

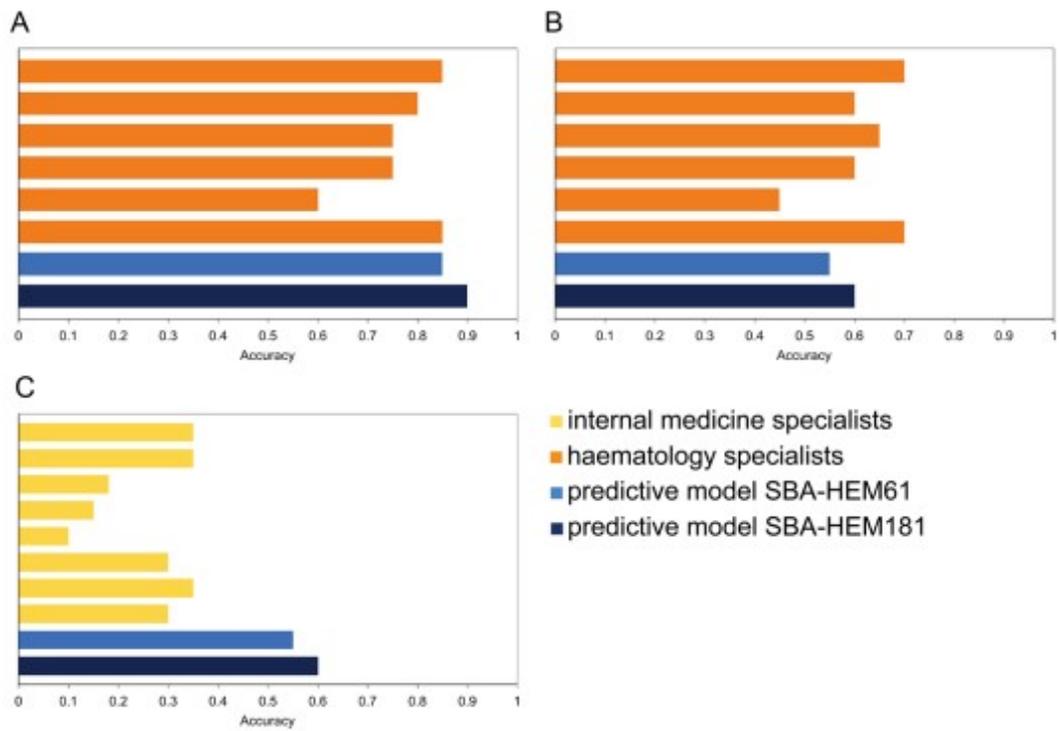
#### Algorithme Smart Blood Analytics :

1. **Acquisition de données** : acquisition de données brutes à partir de la base de données.
2. **Filtrage des données** : sélectionner uniquement les tests sanguins effectués au début du traitement et au diagnostic final comme sous-ensemble d'apprentissage automatique.
3. **Prétraitement des données** : canoniser les paramètres sanguins (les faire correspondre avec notre base de données de paramètres de référence, filtrer les valeurs erronées, et traiter les valeurs manquantes).
4. **Modélisation des données** : construction du modèle prédictif.
5. **Évaluation** : évaluation du modèle prédictif.



**Figure 2.13:** Représentation schématique du processus de l'algorithme Smart Blood Analytics (SBA).

L'utilité des modèles prédictifs pour le diagnostic des maladies hématologiques a été prouvée par un test clinique dans lequel les deux modèles prédictifs ont pu diagnostiquer des maladies hématologiques comme l'auraient fait des spécialistes en hématologie expérimentés et de manière beaucoup plus efficace que les spécialistes en médecine interne, ces résultats sont illustrés dans la Figure. 2.14. Les précisions de prédiction de SBA-HEM061 et SBA-HEM181, étaient de 0,55 et 0,60 respectivement, tandis que les spécialistes en hématologie ont atteint une précision de 0,60. Ces résultats ne sont pas significativement différents. Cependant, en considérant les cinq diagnostics les plus probables proposés par SBA-HEM181 et les spécialistes en hématologie, les précisions ont atteint les valeurs 0,90 et 0,77, respectivement, En conséquence, la SBA pourrait être utilisée pour aider les médecins non spécialisés en hématologie en facilitant la procédure de diagnostic et en suggérant une référence appropriée et précoce des patients.



**Figure 2.14:** Comparaison des résultats de prédiction des maladies hématologiques obtenus par les algorithmes de prédiction SBA avec ceux obtenus par des spécialistes en termes de précision.

## 2.3 Étude comparative

Dans la section précédente, les cinq travaux avec leurs caractéristiques ont été décrits. Après les avoir présentés, nous allons à présent établir une étude comparative qui sera résumée dans le Tableau. 2.1.

**Tableau 2.1:** Tableau comparatif des travaux

Intitulé du travail	Catégorie de l'approche	Quelques techniques et outils utilisés	Type de classification	Résultats optimaux
Une architecture intelligente pour la surveillance des patients diabétiques à l'aide d'algorithmes d'apprentissage automatique [37]	Une approche visant à la surveillance continue et intelligente des patients diabétiques et à l'obtention de recommandation afin d'améliorer leur quotidien.	Naive Bayes, SMO, OneR, forêt aléatoire, logistique simple	Apprentissage supervisé	Le meilleur algorithme est le SMO avec une exactitude de 99,66%, une sensibilité de 99,85% et une précision de 99,66%.
Une étude du diagnostic des maladies à l'aide de l'apprentissage automatique [39]	Une approche visant à trouver la méthode de ML la plus adéquate afin de fournir un diagnostic de pointe	Naïve Bayes, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Perceptron, des réseaux de neurones profonds robustes : LightGBM, XGBoost, MXNet	Apprentissage supervisé et non supervisé	Le meilleur algorithme a été le Naïve Bayes avec un taux de précision de 80% et une exactitude de 77%.

Suite à la page suivante

Intitulé du travail	Catégorie de l'approche	Quelques techniques et outils utilisés	Type de classification	Résultats optimaux
<p>Progrès récents dans la détection du cancer à l'aide de l'apprentissage automatique, étude systématique des décennies, comparaisons et défis [38]</p>	<p>Une approche qui aborde les différents types de détection et de classification du cancer à l'aide de machines pour la détection précoce du cancer.</p>	<p>K-means clustering, l'architecture U-Net, Les réseaux de neurones multi-convolution (MCNN), SVM, KNN</p>	<p>Apprentissage supervisé et non supervisé</p>	<p>Le meilleur algorithme est le CNN avec des performances comme suit : une précision de 82.29% pour les tumeurs cérébrales, une précision de 86.42% et 73.4% de sensibilité pour le cancer des poumons, 98.4% de précision pour le cancer de la peau, 97.74% de précision dans les Leucémies, 82.43% de précision, 81% de sensibilité et 72,26% pour la spécificité au cancer du sein, et enfin 92% de précision, 92% de sensibilité et 98% de spécificité pour le cancer du foie.</p>

Suite à la page suivante

Intitulé du travail	Catégorie de l'approche	Quelques techniques et outils utilisés	Type de classification	Résultats optimaux
Modèles de détection des maladies cardiaques basés sur la technologie d'apprentissage automatique [40]	Une approche ayant pour but la détection précoce des maladies cardiaques.	Naive Bayes avec une prédiction basée sur une approche pondérée, 2 SVM avec une prédiction basée sur XGBoost, une SVM améliorée (ISVM), une prédiction basée sur XGBoost	Apprentissage supervisé	Le meilleur algorithme est le XGBoost avec une précision de 97.1% et 95.9% d'exactitude.
Une application de l'apprentissage automatique au diagnostic hématologique [41]	Une approche basée sur des modèles prédictifs pour le diagnostic des maladies hématologiques.	SBA-HEM061, SBA-HEM181	Apprentissage supervisé	Le meilleur algorithme est le SBA-HEM181 qui a résulté d'une précision allant de 77 à 90% (selon la taille de l'échantillon étudié).

## 2.4 Conclusion

Durant ce chapitre, nous avons fait l'étude de cinq travaux, dont la problématique commune était de prouver l'efficacité du machine learning en matière d'aide au diagnostic médical. Tous ont donné d'excellents résultats démontrant ainsi la capacité du Machine Learning à détecter diverses pathologies à partir de simples analyses médicales. Nous avons également établi un tableau comparatif dans lequel nous avons spécifié pour chaque étude : la catégorie de l'approche, les techniques et outils utilisés, le type de classification ainsi que les résultats obtenus.

La partie suivante portera sur le diabète, ses caractéristiques ainsi que des éléments essentiels à sa prise en charge. Nous proposerons par la suite, l'architecture pour le suivi du diabète sur laquelle repose notre travail.

# Chapitre 3

## Proposition d'une architecture pour le suivi des patients diabétiques

### 3.1 Introduction

De nos jours, le diabète touche 537 millions de personnes à travers le monde, soit 10.5% de la population mondiale. Lorsqu'elle est mal contrôlée, cette pathologie chronique est considérée comme l'une des principales causes de mortalité. Néanmoins, avec une bonne prise en charge, il est parfaitement possible de mener une vie normale et d'éviter toutes les complications éventuelles que pourraient engendrer la maladie. Avec l'avancement de la médecine et de la technologie, il est désormais beaucoup plus facile pour les médecins de veiller au bon équilibre glycémique de leurs patients grâce à certains procédés qui ont révolutionné le monde de la médecine.

Dans ce chapitre, nous aborderons de manière détaillée le diabète : ses symptômes, comment le détecter, mais surtout, comment le contrôler grâce à une méthode adaptée à cet effet. La notion d'hémoglobine glyquée, un critère essentiel au suivi du diabète y est également exposée. Nous avons également fait la description détaillée de l'ensemble des algorithmes utilisés lors de la réalisation de notre travail. Pour finir, nous achèverons cette partie avec l'architecture représentative de notre travail.

### 3.2 Motivations

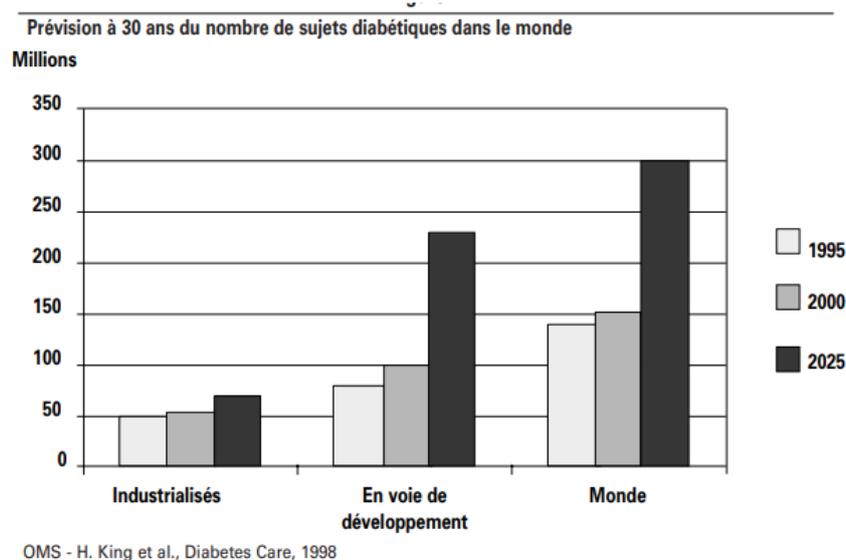
Le diabète est « l'un des principaux tueurs au monde », avec l'hypertension artérielle et le tabagisme, selon l'Organisation Mondiale de la Santé (OMS). Cette maladie constitue un problème de santé publique majeur et malgré les efforts de prévention, la pandémie se poursuit [42].

La prévalence de ce dernier (type 1 et type 2, diagnostiqué et non diagnostiqué) ne cesse d'augmenter depuis une vingtaine d'années. Il touche 537 millions de personnes dans le monde en 2021. Près d'un adulte sur dix pourrait être concerné d'ici 2045, selon la Fédération internationale du diabète.

Le diabète peut être une maladie grave à l'origine de complications s'il n'est pas soigné, il tue près de 2 millions de personnes par an [43]. De ce fait, il est clair que l'ampleur de cette maladie est bien importante, et donc à fin d'avoir le meilleur suivi du taux de glycémie dans le temps et vérifier l'efficacité thérapeutique des traitements mis en place, il est essentiel d'avoir un indicateur fiable et rétrospectif qui permette de signaler tous les écarts sur une période de trois mois. L'indicateur le plus fiable à l'heure actuelle est le dosage de l'hémoglobine glyquée ou HbA1c.

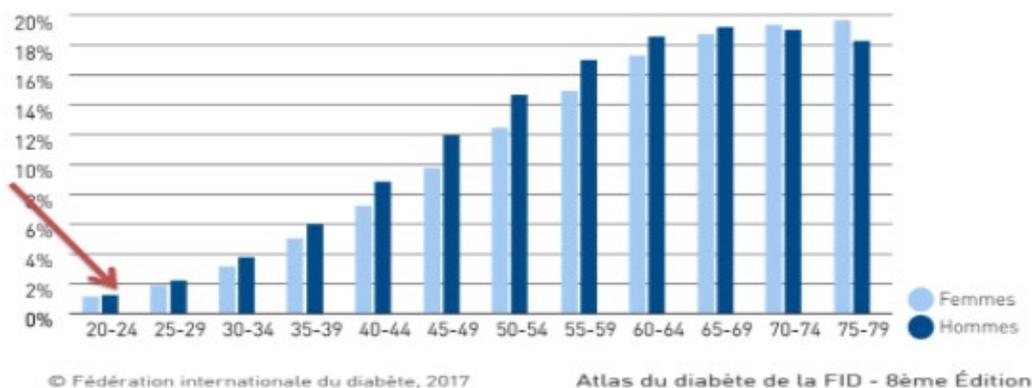
### 3.3 Statistiques

À partir d'une prévalence mondiale du diabète chez l'adulte (à partir de 20 ans) de 4,0 % en 1995, l'OMS prévoit une augmentation à 5,4 % en 2025, sensiblement identique dans les pays industrialisés et dans les pays en voie de développement. Le diabète de type 2 représente la quasi-totalité du diabète de l'adulte à l'échelle planétaire. Cette augmentation de 35 % en 30 ans ne semble pas énorme à première vue. Pourtant, traduite en nombre de diabétiques, le chiffre doit passer de 135 à 300 millions entre 1995 et 2025 [44] comme le montre Dans la Figure. 3.1.



**Figure 3.1:** Statistiques prévisionnelles du nombre de sujets diabétiques sur une durée de 30 ans [44]

Selon la fédération internationale du diabète, environ 3,7 millions d'adultes (âgés de 20 à 79 ans) vivaient avec le diabète en Algérie en 2019, ce qui représente une prévalence de 13,6%. Le diabète de type 2 est la forme la plus courante en Algérie, représentant plus de 90% de tous les cas [45] comme l'illustre la Figure. 3.2.



**Figure 3.2:** Prévalence du diabète selon l'âge et le sexe au niveau de la Wilaya d'Alger (2017)

Le diabète est considéré comme une cause majeure de mortalité et de morbidité prématurées en Algérie. En effet, en 2019 on estimait que les décès liés au diabète représentaient 5,5% de tous les décès dans le pays. Et cela en plus d'être une cause importante de cécité, d'insuffisance rénale, d'amputation des membres inférieurs et de maladies cardiovasculaires.

La prévalence du diabète en Algérie est en augmentation, et cela est attribué à des facteurs tels que l'urbanisation, les changements de mode de vie et le vieillissement de la population du pays. Ces dernières années, le gouvernement algérien a mis en place plusieurs initiatives pour lutter contre la prévalence croissante du diabète, notamment la création de programmes d'éducation et de dépistage du diabète, ainsi que des efforts pour promouvoir une alimentation saine et une activité physique.

### 3.4 Définition du diabète

Le diabète est un trouble de l'assimilation, de l'utilisation et du stockage des sucres apportés par l'alimentation. C'est une maladie chronique qui apparaît lorsque le pancréas ne produit pas suffisamment d'insuline ou que l'organisme n'utilise pas correctement l'insuline qu'il produit. L'insuline étant une hormone qui régule la concentration de sucre dans le sang [46].

Par conséquent, le sucre ne peut pas entrer dans les cellules et il s'accumule dans le sang, ce qui entraîne une hyperglycémie, c'est-à-dire une augmentation de la glycémie au-dessus des valeurs normales.

En présence de diabète, l'un et/ou l'autre des phénomènes suivants se produisent :

- Une résistance du corps à l'action de l'insuline.
- Une diminution ou une absence de production d'insuline.

À long terme, l'hyperglycémie peut mener au développement de certaines complications, notamment aux yeux, aux reins, aux nerfs, au cœur et aux vaisseaux sanguins [47].

Il existe différents types de diabète : le diabète de type 1, le diabète de type 2, le diabète gestationnel ainsi que certains diabètes rares.

### 3.4.1 Symptômes et diagnostic du diabète

Les symptômes du diabète sont le reflet d'une hyperglycémie, autrement dit un taux de sucre dans le sang au-dessus des valeurs normales. Ils peuvent être présents ou non au moment du diagnostic de la maladie et peuvent également survenir lorsque le diabète n'est pas bien géré.

Le diabète de type 1 est le plus souvent diagnostiqué en présence de symptômes qui apparaissent rapidement, tandis que le diagnostic du diabète de type 2 est fréquemment identifié par hasard. Une prise de sang confirme le diagnostic [48].

#### 1. Le diabète de type 1

##### • Ses symptômes

Les symptômes du diabète de type 1 apparaissent brutalement, en quelques jours ou quelques semaines. Les principaux signes caractéristiques de l'hyperglycémie sont :

- une augmentation du besoin d'uriner, qui devient fréquent et survient nuit et jour, avec des urines abondantes (polyurie).
- une augmentation de la soif (polydipsie).
- une perte de poids malgré un appétit qui augmente.
- une fatigue importante ou une vision trouble.

Et si le diabète n'est pas traité, d'autres signes apparaissent :

- **Son diagnostic** Le diabète de type 1 est le plus souvent diagnostiqué en présence de symptômes, il est rarement découvert au cours d'un bilan médical, pour un autre motif.

Le diagnostic de diabète est établi grâce à une prise de sang qui dose le taux de sucre (glycémie) dans le sang. Le diagnostic est posé lorsque :

- La glycémie, à n'importe quel moment de la journée, est supérieure à 2 g/l en présence de symptômes
- La glycémie à jeun est supérieure ou égale à 1,26 g/l, contrôlée à deux reprises en l'absence de symptômes.

Afin de confirmer le résultat du dosage de la glycémie à jeun, le médecin prescrit une seconde prise de sang. Lorsque le diagnostic est posé, un bilan initial permettant le suivi médical du diabète est prescrit par le médecin traitant ou un diabétologue, dont un bilan sanguin visant le dosage de l'HbA1c, ainsi que d'autres examens qui peuvent être nécessaires selon chaque cas.

## 2. Le diabète de type 2

- **Ses symptômes** Le diabète de type 2 évolue souvent en silence (sans aucun symptôme) pendant plusieurs années, car la glycémie augmente très progressivement. Le plus souvent, le diagnostic est fait par hasard, à l'occasion d'une analyse de sang ou lors d'un dépistage du diabète. Occasionnellement, une complication du diabète se manifeste en premier et va permettre le diagnostic.

Des symptômes existent parfois et sont susceptibles d'alerter la personne. Ils peuvent apparaître progressivement, après plusieurs années d'évolution :

- une augmentation du besoin d'uriner.
- une augmentation de la soif.
- une diminution du poids de manière inexplicable alors que l'appétit augmente.
- de la fatigue.
- des démangeaisons au niveau des organes génitaux.
- une cicatrisation très lente d'une plaie
- une vision trouble
- des infections plus fréquentes

- **Son diagnostic** Afin de diagnostiquer un diabète de type 2, une prise de sang réalisée à jeun est nécessaire. Elle permet de mesurer la glycémie. Pour que les résultats soient justes, il faut auparavant respecter un jeûne de huit heures. Le diagnostic est posé lorsque cette glycémie à jeun est égale ou supérieure à 1,26 g/l (ou 7 mmol/l) et est constatée à deux reprises.

- ## 3. Le diabète gestationnel
- Le diabète gestationnel, appelé aussi "diabète de grossesse", survient chez la femme enceinte vers la fin du 2e trimestre. Il peut durer le temps de la grossesse ou être révélateur d'un diabète antérieur. C'est un trouble de la tolérance glucidique conduisant à une hyperglycémie de sévérité variable, débutant ou diagnostiqué pour la première fois pendant la grossesse.

- **Ses symptômes** Le diabète gestationnel peut passer inaperçu, être asymptomatique (sans symptômes) ou présenter des symptômes similaires à ceux des autres types de diabète : soit intense, mictions (urines) fréquentes et abondantes, fatigue importante, etc.

Les principales complications pour la mère et pour l'enfant se situent essentiellement dans la période périnatale. Une fois le diabète découvert, le médecin traitant vous examine et demande plusieurs examens complémentaires dont le bilan biologique qui comporte le dosage de l'hémoglobine glyquée (ou HbA1c) ainsi que d'autres examens selon le cas du patient.

## 3.5 Qu'est-ce que l'hémoglobine glyquée ?

L'hémoglobine glyquée (ou HbA1c) est le reflet de la glycémie. Alors que la glycémie capillaire et la glycémie à jeun sont des instantanés de l'état glycémique, l'HbA1c permet, par un dosage sanguin, d'évaluer l'équilibre glycémique sur une plus longue période (environ deux à trois mois).

Associée à la lecture et à l'interprétation des résultats d'analyses complémentaires, l'HbA1c est un marqueur de risques de complications du diabète à long terme. En effet, chez une personne non diabétique, le glucose se fixe en petite quantité, plus la glycémie est élevée, plus la quantité de glucose fixée sur l'hémoglobine est importante. Une prise de sang réalisée en laboratoire d'analyses tous les trois mois permet de mesurer le taux d'hémoglobine des globules rouges ayant fixé du glucose pendant toute leur durée de vie.

### 3.5.1 Pourquoi l'hémoglobine glyquée ?

Le résultat de cet examen est important, car il permet d'avoir une vision de l'équilibre de votre diabète. Exprimée en pourcentage, puisque cette donnée dépend fortement de la glycémie moyenne, si cette dernière est basse, l'HbA1c sera basse, si elle est élevée, l'HbA1c le sera également. Généralement, un diabète est considéré comme équilibré si le taux d'HbA1c est inférieur ou égal à 7%. Au-delà, le risque de développer des complications à long terme augmente.

### 3.5.2 L'importance de l'HbA1c dans le suivi du diabète

L'objectif de notre travail étant de calculer l'hémoglobine glyquée des patients à partir de leurs glycémies capillaires via une application mobile, il serait intéressant d'en savoir plus au sujet des réels bénéfices que ce projet pourrait apporter aux utilisateurs autrement

dit, aux individus atteints de diabète, mais aussi à leurs médecins traitant qui eux aussi bénéficient de certains avantages :

- **Pour le médecin** : L'ajustement du traitement du patient diabétique est essentiellement basé sur les résultats de l'HbA1c obtenus lors d'analyses sanguines. Comme nous le savons tous, ces dernières peuvent prendre beaucoup de temps en laboratoire et peuvent même, dans certains cas, faire l'objet d'erreurs ce qui retarde voir empêche la prise en charge du patient dans les délais escomptés. Pour cela, la solution que nous proposons aujourd'hui permet non seulement d'éviter ces désagréments, mais aussi d'avoir la possibilité de recevoir les taux d'hémoglobine glyquée par courrier électronique faisant ainsi gagner au médecin, un temps considérable dans l'exercice de sa profession.
- **Pour le patient** : Grâce à cette application mobile facile d'utilisation, il suffit à l'utilisateur d'entrer quotidiennement ses glycémies capillaires pour obtenir son score HbA1c rapidement sans avoir recours à des analyses sanguines coûteuses et peu agréables, lui permettant ainsi d'économiser son temps et son argent en plus du fait qu'il ne soit plus dans l'obligation de se déplacer pour sa consultation. Il est alors pris en charge en ligne, dans les meilleurs délais.

### 3.5.3 Calcul de l'HbA1c :

Afin d'assurer le meilleur suivi possible à ses patients, le médecin se doit de connaître avec exactitude le taux d'hémoglobine glyquée des malades dans le but d'éviter le plus possible, d'éventuelles complications. Fort heureusement, il existe aujourd'hui, une multitude de méthodes mathématiques et statistiques capables de donner des résultats fiables et précis exprimés en pourcentage (%) dont l'une d'elles est donnée par l'équation suivante [49] :

$$HbA1c = \frac{GlycemieMoyenne + 46.7}{28.7} \quad (3.1)$$

Par exemple, si la glycémie moyenne est de 153 mg/dl alors, en appliquant la formule précédemment décrite, nous obtenons :

$$HbA1c = \frac{153 + 46.7}{28.7} = 6.95\% \quad (3.2)$$

Le résultat est  $> 6.5\%$  (ce qui équivaut à un taux se situant dans la plage du diabète), comme le montre la Figure. 3.3 et la Figure. 3.4.



Figure 3.3: Carte de contrôle de l'hémoglobine glyquée [50]

A1C	eAG	
	mg/dl	mmol/l
%		
6	126	7.0
6.5	140	7.8
7	154	8.6
7.5	169	9.4
8	183	10.1
8.5	197	10.9
9	212	11.8
9.5	226	12.6
10	240	13.4

Figure 3.4: Tableau de comparaison du A1C et du eAG [50]

### 3.6 Architecture proposée

Durant cette section, nous allons présenter notre architecture proposée afin de permettre au patient diabétique de contrôler de façon fiable et précise leur hémoglobine glyquée, à partir des glycémies capillaires obtenues chaque matin pendant 90 jours et au médecin traitant le suivi à distance de ses patients. Notre architecture est composée de trois phases à savoir : (1) **Collecte des données**, (2) **Traitement, analyse et classification de données**, (3) **Diagnostic et interprétation des résultats**, comme illustré dans la Figure. 3.5.

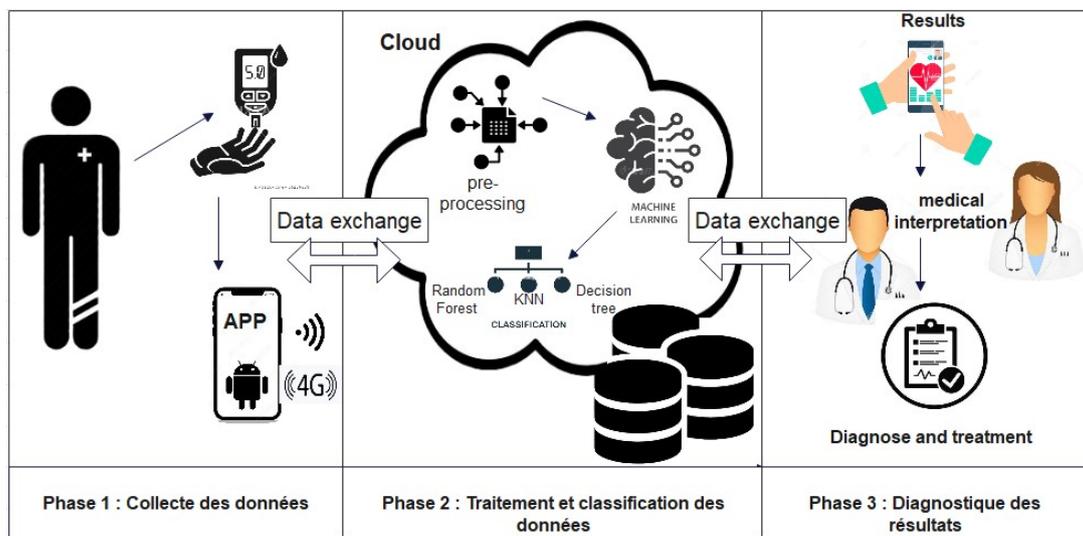


Figure 3.5: Architecture globale du système

Dans ce qui suit, nous présentons une description des trois phases de l'architecture proposée.

- Phase 1 : Collecte des données** Cette étape consiste à collecter les glycémies du patient à l'aide d'un dispositif médical spécialement conçu à cet effet appelé glucomètre. Cet appareil permet à son utilisateur de prélever à l'aide d'un stylo auto-piqueur une goutte de son sang (à partir de l'un de ses doigts), celle-ci sera ensuite contenue dans une bandelette qu'il va ensuite insérer à l'intérieur d'un appareil appelé lecteur de glycémie qui lui indiquera son taux de sucre dans le sang au moment de la prise. Ce type de glycémie est appelé glycémie capillaire. Ce dernier devra alors entrer quotidiennement son taux de glycémie sur son smartphone via notre application mobile pendant 90 jours.
- Phase 2 : Traitement et classification des données** Les informations collectées puis saisies par le patient seront ensuite stockées comme base de données dans le Cloud pour être traitées par divers algorithmes de Machine Learning, dans le but d'effectuer une classification des données du patient. Dans notre cas, nous avons utilisé trois algorithmes très utilisés dans la classification des données : l'algorithme K-NN (K- Nearest Neighbors), Random Forest et Decision Tree.
- Phase 3 : Diagnostic des résultats** Les résultats obtenus après l'exécution des algorithmes sur nos données seront alors communiqués au médecin traitant du patient dans le cadre d'un suivi médical rigoureux.

### 3.7 Les algorithmes utilisés :

Dans le cadre de notre travail, nous avons fait usage de quatre algorithmes différents : le premier est un algorithme conçu pour le traitement de données que nous avons appliqué à notre dataset afin d'obtenir les valeurs de l'hémoglobine glyquée de chaque patient à partir de la formule (1) pour le calcul de l'hémoglobine glyquée, puis de les insérer par la suite à notre dataset. Pour la classification, nous avons choisi trois algorithmes de classification très répandus dans le domaine du machine learning, les algorithmes K-Nearest neighbors (K-NN), forêt aléatoire, ainsi que les arbres de décisions.

L'objectif visé lors de l'utilisation de ces trois algorithmes, est d'effectuer une classification des valeurs de l'hémoglobine glyquée (HbA1c) des patients en trois catégories différentes (Excellent, Bon, Mauvais) à partir du score HbA1c du patient. Avant toute chose, il faut d'abord obtenir la donnée essentielle à la réalisation de cette classification qui est l'hémoglobine glyquée. Pour ce faire, nous avons conçu un algorithme qui consiste à manipuler les données figurant dans notre dataset dans le but de calculer l'hémoglobine glyquée de chaque patient à partir de sa glycémie moyenne.

Le pseudo-code de la Figure. 3.6 expose le raisonnement sur lequel se fonde notre algorithme

**Début Algorithme :**

Données en entrée :  
colonne GM (glycémie moyenne) de notre dataset

Sur la colonne GM faire :

1. Convertir la colonne de valeurs **GM** du dataset en liste de valeurs
2. Créer une nouvelle liste nommée **HbA1c**
3. Pour chaque valeur **X** de la liste **GM** faire  $(X+46.7) / 28.7$
4. Affecter chaque résultat obtenu dans l'étape 3 à la liste vide **HbA1c**
5. Convertir la nouvelle liste obtenue **HbA1c** en colonne de valeurs
6. Insérer la nouvelle colonne de valeurs **HbA1c** à notre dataset

**Fin Algorithme**

**Figure 3.6:** Pseudo-code explicatif de l'algorithme de calcul de l'HbA1c

Après avoir minutieusement étudié l'algorithme permettant de calculer l'hémoglobine glyquée à partir de la glycémie moyenne, que nous avons employé lors de la phase de traitement, il est temps d'aborder de manière détaillée, les trois algorithmes de classification que nous avons choisi :

1. **L'algorithme des k plus proches voisins (K-Nearest Neighbors, KNN) :**  
l'algorithme K-NN est une méthode de machine learning supervisé. Il peut être

utilisé aussi bien pour la régression que pour la classification. Son fonctionnement peut être assimilé à l'analogie "dis-moi qui sont tes voisins, je te dirais qui tu es. . ."

Pour effectuer une prédiction, l'algorithme K-NN se base sur le jeu de données en entier. En effet, pour une observation, qui ne fait pas partie du jeu de données, qu'on souhaite prédire, l'algorithme va chercher les  $K$  instances du jeu de données les plus proches de notre observation.

Ensuite pour ces voisins, l'algorithme se basera sur leurs variables de sortie pour calculer la valeur de la variable de l'observation que l'on souhaite prédire [51]. On peut schématiser le fonctionnement de K-NN avec le pseudo-code de la Figure. 3.7.

#### Début Algorithme

Données en entrée :

- un ensemble de données  $D$ .
- une fonction de définition distance  $d$ .
- Un nombre entier  $K$

Pour une nouvelle observation  $X$  dont on veut prédire sa variable de sortie  $y$  Faire :

1. Calculer toutes les distances de cette observation  $X$  avec les autres observations du jeu de données  $D$
2. Retenir les  $K$  observations du jeu de données  $D$  les proches de  $X$  en utilisation le fonction de calcul de distance  $d$
3. Prendre les valeurs de  $y$  des  $K$  observations retenues :
  1. Si on effectue une régression, calculer la moyenne (ou la médiane) de  $y$  retenues
  2. Si on effectue une classification, calculer le mode de  $y$  retenues
4. Retourner la valeur calculée dans l'étape 3 comme étant la valeur qui a été prédite par K-NN pour l'observation  $X$ .

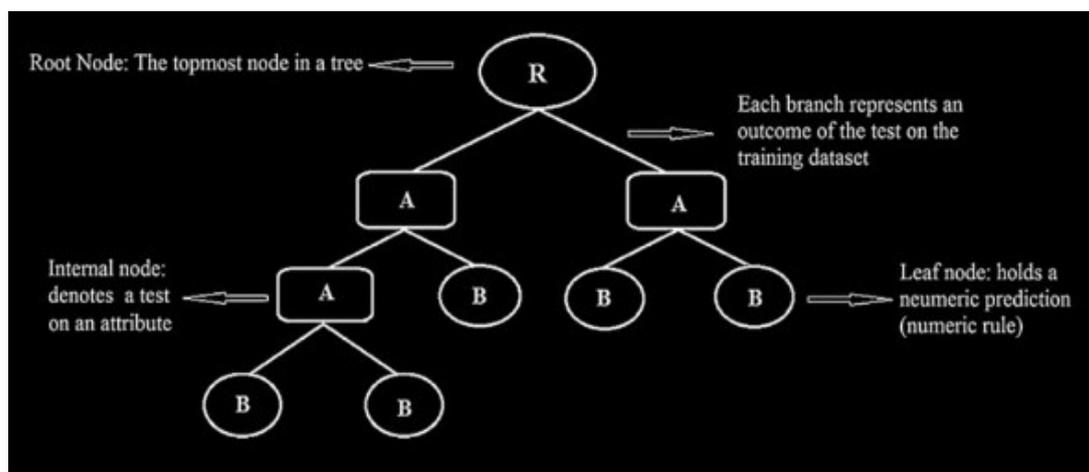
#### Fin Algorithme

**Figure 3.7:** Fonctionnement de l'algorithme K-Nearest Neighbors

Avantages	Inconvénients
Facile à comprendre et à implémenter	Nécessite un calcul de distance pour chaque point de données, ce qui peut être coûteux en termes de temps de calcul pour les ensembles de données volumineux
Peu de paramètres à ajuster	Peut être sensible aux valeurs aberrantes, car il est basé sur la distance entre les points de données
Peut être utilisé pour la classification et la régression	Peut être sensible à la sélection des caractéristiques, car les caractéristiques ayant des échelles de valeurs différentes peuvent influencer de manière disproportionnée la distance entre les points de données
Adapté aux données non linéaires et aux frontières de décision complexes	Le choix du nombre de voisins (valeur de k) peut avoir un impact significatif sur les performances de l'algorithme
Peut être utilisé pour des problèmes multi-classe	Ne fonctionne pas bien avec des ensembles de données de grande dimensionnalité

**Tableau 3.1:** Avantages et inconvénients de l'algorithme KNN

2. **Arbres de décisions (Decision Tree)** : Un arbre de décisions est un algorithme de machine learning supervisé non paramétrique, qui est utilisé à la fois pour les tâches de classification et régression. Il a une structure hiérarchique, une structure arborescente, qui se compose d'un nœud racine, de branches, de nœuds internes et de nœuds feuille. Le mode de fonctionnement de ce type d'algorithme d'apprentissage est expliqué en plus amples détails dans la Figure. 3.8.



**Figure 3.8:** Schéma explicatif du fonctionnement de l'algorithme Decision Tree

Un arbre de décision commence par un nœud racine, qui n'a pas de branches

entrantes. Les branches sortant du nœud racine vont dans les nœuds internes, également connus comme nœuds de décision. Sur la base des caractéristiques disponibles, les deux types de nœud effectuent des évaluations sur des sous-ensembles homogènes, qui sont désignés comme nœuds feuille ou nœuds terminaux. Les nœuds feuilles représentent tous les résultats possibles au sein du fichier.

Ce type de structure en organigramme crée également une représentation facile à assimiler pour la prise de décision, permettant aux différents groupes d'une organisation de mieux comprendre pourquoi une décision a été prise [52].

Avantages	Inconvénients
Facilité d'interprétation : Les arbres de décision sont faciles à comprendre et à interpréter pour les humains. L'arbre peut être visualisé graphiquement, ce qui facilite la compréhension des règles de classification et des décisions prises par l'algorithme	Sensibilité aux données : Les arbres de décision sont sensibles aux données d'entrée et peuvent facilement surajuster (overfitting) aux données d'apprentissage. Cela signifie que l'arbre de décision peut être très précis sur les données d'apprentissage, mais ne pas généraliser correctement sur les données de test
Rapidité : Les arbres de décision sont rapides à entraîner et à prédire, même sur de grands ensembles de données. Ils peuvent également être utilisés pour la classification en temps réel	Instabilité : Les arbres de décision peuvent être instables car de petites variations dans les données d'entrée peuvent entraîner de grands changements dans la structure de l'arbre
Robustesse aux données manquantes : Les arbres de décision sont robustes aux données manquantes et peuvent facilement gérer les données manquantes sans avoir besoin de remplacer les valeurs manquantes	Biais de classe : Les arbres de décision ont tendance à être biaisés en faveur des classes majoritaires, ce qui peut conduire à une classification incorrecte des classes minoritaires
Adaptabilité : Les arbres de décision peuvent être utilisés pour différents types de données (numériques, catégorielles, binaires, etc.) et pour différents types de problèmes (classification, régression, clustering, etc.)	Risque d'overfitting : Les arbres de décision peuvent facilement surajuster les données d'apprentissage, en créant des règles trop complexes pour représenter les données. Cela peut rendre l'arbre de décision moins utile pour la prédiction de nouvelles données
Performances compétitives : Les arbres de décision peuvent atteindre des performances compétitives par rapport à d'autres algorithmes de classification et de prédiction	Interprétation difficile : Les arbres de décision peuvent devenir très grands et complexes, ce qui peut rendre difficile leur interprétation et leur compréhension pour les utilisateurs

**Tableau 3.2:** Avantages et inconvénients de l'algorithme arbre de décision

### 3. L'algorithme de forêt aléatoire (Random Forest) : Random Forest est un

algorithme de machine learning robuste qui peut être utilisé pour une variété de tâches, y compris la régression et la classification. Il s'agit d'une méthode d'ensemble, ce qui signifie qu'un modèle de forêt aléatoire est composé d'un grand nombre de petits arbres de décision, appelés estimateurs, qui produisent chacun leurs propres prédictions. Le modèle de forêt aléatoire combine les prédictions des estimateurs pour produire une prédiction plus précise.

Les classificateurs d'arbre de décision standard ont l'inconvénient d'être enclins à s'adapter à l'ensemble d'apprentissage. La conception d'ensemble de la forêt aléatoire permet à la forêt aléatoire de compenser cela et de bien généraliser les données invisibles, y compris les données avec des valeurs manquantes. Les forêts aléatoires idéales pour les problèmes de classification, mais sont moins performantes pour les problèmes de régression.

Parce qu'elles sont extrêmement robustes, faciles à démarrer, adéquates pour les types de données hétérogènes et ont très peu d'hyper-paramètres, les forêts aléatoires sont souvent le premier port d'escale lors du développement d'un nouveau système d'apprentissage automatique, car elles permettent d'obtenir un aperçu rapide du type de précision qui peut raisonnablement être atteint sur un problème. Le principe de fonctionnement de l'algorithme de Random Forest est défini comme dans la Figure. 3.9.

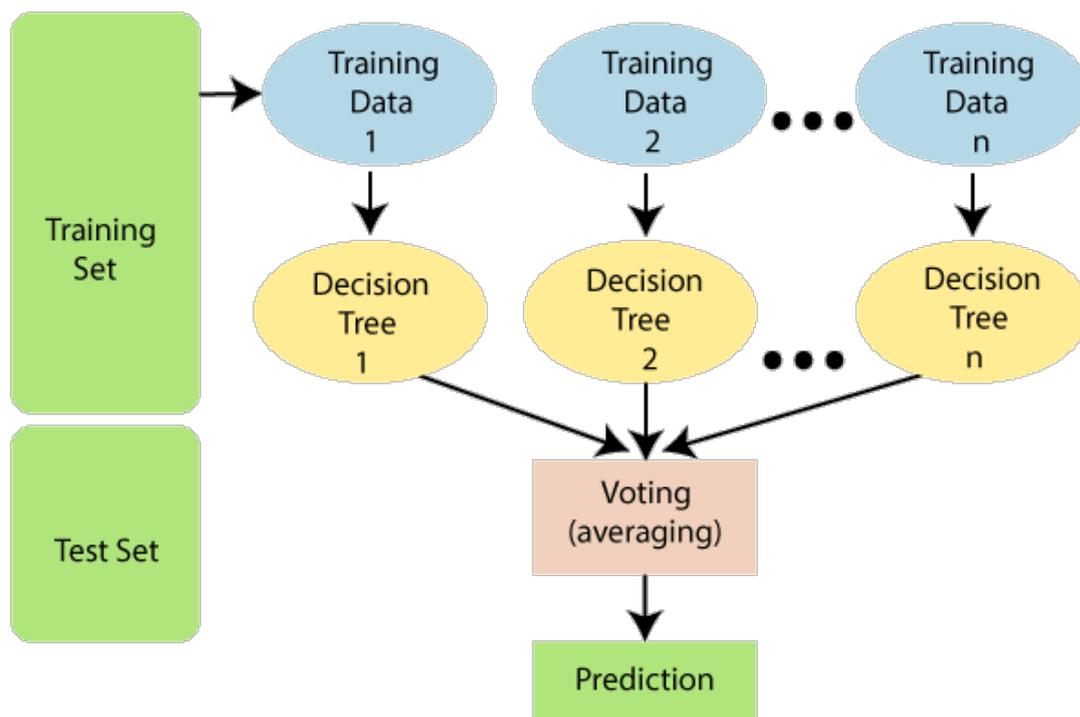


Figure 3.9: Fonctionnement de l'algorithme Random Forest

Avantages	Inconvénients
Bonne précision de prédiction : Les Random Forest sont capables de fournir une bonne précision de prédiction dans une variété de situations, en particulier lorsque les données sont bruyantes ou contiennent des valeurs manquantes	Surapprentissage : Les Random Forest peuvent souffrir de sur-apprentissage si le nombre d'arbres dans la forêt est trop élevé ou si les données sont bruyantes
Facilité d'utilisation : Les Random Forest sont relativement simples à mettre en œuvre et ne nécessitent pas beaucoup de préparation de données ou de réglages de paramètres	Interprétabilité : Les Random Forest peuvent être difficiles à interpréter en raison du grand nombre d'arbres et des relations complexes qu'ils capturent
Robustesse aux données manquantes : Les Random Forest peuvent fonctionner avec des données incomplètes, ce qui est un avantage par rapport à de nombreuses autres techniques d'apprentissage automatique	Taille de l'ensemble de données : Les Random Forest peuvent ne pas fonctionner efficacement avec des ensembles de données très grands en raison du temps de traitement requis pour construire de nombreux arbres
Traitement des données non linéaires : Les Random Forest peuvent capturer des relations non linéaires entre les caractéristiques et les résultats, ce qui les rend adaptés aux problèmes de classification et de régression non linéaires	Temps de traitement : Les Random Forest peuvent être lents à traiter en raison du grand nombre d'arbres nécessaires pour obtenir une précision élevée

**Tableau 3.3:** Avantages et inconvénients de l'algorithme Forêt aléatoire

### 3.8 Déroulement de l'architecture

Après avoir décrit notre architecture, nous allons donner le déroulement de cette dernière sous forme d'étapes, allant de l'entrée des informations par l'utilisateur jusqu'à l'obtention des résultats. Ces étapes sont représentées dans la Figure. 3.10.

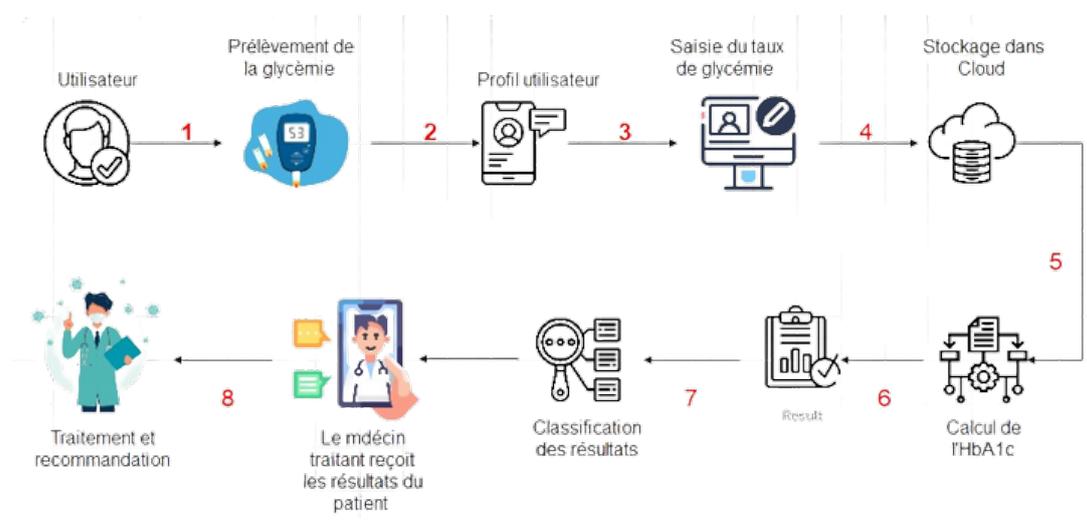


Figure 3.10: Schéma décrivant l'ensemble des étapes détaillées du fonctionnement du système

Cette opération consiste à suivre les étapes suivantes :

- **Étape 1** : À l'aide de son glucomètre, le patient procède à un prélèvement d'une goutte de sang de son doigt pour l'introduire à l'intérieur de l'appareil. Le lecteur de glycémie affiche ensuite la valeur de la glycémie du patient à l'instant "t", ce type de prélèvement est appelé glycémie capillaire.
- **Étape 2** : Le patient accède à la première interface de l'application afin de se créer un profil utilisateur en introduisant ses informations personnelles. Cette étape nous permet de récupérer les données de l'utilisateur dans le but de les utiliser ultérieurement.
- **Étape 3** : Après s'être identifié, il est alors demandé à l'utilisateur d'introduire sa glycémie capillaire quotidiennement durant trois mois.
- **Étape 4** : La glycémie sera ensuite stockée comme donnée de traitement dans le but de constituer une base de données.
- **Étape 5** : Après avoir récupéré l'ensemble des valeurs stockées, notre algorithme utilisé procède alors au calcul de l'HbA1c conformément à la formule (1)
- **Étape 6** : Notre algorithme retourne la valeur de l'HbA1c obtenue qui servira de valeur d'entrée à notre algorithme de classification.
- **Étape 7** : L'algorithme de classification (K-NN, Random Forest, Arbre de décision) se charge de classer chaque valeur d'HbA1c recueillie dans l'une des trois classes suivantes (Excellent, Bon, Mauvais)
- **Étape 8** : Les informations du patient sont alors envoyées à son médecin traitant pour garantir le meilleur suivi possible selon son cas.

## 3.9 Conclusion

Durant ce chapitre, diverses notions ont été abordées : Tout d'abord, nous avons commencé par introduire le diabète en citant sa définition, ses caractéristiques ainsi que la manière d'assurer sa bonne prise en charge. Ensuite, la définition de l'hémoglobine glyquée suivie de sa méthode de calcul ont été expliquées afin d'apprendre à distinguer les différents états possibles d'un patient diabétique. Pour finir, nous terminons ce chapitre en présentant l'architecture de notre travail par le biais d'un schéma explicatif résumant les différentes étapes nécessaires à l'aboutissement de notre projet.

Après avoir proposé une architecture représentative de notre travail, l'étape suivante consistera à l'implémenter dans le but de permettre aux patients atteints de diabète d'avoir un suivi rapide et facile de leurs pathologies, et ce via une application mobile qui retourne leurs scores d'hémoglobine glyquée. Cette étape sera mise en exergue en plus amples détails lors du chapitre suivant.

# Chapitre 4

## Implémentation d'un système basé sur une architecture pour le suivi des patients diabétiques

### 4.1 Introduction

Au cours du chapitre précédent, nous avons pu concevoir l'architecture représentant notre travail et identifier les diverses étapes nous permettant de réaliser nos objectifs, et ce, dans un cadre purement théorique.

Lors de ce quatrième et dernier chapitre, il sera question de mettre en pratique l'aspect théorique de notre travail mis en œuvre jusque-là. Pour ce faire, l'ensemble des technologies et environnements de développement nécessaires à la réalisation de notre projet seront exposés, suivi de la présentation du prototype de notre application mobile : interface, interaction avec l'utilisateur, récupération des données à partir de la base de données. Par la suite, la partie "data" représentant le corps de notre travail est abordée : présentation du dataset utilisé, data pre-processing, suivi des résultats obtenus par chaque algorithme de classification employé. Pour finir, une étude comparative entre les différents algorithmes utilisés viendra clôturer ce chapitre.

### 4.2 Environnement de développement

Durant cette partie, l'ensemble des outils utilisés pour la mise en œuvre de notre système, ainsi que ceux utilisés lors de la création de notre application seront exposés.

### 4.2.1 Anaconda

Anaconda est un outil en distribution libre et open source destiné à la programmation dans différents environnements comme Python et R. Ces environnements, également appelés environnements de développement intégrés (IDE), sont des plates-formes ou des applications qui facilitent grandement le développement du code.

Il est utilisé dans de nombreux domaines tels que la science de données, le machine learning et l'intelligence artificielle, et ceci revient au fait qu'il contient plusieurs packages nécessaires dans ces domaines, notamment Python, Numpy, Panda, Jupyter, etc. Et, tout comme le langage Python, anaconda est multiplateforme.

Cet outil qui recense plus de 20 millions d'utilisateurs dans le monde comprend entre autres :

- une installation de l'environnement Python.
- des IDE (environnement de développement intégré) de dernière génération à l'instar de Jupyter ou de Spyder
- des packages de Data Science comme Panda, Numpy, Scikit-Learn. . .
- L'outil Conda pour la gestion des environnements et des répertoires de package.

Anaconda assure grâce à ses différents outils et à l'environnement Python, une collecte et une transformation à grande échelle des données. Jupyter, l'un des IDE présents dans Anaconda, prend en charge désormais plus de 40 langages de programmation [53].

Anaconda est l'une des meilleures façons pour installer et utiliser python. Cela se fait à travers GUI (interface utilisateur graphique) qui est simple et bien prise en charge en plus de comporter les bibliothèques les plus importantes et IDE dans son installation. Elle simplifie également le processus de mise à jour de toutes ces bibliothèques. Ainsi, plutôt que d'installer Python séparément avec différents IDE, bibliothèques et fonctionnalités, Anaconda le fait dans une seule installation. La Figure. 4.1 illustre l'interface du navigateur anaconda.

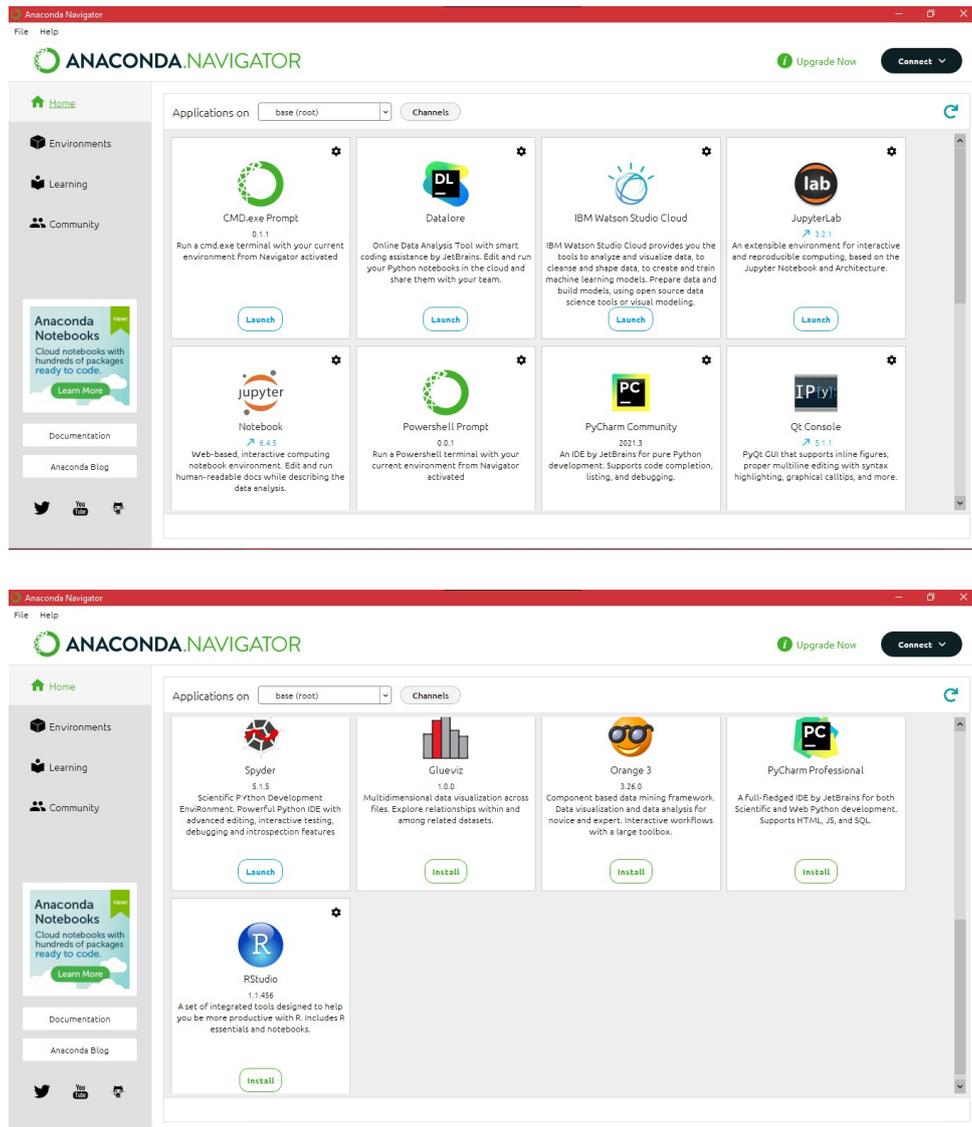


Figure 4.1: Page d'accueil du Navigateur Anaconda [53]

## 4.2.2 Spyder

Spyder est un acronyme signifiant « Scientific PYTHON Development EnviRONment ». C'est un environnement scientifique gratuit et open source écrit en Python, pour Python, et conçu par et pour des scientifiques, des ingénieurs et des analystes de données.

Il présente une combinaison unique des fonctionnalités avancées d'édition, d'analyse, de débogage et de profilage d'un outil de développement complet avec l'exploration de données, l'exécution interactive, l'inspection approfondie et les belles capacités de visualisation d'un package scientifique. La Figure. 4.2 montre l'interface d'accueil de l'IDE [54].

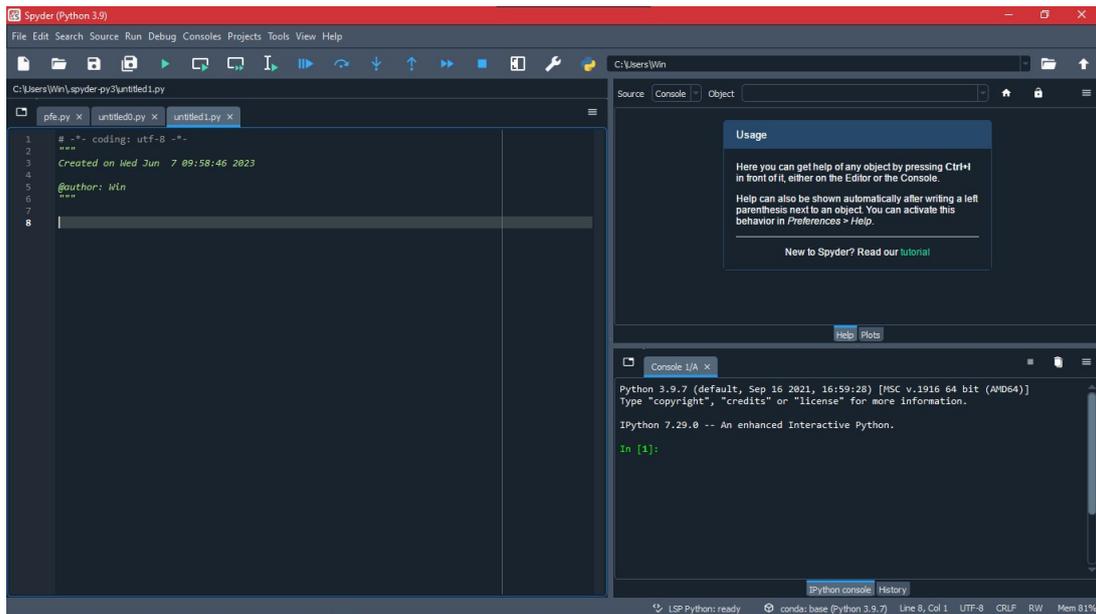


Figure 4.2: Interface d'accueil de l'environnement de développement Spyder

### 4.2.3 kodular

Fondé le 6 juillet 2017 par Conor Shipp, Diego Barreiro, Mika, Pavitra Golchha, Sander Jochems, Sivagiri Visakan et Vishwas Adiga, Kodular est un site en ligne qui fournit principalement une application web en ligne permettant de développer des applications sans avoir besoin de coder. Cependant, ce n'est pas un projet open source et le code source est hébergé en privé sur GitHub. Kodular fournit également une boutique d'applications en ligne comme alternative au Google Play Store et un IDE d'extensions permettant aux utilisateurs avancés de coder de nouveaux composants pour Kodular. la page d'accueil est illustrée par la Figure. 4.3.

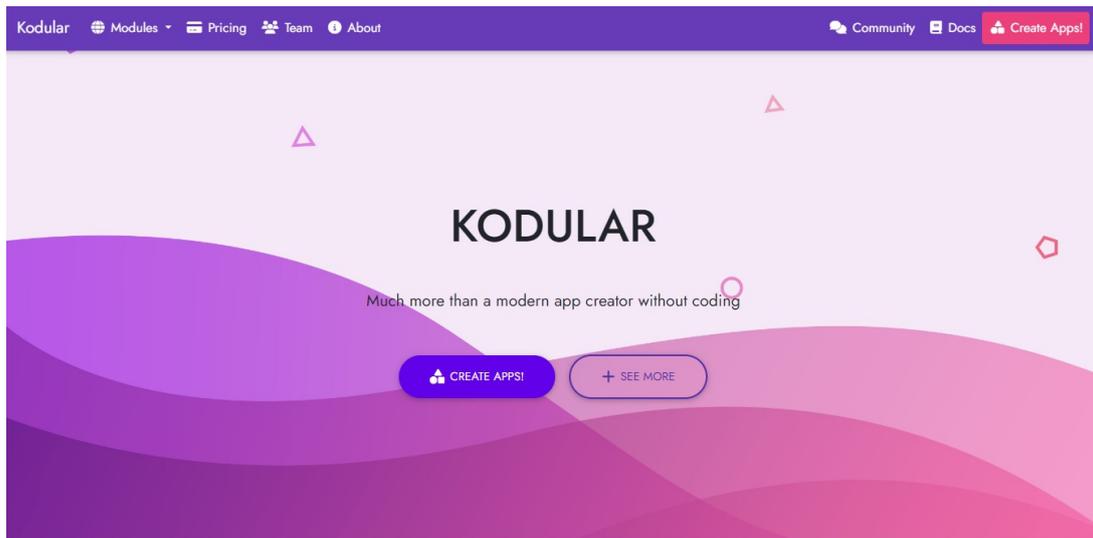


Figure 4.3: Page d'accueil de la plateforme en ligne KODULAR

#### 4.2.4 Firebase

Firebase est une plateforme de développement d'applications mobiles de Google dotée de puissantes fonctionnalités pour le développement, la manipulation et l'amélioration des applications. Firebase est fondamentalement un ensemble d'outils sur lesquels les développeurs peuvent compter, créant des applications et les développant en fonction de la demande.

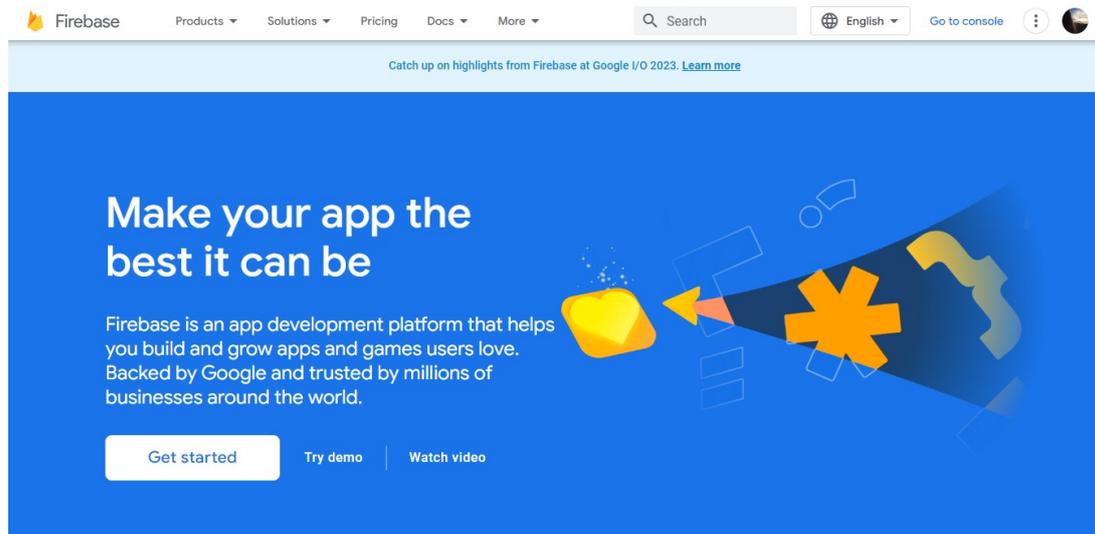
Firebase vise à résoudre trois problèmes principaux pour les développeurs :

- Créer une application, rapidement.
- Publier et superviser une application en toute confiance.
- Faire participer les utilisateurs.

Les développeurs qui s'appuient sur cette plateforme ont accès à des services qu'ils devraient développer eux-mêmes, et cela leur permet de se concentrer sur la fourniture d'expériences d'applications robustes.

Parmi les caractéristiques les plus remarquables de la plateforme Google Firebase, citons les bases de données, l'authentification, les messages « push », l'analyse, le stockage de fichiers, et bien plus encore.

Comme les services sont hébergés dans le cloud, les développeurs peuvent effectuer une mise à l'échelle à la demande sans aucun problème. Firebase est actuellement l'une des principales plateformes de développement d'applications sur lesquelles s'appuient les développeurs du monde entier [55]. la Figure. 4.4 représente l'interface d'accueil de la plateforme.



**Figure 4.4:** Page d'accueil de la plateforme Firebase

## 4.2.5 Langage python

Python est un langage de programmation puissant et facile à apprendre. Il dispose de structures de données de haut niveau et permet une approche simple, mais efficace de la programmation orientée objet. Il dispose d'une syntaxe élégante, un typage dynamique et il est facilement interprété, Python est un langage idéal pour l'écriture de scripts et le développement rapide d'applications dans de nombreux domaines et sur la plupart des plateformes [56].

Ce langage de programmation présente de nombreuses caractéristiques intéressantes :

- Il est multiplateforme : il fonctionne sur de nombreux systèmes d'exploitation : Windows, Mac OS X, Linux, Android, iOS, depuis les mini-ordinateurs Raspberry Pi jusqu'aux supercalculateurs.
- Il est gratuit : vous pouvez l'installer sur autant d'ordinateurs que vous voulez.
- C'est un langage de haut niveau : Il demande relativement peu de connaissance sur le fonctionnement d'un ordinateur pour être utilisé.
- C'est un langage interprété : Un script Python n'a pas besoin d'être compilé pour être exécuté, contrairement à des langages comme le C ou le C++.
- Il est orienté objet : C'est-à-dire qu'il est possible de concevoir en Python des entités qui miment celles du monde réel (une cellule, une protéine, un atome, etc.) avec un certain nombre de règles de fonctionnement et d'interactions.
- Il est relativement simple à prendre en main.
- Enfin, il est très utilisé en bio-informatique et plus généralement en analyse de données.

#### 4.2.5.1 Le python et l'IA

Python est également très populaire dans le domaine de l'IA en raison de ses bibliothèques de haute qualité et de sa puissance de calcul. Beaucoup de chercheurs et de développeurs utilisent Python pour implémenter et expérimenter avec de nouveaux algorithmes d'apprentissage automatique et de deep learning, en particulier grâce aux frameworks tels que TensorFlow et PyTorch. En utilisant Python, il est possible de développer rapidement des modèles complexes et puissants, ce qui en fait un choix idéal pour les projets d'IA de toutes tailles [57].....

#### 4.2.5.2 Les bibliothèques utilisées

Les bibliothèques ne sont rien d'autre que des collections de modules avec du code pré-écrit et peuvent être facilement importées et utilisées par les développeurs pour implémenter n'importe quelle fonctionnalité. Le ML nécessite un traitement de données régressif et continu, et Python répond à cette exigence en accédant aux données, en les manipulant et en les transformant.

Les bibliothèques utilisées sont [58] :

- **Pandas** : Pandas est une bibliothèque Python de bas niveau construite sur NumPy. Elle comporte plusieurs avantages qu'on peut citer :
  1. **Des dataframes puissants** :Pandas est surtout utilisé pour l'analyse et la manipulation des données, ainsi que pour les opérations de machine learning sous la forme de dataframes. En utilisant les dataframes, les développeurs peuvent facilement avoir une vue d'ensemble des données pour garantir une meilleure qualité du produit.
  2. **Traitement flexible des données** : Les développeurs utilisent cette bibliothèque pour structurer, remodeler et filtrer de grands ensembles de données.
- **Matplotlib** : Une unité de NumPy, SciPy, et Matplotlib censée remplacer le besoin d'utiliser le langage statistique propriétaire MATLAB. Cette bibliothèque est surtout appréciée pour le fait qu'elle comporte un ensemble complet d'outils de traçage. En effet, on peut trouver des diagrammes, des schémas 2D et 3D, des graphiques et d'autres outils de visualisation permettent aux scientifiques d'effectuer une analyse détaillée des données. Sur la base de cette analyse, un programmeur peut construire des modèles de machine learning fiables.
- **Scikit-learn** : Scikit-learn est la principale bibliothèque d'outils dédiés au machine learning et à la data-science dans l'univers Python. Celle-ci est considérée comme la bibliothèque par excellence pour les algorithmes de machine learning classique,

ceci est dû au fait qu'elle comprend des algorithmes de classification pour la détection des spams (pourriels) et la reconnaissance d'images, des algorithmes de régression pour la prédiction, des algorithmes de regroupement pour la segmentation des clients et des opérations similaires, la sélection de modèles pour améliorer la précision des calculs, etc.

## 4.3 Implémentation et résultats

### 4.3.1 Présentation du prototype de l'application mobile

Nous avons développé un prototype d'application mobile permettant de collecter les données des patients diabétiques (signes vitaux + informations personnelles) puis de les stocker localement de façon temporaire. Cette application joue le rôle de passerelle, elle récupère les données du patient pour ensuite les envoyer au cloud pour un stockage permanent. Avant d'aborder les différentes interfaces qui composent notre application mobile, il est nécessaire de justifier l'appellation de cette dernière.

**Pourquoi Glyco-Wise ?** Le nom "Glyco-Wise" est un mot composé du diminutif de "glycogène" ou "glycogen" en anglais qui représente les réserves de sucre de l'organisme, et du mot anglais "wise" qui signifie prudence et sagesse. Dans sa globalité, "Glyco-Wise" signifie surveillance du taux de sucre, un élément essentiel à la prise en charge du diabète.

#### 4.3.1.1 Écran de démarrage

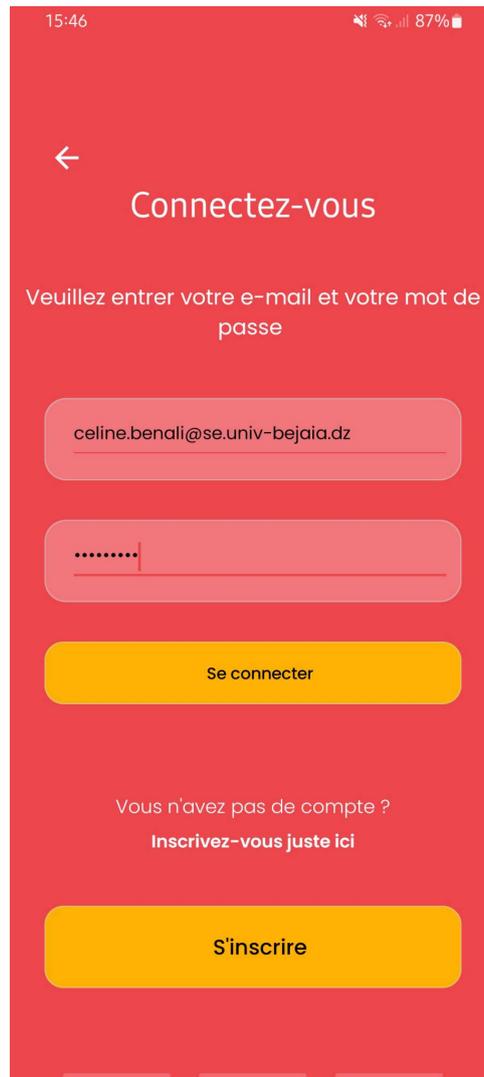
Au lancement de l'application, nous rencontrons la première interface illustrée dans la figure Figure. 4.5. Cette interface comporte le logo de l'application ainsi qu'un message introductif et un seul bouton "Continuer" qui nous permet de passer à l'écran suivant.



Figure 4.5: Écran de démarrage de l'application

#### 4.3.1.2 Espace connexion

La seconde interface qui apparaît après l'écran de démarrage est l'interface réservée à la connexion, si l'utilisateur dispose déjà d'un compte sur l'application alors, il sera invité à saisir son adresse mail ainsi que son mot de passe dans les champs dédiés à ces derniers, et cela, afin d'accéder à l'application comme le montre la Figure. 4.6.

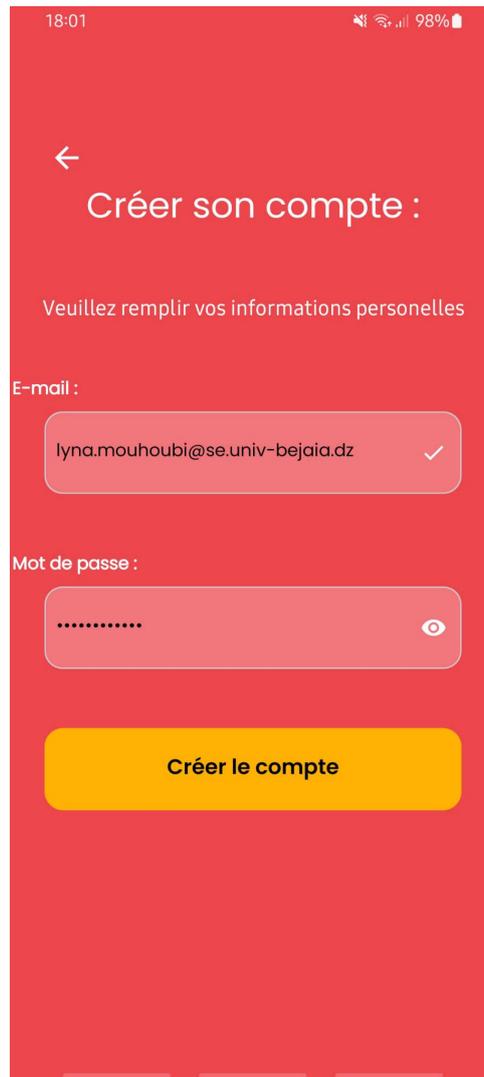


**Figure 4.6:** Interface de connexion à l'application

Dans le cas où l'utilisateur ne dispose pas d'un compte et que c'est sa première interaction avec l'application, il pourra cliquer sur le bouton inscription et sera dirigé vers l'écran conçu à cet effet.

#### 4.3.1.3 Espace inscription

Une fois que l'utilisateur clique sur le bouton "S'inscrire", un nouvel écran se présente à lui, celui-ci comporte deux champs, l'un pour saisir son adresse e-mail, l'autre pour le mot de passe. Ceci est illustré par la Figure. 4.7.



**Figure 4.7:** Interface d'inscription à l'application

Dès que l'utilisateur aurait cliqué sur "Créer le compte" un nouvel écran se présente à lui et celui-ci se présente sous forme de fiche contenant les différentes informations personnelles que l'utilisateur devrait saisir à fin de compléter son profil, telles que : le nom, le prénom, l'âge, le poids, la taille, le sexe ainsi que son type de diabète (type 1 ou type 2) comme le montre la Figure. 4.8.

12:42 100%

← Complétez votre fiche personnelle :

Prénom :

Nom :

Taille :

Poids :

Âge :

Sexe :  Femme  Homme

Type de diabète :  Type 1  Type 2

**Continuer**

(a) Fiche personnelle vide

12:42 100%

← Complétez votre fiche personnelle :

Prénom :

Nom :

Taille :

Poids :

Âge :

Sexe :  Femme  Homme

Type de diabète :  Type 1  Type 2

**Continuer**

(b) Fiche personnelle remplie

Figure 4.8: Fiche personnelle

#### 4.3.1.4 Menu principal

Une fois que l'utilisateur a terminé son inscription, ou s'est connecté à son compte déjà existant, ce dernier se trouve face à l'écran comportant le menu principal comme le montre la Figure. 4.9. Le menu principal se compose d'un champ permettant à l'utilisateur de saisir sa glycémie du jour, mais également un champ sur lequel il pourrait mentionner des symptômes qu'il aurait rencontrés qui pourraient aider son médecin traitant par la suite.

Le menu principal se compose également de deux boutons situés en haut de l'écran, l'un sert à ouvrir le menu latéral tandis que le deuxième bouton servira à mener l'utilisateur vers l'interface qui englobe ses informations personnelles.



**Figure 4.9:** Le menu principal

Le menu latéral est illustré dans la Figure. 4.10. Ce dernier se compose de deux boutons “Accueil” et “Historique des glycémies”.

Le bouton “accueil” va servir à revenir à la page du menu principal, tandis que le bouton “Historique des glycémies” une fois sensibilisé, affichera un historique des glycémies saisies par l'utilisateur.



**Figure 4.10:** Le menu latéral

#### 4.3.1.5 Profil

L'interface représentant le profil de l'utilisateur est représentée par la Figure. 4.11. Celle-ci représente un sommaire des informations personnelles saisies précédemment.



**Figure 4.11:** Le menu latéral

#### 4.3.1.6 Historique

L'utilisateur, après avoir saisi son taux de glycémie ainsi que les potentiels symptômes qu'il aurait pu présenter ce jour-là, peut accéder à l'écran présentant l'historique des glycémies et des symptômes, en cliquant sur le bouton "Historique des glycémies" se trouvant sur le menu latéral du menu principal. La Figure. 4.12 montre un exemple d'historique, tel que la Figure. 4.12a montre un exemple d'un nouveau compte avec une seule saisie de glycémie et de symptômes, tandis que la Figure. 4.12b montre un historique d'un compte d'un patient qui a déjà saisi ces informations à plusieurs reprises.

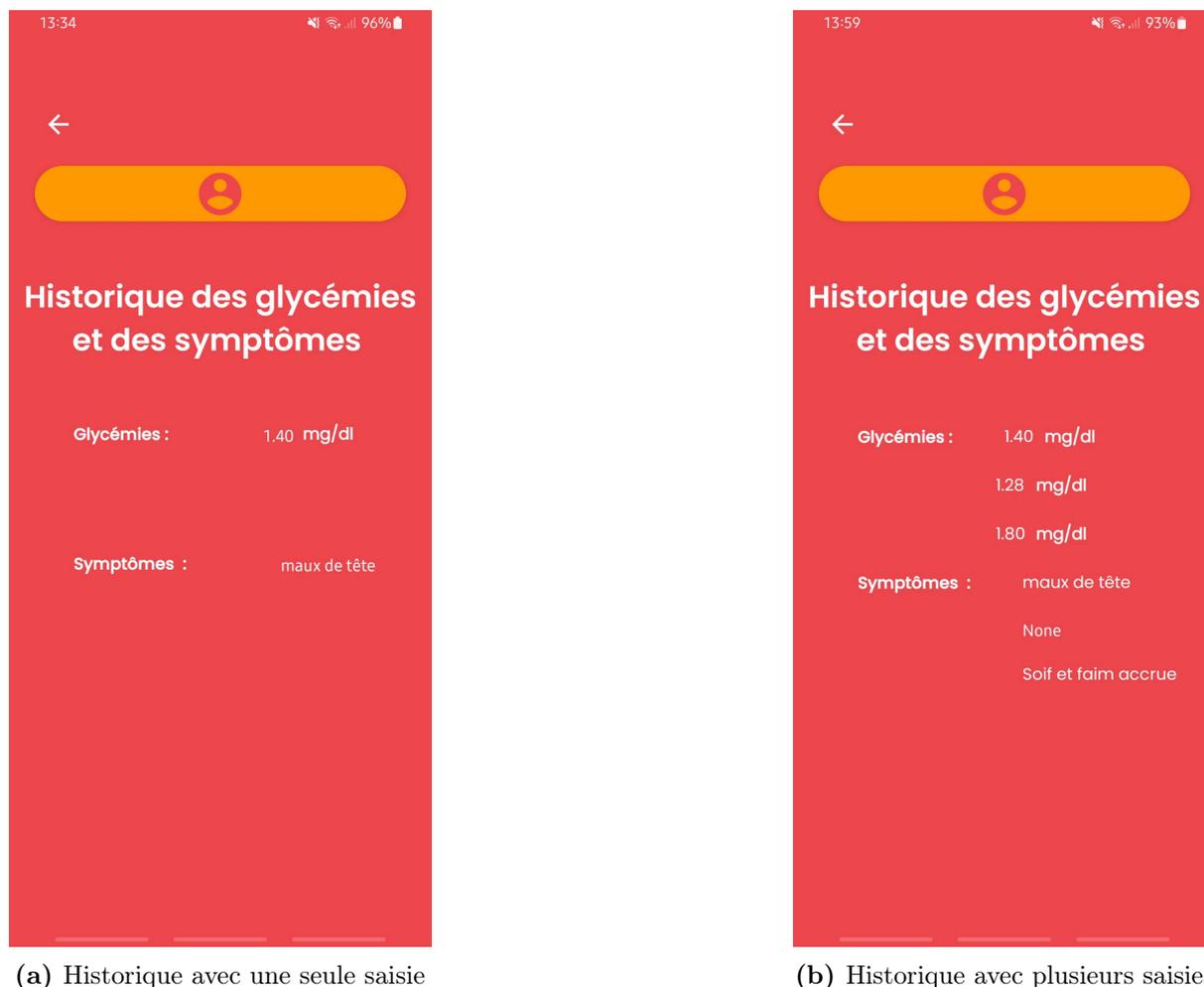


Figure 4.12: Historique des glycémies et des symptômes

## 4.3.2 Description du dataset

### 4.3.2.1 Présentation des données

Cette section portera sur l'aspect traitement des données qui reste l'objectif principal de ce projet et qui se compose de deux étapes clés qui feront l'objet d'une description complète et détaillée : le prétraitement et la classification des données.

Pour ce travail, nous avons utilisé un dataset intitulé 'Diabetes data set' provenant du site de l'Université de Californie (UC Irvine) machine learning repository, Center for machine learning and intelligent systems [59].

Les dossiers des patients diabétiques ont été obtenus à partir de deux sources : la première est un dispositif d'enregistrement électronique automatique, la deuxième constitue l'ensemble des dossiers en version papier. Le fichier de base se compose de quatre champs qui représentent respectivement la date au format MM-JJ-AAAA, l'heure au

format XX :AA, un code ainsi que la valeur de la glycémie.

Le dataset regroupe les dossiers de 70 patients où la glycémie quotidienne de chacun a été enregistrée et ceci représente l'unique champ qui a été pris en considération lors de notre étude.

#### 4.3.2.2 Prétraitement des données

Dans le cadre de la réalisation de ce travail, nous avons été capables de sélectionner certaines données figurant sur un dataset initial afin d'y extraire uniquement les colonnes de valeurs qui nous intéressent. Ensuite, une opération de filtrage a été effectuée dans le but d'éliminer les valeurs aberrantes pouvant altérer de manière significative les prochaines étapes, ce processus est appelé nettoyage de données.

Après avoir calculé la glycémie moyenne de chaque patient, nous les avons regroupées dans un fichier (type .csv) comportant deux champs :

1. Le champ patient représente L'id du patient.
2. Le champ GM représente la glycémie moyenne du patient.

Un aperçu d'une partie de notre dataset après le nettoyage de données est illustré par la figure Figure. 4.13.



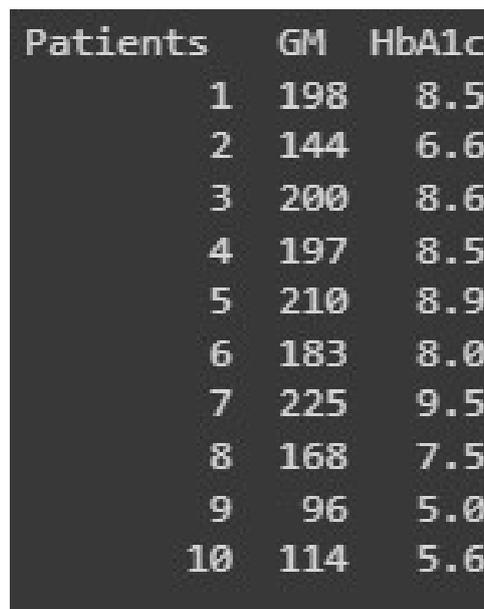
Patients	GM
1	198
2	144
3	200
4	197
5	210
6	183
7	225
8	168
9	96
10	114
11	278

**Figure 4.13:** Aperçu du dataset après nettoyage des données

Après avoir éliminé les valeurs aberrantes de la colonne de valeurs des glycémies

moyennes (GM) de notre dataset, la prochaine étape consiste à calculer à partir des valeurs de la colonne GM, les valeurs de l'hémoglobine glyquée (HbA1c) propres à chacune de ces valeurs en appliquant la formule (1) (Voir section 3.6).

Les résultats obtenus sont alors regroupés en une seule colonne de valeurs nommée HbA1c qui représente le taux d'hémoglobine glyquée correspondant à chaque valeur de la colonne GM que nous avons ajouté à notre dataset déjà existant comme le montre la Figure. 4.14.



Patients	GM	HbA1c
1	198	8.5
2	144	6.6
3	200	8.6
4	197	8.5
5	210	8.9
6	183	8.0
7	225	9.5
8	168	7.5
9	96	5.0
10	114	5.6

**Figure 4.14:** Aperçu du dataset après l'ajout de la nouvelle colonne de valeurs HbA1c

Maintenant que notre dataset comporte l'ensemble des données dont nous avons besoin, nous pouvons passer à la prochaine étape qui est la classification des données.

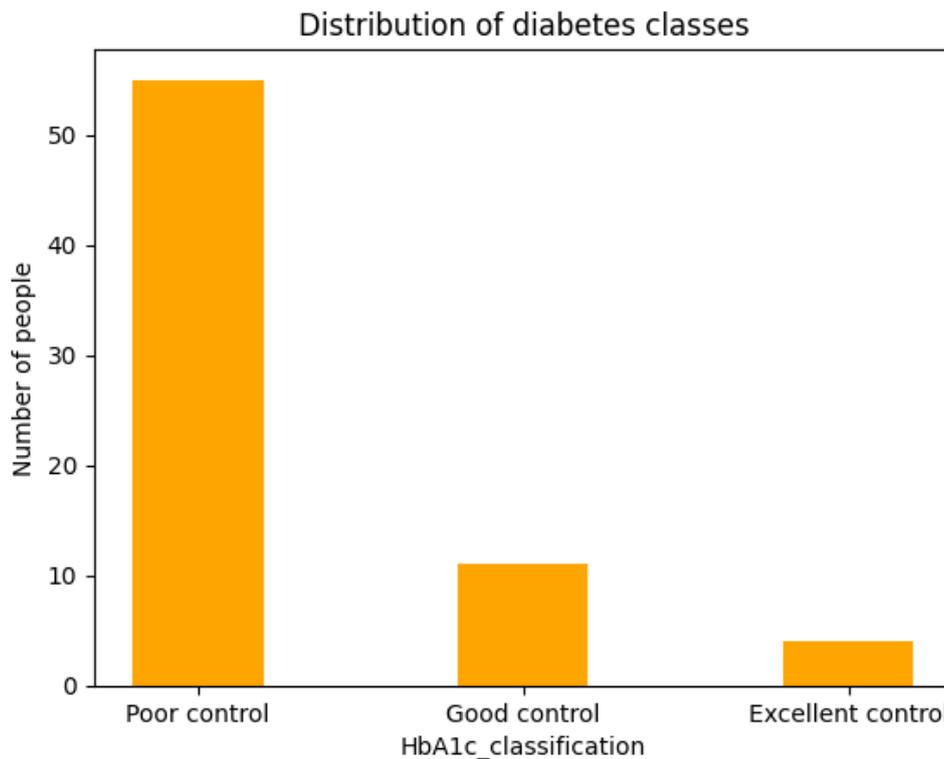
## 4.4 Résultats et évaluation

### 4.4.1 Classification des données

Une fois le score HbA1c de chaque patient calculé, nous passons à la dernière étape de notre travail, la classification. Dans notre étude, nous avons choisi d'appliquer à notre Dataset trois types d'algorithmes de classification différents : K-Nearest Neighbors (KNN), Arbre de décisions (DT) ainsi que Random Forest (RF). Les résultats de cette classification sont mis en évidence par la Figure. 4.15 :

Patients	GM	HbA1c	HbA1c_classification
1	198	8.5	Poor control
2	144	6.6	Poor control
3	200	8.6	Poor control
4	197	8.5	Poor control
5	210	8.9	Poor control
6	183	8.0	Poor control
7	225	9.5	Poor control
8	168	7.5	Poor control
9	96	5.0	Excellent control
10	114	5.6	Excellent control
11	278	11.3	Poor control
12	303	12.2	Poor control
13	164	7.3	Poor control
14	219	9.3	Poor control
15	124	5.9	Good control
16	80	4.4	Excellent control
17	276	11.2	Poor control
18	158	7.1	Poor control
19	84	4.6	Excellent control
20	183	8.0	Poor control

(a) Résultats de la classification des patients selon leurs scores HbA1c



(b) Histogramme illustrant la répartition des patients sur les trois classes (Excellent control, Good control, Bad control) à partir de leurs scores HbA1c

**Figure 4.15:** Résultats de la classification

Afin d'établir la meilleure comparaison possible entre ces trois algorithmes, une éva-

luation de leurs performances, basée sur quatre métriques d'évaluation différentes est effectuée, ces quatre métriques sont Accuracy, Precision, Recall, et F1 Score.

Les valeurs obtenues lors du calcul des métriques varient en fonction des algorithmes de classification employés ainsi que du dataset étudié.

La Figure. 4.16 regroupe l'ensemble des formules permettant de calculer ces quatre métriques.

- TP** : Vrai positif
  - TN** : Vrai négatif
  - FP** : Faux positif
  - FN** : Faux négatif
- Un **vrai positif (TP)** est un résultat où le modèle prédit correctement la classe positive.
- Un **vrai négatif (TN)** est un résultat où le modèle prédit correctement la classe négative.
- Un **faux positif (FP)** est un résultat où le modèle prédit de manière incorrecte la classe positive.
- Un **faux négatif (FN)** est un résultat où le modèle prédit de manière incorrecte la classe négative.
- Classe positive** : représente le résultat recherché
- **Classe négative** : représente l'autre possibilité
- Les formules des quatre métriques d'évaluation Precision, Accuracy, Recall, F1 Score :
1. **Precision** =  $TP / (TP + FP)$
  2. **Recall** =  $TP / (TP + FN)$
  3. **Accuracy** =  $TP + TN / \text{Total Observations}$
  4. **F1 Score** =  $2 * (\text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall}))$

**Figure 4.16:** Formules pour le calcul des métriques d'évaluation de la Precision, Accuracy, Recall et F1 Score

Les résultats obtenus sont observables dans le Tableau. 4.1 :

Algorithm	Accuracy	Precision	Recall	F1 score
K-Nearest Neighbors (K-NN)	1.0	1.0	1.0	1.0
Random Forest (RF)	0.92	1.0	0.5	0.55
Decision Tree (DT)	1.0	1.0	1.0	1.0

**Tableau 4.1:** Performances des algorithmes utilisés selon les quatre critères Accuracy, Precision, Recall et F1 Score

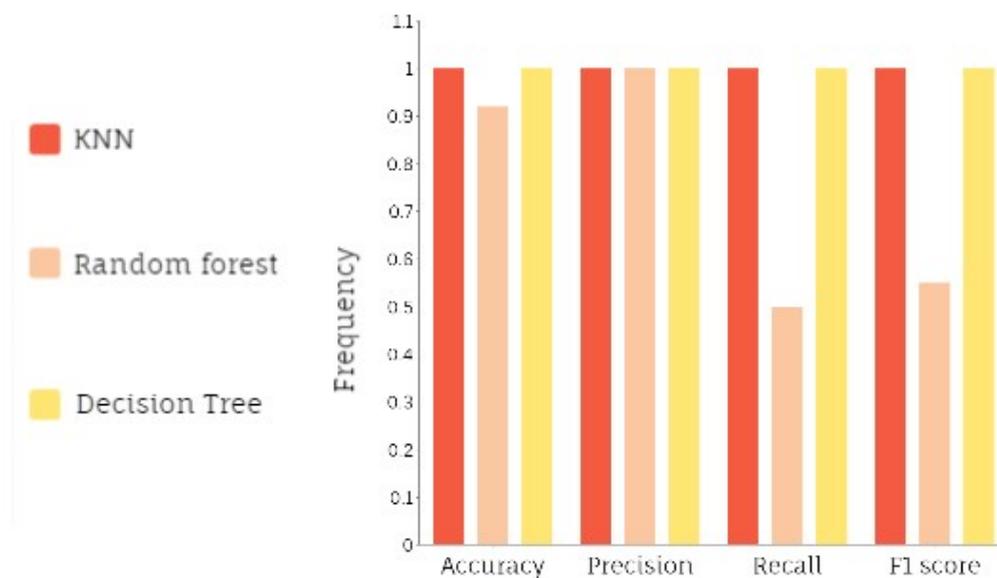
#### 4.4.2 Interprétation des résultats

Selon les résultats obtenus à partir de nos données, la première remarque est que les algorithmes K-NN et Decision Tree ont tous deux obtenu les meilleurs scores possibles

à tous les niveaux (100%). L'algorithme de Random Forest obtient également un score parfait en termes de précision (1.0 qui équivaut à 100%) et un excellent score en termes d'exactitude (avec 92.86%).

Cependant, Random Forest a un score de 0.5 seulement selon les critères F1 Score et le rappel contre 1.0 pour K-NN et Decision Tree.

ces résultats sont illustrés par les histogrammes de la Figure. 4.17.



**Figure 4.17:** Histogrammes illustrant les performances des algorithmes K-NN, Random Forest et Decision Tree

Il est maintenant évident que les meilleurs algorithmes de classification appliqués à notre cas sont K-NN et Decision Tree . Néanmoins, l'un de ces deux algorithmes est forcément plus adapté à la situation. En effet, lors de notre étude des algorithmes utilisés, il a été démontré que les performances de K-NN et Decision Tree peuvent varier en fonction de la taille du dataset.

L'algorithme K-NN dépend de la quantité de données disponibles, car il utilise une approche de voisinage local. Si le dataset est trop grand, cela peut augmenter le temps de calcul et rendre l'algorithme inefficace en termes de temps. Quant à Decision Tree, il s'avère que c'est plutôt le contraire qui se produit : Les arbres de décision sont capables de traiter des ensembles de données nettement plus importants que K-NN, car ils utilisent une approche de division descendante plutôt que de voisinage local, par conséquent, leurs performances baissent si la taille des données est trop petite.

Compte tenu de la taille de notre dataset (70,3), l'algorithme K- Nearest Neighbors est définitivement le plus adapté à la situation.

## 4.5 Conclusion

Dans ce dernier chapitre, nous avons abordé de façon détaillée l'aspect pratique de notre travail : Tout d'abord, nous nous sommes familiarisés avec les divers environnements de développements utilisés lors de la réalisation de ce projet. Ensuite, les différentes interfaces de notre application mobile Glyco-Wise ont été présentées. Puis, le traitement de données (prétraitement et classification) appliqué à notre dataset ainsi que les résultats obtenus après le traitement ont été exposés. Enfin, une étude comparative suivie d'une analyse des résultats obtenus par les trois algorithmes de classification choisis ont été faites dans le but d'identifier l'algorithme le plus approprié à notre dataset.

# Conclusion générale et perspectives

Comme nous le savons maintenant, l'AIoT est une technologie hybride issue de l'union de l'intelligence artificielle (IA) et de l'internet des objets (IoT). Depuis sa mise au point, elle est sans cesse utilisée dans divers domaines. L'atout principal de cette technologie réside dans le fait qu'elle soit facile d'utilisation et adaptée à des personnes de tout âge, et ne requiert donc aucune expérience technique particulière. Elle permet alors aux utilisateurs d'interagir de manière régulière avec leurs smartphones et autres dispositifs connectés à Internet, dans le but d'acquérir divers services que proposent ces outils révolutionnaires.

Ce domaine en constante expansion a connu un véritable rebond ces dernières années, et voit ses domaines d'utilisation se multiplier sans cesse. Cependant, l'un de ces domaines en particulier, connaît une évolution et un succès plus retentissant que les autres : le domaine de la santé.

Ayant pour but de contribuer au développement de ces innovations, nous nous sommes engagées par le biais de ce mémoire de fin d'études, à approfondir nos connaissances en AIoT ainsi qu'en Machine Learning appliqué au secteur de la santé, en réalisant une étude minutieuse de plusieurs articles abordant cette thématique. Nous avons également conçu une application mobile sous Android pour le suivi des patients diabétiques directement à partir d'un smartphone. Pour finir, nous avons été capables de manipuler et traiter des données à partir d'un dataset puis de les classifier en faisant appel à plusieurs algorithmes de Machine Learning que nous avons comparé par la suite à travers une étude comparative.

Au cours de ce projet, nous avons abordé deux aspects pratiques différents : le développement mobile, mais aussi l'aspect traitement de données qui reste l'objectif principal de notre travail. Les nombreuses étapes par lesquelles nous avons dû passer nous ont appris combien la réalisation ainsi que l'aboutissement de ces travaux étaient difficiles et gourmands de savoir-faire et de travail acharné nous permettant ainsi de nous surpasser et d'acquérir de nouvelles notions dans plusieurs disciplines.

Ce travail qui comporte aussi bien l'aspect pratique que théorique contribue à un autre aspect introduit implicitement par l'intermédiaire de sa thématique qui est l'aspect humain. Effectivement, l'objectif principal de ce projet est de permettre aux patients

diabétiques de contrôler leur pathologie sans avoir à se déplacer ou à effectuer de prise de sang.

Avec Glyco-Wise, vous profiterez d'un suivi rigoureux et agréable grâce à sa facilité d'utilisation ainsi que son accessibilité à toute heure de la journée ou de la nuit

Pour ce qui est des perspectives, plusieurs améliorations sont possibles. Quelques fonctionnalités peuvent être ajoutées dans le but de rendre l'expérience utilisateur plus satisfaisante et plus intuitive, comme l'ajout du facteur IMC (rapport poids-taille), ainsi que la saisie du rythme cardiaque afin d'estimer les risques encourus par les patients diabétiques de développer d'autres pathologies concomitantes comme les maladies cardio-vasculaires, les cancers, ou d'autres troubles métaboliques touchant les personnes en surpoids.

# Bibliographie

- [1] IONOS Digital Guide. Quelles sont les différences entre le Deep learning et le Machine learning? Available at : <https://www.ionos.fr/digitalguide/web-marketing/search-engine-marketing/deep-learning-vs-machine-learning/>. (Accédé : 11 Juin 2023).
- [2] OVHcloud. Qu'est-ce que l'internet des objets? Available at : [https://www.ovhcloud.com/fr/learn/what-is-iot/#:~:text=L\T1\textquoterightInternet%20des%20objets%20\(ou,montre%20de%20fitness%2C%20etc.\)](https://www.ovhcloud.com/fr/learn/what-is-iot/#:~:text=L\T1\textquoterightInternet%20des%20objets%20(ou,montre%20de%20fitness%2C%20etc.)). (Accédé : 11 Juin 2023).
- [3] M Ménard. Le rôle du Cloud dans l'IoT. Available at : <https://www.linkedin.com/pulse/le-r%C3%B4le-du-cloud-dans-liot-maxence-m%C3%A9nard/?originalSubdomain=fr>. (Accédé : 11 Juin 2023).
- [4] TIBCO Software. Qu'est ce que l'internet des objets (IoT). Available at : <https://www.tibco.com/fr/reference-center/what-is-the-internet-of-things-iot>. (Accédé : 11 Juin 2023).
- [5] B Lepine. Cloud Computing – Définition, Avantages et Exemples d'utilisation. Available at : <https://www.lebigdata.fr/definition-cloud-computing>, 2017. (Accédé : 11 Juin 2023).
- [6] DataScientest. Cloud Computing : qu'est-ce que c'est et comment devenir expert. Available at : <https://datascientest.com/cloud-computing>, 2022. (Accédé : 05 Juin 2023).
- [7] cloudflare. What is the cloud? | Cloud definition | Cloudflare. Available at : <https://www.cloudflare.com/learning/cloud/what-is-the-cloud/>, 2017. (Accédé : 09 Juin 2023).
- [8] Intel. Qu'est-ce que le Cloud Computing? Available at : <https://www.intel.fr/content/www/fr/fr/cloud-computing/what-is-cloud-computing.html>. (Accédé : 01 Juin 2023).
- [9] Global SP - Hébergement Cloud France. Private cloud, public cloud et hybrid cloud. Available at : <https://www.globalsp.com/>

- [private-cloud-public-cloud-et-hybrid-cloud/](#), 2022. (Accédé : 09 Juin 2023).
- [10] Technoname. What is Cloud Computing? Everything you need to know is explained. Available at : <https://technoname.com/index.php/2021/02/22/what-is-cloud-computing-everything-you-need-to-know-is-explained/>, 2021. (Accédé : 10 Juin 2023).
- [11] Blue Soft Group. L'IA et le Cloud : combo gagnant des entreprises de demain. Available at : <https://www.bluesoft-group.com/lia-et-le-cloud-combo-gagnant-des-entreprises-de-demain/>, 2023. (Accédé : 12 Juin 2023).
- [12] Rcarre. L'intelligence Artificielle Auto-Apprenante pour le Cloud and Saas. Available at : <https://www.rcarre.com/blog/intelligence-artificielle-auto-apprenante-pour-le-cloud-and-saas/>, 2022. (Accédé : 10 Juin 2023).
- [13] NetApp. Data Storage Data Management Solutions. Available at : <https://www.netapp.com/fr/artificial-intelligence/what-is-artificial-intelligence/>. (Accédé : 10 Juin 2023).
- [14] Neovision. L'IA c'est quoi ? Découvrez la réponse de nos experts. Available at : <https://neovision.fr/definition-intelligence-artificielle/>, 2022. (Accédé : 10 Juin 2023).
- [15] Talend. Machine learning : définition et fonctionnement. Available at : <https://www.talend.com/fr/resources/what-is-machine-learning/>. (Accédé : 08 Juin 2023).
- [16] Machine Learnia. Apprentissage Supervisé : Introduction. Available at : [https://machinelearnia.com/apprentissage-supervise-4-etapes/#:~:text=L'apprentissage%20supervis%C3%A9%20\(en%20anglais,t%C3%A2che%20qu'elle%20doit%20r%C3%A9alis%C3%A9e.,2020](https://machinelearnia.com/apprentissage-supervise-4-etapes/#:~:text=L'apprentissage%20supervis%C3%A9%20(en%20anglais,t%C3%A2che%20qu'elle%20doit%20r%C3%A9alis%C3%A9e.,2020). (Accédé : 10 Juin 2023).
- [17] A Cutler L Breiman. Random forest : An ensemble learning algorithm. *Machine Learning*, 45(1) :05–32, 2001.
- [18] L Breiman, J Friedman, and Olshen, Richard A. and Stone, Charles. *Classification and Regression Trees*. Chapman and Hall/CRC, 1984.
- [19] E Fix, J Hodges. Discriminatory Analysis. Nonparametric Discrimination : Consistency Properties. *Technical Report*, 1951.
- [20] C Gauss. Theory of the Motion of Heavenly Bodies Moving about the Sun in Conic Sections. *Memoires de l'Academie royale des sciences de Gottingen*, 1809.

- [21] E.T. Jaynes, J Pearl. The Bayesian Classification Algorithm, Naive Bayes. *International Journal of Approximate Reasoning*, 2(3) :297–326, 1965.
- [22] V Vapnik. Support-Vector Networks. *Machine Learning*, 20(3) :273–297, 1995.
- [23] D Cox. The Logistic Regression Model. *Journal of the Royal Statistical Society*, 20(1) :1–45, 1958.
- [24] J Friedman. Greedy Function Approximation : A Gradient Boosting Machine. *Annals of Statistics*, 29(5) :1189–1232, 2001.
- [25] SP loyd. Least squares quantization in pcm. *Bell Telephone Laboratories Paper*, 1957.
- [26] R Sibson. Slink : An optimally efficient algorithm for the single-link cluster method. *The Computer Journal*, 16(1) :30–34, 1973.
- [27] A Elisseeff O Bousquet. Stability and generalization. *Journal of machine learning research*, pages 499–526, 2002.
- [28] ProjectPro. A-Z Guide to the Types of Machine Learning Problems. Available at : <https://www.projectpro.io/article/types-of-machine-learning/623>. (Accédé : 11 Juin 2023).
- [29] LEBIGDATA.FR. Deep Learning ou apprentissage profond : définition, concept. Available at : <https://www.lebigdata.fr/deep-learning-definition>, 2022. (Accédé : 01 Juin 2023).
- [30] GeeksforGeeks. Introduction to Deep Learning. Available at : <https://www.geeksforgeeks.org/introduction-deep-learning/>, 2023. (Accédé : 07 Juin 2023).
- [31] Mobiskill. Quels sont les algorithmes de deep learning? Available at : <https://mobiskill.fr/blog/conseils-emploi-tech/quels-sont-les-algorithmes-de-deep-learning/>, 2021. (Accédé : 02 Juin 2023).
- [32] Medium. Convolutional Neural Networks, Explained. Available at : <https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939>, 2020. (Accédé : 12 Juin 2023).
- [33] AILabPage. Deep Learning – Introduction to Recurrent Neural Networks. Available at : <https://ailabpage.com/2019/01/08/deep-learning-introduction-to-recurrent-neural-networks/>, 2019. (Accédé : 12 Juin 2023).

- [34] Jedha. Quels sont les différents algorithmes de Deep Learning ? Available at : <https://www.jedha.co/formation-ia/algorithmes-deep-learning>, 2020. (Accédé : 10 Juin 2023).
- [35] GAN Explained | Papers With Code. Papers with Code - GAN Explained. Available at : <https://paperswithcode.com/method/gan>. (Accédé : 11 Juin 2023).
- [36] lawtomated. A.I. technical - Machine Learning vs. Deep Learning. Available at : <https://lawtomated.com/a-i-technical-machine-vs-deep-learning/>, 2020. (Accédé : 08 Juin 2023).
- [37] Rghioui et al. A Smart Architecture for Diabetic Patient Monitoring Using Machine Learning Algorithms. *Healthcare*, 2020.
- [38] T Saba. Recent advancement in cancer detection using machine learning : Systematic survey of decades, comparisons and challenges. *Journal of Infection and Public Health*, 2020.
- [39] S Poudel. A Study of Disease Diagnosis using Machine Learning. *The 2nd International Electronic Conference on Healthcare*, 2022.
- [40] Nagavelli et al. Machine Learning Technology-Based Heart Disease Detection Models. *Journal of Healthcare Engineering*, 2022.
- [41] Gunčar et al. An application of machine learning to haematological diagnosis. *Scientific Reports*, 2018.
- [42] Centre européen d'étude du Diabète. Les chiffres du diabète. Available at : <https://ceed-diabete.org/fr/le-diabete/les-chiffres/>, 2021. (Accédé : 10 Juin 2023).
- [43] Le Monde.fr. Le diabète, une maladie qui progresse. Available at : [https://www.lemonde.fr/sciences/article/2022/05/30/le-diabete-une-maladie-qui-progresse\\_6128229\\_1650684.html](https://www.lemonde.fr/sciences/article/2022/05/30/le-diabete-une-maladie-qui-progresse_6128229_1650684.html), 2022. (Accédé : 09 Juin 2023).
- [44] Sante publique france. Données épidémiologiques sur le diabète de type 2. Available at : <https://www.santepubliquefrance.fr/>. (Accédé : 10 mai 2023).
- [45] Sanofi Algérie. Communiqué de presse Algérie - Sanofi. Available at : , 2019. (Accédé : 08 Juin 2023).
- [46] Organisation panaméricaine de la santé. Diabète | OPS/OMS | Organisation panaméricaine de la santé - PAHO. Available at : <https://www.paho.org/fr/sujets/diabete>. (Accédé : 09 Juin 2023).
- [47] Diabète Québec. Qu'est-ce que le diabète? Available at : <https://www.diabete.qc.ca/fr/comprendre-le-diabete/tout-sur-le-diabete/>

quest-ce-que-le-diabete-2/quest-ce-que-le-diabete/#:~:text=%C3%80%20long%20terme%2C%201%27hyperglyc%C3%A9mie,d%27autres%20types%20plus%20rares, 2021. (Accédé : 12 Juin 2023).

- [48] ameli.fr . Symptômes et diagnostic du diabète. Available at : <https://www.ameli.fr/assure/sante/themes/diabete/diabete-symptomes-evolution/diagnostic-diabete>. (Accédé : 08 Juin 2023).
- [49] Nathan et al. Translating the A1C Assay Into Estimated Average Glucose Values. *Diabetes Care*, 2008.
- [50] American diabetes association. Understanding A1C. Available at : <https://diabetes.org/diabetes/a1c>. (Accédé : 11 Juin 2023).
- [51] Y Benzaki . Introduction à l’algorithme K Nearest Neighbors (K-NN). Available at : <https://mrmint.fr/introduction-k-nearest-neighbors>, 2018. (Accédé : 12 Juin 2023).
- [52] IBM. Qu’est-ce qu’un arbre de décisions. Available at : <https://www.ibm.com/fr-fr/topics/decision-trees>. (Accédé : 08 Juin 2023).
- [53] Jedha. Anaconda pour Python - Présentation et installation. Available at : <https://www.jedha.co/formation-python/ananconda-python>. (Accédé : 08 Juin 2023).
- [54] Spyder. Home - Spyder IDE. Available at : <https://www.spyder-ide.org/>. (Accédé : 013 Juin 2023).
- [55] G Batschinski. Qu’est-ce que Firebase? Tous les secrets dévoilés. Available at : <https://blog.back4app.com/fr/quest-ce-que-firebase/>, 2021. (Accédé : 11 Juin 2023).
- [56] Python documentation. The Python Tutorial. Available at : <https://docs.python.org/3/tutorial/>. (Accédé : 04 Juin 2023).
- [57] Apprendre le langage de programmation python. Available at : <https://python.doctor/>. (Accédé : 09 Juin 2023).
- [58] Mobiskill. Les bibliothèque Python à utiliser pour le machine learning. Available at : <https://mobiskill.fr/blog/conseils-emploi-tech/les-bibliotheque-python-a-utiliser-pour-le-machine-learning/>, 2021. (Accédé : 08 Juin 2023).
- [59] Kahn,Michael. Diabetes. UCI Machine Learning Repository, Available at : <https://archive.ics.uci.edu/dataset/34/diabetes>. DOI : <https://doi.org/10.24432/C5T59G>.

---

## Résumé

L'union de l'internet des objets (IoT) avec l'intelligence artificielle (IA), est un parfait mélange entre la technologie accessible via ses nombreux objets connectés (smartphone, smartwatch ...) et le monde abstrait des algorithmes de machine learning et de deep learning qui est plutôt destiné aux experts de ce domaine. Cette nouvelle technologie appelée intelligence artificielle des objets, plus connue sous son acronyme (AIoT), a grandement contribué au développement de nombreux domaines dont la gestion reste jusqu'à aujourd'hui extrêmement complexe, notamment celui de la santé. Ayant pour objectif d'apporter notre contribution, nous proposons par le biais de ce travail, une architecture permettant le suivi des patients diabétiques en mettant à leur disposition un système via leur smartphone qui consiste à enregistrer quotidiennement leurs données vitales, plus particulièrement celles liées à leurs pathologie et de les communiquer directement à leurs médecin traitant sans avoir à se déplacer, offrant ainsi un gain de temps et une meilleure qualité de vie au quotidien aux patients, mais aussi aux médecins qui sont surchargés. Afin de concrétiser notre système, nous avons conçu Glyco-Wise, une application Android pour le suivi à distance des patients atteints de diabète.

**Mots Clés :** IoT, IA, AIoT, diabète, suivi, application Android.

## Abstract

The union of the Internet of Things (IoT) with Artificial Intelligence (AI), is a perfect mix between the technology that is accessible through its many connected objects (smartphones, smartwatches...) and the abstract world of machine learning and deep learning algorithms which are more intended for experts. This new technology called artificial intelligence of objects, better known by its acronym (AIoT), has greatly contributed to the development of many fields whose management remains extremely complex until today, in particular the health field. Aiming to make our contribution, we propose through this work, an architecture allowing the monitoring of diabetic patients by providing them a system via their smartphone which consists in recording their vital data daily, more particularly those related to their pathologie and to communicate them directly to their doctor, thus offering time savings and a better daily life quality for patients, but also for doctors who are overworked. In order to realize our system, we designed Glyco-Wise, an Android application for the remote monitoring of patients with diabetes.

**Keywords:** IoT, IA, AIoT, diabetes, monitoring, Android application.

---