République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Kasdi Merbah Ouargla

Faculté des Mathématiques et des Sciences de la Matière

Département de Physique

Mémoire

Master Professionnel

Présenté en vue de l'obtention du diplôme de Master

Domaine : Sciences de la Matière

Filière : Physique

Spécialité : Physique Médical

Réalisé par :

BEN ACHOURA Wissal et KAFI Malak

Thème :

Collecte et analyse de données de Diabète utilisant les techniques d'Apprentissage Profond : Réseaux de Neurones (ANN)

Soutenu publiquement, le 18 / 06 / 2026

Devant le jury composé de :

Dr AYAT Zahia	*MCA*	*Président*	*UKM Ouargla*
Pr KHELFAOUI Fethi	*Professeur*	*Rapporteur*	*UKM Ouargla*
Dr ZENKHRI Djamel Eddine	*MCB*	*Examinateur*	*UKM Ouargla*

Année universitaire : 2025 – 2026

﴿وَآخِرُ دَعْوَاهُمْ أَنِ الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ﴾

Dédicace

Après un parcours de dix-sept années de travail, de persévérance et de patience, je me tiens aujourd'hui au terme de cette aventure, présentant le fruit de mes efforts, couronné par la grâce et l'aide d'Allah Tout-Puissant en premier lieu, puis par les prières et le soutien de ceux que j'aime et qui m'ont accompagnée tout au long de ce chemin.

À ma mère, ma première enseignante dans la vie, qui n'a pas seulement été une mère, mais également mon enseignante à l'école, et qui a semé en moi l'amour du savoir, des valeurs et de la persévérance dès mon plus jeune âge. Qu'Allah bénisse ta vie et fasse de ton travail une source de récompense dans la balance de tes bonnes actions.

À mon cher père, mon guide et mon soutien, merci pour ta patience, ton encouragement constant et la confiance que tu as toujours placée en moi. Qu'Allah te protège, te bénisse et te garde toujours à mes côtés.

À mes sœurs Nesima et Chaima, ainsi qu'à mes frères Charafeddine et Mohamed Amine et à leurs épouses, qui ont été derrière chacun de mes pas et un soutien dans les moments difficiles, je prie Allah de préserver l'amour et l'affection qui nous unissent et de nous permettre de rester toujours unis et solidaires.

À nos petits trésors de la famille : Abrar, Sif Eddine, Ahmed Chahine et Ahmed Ayan, qu'Allah vous protège, vous guide et fasse de vous des personnes vertueuses.

À ma grande famille Kafi et Dibouna, chacun de vous occupe une place particulière dans mon cœur. Merci pour votre affection et votre soutien.

À ma collègue de projet Wissal, je te souhaite beaucoup de réussite et de succès dans ton parcours.

À mon amie Ouissal, compagne des moments difficiles et des instants heureux, puisse notre amitié durer toujours. Et à toutes mes amies, sans exception, merci pour votre affection, votre soutien et votre présence.

Enfin, je demande à Allah Tout-Puissant de m'accorder la réussite dans mon avenir, de bénir mon parcours et de réaliser mes ambitions.

Malak Kafi

﴿وَآخِرُ دَعْوَاهُمْ أَنِ الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ﴾

اهداء

الحمد لله الذي وفقني لإتمام هذا العمل المتواضع.

أهدي ثمرة جهدي إلى أعز الناس على قلبي، إلى والدي عبد الحميد و امي نصيرة ، عرفانًا بفضلهما ودعمهما ودعواتهما التي كانت سندًا لي في كل خطوة. وإلى عائلة بن عشورة وعلى راسهم أخواتي(سميحة الهام والمازوزية نريمان) وإخوتي(ساسي عبد الرحمان عبد المطلب فاتح سامي سمير) مع زوجاتهم وأبنائهم(ردينة نور الدين ولاء الدين رفيف تقوى سندس) الذين شاركوني لحظات التعب والأمل.

إلى صديقتي التي تقاسمت معها هذا الجهد و العمل (ملاك) وصديقاتي(سماح نور سلسبيل راوية فاطمة ليندة اكرام ....) وكل من رافقني وشجعني خلال مشواري الدراسي، وإلى أساتذتي الأفاضل الذين أناروا لي طريق العلم خاصة معلمي طرودي محمد الطاهر.

وإلى كل من ساهم من قريب أو بعيد في إنجاز هذه المذكرة، أهدي هذا العمل عربون محبة وتقدير.

وأخيراً، أهدي هذا الإنجاز إلى نفسي، لأنها صبرت وإجتهدت وآمنت بأن الأحلام تستحق السعي حتى تتحقق

وصال بن عشورة

﴿وَمَا تَوْفِيقِي إِلَّا بِاللَّهِ عَلَيْهِ تَوَكَّلْتُ وَإِلَيْهِ أُنِيبُ﴾

Remerciements

Louange à Allah, par la grâce duquel les bienfaits se réalisent et par la guidance duquel les objectifs sont atteints, qui a facilité nos démarches et nous a permis de mener à bien ce travail. À Lui appartient toute la louange, au commencement comme à la fin.

«Celui qui ne remercie pas les gens ne remercie pas Allah» (Rapporté par At-Tirmidhi).Nous tenons à exprimer notre profonde gratitude et notre sincère reconnaissance à toutes les personnes qui nous ont soutenus et accompagnés dans la réalisation de ce travail, de près ou de loin.

Nous adressons tout particulièrement nos remerciements à notre encadreur, le Professeur KHELFAOUI Fethi, enseignant à l'Université Kasdi Merbah d'Ouargla, qui, par ses précieux conseils, sa disponibilité et son suivi constant, nous a permis d'apprendre, de progresser et de mener à bien ce travail.Nous lui témoignons notre plus profonde reconnaissance et notre grand respect, et lui souhaitons santé, bonheur et réussite.

Nous adressons également nos plus sincères remerciements et notre profonde reconnaissance aux honorables membres du jury pour avoir accepté d'examiner et d'évaluer notre travail : en qualité de Président Dr AYAT Zahia (MCA) et en qualité d'Examinateur Dr ZENKHRI Djamel Eddine (MCB). Nous les remercions par avance pour l'intérêt qu'ils lui porteront ainsi que pour leurs précieuses remarques et suggestions, qui contribueront sans aucun doute à son enrichissement et à son amélioration.

Nous tenons également à exprimer notre profonde gratitude au personnel du Centre du diabète d'Ouargla pour son accueil, sa disponibilité et son aide précieuse tout au long de notre stage.

Nous adressons également nos sincères remerciements au Laboratoire de Physique des Rayonnements et Plasmas et Physique des Surfaces (LRPPS) pour les moyens mis à notre disposition et le soutien apporté à la réalisation de ce travail.Que Dieu les récompense pour leur générosité et leur bienveillance.

Enfin, nous remercions l'administration du Département de physique pour son aide et sa coopération.

الملخص:

يعد مرض السكري أحد أكثر الأمراض المزمنة شيوعاً، حيث بلغت نسبة انتشاره في الجزائر 10.0%. وفي هذا العمل، نود أن نقترح نهجاً جديداً يعتمد على فرع من فروع الذكاء الاصطناعي، وهو التعلم العميق باستخدام الشبكات العصبية الاصطناعية(ANN) . يهدف البحث إلى مساعدة الأطباء في تصنيف البيانات والتنبؤ المبكر بالمرض. تم الاعتماد في هذه الدراسة على قاعدتي بيانات؛ الأولى عالمية تضم 100,000 سجل بنسبة إصابة 8.82%، والثانية محلية جُمعت من "دار السكري" بمستشفى محمد بوضياف بورقلة وشملت 700 ملف طبي لمرضى السكري من النوعين الأول (DT1) والثاني (DT2).

للمساهمة في رفع كفاءة التنبؤ، قمنا بمعالجة وتنظيف البيانات وإجراء التحليلات الإحصائية والمدرجات التكرارية لمعرفة تأثير المؤشرات الحيوية. لتحقيق ذلك، قمنا ببناء وتقييم نموذج شبكة عصبية اصطناعية من نوع (MLP) وتدريبها عبر بيئة بايثون. وأظهرت النتائج التجريبية كفاءة عالية للنموذج المقترح بدقة تصنيف إجمالية تجاوزت 93%، وقدرة ممتازة على الفصل الدقيق بين الحالات. تفتح هذه النتائج آفاقاً واعدة لتطوير أنظمة دعم القرار الطبي ورقمة الملفات الصحية لتخفيف العبء عن الأطقم الطبية.

الكلمات المفتاحية :الذكاء الاصطناعي، الشبكات العصبية (ANN) ، مرض السكري، التحليل الإحصائي، ورقلة.

Abstract:

Diabetes is one of the most prevalent chronic diseases, with a prevalence rate reaching 10.0% in Algeria. In this work, we propose a new approach based on deep learning, a branch of artificial intelligence, using Artificial Neural Networks (ANN). This research aims to assist physicians in data classification and early prediction of the disease. This study relies on two datasets: the first is a global dataset comprising 100000 records with an infection rate of 8.82%, and the second is a local dataset collected from the "Diabetic's House" at Mohamed Boudiaf Hospital in Ouargla, including 700 medical files of patients with type 1 and type 2 diabetes (T1D and T2D).

To enhance prediction efficiency, we preprocessed and cleaned the data, and conducted statistical analyses and histograms to determine the impact of biomarkers. To achieve this, a Multi-Layer Perceptron (MLP) neural network model was built, evaluated, and trained using the Python environment. The experimental results demonstrated high efficiency for the proposed model, with an overall classification accuracy exceeding 93% and an excellent ability to precisely distinguish between cases. These findings open promising prospects for developing clinical decision support systems and digitizing health records to ease the burden on medical staff.

Keywords: Artificial Intelligence, Neural Networks (ANN), Diabetes, StatisticalAnalysis, Ouargla.

Résumé:

Le diabète est l'une des maladies chroniques les plus répandues, avec un taux de prévalence atteignant 10,0% en Algérie. Dans ce travail, nous proposons une nouvelle approche basée sur l'apprentissage profond (Deep Learning), une branche de l'intelligence artificielle, en utilisant les Réseaux de Neurones Artificiels (ANN). Cette recherche vise à aider les médecins dans la classification des données et la prédiction précoce de la maladie. Cette étude s'appuie sur deux bases de données : la première est mondiale et comprend 100 000 enregistrements avec un taux d'infection de 8,82 %, tandis que la seconde est locale, collectée auprès de la "Maison du Diabète" de l'hôpital Mohamed Boudiaf de Ouargla, englobant 700 dossiers médicaux de patients atteints de diabète de types 1 et 2 (DT1 et DT2).

Pour contribuer à l'amélioration de l'efficacité de la prédiction, nous avons traité et nettoyé les données, puis effectué des analyses statistiques et des histogrammes afin de déterminer l'impact des biomarqueurs. Pour ce faire, un modèle de réseau de neurones de type Perceptron Multicouche (MLP) a été construit, évalué et entraîné via l'environnement Python. Les résultats expérimentaux ont démontré une haute efficacité du modèle proposé, avec une précision globale de classification dépassant 93 % et une excellente capacité à distinguer précisément les cas. Ces résultats ouvrent des perspectives prometteuses pour le développement de systèmes d'aide à la décision médicale et la numérisation des dossiers de santé afin d'alléger la charge de travail du personnel médical.

Mots-clés : Intelligence Artificielle, Réseaux de Neurones (ANN), Diabète, Analyse Statistique, Ouargla.

Table des Matières

Liste des figures

Liste des tableaux

Introduction générale

Introduction Générale

L’ère actuelle a été témoin d’une révolution technologique et informationnelle massive qui a modifié les caractéristiques et le cours de nombreux secteurs vitaux, et l’essor qualitatif de l’informatique et des techniques d’intelligence artificielle. L’intelligence artificielle n’est plus seulement un concept théorique ou un luxe technique. Au contraire, elle est devenue un outil essentiel qui contribue à résoudre des problèmes complexes que les méthodes traditionnelles sont incapables de résoudre. Parmi les branches les plus développées et les plus influentes dans ce domaine, se distinguent les technologies d'apprentissage automatique et d'apprentissage profond, qui ont prouvé leur efficacité supérieure dans l'analyse des mégadonnées, l'extraction de modèles cachés et la fourniture de prédictions très précises basées sur des algorithmes de réseaux neuronaux artificiels (ANN) qui imitent la façon dont ils fonctionnent le cerveau humain.

Avec ce développement rapide, le secteur médical et de la santé a considéré les technologies d’apprentissage profond comme un partenaire stratégique indispensable pour développer le système de santé. Les systèmes médicaux d'aujourd'hui sont confrontés à un afflux massif et continu de données et d'informations sur les patients, ce qui impose une lourde charge au personnel médical dans les domaines du diagnostic, du suivi et de la prévision de l'évolution des problèmes de santé. Dans ce contexte, les réseaux de neurones artificiels permettent de traiter des variables cliniques multiples et complexes et de les classer avec précision, ce qui favorise une prise de décision médicale judicieuse, réduit les taux d'erreur humaine et contribue à sauver des vies et à économiser du temps et des efforts au sein des établissements hospitaliers

Cependant, le succès de tout modèle basé sur le deeplearning dans le domaine médical reste dépendant de la qualité de la méthodologie utilisée et de l’environnement de travail logiciel utilisé. L'environnement Google Colab et les bibliothèques de programmation avancées qu'il fournit dans le langage Python telles que (Pandas, NumPy, TensorFlow, Scikit-learn) constituent le terrain fertile qui permet aux chercheurs et aux développeurs de construire, nettoyer et entraîner ces modèles complexes avec une grande efficacité et de tester l'étendue de leur capacité prédictive.

La principale lacune réside dans la manière d’exploiter de manière optimale la grande quantité de données cliniques et de les transformer en indicateurs prédictifs précis qui servent à la fois le médecin et le patient. Par conséquent, le problème principal de cette étude se porte sur l’étendue du succès et de l’efficacité des réseaux de neurones artificiels (ANN) dans la modélisation et la classification des données médicales.

Les sous-questions suivantes émergent de ce problème:

- ✓ Quels sont les fondements théoriques et méthodologiques qui régissent le travail des techniques de machine et de deeplearning dans le domaine médical ?
- ✓ Comment les données cliniques brutes collectées sur le terrain peuvent-elles être nettoyées et traitées pour les rendre adaptées aux modèles de formation?
- ✓ Quel est le pourcentage de précision et de prédiction que le réseau neuronal conçu peut atteindre lorsqu'il est appliqué aux bases de données étudiées?

L'importance de cesrecherchesrésidedans la tentative de fournir des solutions technologiques pratiques qui soutiennent le système de santé. Les objectifs fondamentaux sont :

- ✓ Développer et construire un modèle intégré de réseau neuronal artificiel (ANN) capable de classer avec précision les données médicales.
- ✓ Mettre en évidence le rôle du traitement préliminaire et du nettoyage des données dans l'augmentation de l'efficacité des modèles intelligents.
- ✓ Incarner l'intégration entre l'université et le terrain en s'appuyant sur des données réelles collectées en coordination avec du personnel médical spécialisé.
- ✓ Évaluer les performances du modèle logiciel et déterminer sa fiabilité dans l'environnement de diagnostic.

Pour répondre à la problématique soulevée et atteindre quelques objectifs, ce mémoire de recherche a été divisé en trois chapitres.

Nous avons consacré le premier chapitre à revoir les intérêts et concepts fondamentaux du diabète : types, aspects cliniques et biologiques, diagnostics et indicateurs, traitements et prévetion.

Dans le deuxième chapitre, nous nous sommes concentrés sur les fondements théoriques et méthodologiques, en expliquant l'environnement de travail du logiciel et les bibliothèques utilisées. Les modèles et algorithmes pour le deep-Learning et les réseaux neuronaux aisi que les calculs statistiques. Le chapitre présente nos activités du stage sur terrain et des étapes de collecte de données du Centre de diabète de Ouargla. On présente aussi les étapes de traitement statistique et par ANN pour ce mémoire.

Enfin, le troisième et dernier chapitre présente les résultats expérimentaux et les graphiques obtenues, sur deux bases de données internationale et locales, discutant de la précision du modèle et évaluant ses performances globales.

La conclusion générale du mémoire résume les conclusions les plus importantes et les recommandations futures.

Chapitre I :

Généralités sur le diabète

Chapitre I : Généralités sur le diabète

Introduction

Au cours des dernières décennies, l'expansion des maladies chroniques a connu une croissance exponentielle à l'échelle mondiale. Ces pathologies représentent un enjeu majeur pour les systèmes de santé, en raison de leur impact direct sur la santé des individus et des sociétés. Cette augmentation est liée à plusieurs facteurs, notamment les changements du mode de vie, la diminution de l'activité physique, une alimentation déséquilibrée ainsi que l'urbanisation croissante, ce qui a contribué à l'augmentation des taux de nombreuses maladies chroniques. Parmi ces pathologies, le diabète se distingue par son importance, en raison de sa prévalence élevée à l'échelle mondiale et des complications graves qu'il peut engendrer au niveau de divers organes. Cette situation a induit un intérêt grandissant pour l'étude de cette pathologie, ainsi que pour le développement de méthodes efficaces de diagnostic, de suivi et de prévention de ses complications.

Dans le cadre de ce chapitre, une exposition des principaux aspects théoriques relatifs au diabète est proposée. Cette exposition aborde la définition de la pathologie, ses différents types, ses causes, ses symptômes, ses complications, ainsi que les méthodes de diagnostic et de traitement. En outre, il met en exergue les principaux indicateurs médicaux et biologiques employés dans le cadre du suivi et de l'évaluation de l'état des patients atteints de diabète.

1. Diabète sucré

1.1. Définition du diabète

Le diabète est une maladie chronique dont la gravité et la prévalence ne cessent d'augmenter à l'échelle mondiale. Son développement est associé à plusieurs facteurs génétiques, environnementaux et liés au mode de vie. Selon l'Organisation mondiale de la Santé, le diabète pourrait devenir l'une des principales causes de décès d'ici 2030 [1] [2] , ce qui souligne son impact majeur sur la santé publique mondiale.

Le diabète constitue un véritable problème de santé publique en raison des nombreuses complications qu'il peut entraîner et qui peuvent affecter différents organes du

corps. Cette maladie se divise également en plusieurs types qui se distinguent par leurs mécanismes de développement et leurs causes[1].

1.2. Physiopathologie du diabète

L'apparition de divers troubles physiques, tels qu'une augmentation de la fréquence des mictions, une soif excessive, de la fatigue, des troubles de la vision, ainsi que des complications neurologiques et cardiovasculaires associées à une hyperglycémie persistante, indique un dysfonctionnement de l'insuline, hormone essentielle à la régulation de la glycémie et au maintien de l'équilibre énergétique de l'organisme.

Ce dysfonctionnement est dû soit à une sécrétion insuffisante d'insuline par les cellules bêta du pancréas, soit à une résistance des cellules à son action. Le glucose ne peut alors plus pénétrer dans les cellules ni être utilisé pour la production d'énergie, ce qui entraîne son accumulation dans le sang et provoque une hyperglycémie.

Malgré cette hyperglycémie, le foie continue de produire et de libérer du glucose, tandis que les muscles et les tissus adipeux ne parviennent pas à l'absorber normalement, entraînant ainsi une diminution de l'énergie et une sensation de fatigue. Lorsque la glycémie devient très élevée, les reins éliminent l'excès de glucose dans les urines, provoquant une déshydratation et une soif intense.

En cas de déficit énergétique, l'organisme puise son énergie dans les graisses, ce qui conduit à la formation de corps cétoniques et à l'apparition de divers troubles métaboliques, constituent ainsi la physiopathologie du diabète[3].

2. Épidémiologie et impact socio-économique du diabète

2.1. Données épidémiologiques du diabète

2.1.1. Dans le monde

Les données épidémiologiques internationales révèlent une croissance exponentielle et persistante des cas de diabète à l'échelle mondiale. En 1980, on estimait à environ 108 millions le nombre d'adultes diabétiques, soit une prévalence de 4,7 %. Ce chiffre est passé à 422 millions en 2014, soit une augmentation de 8,5 %. Les données les plus récentes de l'étude '*Global Burden of Disease 2021'* confirment cette augmentation significative : le nombre de personnes diabétiques dans le monde atteint environ 529 millions, soit environ 6,1 % de la population mondiale. Cette augmentation, particulièrement marquée dans les pays à revenu faible et intermédiaire, est attribuée à la

hausse des taux d'obésité et de surpoids. En outre, les rapports indiquent que le diabète de type 2 représente la grande majorité des cas, soit environ 96 % de tous les cas dans le monde, avec une augmentation significative de sa prévalence chez les enfants. Cette dynamique s'inscrit dans un contexte de croissance démographique et d'augmentation de l'espérance de vie, dont les projections font état d'une poursuite de cette tendance. Selon les estimations, le nombre de personnes atteintes de diabète devrait atteindre environ 1,31 milliard d'ici 2050 [4] [5].

2.1.2. En Afrique

Depuis 1990, le fardeau du diabète connaît une augmentation continue dans le monde, notamment en Afrique. Selon les études de "The Lancet" et "Global Burden of Disease Study", la région de l'Afrique du Nord et du Moyen-Orient figure parmi les régions les plus touchées par le diabète, avec un taux de prévalence atteignant 9,3 % en 2021, supérieur à la moyenne mondiale[5].L'Afrique a également enregistré une hausse importante des taux d'incidence, de prévalence et de mortalité liés au diabète entre 1990 et 2021, ce qui reflète l'aggravation de l'impact de cette maladie dans la région. Cette augmentation est principalement liée à l'obésité, aux changements alimentaires, à la diminution de l'activité physique ainsi qu'à l'urbanisation rapide. Toutefois, certaines régions d'Afrique subsaharienne orientale restent parmi les moins touchées, avec un taux d'environ 2,9 % en 2021 [4] [5].

2.1.3. En Algérie

Après avoir présenté la situation générale du diabète, nous aborderons brièvement la situation épidémiologique de cette maladie en Algérie. L'analyse des statistiques mondiales et africaines révèle que le diabète représente un défi sanitaire croissant en Algérie, avec un taux de prévalence standardisé selon l'âgea atteint 10,0 % en 2021, supérieur à la moyenne régionale pour l'Afrique du Nord et le Moyen-Orient, estimée à 9,3 %. Ce taux place l'Algérie parmi les 11 pays de la région où la prévalence du diabète a dépassé le seuil de 10 %. De plus, les projections futures indiquent une augmentation significative du taux de prévalence régional, qui devrait atteindre 16,8 % d'ici 2050, une hausse accélérée principalement due aux évolutions démographiques et au vieillissement de la population[5]

La figure I.1 nous présentons le taux de diabète par population dans le monde, en Afrique et en Algérie en 2021.

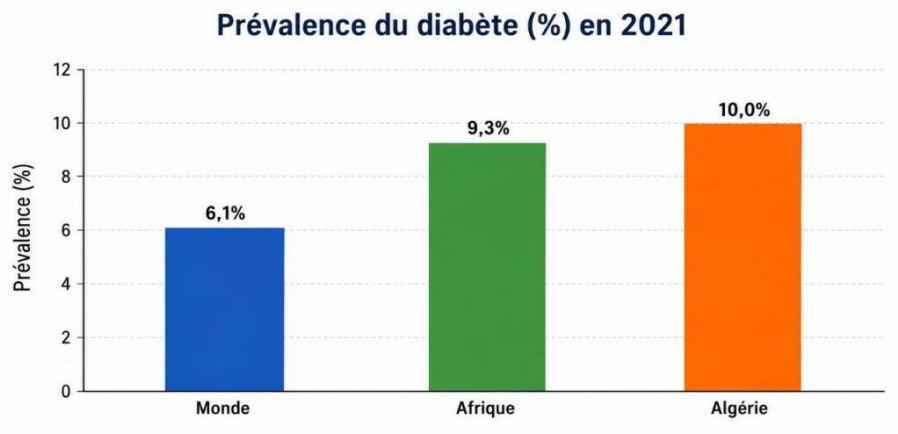


Figure I.1 : Taux de diabète par population Mondiale, en Afrique et en Algérie [18]

2.2. Impact socio-économique du diabète

La prévalence croissante du diabète et ses graves complications en font l'une des maladies chroniques les plus invalidantes pour les individus et les sociétés. Ses effets dépassent le simple cadre de la santé et englobent les dimensions économiques, sociales et psychologiques. Sur le plan sanitaire, le diabète peut entraîner de nombreuses complications chroniques, telles que les maladies cardiovasculaires, l'insuffisance rénale, la neuropathie et la rétinopathie diabétique, nécessitant un suivi médical régulier et un traitement au long cours. D'un point de vue économique, cette maladie peut engendrer une augmentation des coûts de santé liés aux examens médicaux, aux analyses biologiques, aux médicaments et aux hospitalisations. Elle contribue également à une baisse de la productivité individuelle due à l'absentéisme et aux limitations fonctionnelles résultant de ses complications. De plus, le diabète a un impact social et psychologique important, pouvant causer aux patients stress, anxiété et difficultés dans la gestion quotidienne de la maladie, en plus du respect d'un traitement et d'un régime alimentaire stricts. Cela représente un véritable défi pour les systèmes de santé publique, nécessitant une prévention renforcée, un dépistage précoce et une sensibilisation accrue afin de réduire sa prévalence et la gravité de ses complications[6]

3. Les types de diabète

3.1. Diabète de type1 (Diabète juvénile)

Le diabète de type 1 **(DT1)**, également appelé diabète sucré insulinodépendant (DID), affecte tous les âges, mais est le plus souvent diagnostiqué chez les enfants et les jeunes adultes. Ce type survient lorsque les cellules bêta productrices d'insuline dans le

pancréas sont détruites (figure I.2), conduisant à une absence complète de sécrétion d'insuline (Muhammad Mika'ilu Yabo, 2022)[1]. Le diabète de type 1 est une maladie auto-immune résultant d'une réponse immunitaire anormale dirigée contre les cellules β des îlots de Langerhans. Cette destruction est principalement médiée par les lymphocytes T autoréactifs. Les lymphocytes CD4+ activent les macrophages et provoquent des lésions tissulaires, tandis que les lymphocytes CD8+ cytotoxiques détruisent directement les cellules β pancréatiques. Plusieurs cytokines inflammatoires, notamment le TNF-α, l'IL-1 et l'interféron-γ, participent également à cette réaction immunitaire et induisent l'apoptose des cellules β, entraînant ainsi l'arrêt de la production d'insuline. Dans environ 80 % des cas, des auto-anticorps dirigés contre l'enzyme Glutamic Acid Decarboxylase (GAD) sont détectés. En raison de l'absence d'insuline, le glucose ne peut plus pénétrer normalement dans les cellules et s'accumule dans le sang, provoquant une hyperglycémie persistante. L'organisme utilise alors les lipides comme source d'énergie, ce qui entraîne une production excessive de corps cétoniques pouvant conduire à une acidocétose diabétique. Certains virus, tels que les Coxsackievirus, les Rotavirus et les Cytomégalovirus, peuvent également favoriser le déclenchement de cette réponse auto-immune chez les individus génétiquement prédisposés [3] Les personnes atteintes de DT1 doivent recevoir des injections d'insuline à vie pour contrôler leur glycémie[1].

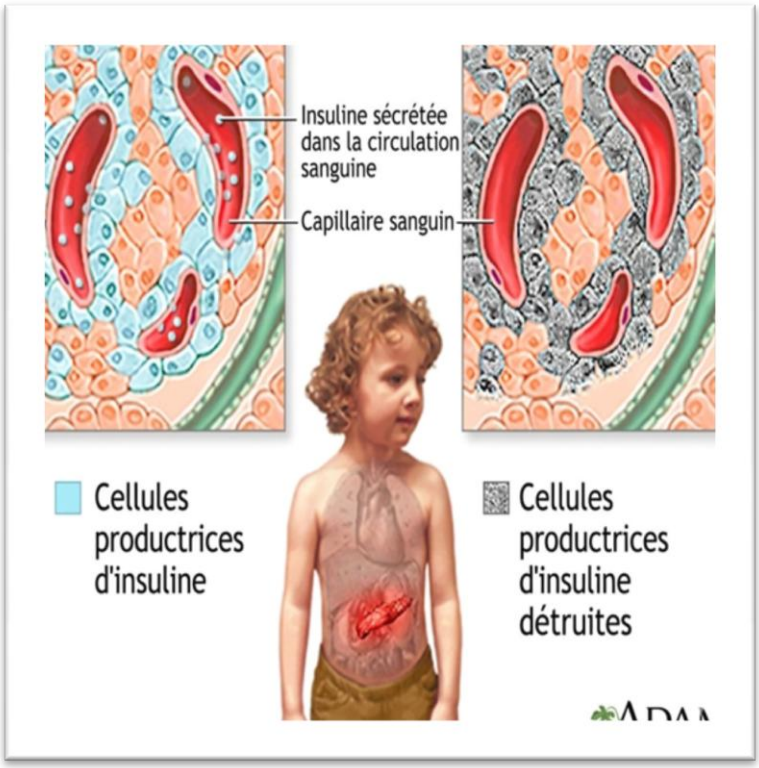


Figure I.2: DT1, Cellules productrices d'insuline saines et détruites [19].

3.2. Diabète de type 2 (Diabète de la maturité)

Le diabète de type 2 **(DT2)**, anciennement appelé diabète non insulinodépendant, représente la forme la plus fréquente du diabète et touche principalement les adultes. Il se caractérise par une incapacité de l'organisme à utiliser correctement l'insuline, appelée insulinorésistance, accompagnée d'une diminution progressive de la sécrétion d'insuline [1]. Cette résistance affecte principalement les muscles, le foie et le tissu adipeux, ce qui entrave l'entrée du glucose dans les cellules et entraîne une augmentation de son taux dans le sang. De plus, la production hépatique du glucose augmente en raison d'une élévation de l'activité du glucagon, contribuant ainsi à l'aggravation de l'hyperglycémie.

Dans les premières phases, les cellules β pancréatiques tentent de compenser ce dysfonctionnement par une augmentation de la sécrétion d'insuline. Cependant, la persistance du stress fonctionnel entraîne progressivement une diminution de leur capacité sécrétoire ainsi qu'une altération progressive de leur fonction. L'évolution de ce type de diabète est associée à plusieurs perturbations métaboliques, notamment l'augmentation des acides gras libres, le stress oxydatif causé par l'accumulation des espèces réactives de l'oxygène (ROS), ainsi qu'un état d'inflammation chronique de bas grade affectant la sensibilité cellulaire à l'insuline.

Par ailleurs, l'obésité, en particulier l'accumulation du tissu adipeux, provoque une perturbation de la sécrétion des adipokines et de la résistine, ce qui favorise l'augmentation de l'insulinorésistance [3] Ce type de diabète est étroitement lié à la sédentarité et à un mode de vie malsain. Les facteurs génétiques jouent également un rôle important dans son apparition. Il représente plus de 90 % des cas de diabète dans le monde et peut, en l'absence d'une prise en charge adéquate, entraîner des complications graves telles que les maladies cardiovasculaires, la néphropathie diabétique, la rétinopathie diabétique et la neuropathie diabétique [1].

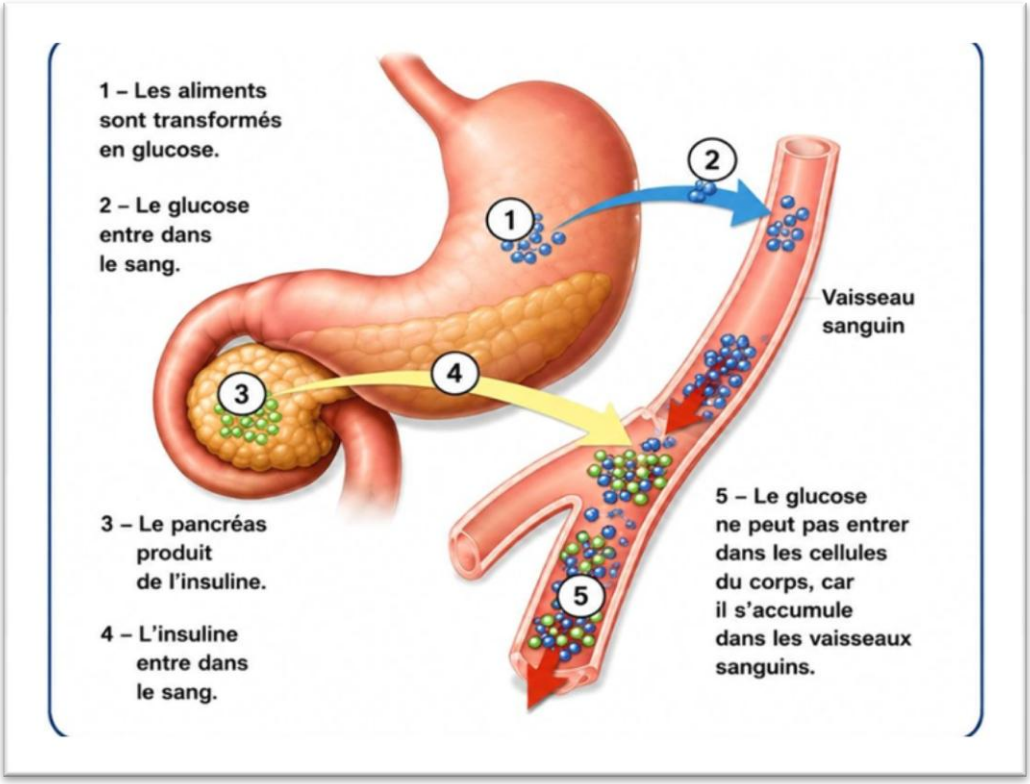


Figure I.3: DT2, Interaction insuline - glucose dans un vaisseau sanguin [11].

3.3. Diabète gestationnel

L'hyperglycémie gestationnelle, ou hyperglycémie pendant la grossesse, est un état pathologique caractérisé par une augmentation du taux de sucre dans le sang. Elle peut conduire à une forme de diabète de type 2, caractérisée par une variabilité du taux de sucre dans le sang. Le diagnostic de cette pathologie est posé soit avant la naissance de l'enfant, soit dans les mois suivant l'accouchement, et ce, indépendamment du type de prise en charge médicale précédemment suivie. Cette étude concerne deux populations de femmes distinctes. La première population est composée de femmes ayant un déséquilibre glucidique préalable à la grossesse, un problème qui n'est pas encore bien compris. La seconde population est composée de femmes qui développent un déséquilibre glucidique pendant la grossesse. Il est important de noter que la prévalence de la division due à la prévalence globale de l'émail n'est pas connue. Cependant, il est établi que la prévalence de l'hyperglycémie gestationnelle (de 1 à 14 % selon les populations) est plus élevée que celle de l'hyperglycémie métaphysique de type 2 en ethnie [7]

4. Aspects cliniques du diabète

4.1. Symptômes du diabète

Les signes symptômes du diabète varient selon le type, mais certains signes cliniques sont communs aux diabètes de type 1 et de type 2. Les plus importants sont [8]:

- ***Troubles visuels*** : vision floue ou trouble due à la sécheresse du cristallin;
- ***Fatigue extrême :*** sensation persistante d'épuisement physique et de faiblesse générale;
- ***Perte de poids :*** perte de poids inexpliquée malgré une alimentation équilibrée;
- ***Infections fréquentes :*** infections urinaires ou mycoses, par exemple;
- ***Mictions fréquentes :*** augmentation de la fréquence et du volume des urines;
- ***Soif excessive :*** sensation constante de soif et bouche sèche.

4.2. Causes et facteurs de risque

De multiples facteurs contribuent au développement du diabète et influencent son type ainsi que son degré de gravité. La compréhension de ces facteurs est essentielle afin d'identifier les principaux risques associés à l'apparition de cette maladie. Parmi ces facteurs[8].

a. Facteurs génétiques :

Les études portant sur les antécédents familiaux et les cas de jumeaux ont démontré l'importance des facteurs génétiques dans le développement du diabète. Dans le diabète de type 1, la probabilité que l'autre jumeau développe la maladie atteint environ 50 % lorsqu'un des deux est atteint, tandis que ce risque est encore plus élevé dans le diabète de type 2. Certains gènes jouent un rôle déterminant dans la prédisposition à la maladie, contrairement à d'autres maladies génétiques, telles que la mucoviscidose, qui résultent d'une anomalie touchant un seul gène.

b. Infections virales:

Le diabète de type 1 apparaît plus fréquemment durant l'hiver. Certains virus, tels que le virus des oreillons et le virus Coxsackie, peuvent contribuer à la destruction des cellules pancréatiques, favorisant ainsi l'apparition de la maladie, bien que cette éventualité demeure relativement rare.

c. Environnement et mode de vie:

Les facteurs environnementaux et les habitudes de vie jouent un rôle majeur dans le développement du diabète. Le risque de développer la maladie peut augmenter lors du passage d'un pays à faible prévalence vers un pays où celle-ci est élevée. De plus, les personnes souffrant d'obésité ou adoptant une alimentation déséquilibrée présentent un

risque accru de développer un diabète, notamment de DT2. Il convient toutefois de préciser que la consommation de sucre, à elle seule, ne constitue pas la principale cause du diabète, tandis que l'obésité représente un facteur de risque majeur.

4.3. Complications du diabète

Le diabète peut entraîner de nombreuses complications, certaines apparaissant précocement, tandis que d'autres se développent après plusieurs années de maladie.

4.3.1. Complications aiguës (soudaines)

Un déséquilibre soudain de la glycémie entraîne deux dangers: Une augmentation brutale qui provoque une soif et une fatigue extrêmes, et peut conduire à l'accumulation de « corps cétoniques » toxiques ou à un coma hyperosmolaire et une dépression sévère (due à un excès de médicaments ou au fait de sauter un repas) qui provoque des tremblements et des étourdissements. Le plus grand risque dans les deux cas est la possibilité de perdre connaissance et de tomber dans le coma.

4.3.2. Complications chroniques (à long terme)

Il apparaît après des années en raison de la négligence du contrôle du sucre, qui endommage les vaisseaux sanguins, et se divise en :

- **Petits vaisseaux :** Les yeux sont touchés (mauvaise vision pouvant conduire à la cécité), les reins (insuffisance pouvant évoluer vers une insuffisance rénale) et les nerfs (douleurs, engourdissements et perte de sensation dans les extrémités) ;
- **Gros vaisseaux :** le risque de tension artérielle, et stable est la seule ligne de défense pour prévenir ces complications[3]. [9].
- **Pied diabétique :** Il s'agit d'une maladie complexe qui commence par des lésions nerveuses et une mauvaise circulation sanguine vers la jambe. Le patient perd la sensation de douleur et peut développer de graves ulcères pouvant conduire à une amputation. Cependant, une fois l'ulcère cicatrisé, il réapparaît souvent dans plus de 60 % des cas. Par conséquent, les médecins considèrent le site d'une plaie cicatrisée simplement comme une « trêve temporaire » et non comme une guérison complète. Il n'existe pas de réponse préventive unique autre que contrôler la glycémie, s'engager à un suivi régulier auprès du médecin et porter des chaussures médicales spécialement conçue[10]

5. Méthodes de diagnostic du diabète

Pour diagnostiquer le pré-diabète et le diabète, les médecins s'appuient sur quatre principaux tests sanguins [11]:

1. La glycémie aléatoire (GA);

2. La glycémie à jeun (GAJ);
3. Le test d'hyperglycémie provoquée par voie orale (HGPO);
4. Le dosage de l'hémoglobine glyquée (HbA1c).

5.1. Test de glycémie capillaire et glycémie aléatoire(GA)

La glycémie peut être mesurée à tout moment de la journée, sans nécessité de jeûne. Elle correspond à la glycémie dite « aléatoire ». Ce dosage peut être réalisé soit par une analyse sanguine veineuse en laboratoire, soit de manière rapide à l'aide d'un glucomètre, qui permet d'évaluer la glycémie capillaire à partir d'une petite goutte de sang.

Le glucomètre est principalement utilisé pour le suivi quotidien des patients diabétiques, tandis que l'analyse veineu0se est la méthode de référence pour le diagnostic[11].

Procédure de mesure avec glucomètre :

1) **Préparation du dispositif :** insertion de la bandelette réactive dans le glucomètre.
2) **Piqûre du doigt :** utilisation d'une lancette pour effectuer une légère piqûre au bout du doigt.
3) **Prélèvement de l'échantillon :** dépôt d'une goutte de sang sur la bandelette.
4) **Mesure :** le glucomètre analyse la glycémie et affiche le résultat en quelques secondes.

Une glycémie **≥ 200 mg/dl (11,1mmol/l)** associée à des symptômes de diabète est considérée comme un indicateur diagnostique du diabète.

5.2. Test de glycémie à jeun(GAJ)

La glycémie à jeun (GAJ) correspond à la mesure de la concentration de glucose dans le sang après une période d'abstinence calorique d'au moins 8 heures. Contrairement à la glycémie aléatoire, ce dosage est l'examen de première intention standardisé et privilégié pour le dépistage et le diagnostic du diabète en pratique clinique. Bien qu'elle puisse être évaluée de manière indicative par un lecteur de glycémie capillaire, la méthode de référence (Gold Standard) repose sur un prélèvement de sang veineux analysé en laboratoire par méthode enzymatique (par exemple, la méthode à l'hexokinase)[11]

Procédure de mesure en laboratoire (méthode de référence) :

1. **Conditionnement du patient :** vérification du respect strict du jeûne hydrique et calorique (8 à 12 heures).

2. **Prélèvement veineux :** réalisation d'une ponction veineuse, généralement au pli du coude, dans un tube contenant un inhibiteur de la glycolyse (comme le parahydroxybenzoate ou le fluorure de sodium) pour stabiliser le taux de glucose.
3. **Centrifugation et analyse :** séparation du plasma et dosage automatisé de la concentration de glucose par le système analytique du laboratoire.
4. **Interprétation des résultats :** transmission de la valeur quantitative exprimée en mg/dl ou en mmol/l.

Selon les critères de l'Organisation Mondiale de la Santé (OMS) et de l'ADA, les seuils d'interprétation pour la GAJ sont définis comme suit :

- **Glycémie normale :**< 100 mg/dl (5,6 mmol/l).
- **Prédiabète (anomalie de la glycémie à jeun) :** comprise entre 100 et 125 mg/dl (5,6 à 6,9 mmol/l).
- **Diabète :** une valeur > 126 mg/dl (7,0 mmol/l), constatée à deux reprises, confirme le diagnostic de diabète.

5.3. Test de tolérance au glucose (HGPO)

Ce test est un moyen de déterminer comment le corps traite le sucre, dans lequel le patient boit un liquide contenant 75 grammes de glucose après un jeûne d'une nuit. Les résultats sont mesurés après deux heures, S'il est inférieur à 140 mg/dl, le résultat est normal, mais s'il est compris entre 140 et 200 mg/dl, il indique un « diabète latent » ou pré-diabète, et il est considéré comme un diabète clair s'il dépasse 200 mg/dl(Sirsath & Dahiwal, 2023[11].

5.4. Test de glycémie cumulative (HbA1c)

Ce test permet d’évaluer le pourcentage de sucre présent dans le sang sur une période de deux à trois mois et constitue un indicateur important pour mesurer le niveau de contrôle du diabète chez le patient. Le taux normal chez les personnes non diabétiques se situe entre 4 % et 6 %, tandis qu’il est recommandé aux patients diabétiques de maintenir un taux inférieur à 7 % ou à 6,5 % afin de réduire le risque de complications oculaires, rénales et neurologiques. Une diminution de 1 % du taux d’HbA1c réduit le risque de complications d’environ 12 %.

Il existe une relation étroite entre le taux d’HbA1c et la glycémie moyenne dans le sang. Un taux de 4 % correspond à une glycémie moyenne de 60 mg/dl, tandis qu’un taux de 6 % représente environ 120 mg/dl, et un taux de 8 % correspond à une moyenne de 180

mg/dl. En général, chaque augmentation de 1 % du taux d'HbA1c correspond à une augmentation d'environ 30 mg/dl de la glycémie moyenne.

Recommandations de suivi périodique : Une bonne prise en charge de la maladie nécessite un contrôle régulier de la glycémie, au moins deux fois par année, pour les patients de DT2 dont l'état c'est stabilisé. En cas de changement de plan de traitement ou lorsque le taux de sucre reste constamment élevé, l'examen doit être effectué tous les 3 mois pour s'assurer que des niveaux acceptables sont atteints et le traitement ou le régime alimentaire doit être modifié si nécessaire. Il convient également de noter que l'analyse d'urine n'est généralement pas utilisée pour le diagnostic final car les analyses de sang sont plus précises et plus détaillées[12].

6. Indicateurs cliniques et biologiques

Les indicateurs cliniques du diabète sont divisés en plusieurs catégories principales utilisées dans le diagnostic, le suivi médical et l'évaluation des complications. Parmi les plus importants figurent les indicateurs biochimiques comprenant la glycémie à jeun (Fasting Blood Glucose), la glycémie aléatoire, le test oral de tolérance au glucose (OGTT) et l'hémoglobine glyquée (HbA1c).

Les indicateurs anthropométriques et physiologiques comprennent l'indice de masse corporelle (BMI), le poids, la taille, le tour de taille et la pression artérielle. Ces indicateurs permettent d'évaluer l'obésité et le risque de complications associées au diabète.

Les indicateurs lipidiques (Lipid Profile) regroupent le cholestérol total, les lipoprotéines de basse densité (LDL), les lipoprotéines de haute densité (HDL) ainsi que les triglycérides (TG). Ils sont utilisés pour évaluer le risque de maladies cardiovasculaires chez les patients diabétiques.

Les indicateurs hormonaux et métaboliques incluent le taux d'insuline et l'indice de résistance à l'insuline (HOMA-IR), utilisés pour évaluer la sensibilité de l'organisme à l'insuline et la fonction pancréatique.

Enfin, les indicateurs de la fonction rénale comprennent la créatinine (Creatinine), le débit de filtration glomérulaire estimé (EGFR) et la microalbuminurie (Microalbuminuria), qui servent au dépistage précoce de la néphropathie diabétique[15]. [16] [17]

Le tableau I.1 présente quelques indicateurs cliniques et biologiques utilisés par les spécialistes pour la prise en charge de la maladie du diabète.

Tableau I.1: Indicateurs cliniques et biologiques du diabète [15]. [16] [17]

Indicateur	Symbole	Valeurs normales
Glycémie à jeun	FBG / GLY	70 – 99 mg/dl
Hémoglobine glyquée	HbA1c	< 5.7 %
Indice de masse corporelle	BMI	18.5 – 24.9
Pression artérielle	BP	< 120/80 mmHg
Triglycérides	TG	< 150 mg/dl
Cholestérol HDL	HDL	> 40 mg/dl chez les hommes > 50 mg/dl chez les femmes
Cholestérol (LDL)	LDL	< 100 mg/dl
Cholestérol total	TC	< 200 mg/dl
Insuline	Insulin	2 – 25 µIU/ml
Tour de taille	WC	< 102 cm homes < 88 cm femmes

7. Traitement du diabète

7.1. Traitement médicamenteux

Il nécessite de contrôler la glycémie pour éviter les complications, et cela se produit en combinant des médicaments avec des changements de mode de vie. La metformine est le traitement primitif et le plus courant pour les patients de DT2. Il existe également de nouvelles préférences telles que les inhibiteurs du SGLT-2 qui protègent le cœur et les reins, et les agonistes des récepteurs GLP-1 qui aident à réduire le poids et à réguler le sucre. Quant aux cas de DT1 ou avancés de DT2, la seule solution est l'insuline pour le traitement[13]

7.2. Traitement non pharmacologique

7.2.1. Piliers de base du mode de vie

Le traitement du diabète de DT1 nécessite une modification du mode de vie, comme le fait de consommer des aliments sains, de faire de l'exercice et de favoriser le sommeil. Même le soutien psychologique et social joue un rôle important pour contrôler le taux de sucre et éviter les complications [14].

7.2.2. Le protocole thérapeutique en fonction du taux d'HbA1c

La mise en route du traitement non médicamenteux seul est ajustée en fonction du taux d'hémoglobine glycosylée (cumulatif). Si le pourcentage d'HbA1c est inférieur à 7 %, le traitement commence par une réforme du mode de vie, y compris l'alimentation et l'exercice. Le patient dispose d'un délai de 3 à 6 mois pour évaluer son état avant d'introduire un traitement médicamenteux tel que la metformine, mais si le taux est compris entre 7 % et 9 %. Nous appliquons des procédures non pharmacologiques en association avec des médicaments dès le premier instant[14].

7.2.3. Intervention chirurgicale

L'intervention chirurgicale (chirurgie bariatrique) est considérée comme une option essentielle pour les patients souffrant d'obésité sévère, lorsque l'indice de masse corporelle dépasse (BMI > 35 kg/m²). Cette procédure vise à contrôler l'appétit et à améliorer la réponse de l'organisme à l'insuline, et est intégrée dans un plan qui comprend des médicaments modernes (tels que l'arGLP1 ou l'arGLP/GIP) pour assurer une réduction de poids et des niveaux cumulatifs[14]

8. La Prévention du diabète

La prévention du diabète, en particulier du diabète de DT2, repose sur l'adoption d'un mode de vie sain. Le risque de développer cette maladie peut être réduit en modifiant certains facteurs de risque tels que l'obésité, une mauvaise alimentation, la sédentarité et les mauvaises habitudes de vie. Des études ont montré que l'amélioration des habitudes alimentaires et la pratique régulière d'une activité physique permettent de prévenir le diabète ou d'en retarder l'apparition.

L'Organisation Mondiale de la Santé (OMS)[2] et la Fédération Internationale du Diabète (FID)[7] recommandent les mesures suivantes :

- Réduire la consommation de sucres ;
- Réduire la consommation de graisses saturées ;
- Augmenter la consommation de fibres, de légumes et de fruits ;
- Pratiquer une activité physique régulière de 30 à 45 minutes, 3 à 5 fois par semaine ;
- Ces recommandations visent à maintenir une glycémie normale et à diminuer le risque de développer le diabète.

Conclusion

Dans ce chapitre nous avons présenté une description de la maladie du diabète : ces différents types, les méthodes de diagnostic du diabète, le traitement et la prévention.

Différents indicateurs cliniques et biologiques sont utilisés par les spécialistes pour la prise en charge de la maladie.

Références:

[1] M. M. Yabo, A. B. Garko, A. A. Muslim, & H. U. Suru. (2022). A Review of Diabetes Datasets. Journal of Computer Sciences and Applications, 10(1), 6–15. https://doi.org/10.12691/jcsa-10-1-2

[2] Organisation mondiale de la Santé, Rapport mondial sur le diabète, Genève, OMS, 2016, «La charge de morbidité mondiale du diabète».

[3] S.Alam ;R.Hasan ,S.Begum;M.A.Kabir,and M.S.Ahsan. Journal of Diabetology : Diabetes Mellitus: Insights from Epidemiology, Biochemistry, Risk Factors, Diagnosis, Complications and Comprehensive Management, 2021, **2**, pp. 36–50.

[4] Q.-Y Wang ,E. Mensah , Li Z.-C., D.-G. Wang ,C.-W. Zhang , B. M. Baako , L Zha and X Kong ; African regional and national burden of diabetes mellitus and its attributable risk factors from 1990 to 2021; Frontiers in Endocrinology, 2025 , **16**.https://doi.org/10.3389/fendo.2025.1643999

[5] K. L .Ong,L. K Stafford, S. A .McLaughlin , E. J Boyko , S. E.Vollset , A. E Smith , **et al.** (2023). *Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: A systematic analysis for the Global Burden of Disease Study 2021*. *The Lancet*, ***402***(10397), 203–234. https://doi.org/10.1016/S0140-6736(23)01301-6

[6] Fédération Internationale du Diabète (IDF), IDF Diabetes Atlas, 11e édition, International Diabetes Federation, Bruxelles, 2025, p. 28.

[7] HAS. Rapport de synthèse : dépistage et diagnostic du diabète gestationnel. Recommandation en santé publique. 2006

[8] R .Bilous , &R. Donnelly, (2010). *Handbook of Diabetes* (4th ed.). Wiley-Blackwell

[9] محمود, بابللي. ضحى.(2002). حقائق عن داء السكري. الرياض: مكتبة العبيكان

[10] D. G .Armstrong., A. J. M .Boulton, and S. A .Bus. Journal of Medicine : Diabetic foot ulcers and their recurrence. New England; (2017), 376(24), 2367-2375. https://doi.org/10.1056/NEJMra1615439

[11] S. U .Sirsath , &S. S. Dahiwal . (2023). Review on diabetes mellitus disease. *International Research Journal of Modernization in Engineering Technology and Science*, **5**(4), 4424–4433. https://doi.org/10.56726/IRJMETS36867

[12] نسرين كمال عبد الله . (2016). الداء السكري. الكويت :المركز العربي لتأليف وترجمة العلوم الصحية.

[13] R. Weinberg Sibony, O. Segev, S. Dor, and I. Raz, (2023). Drug Therapies for Diabetes. International Journal of Molecular Sciences, 24(24), 17147. https://doi.org/10.3390/ijms242417147

[14] F. J García Soidán, &J Riveiro Villanueva. (2025). Tratamiento farmacológico de la diabetes mellitus tipo 2. Atención Primaria, 57, 103143. https://doi.org/10.1016/j.aprim.2024.103143

[15] World Health Organization (WHO), "Diabetes", Geneva, Switzerland, 2023.

[16] American Diabetes Association (ADA), “Standards of Care in Diabetes—2025,” Diabetes Care, vol. 48, Supplement 1, 2025.

[17] International Diabetes Federation (IDF), IDF Diabetes Atlas, 10th ed., Brussels, Belgium: International Diabetes Federation, 2021.

[18] G. D. C. F. D. C. P. Project, "Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021," The Lancet, vol. 402, no. 10397, pp. 203-234, 2023.

[19] MedlinePlus Medical Encyclopedia – Type 1 Diabetes Source: U.S. National Library of Medicine medlineplus.gov

Chapitre II :

Outils et méthodologies pour l'étude des données du diabète

Chapitre II : Outils et méthodologies pour l'étude des données du diabète

1. Introduction

L'étude des indicateurs (cliniques, biologiques et autres) relatifs à la maladie du diabète a suscité l'utilisation de plusieurs outils. Les calculs de la statistique descriptive présentent le plus des travaux.

Aujourd'hui, l'Intelligence Artificielle (IA) est sollicitée pour analyser les données et pour prédire plusieurs phénomènes et situations.

Dans ce chapitre nous rappellerons les notions de base de l'Intelligence Artificielle et les différentes méthodes d'apprentissage profond pour développer de modèles d'analyse et de prédiction. Nous rappellerons les notions de base du calcul statistique nécessaire pour l'analyse de données relative à la maladie.

2. Importance de l'Intelligence Artificielle

2.1. Définition d'intelligence artificielle

Dernièrement, les technologies de l'intelligence artificielle ont connu un développement accéléré et une diffusion à grande échelle dans divers secteurs, au point de devenir un pilier fondamental dans la conception de solutions technologiques avancées et l'amélioration de la performance des systèmes modernes. L'intelligence artificielle est définie comme un ensemble de modèles et d'algorithmes informatiques capables de simuler les processus cognitifs humains, notamment l'apprentissage, le raisonnement et la prise de décision. Ces systèmes se distinguent par leurs capacités adaptatives avancées, reposant sur l'analyse de vastes volumes de données et l'extraction de schémas pertinents, ce qui leur permet d'améliorer leur performance de manière dynamique et continue. Par conséquent, ces technologies offrent la possibilité de développer des modèles prédictifs à haute précision, de fournir des recommandations intelligentes et personnalisées, ainsi que de concevoir des solutions évolutives aptes à traiter des problématiques complexes dans des environnements variés [1].

2.2. Applications de l'Intelligence Artificielle dans divers domaines

L'intelligence artificielle est largement utilisée dans plusieurs domaines.

2.2.1. Domaine scientifique

L'intelligence artificielle, soutenue par des techniques d'apprentissage automatique, représente un moteur fondamental pour le développement des sciences fondamentales telles que les mathématiques, la physique, la chimie et les sciences de la terre. Son importance scientifique est évidente dans sa capacité supérieure à traiter et analyser des données volumineuses à haut débit, à les classer et à créer des modèles prédictifs précis qui fournissent de nouvelles informations cognitives. Ce rôle recoupe également étroitement la recherche biologique et la biologie, ce qui ouvre des horizons prometteurs pour le diagnostic intelligent et la découverte de bio marqueurs de maladies [2].

2.2.2 Domaine financier (Finance)

Dans le secteur financier, l'intelligence artificielle est utilisée pour analyser les marchés et exécuter des transactions automatisées. Il permet de détecter les fraudes en analysant les opportunités en temps réel. De plus, il permet d'identifier les risques et d'améliorer la gestion de la dette des clients, rendant ainsi les décisions financières plus fiables et efficaces [3].

2.2.3 Domaine industriel (Manufacturing)

Dans le secteur industriel, l'intelligence artificielle contribue à améliorer la maintenance prédictive, permettant de prédire les pannes des machines, et est également utilisée pour surveiller la qualité des produits et détecter rapidement les défauts. En outre, il s'efforce d'améliorer les chaînes d'approvisionnement, ce qui contribue à réduire les coûts et à augmenter l'efficacité du travail [3].

2.2.4 Domaine du commerce

Dans le domaine du commerce, l'intelligence artificielle améliore l'expérience utilisateur en proposant des nominations de produits personnalisées, mais permet également une gestion optimale des stocks et une mise à jour dynamique des prix en fonction du volume des commandes [3].

2.3. Applications de L'intelligence artificielle dans le domaine de la santé et la médecine

L'intelligence artificielle est largement utilisée dans le secteur de la santé, permettant de diagnostiquer des maladies grâce à l'analyse d'images médicales. Il contribue également à l'identification et au développement de sites de médicaments personnalisés. De plus, il fournit des informations précieuses dans le domaine de la médecine personnalisée en proposant des

traitements adaptés à chaque patient en fonction de ses conditions médicales et génétiques [3].

Ainsi, l'Intelligence Artificielle est devenue nécessaire en médecine pour analyser les données. Historiquement, cela a commencé dans les années 1970 avec les systèmes experts, puis s'est développé vers l'apprentissage automatique dans les années 1990, conduisant à la révolution actuelle de l'Apprentissage Profond (Deep Learning). De nombreux chercheurs ont utilisé ces techniques pour prédire le diabète, comme en témoignent plusieurs travaux.

Les travaux de Gulshan V. et al. (2018) [4], intitulée '*Deep Learning for Diabetic Retinopathy'*, s'est concentrée sur l'utilisation de techniques d'apprentissage profond pour diagnostiquer la maladie du diabète grâce à l'analyse d'images médicales. Les travaux étaient basées sur l'utilisation de réseaux neuronaux convolutifs (CNN) pour examiner les images du fond d'œil de patients diabétiques. Les résultats ont montré la capacité supérieure de l'intelligence artificielle à surveiller les complications visuelles de la maladie et à les diagnostiquer précocement [4].

En 2021, Hima K., dans son mémoire de Master intitulée '*La prédiction du diabète utilisant le deeplearning*' [5], a présenté un modèle de prédiction du diabète à l'aide de réseaux de neurones artificiels (ANN). Le mémoire s'est appuyé sur une base de données extraite de l'hôpital de Francfort en Allemagne qui comprenait 9 variables. Les résultats ont conclu que le modèle a atteint une précision de prédiction de 82,08 %.

En 2022, Alex S. A. et al. [6] ont mené une étude intitulée 'Deep LSTM Model for DiabetesPredictionwith Class Balancing by SMOTE', dans le but d'améliorer la prédiction du diabète à l'aide du modèle LSTM. Les résultats ont montré que le modèle proposé a atteint une précision élevée de 99,64 % après avoir résolu le problème du déséquilibre des données.

En 2023, Kowsher Md. et al. [7] ont mené une étude intitulée *'Prognosis and TreatmentPrediction of Type-2 DiabetesUsingDeep Neural Network and Machine Learning Classifiers'*, qui visait à comparer plusieurs algorithmes pour prédire le diabète et suggérer un traitement approprié. Les résultats ont montré la supériorité des réseaux neuronaux profonds, atteignant une précision de 95,14 % dans la prévision de la maladie.

D'autres mémoires de master sur la maladie du cancer du sein, utilisant l'Apprentissage Profond, ont été soutenus dans notre département. Le mémoire de Debba N.-H et Belaid S. en 2023 [8] et le mémoire de Berguiga I. et Laib S. en 2024 [9]

3. Apprentissage automatique (Machine Learning)

3.1.Apprentissage automatique

C'est une science qui consiste à étudier les ordinateurs pour acquérir des expériences. Il s'agit d'un type d'intelligence artificielle qui entraîne des algorithmes informatiques à l'aide de données, de sorte qu'une fois que nous y travaillons, ils produisent des données et des résultats pour nous. L'apprentissage automatique prend également des données énormes et volumineuses, les traite, fait des prédictions, classe, collecte et réduit les dimensions des données fournies [10]. Il est divisé en trois types importants, à savoir [11]:

3.1.1. Apprentissage par renforcement

C'est l'une des principales branches de l'apprentissage automatique, dans laquelle le modèle informatique (agent intelligent) apprend à prendre des décisions par interaction directe avec l'environnement par essais et erreurs. Cette approche repose sur l'idée de « récompense et punition ». Chaque fois que le système prend une décision correcte qui le rapproche de l'objectif, il reçoit une récompense positive (+), et chaque fois qu'il commet une erreur, il reçoit une punition ou une récompense négative (-). La fonction de l'algorithme consiste à développer une politique autonome garantissant la maximisation des récompenses totales à long terme, ce qui le rend idéal pour les applications robotiques, les jouets intelligents et les systèmes autonomes [12].

3.1.2. Apprentissage semi-supervisé

Il s'agit d'une combinaison d'apprentissage dirigé et non dirigé, ce qui signifie qu'il repose sur un ensemble de données contenant une partie des données marquées et une autre partie non marquée, et qu'il utilise des techniques d'apprentissage par renforcement profond (DRL) et (GAN) comme méthodes d'apprentissagesemi-supervisé.

3.1.3. Apprentissage non supervisé

Il s'agit de systèmes capables d'apprendre sans étiquettes ni marqueurs sur les données, et le clustering, la réduction de dimensionnalité et les techniques génératives sont considérés comme des méthodes d'apprentissage non supervisé.

3.1.4. Apprentissage supervisé

L'apprentissage dirigé est une composante essentielle de l'apprentissage automatique. Il repose sur des données paramétrées utilisées pour entraîner les modèles d'IA à classifier correctement de nouvelles données. Son environnement comporte un ensemble d'entrées et de sorties parallèles x_{t,Y_t} Plusieurs méthodes existent pour l'apprentissagesupervisé: CNN, RNN,

LSTM, ANN. Ces réseaux seront expliqués dans les sections suivantes.

3.2. Apprentissage Profond(deeplearning)

L'Apprentissage Profond est une branche de l'apprentissage automatique, une méthode d'apprentissage complète capable de résoudre presque tous les types de problèmes dans divers domaines. Il repose sur des réseaux de neurones artificiels à plusieurs couches cachées. Ces couches permettent au modèle d'extraire automatiquement des caractéristiques complexes à partir d'énormes quantités de données, améliorant ainsi sa capacité à produire des résultats précis et efficaces. Ce réseau est composé de trois types couches principales [10].

4. Les réseaux de neurones artificiels / Artificial Neural Networks (ANN)

Les réseaux de neurones artificiels, qui constituent l'élément fondamental de l'Apprentissage Profond, sont inspirés du fonctionnement des neurones du cerveau humain. En effet, le cerveau contient un grand nombre de neurones interconnectés capables de recevoir des entrées, de les traiter, puis de transmettre des sorties à d'autres cellules. De manière similaire, le traitement des données informatiques s'appuie sur des algorithmes qui imitent ce mécanisme. Ces réseaux sont structurés en plusieurs couches, comprenant une couche d'entrée qui reçoit les données, une ou plusieurs couches cachées chargées du traitement de l'information et de l'extraction des caractéristiques, ainsi qu'une couche de sortie qui fournit le résultat final.Ce réseau est composé de trois types de couches principales [13].

4.1 Couche d'entrée (Input Layer)

La couche d'entrée est la première couche du réseau de neurones. Elle est composée d'un ensemble de neurones qui représentent les données d'entrée ou les caractéristiques (features) du problème. Cette couche n'effectue pas de calculs complexes, mais elle représente le vecteur initial des données, qui s'écrit sous la forme :

$$X = [x_1, x_2, x_3, \dots, x_n]$$

où chaque élément x_i correspond à une caractéristique des données (par exemple : l'âge, le poids, la pression…). Dans certains cas, ces valeurs sont transmises telles quelles, ou bien après une étape de normalisation afin de les mettre à la même échelle, comme suit :

$$x_i^{'} = \frac{x_i - \mu}{\sigma}$$

x'_i*:* représente la valeur normalisée (ou standardisée) de la caractéristique $x_i$après la mise à l'échelle.

x_i *: représente la valeur initiale (brute) de la caractéristique avant traitement.*

μ:représente la moyenne arithmétique (Mean) de la variable x sur l'ensemble des données.

σ: représente l'écart-type (Standard Deviation) de la variable x, qui mesure la dispersion des données autour de la moyenne.

Ainsi, le rôle principal de la couche d'entrée est de recevoir les valeurs initiales, de les organiser sous forme de vecteur, puis de les transmettre à la couche cachée sans effectuer de transformation complexe, ce qui en fait le point de départ du processus d'apprentissage dans le réseau.

4.2 Couche cachée (Hidden Layer)

La couche cachée est la deuxième couche du réseau neuronal. Elle est composée d'un ensemble de neurones qui traitent les données reçues de la couche d'entrée, et c'est là que commence la véritable "intelligence" du réseau. Dans cette couche, chaque neurone effectue d'abord une combinaison linéaire en multipliant les entrées par les poids puis en ajoutant le biais selon la relation :

$$z = \sum_{i=1}^{n} wix_i + b$$

- z **:** représente la somme pondérée (appelée aussi l'entrée nette ou la combinaison linéaire) du neurone.
- ∑: représente le symbole mathématique de la somme (Sommation).
- i : représente l'indice de l'entrée actuelle (allant de 1 jusqu'à n).
- n : représente le nombre total d'entrées connectées au neurone.
- W_i : représente le poids synaptique (Weight) associé à l'entrée i, qui détermine l'importance de ce signal.
- x_i : représente la valeur de la donnée d'entrée i reçue par le neurone.
- b : Représente le biais (Bias),

Ensuite, le résultat est passé à travers une fonction d'activation :

$$a = f(z)$$

Ce qui introduit la non-linéarité et permet d'extraire des caractéristiques complexes des données. Parmi les fonctions d'activation les plus importantes, on trouve la fonction

Sigmoid qui donne des valeurs entre 0 et 1, la fonction ReLU définie par $max(0, z)$,et la fonction Tanh qui donne des valeurs entre -1 et 1. De manière générale, la sortie d'un neurone dans cette couche peut s'écrire comme :

$$a_i = f(\sum_i w_{ij} x_i + b_j).$$

Le rôle principal de cette couche est d'extraire des relations complexes et non linéaires à partir des données, et elle est directement reliée à la couche de sortie.

A noter que pour la modélisation numérique, ***on peut proposer de travailler avec plusieurs couches cachées.***

4.3 Couche de sortie (Output Layer)

La couche de sortie (*Output Layer*) est la dernière couche du réseau neuronal et représente le résultat final du modèle. Elle fonctionne selon le même principe que la couche cachée, en recevant les valeurs provenant des couches précédentes et en les traitant à l'aide des poids (*Weights*), du biais (*Bias*) et de la fonction d'activation (*Activation Function*) afin de produire les sorties finales. Son fonctionnement peut être exprimé par l'équation générale suivante :

$$y = f\left(\sum_j w_j a_j + b\right)$$

Selon le type de problème, la fonction d'activation utilisée dans cette couche varie. Dans le cas de la **classification binaire** "*Binary Classification*", la fonction **Sigmoid** est utilisée pour fournir une valeur comprise entre 0 et 1 représentant la probabilité d'appartenance à une classe. Pour la classification multi-classes, on utilise la fonction **Softmax**, qui transforme les valeurs en probabilités dont la somme est égale à 1 pour toutes les classes possibles. En revanche, dans les problèmes de régression (*Regression*), aucune fonction d'activation n'est généralement utilisée, et la sortie est directement :$\hat{y} = z$

Le rôle principal de cette couche est de fournir la décision finale du modèle sur la base des données traitées par les couches précédentes, ce qui en fait l'étape décisive du réseau neuronal.

4.4 Neurones artificiels

Les neurones artificiels (ou nœuds) sont les unités de base pour la construction du réseau numérique. Ils s'inspirent des cellules biologiques du cerveau humain. Chaque cellule reçoit des signaux d'entrée (provenant de variables externes ou d'autres cellules) et les combine mathématiquement en calculant la « somme pondérée » pour produire une seule

sortie. Malgré la faiblesse de la capacité d'une seule cellule individuellement, son interconnexion dense en couches successives donne au réseau sa force pour résoudre des problèmes complexes[14].

Poids et biais : Ce sont les paramètres mathématiques ajustables responsables de l'apprentissage du réseau [14]:

-Poids (w) : Ils agissent comme des multiplicateurs qui ajustent la force du signal entre les cellules (simulant les synapses biologiques). Ils amplifient ou affaiblissent le signal.

-Biais (b) : agit comme un seuil spécifique à la cellule, et s'ajoute à la somme pondérée pour déterminer la facilité avec laquelle il peut être activé. L'algorithme d'apprentissage ajuste constamment ces valeurs pour réduire les erreurs de prédiction.

Fonctions d'activation : Ce sont des formules mathématiques (somme + pondération + biais) utilisées pour déterminer la valeur finale. Cela permet au modèle de comprendre et de représenter les relations complexes et non linéaires présentes dans les données [14].Voici quelques exemples parmi les plus courants [15]:

- **ReLU :** Renvoie la même valeur si elle est positive, et zéro si elle est négative :

$$f(x) = \max(0, x)$$

- **Sigmoïde (Sigmoid**): Convertit les valeurs en un intervalle entre0 et 1 :

$$f(x) = \frac{1}{1 + e^{-x}}$$

- **Tanh** : Convertit les valeurs en un intervalle entre -1 et 1 :

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

D'autres variantes spécifiques de la fonction Sigmoïde peuvent également être citées :

- La Sigmoïde Ajustée (Scaled Sigmoid) : Une version modifiée pour ajuster la sensibilité ou l'échelle de la courbe standard grâce à un paramètre k :

$$f(x) = \frac{1}{1 + e^{-k*x}}$$

- La Hard-Sigmoïde : Une approximation linéaire de la fonction sigmoïde, très utilisée pour accélérer le temps de calcul dans les modèles profonds :

$$f(x) = \max(0, \min(1, 0.2x + 0.5))$$

- La Log-Sigmoïde : Souvent utilisée dans le calcul de certaines fonctions de coût, elle renvoie le logarithme de la valeur sigmoïdale :

$$f(x) = \ln\left(\frac{1}{1+e^{-x}}\right)$$

4.5 Évaluation des performances du modèle

Pour mesurer l'efficacité des modèles d'IA utilisés pour prédire les données médicales, nous nous sommes appuyés sur un ensemble de mesures de performance de base. Le processus d'évaluation s'appuie principalement sur la ***Matrice de Confusion***, qui fait correspondre les vraies valeurs des données avec les attentes émises par le modèle selon quatre possibilités :

- ***Vrais positifs (TP):*** lorsque le modèle prédit correctement la présence de la maladie;
- ***Vrais Négatifs (TN)*** : Lorsque le modèle prédit correctement l'absence de la maladie;
- ***Faux Positifs (FP) :*** Lorsque le modèle prédit de manière incorrecte la présence d'une maladie (erreur de type I);
- ***Faux Négatifs (FN) :*** Lorsque le modèle prédit de manière incorrecte l'absence d'une maladie (erreur de type II).

Normes mathématiques approuvées :

Sur la base de la matrice de confusion, la précision globale (Accuracy) est calculée, qui représente le pourcentage de prédictions complètement correctes, en plus d'autres mesures cruciales dans l'aspect médical [16]:

a) Précision globale(Accuracy):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

b) Sensibilité/Rappel : Elle est très nécessaire en médecine, car elle mesure la capacité du modèle à détecter tous les patients réellement infectés sans manquer aucun cas.

$$\text{Sensibilité} = \frac{TP}{TP + FN}$$

c) Précision : Cela détermine la fiabilité du modèle lorsqu'il émet une alerte d'infection.

$$\text{Précision} = \frac{TP}{TP + FP}$$

d) F1-Score : il représente la moyenne cohérente entre la précision et la sensibilité spécifiées et constitue une norme idéale pour évaluer les données déséquilibrées.(Fawcett, 2006)0

$$\text{F1} - \text{Score} = 2 * \frac{\text{Précision} * \text{Sensibilité}}{\text{Précision} + \text{Sensibilité}}$$

5. Autres types de réseaux neuronaux

Il existe également différents types de réseaux neuronaux, tels que les réseaux neuronaux convolutifs (CNN) et les réseaux neuronaux récurrents (RNN), ainsi que les réseaux de mémoire à long terme (Long Short-Term Memory, LSTM) [11] :

5.1 Les réseaux neuronaux convolutifs (CNN)

Ce réseau, principalement utilisé pour le traitement d'images, fonctionne en combinant des couches de convolution et de clustering pour extraire automatiquement les caractéristiques importantes de l'image avant l'étape de classification, et a obtenu des résultats exceptionnels en matière de reconnaissance d'images médicales et de détection médicale.

5.2 Les réseaux neuronaux récurrents (RNN)

Un réseau dédié au traitement des données séquentielles et temporelles, où les étapes précédentes sont stockées pour être utilisées lors de la prédiction et sont utilisées dans le traitement de texte, la traduction automatique et la prédiction de séries temporelles.

5.3 Les réseaux de mémoire à long terme (LSTM)

Il s'agit d'un développement de réseau qui conserve les informations sur de longues périodes afin d'éviter le problème de l'oubli des données anciennes. Il est utilisé dans l'analyse des séries temporelles et le traitement automatique du langage naturel.

6. Rappels de calcul statistique

Les mesures de tendance centrale et de dispersion jouent un rôle dans l'analyse des données, car les mesures de tendance centrale (la médiane ….) sont utilisées pour résumer les données en une seule valeur qui représente son centre, tandis que les mesures de dispersion (variance…) montrent l'étendue de la propagation et de la divergence des valeurs autour de ce

centre. Ce qui permet de comprendre les caractéristiques des données et d'interpréter plus précisément les résultats de l'analyse statistique, à savoir [18]:

6.1. Valeur moyenne, varianceet écart type

a) Moyenne arithmétique

La moyenne arithmétique est la moyenne résultant de la somme de toutes les valeurs de données et de leur division par leur nombre total. Il calcule des données d'intervalle relatif mais est affecté par les valeurs extrêmes. Son équation est :

$$\bar{X} = \frac{\sum Xi}{n}$$

b) Variance

La variance est une mesure qui montre l'étendue de la dispersion des valeurs par rapport à leur moyenne arithmétique en calculant la moyenne des carrés des écarts des valeurs. Elle est exprimée en unités carrées. L'équation de la variance est la suivante:

$$Variance = \frac{\sum(\bar{X} - Xi)^2}{n-1}$$

c) Ecart type

Il s'agit de la racine carrée de la variance ; il est utilisé pour décrire l'étendue de la dispersion des points de données autour de la moyenne de l'échantillon avec la même unité de mesure d'origine. Son équation est :

$$\text{Ecart type} = \sqrt{\frac{\sum(Xi - \bar{X})^2}{n-1}}$$

6.2. Mode

Mode : Il s'agit de la ou des valeurs les plus fréquentes et les plus courantes dans l'ensemble de données et représente la valeur la plus élevée dans la distribution de fréquence et est utile pour les données nominales. Son équation basée sur l'observation directe :

$$Mode = MostFrequentValue$$

6.3. Médiane et quartiles

Médiane

La médiane est la valeur qui se situe exactement au milieu des données (le 50ème percentile). La moitié des valeurs se situent au-dessus et l'autre moitié en dessous et ne sont pas affectées par des événements anormaux. Son équation est :

$$Median = Q_2 = P_{50}$$

QuartilesQ1, Q2 et Q3 :

La Quartile est la valeur qui divise les données classées par ordre croissant en quatre parties égales qui incluent le premier, le deuxième et le troisième quartile (Q1, Q2, Q3) pour décrire la dispersion des données. Son équation symbolique (25%, 50% et 75%):

$$Q_1 = P_{25}\ , \qquad Q_2 = \text{Médiane} = P_{50}\ , Q_3 = P_{75}$$

Ecart interquartile (IQR) :

Il s'agit d'une mesure de dispersion qui calcule la distance entre le premier ($25^{ème}$ centile) et le troisième ($75^{ème}$centile) quartiles, et elle détermine la répartition des 50 % centraux des données. Son équation est :

$$IQR = Q_3 - Q_1$$

7. Utilisation des bases de données de diabète

7.1. Bases de données disponibles sur Internet

Lors de nos recherches en ligne, nous avons pu télécharger une base de données de patients diabétiques, un fichier CSV contenant un ensemble de données. De nombreuses bases de données similaires existent sur des sites web tels que Kaggle, PhysioNet, CDC Open Data et OpenML. Les chercheurs utilisent fréquemment ces bases de données car elles contiennent un grand nombre d'enregistrements, sont claires et bien organisées, faciles d'utilisation, gratuites et accessibles à tous pour l'expérimentation et la comparaison des résultats. La base de données que nous avons trouvée date de 2023 et provient de Kaggle. Elle contient des données cliniques et biologiques, notamment les variables suivantes: âge, genre, IMC, pression artérielle, fonction cardiaque, hémoglobine glyquée (HbA1c), glycémie à jeun et statut tabagique, toutes traitées pour l'analyse.

7.2. Bases de données locales (Maison du diabète)

Lors de notre stage au Maison du Diabète (Centre du Diabète) de l'hôpital Mohamed

Boudiaf de Ouargla, nous avons collecté manuellement les données des dossiers papiers et les comptes rendus médicaux des patients diabétiques pour l'année 2025. Ces documents ont été traités selon les critères établis par les médecins du service, notamment le taux d'hémoglobine glyquée (HbA1c), la glycémie à jeun, ainsi que certains indicateurs lipidiques sanguins (triglycérides, HDL), la pression artérielle, le poids, l'âge et le genre***. Environ 700 dossiers sur près de 900 ont été traités en une dizaine de jours, dans le strict respect de la confidentialité des données de tous les patients***. Ensuite, ces informations ont été saisies et organisées dans des tableurs Excel afin d'en faciliter le traitement et l'analyse.

La section suivanteprésente la Maison du diabète – Ouargla ainsi que des aspects liés au déroulement de notre stage pour la collecte de données.

8. Présentation de la Maison du diabète - Ouargla

8.1. Organisation du service

LaMaison du Diabètede l'Hôpital Public Mohamed Boudiaf de Ouargla est une unité spécialisée dans la prise en charge des patients diabétiques. Ses missions comprennent l'accueil des patients et la dispensation des soins nécessaires, tels que le diagnostic, le traitement, les soins des plaies et l'organisation de conférences sur la maladie à destination des personnes concernées. Les conférences abordent des sujets comme la nature du diabète, sa gestion, les recommandations diététiques, les glucomètres et d'autres informations visant à sensibiliser aux complications et aux méthodes de prévention.

La Maison du Diabèteest ouverte du dimanche au jeudi, de 8h00 à 13h00. La semaine est organisée comme suit: le dimanche est réservé aux femmes enceintes, le lundi et le mercredi aux hommes, et le mardi et le jeudi aux femmes.

L'unité dispose d'une équipe composée de trois médecins, d'une équipe d'infirmières, de personnel administratif et d'un agent de sécurité, qui œuvrent ensemble pour assurer le bon fonctionnement de l'unité et la prise en charge complète des patients.

Des informations complètes ont été recueillies lors de la visite et des séances de travail permettant de recenser un nombre considérable de cas, tout en consultant les archives disponibles depuis 2022.

8.2. Orientation des médecins spécialistes

Des discussions avec les médecins de la Maison du Diabète, ainsi qu'avec le médecin endocrinologue spécialisé dans le diabète (*Dr HARIZ Abdellah,* Clinique Privée à Ouargla), ont permis d'obtenir des conseils et des explications sur:

- La surveillance des patients;

- Les méthodes diagnostiques;
- L'administration des médicaments et des injections d'insuline;
- L'importance du respect des plans de traitement.

Ils ont également exprimé leur inquiétude quant aux complications liées à un suivi insuffisant. Face aux difficultés rencontrées, les médecins ont insisté sur la précision des examens médicaux et l'efficacité des outils de surveillance utilisés pour le diagnostic et le suivi. Ceci souligne l'importance des technologies médicales et de la physique médicale pour améliorer la qualité et garantir des soins de santé précis. Les médecins ont également souligné l'importance de développer de futurs projets intégrant diverses spécialités médicales et techniques afin d'améliorer la prise en charge, de développer des méthodes de surveillance et d'améliorer les statistiques médicales pour faciliter les prédictions, gagner du temps et réduire la pression sur les patients.

8.3. Relevé des données statistiques

Après avoir obtenu l'accord préalable de la direction de l'hôpital pour effectuer notre stage, nous nous sommes rendus à l'unité de diabétologie le dimanche 17 mai 2026(pour une durée de 10jours). Accueillis par le personnel du service, nous avons été conduits aux boîtes d'archives des patients, situées sur des étagères dans une salle. Chaque boîte contient100 dossiers médicaux, classés par année, de 2022 à nos jours. Leur organisation facilite la recherche et le suivi des dossiers. La saisie des données a été faite en garantissant la confidentialité des informations.

Le processus de collecte de statistiques a débuté à partir des dossiers médicaux récents. Ceux de l'année 2025, représente un grand nombre de cas enregistrés. Selon le médecin, l'évaluation et le suivi des patients reposent sur:

- des indicateurs biologiques de base: HbA1c et glycémie à jeun
- d'autres indicateurs liés au bilan lipidique, utilisés en fonction de chaque cas, du contexte médical et de l'avis du médecin spécialiste.

Finalement, nous avons traité les fichiers qui répondaient à nos besoins, malgré de quelques difficultés:

- Le processus nécessite beaucoup de temps, d'efforts et d'aide ;
- L'espace de travail est limité;
- Quelques dossiers sont incomplets et la difficulté à en extraire rapidement les informations.

Cette expérience nous a permis d'acquérir de précieuses connaissances, notamment :

- La gestion des dossiers médicaux et l'organisation des statistiques au sein des établissements de santé.

- L'importance de l'exactitude des données collectées et du respect de la confidentialité des patients.
- La nécessite de numérisation des dossiers des patients ainsi que la gestion des différentes opérations du service.

9. Programmation et implémentation

9.1. Programmation sur Google Colab

Pour la programmation nous avons opté pour ***le Langage Python***. Nous avons opté aussi de travaillé sur ***la plateforme Google Colab***.

En effet la plateforme Google Colab est une plateforme cloud permettant d'écrire et d'exécuter du code Python incorporant plusieurs bibliothèques (Libraries) en particuliers celles de calcul de l'Apprentissage Profond. De la plateforme Google, Google Colab permet aussi:

- d'utiliser de grand espace mémoire et de la rapiditéd'exécution des calculs;

- espace de travail collaboratif entre les membres du projet;

- utilisation des prompts de l'intelligence artificielle collaborative (Gemini) pour l'écriture de quelques syntaxe Python.

Dans notre projet, la plateforme nous a servi d'environnement de développement pour le traitement des données de patients diabétiques et la mise en œuvre d'algorithmes d'apprentissage profond. Ses fonctionnalités incluent :

- Un fonctionnement direct via la plateforme web, sans installation sur l'appareil ;

- Des notebooks composés de cellules : certaines dédiées à l'écriture et à l'exécution de code, d'autres à l'affichage des résultats, et d'autres encore à la rédaction de texte, d'explications et de commentaires, facilitant ainsi l'organisation du travail et la documentation ;

- L'enregistrement automatique des fichiers et des notebooks sur Google Drive ;

- Des résultats rapides grâce à son environnement d'exécution performant.

Google Colab prend également en charge un large éventail de bibliothèques Python et les rend facilement accessibles. Parmi les bibliothèques utilisées dans notre projet figurent:

- ***Pandas*** : utilisé pour traiter et préparer efficacement des données structurées sous forme de tableau [19].

- ***NumPy :*** dédié au calcul scientifique et à la manipulation rapide de matrices mathématiques [20].

- ***TensorFlow :*** utilisé pour créer et entraîner des réseaux de neurones artificiels et des modèles d'apprentissage en profondeur [21].

- ***Matplotlib/Seaborn :*** utilisés ensemble pour générer des graphiques et visualiser des données statistiques [22][23]

- ***Scikit-learn (sklearn) :*** fournit des outils de base pour segmenter les données, classer et appliquer des algorithmes d'apprentissage automatique [24].

Les étapes de programmation sont les suivantes[25][26] :

1. Collecte et lecture des données : Importer la base de données collectée sur le terrain et la lire dans l'environnement logiciel.

2. Nettoyage et traitement des données : gestion des valeurs manquantes, filtrage du bruit et préparation des données pour qu'elles soient prêtes à l'analyse.

3. Diviser les données en ensembles d'entraînement et de test: Diviser la base de données en une partie dédiée à l'entraînement du modèle et une autre partie indépendante pour le tester.

4. Sélection du modèle et formation : Déterminez la structure de réseau neuronal artificiel appropriée et commencez le processus d'apprentissage (Formation).

5. Tester le modèle et évaluer ses performances: exécuter le modèle sur des données de test pour mesurer sa capacité à faire des prédictions correctes.

6.Afficher les résultats et la précision du modèle : affichez les histogrammes environnementaux et les pourcentages de précision obtenus pour le modèle final.

9.2. Organigramme de calcul

Notre programme fonctionnait comme suit:

1. Importation la base de données
2. Conversion de variables : genre, type, ……
3. Séparation X / Y
4. Normalisation
5. Division train/test
6. Construction du modèle ANN
7. Entraînement
8. Évaluation (Accuracy, Precision, Recall, F1-score...)
9. Prédiction

La Figure II.1 montre l'organigramme de la méthodologie de développement du modèle ANN pour la prédiction du diabè

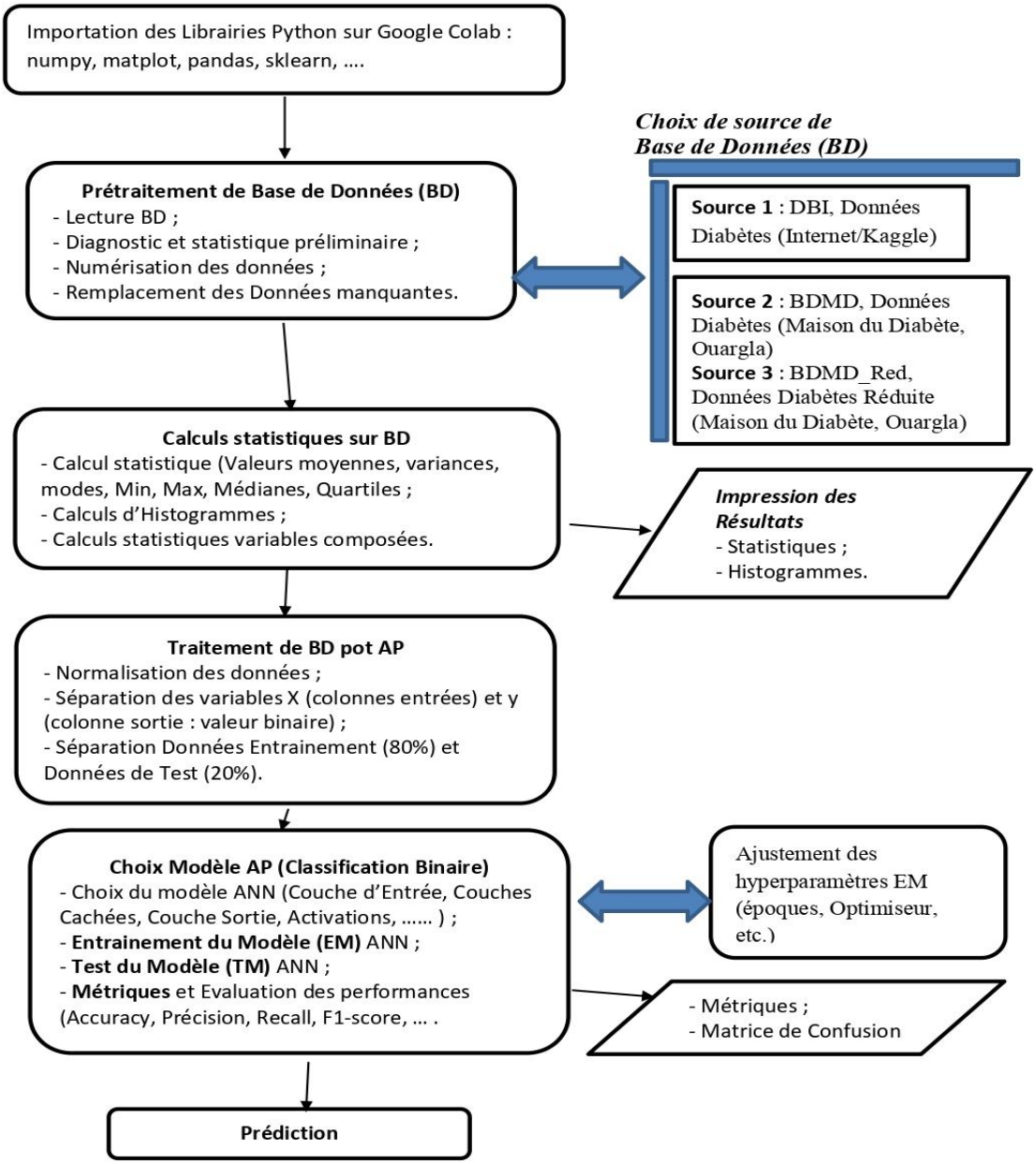


Figure II.1 : Organigramme de la méthodologie de développement du modèle ANN pour la prédiction du diabète

Conclusion

En conclusion, ce chapitre a constitué une base pour l'étude en reliant et en intégrant les parties théoriques et appliquées. Il a encadré les concepts liés à l'intelligence artificielle, aux techniques d'apprentissage automatique et profond, en plus de définir les six étapes méthodologiques adoptées au sein de l'environnement Google Colab. Ce chapitre a également permis d'expliquer le parcours de terrain crucial par lequel les données directes ont été collectées et coordonnées avec le médecin spécialiste et la maison du diabète. Ainsi, le cadre théorique et méthodologique intégré adoptéouvre directement la porte au chapitre suivant, qui sera consacré à la présentation des résultats expérimentaux obtenus par les modèles développés et à la discussion de l'étendue de leur précision et de leur capacité prédictive.

Références :

[1] C. Li, W. Li, Y. Shao, Z. Xu, J. Song, & Y. Weng; A scoping review of artificial intelligence-based health education interventions for patients with type 2 diabetes. Diabetes; MetabolicSyndromeandObesity; (2025; **18**, 645–663.https://doi.org/10.2147/DMSO.S491515

[2] K. Zhou. (2024). "Artificial intelligence for science," Scientific Reports, vol. 14, no. 1, p. 5557. https://doi.org/10.1038/s41598-024-5557-0

[3] Chaudhari, T. (2024). Artificial intelligence, its types and application in various fields. International Journal of Commerce and Management Research, 10(6), 49–51.

[4]V. Gulshan, L. Peng, M. Coram, et al. (2018). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316(22), 2402–2410.

[5]K. Hima. (2021). La prédiction du diabète utilisant le deep learning (Mémoire de master). Université Kasdi Merbah Ouargla.

[6] S. A. Alex, J. Nayahi, & V. Shine. (2022). Deep LSTM Model for Diabetes Prediction with Class Balancing by SMOTE. Journal of Artificial Intelligence and Systems, 4(1), 51–68.

[7]Md. Kowsher, M. M. Rahman, M. A. J. Khan, & M. A. Prottasha. (2023). Prognosis and treatment prediction of type-2 diabetes using deep neural network and machine learning classifiers. International Journal of Cognitive Computing in Engineering, 4, 10-21.

[8]S. Belaid, & N. H. Debba. (2023). Prédiction de la dose de rayonnement du cancer du poumon à l'aide de deep learning dans la région d'Ouargla (Mémoire de Master Professionnel en Physique Médicale). Université de Kasdi Merbah Ouargla

[9]Berguiga, & S. Laib. (2024). Apprentissage automatique pour la prédiction des quelques traitement du cancer du sein dans la région d'Ouargla (Mémoire de Master Professionnel en Physique Médicale). Université de KASDI MERBAH - Ouargla.

[10] R. K. Barbhuiya (2023). Introduction to Artificial Intelligence: Current Developments, Concerns and Possibilities for Education. Indian Journal of Educational Technology, **5**(2).

[11]M. Alom, T. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. Nasrin, M. Hasan, B. Van Essen, A. Awwal et V. Asari. (2019). « A State-of-the-Art Survey on Deep Learning Theory and Architectures », *Electronics*, vol. 8, n° 3.

[12]Schmidhuber (2015). Deep learning in neural networks: An overview. Neural Networks, 61, 85-117.

[13]AseelShakir I. Hilaiwah, Hanan Abed Alwally Abed Allah, BasimAkhudir Abbas, ToleSutikno. (2021). "Live to learn: learning rules-based artificial neural network,"Indonesian Journal of Electrical Engineering and Computer Science,Vol. 21, No. 1, pp. 558–565. DOI: 10.11591/ijeecs.v21.i1.pp558-565.

[14]O. A. Montesinos-López, A. Montesinos-López, & J. Crossa (2022). Multivariate Statistical Machine Learning Methods for Genomic Prediction. Springer, Cham.

[15]E. Haghighat, & R. Juanes (2020). SciANN: A Physics-Informed Deep Learning Framework for Scientific Computing. Computer Methods in Applied Mechanics and Engineering, 372, 113413.

[16]T. Fawcett. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874.

[17]M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, M. Hasan, B. C. Van Essen, A. A. S. Awwal, & V. K. Asari (2019). A State-of-the-Art Survey on Deep Learning Theory and Architectures. Electronics, 8(3), 292.

[18]G. M. Gaddis, & M. L. Gaddis. (1990). Introduction to biostatistics: Part 2, descriptive statistics. Annals of Emergency Medicine, 19(3), 309-315.

[19]W. McKinney. (2010). Data structures for statistical computing in Python. Proceedings of the 9th Python in Science Conference, 51–56.

[20] C. R. Harris, K. J. Millman, S. J. van der Walt, et al. (2020). Array programming with NumPy. Nature, 585(7825), 357–362.

[21]M. Abadi, A. Agarwal, P. Barham, et al. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

[22]J. D. Hunter. (2007). Matplotlib: A 2D graphics environment. Computing in Science &Engineering, 9(3), 90–95.

[23]M. L. Waskom. (2021). Seaborn: statistical data visualization. Journal of Open Source Software, 6(60), 3021.

[24]F. Pedregosa, G. Varoquaux, A. Gramfort, et al; Journal of Machine : Scikit-learn Machine learning in Python; (2011).

[25]I. Goodfellow, Y. Bengio, & A. Courville. (2016). Deep Learning. MIT Press.

[26]F. Pedregosa, G. Varoquaux, A. Gramfort, et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

Chapitre III :

Résultats et discussion

Chapitre III : Résultats et discussion

Introduction

Ce chapitre représente la partie appliquée et expérimentale de l'étude, car il nous amène à construire un modèle de réseaux neuronaux pour classer les données médicales des patients diabétiques. Avec ces objectifs, nous avons proposé de travailler sur deux bases de données. En premier, nous avons commencé à explorer, traiter et analyser une immense base de données internationale de référence (relevée sur internet). Le but est de former et tester l'efficacité du modèle d'intelligence artificielle proposé et d'ajuster sa précision, suivie de l'étape de test de l'efficacité de ce modèle et de sa capacité prédictive sur un échantillon contrôlé. Ensuite, on travaille sur une deuxième base de données de 700 patients. Cette deuxième base est une base de données locale, qui a été collectée manuellement auprès de laMaison du Diabète (Centre de Diabète) affiliée à l'hôpital Mohamed Boudiaf de Ouargla. Le but est d'évaluer les performances du modèle dans l'environnement sanitaire local et de discuter des résultats obtenus.

1. Implémentions sur une Base de Données disponible sur Internet

1.1. Identification de la base de données

Nous avons téléchargé une Base de Données disponible sur la plateforme collaborative ***Kaggle***, sous-domaine ***KaggleDataset***, de Google (https://www.kaggle.com/datasets). Cette base de données, notée ci-après ***DBI***, a les caractéristiques suivantes :

- Nombre de patients ou d'observations:100000;

- Nombre d'indicateurs ou de variables : 8;

-Variable cible : y; relative à deux classes : diabétique / non diabétique.

Après avoir examiné la base de données, nous entamerons les étapes du programme. En commençant par la préparation et le traitement initial des données. Les étapes suivantes sont le calcul statistique et l'application du modèle apprentissage profond.

1.2 Importation et prétraitement de DBI

1.2.1 Importation des données de DBI

a. Dans la première étape, la base de données (diabetes_prediction_dataset) a été importée à partir d'un fichier Excel en utilisant la fonction "pd.red" de la bibliothèque Pandas.
b. Conversion la base de données en "Data-Frame" nommée "df".

1.2.2 Exploration de basse des données

a. L'application d'une fonction *"info()"*, permet d'identifier les types de variables.
b. Affichage du nombre de lignes dupliquées, de données manquantes et vides.

Nous calculons les statistiques initiales des données en appliquant la fonction *"df.describe()"*.

Le Tableau III.1 présente les résultats des statistiques préliminairesde la base de données DBI.

Tableau III.1 : Statistiques préliminaires de la base de données DBI.

Caractéristique	Nombre (Count)	Moyenne (Mean)	Écart-type (Std)	Minimum (Min)	25%	Médiane (50%)	75%	Maximum (Max)
Age	100000	41.89	22.52	0.08	24	43	60	80
Hypertension	100000	0.07	0.26	0	0	0	0	1
Maladie cardiaque	100000	0.04	0.19	0	0	0	0	1
IMC	100000	27.32	6.64	10.01	23.63	27.32	29.58	95.69
Glycémie cumulée (HbA1c)	100000	5.53	1.07	3.5	4.8	5.8	6.2	9
Glycémie à jeun	100000	138.06	40.71	80	100	140	159	300
Diabetes	100000	0.08	0.28	0	0	0	0	1

Le tableau III.1 présente les données relatives à 100000 patients dans DBI avant nettoyage de la base. L'âge moyen est de 42 ans et l'indice de masse corporelle (IMC) est de 27,32 (surpoids). Les valeurs nulles dans les quartiles (25èmeà 75èmepercentiles) indiquent que la majorité de l'échantillon est en bonne santé. D'autres valeurs moyennes sont limitées : au diabète (8 %), à l'hypertension (7 %) et aux maladies cardiaques (4 %). Les valeurs anormales nécessitant une intervention comprennent une glycémie supérieure à 300 mg/dl et un IMC supérieur à 95,69.

1.2.3 Prétraitement et nettoyage des données et DBI

a. En appliquant *"df.drop_duplicates()"*, nous supprimons les lignes en double.

b. Application de la fonction *"df.map()"* permet d'encoder les informations suivantes:

- Genre : (Male : 0;Female : 1; Other : 2)
- Historiquetabagisme : (Aucune information : 0; Jamais : 1; Ancien fumeur : 2; Non-fumeur actuel : 3; Fumeur actuel : 4; Autrefois : 5).

Le Tableau III. 2 présente les données relatives au tabagisme suivant le genre.

Tableau III.2 : Données relatives au tabagismede la base de données DBI.

genre	Historique tabagisme
0	1
0	0
1	1
0	4
1	4
0	1

c. Étant donné que la base de données est hétérogène, en appliquant la fonction (fillna()), nous remplaçons les valeurs manquantes par la médiane (ou la moyenne).

Après traitement des données, nous avons refait les statistiques descriptives afin de garantir la pureté des données et d'obtenir des résultats réalistes, le tableau III.3 présente les résultats.

Tableau III.3 : Statistiques de la base de données DBI après traitement.

Caractéristique	Nombre (Count)	Moyenne (Mean)	Écart-type (Std)	Minimum (Min)	25%	Médiane (50%)	75%	Maximum (Max)
genre	96128	0.4158	0.5	0	0	0	1	1
Age	96128	41.7966	22.5	0.08	24	43	59	80
Hypertension	96128	0.0776	0.3	0	0	0	0	1
Maladie cardiaque	96128	0.0408	0.2	0	0	0	0	1
Historique tabagisme	96128	1.3403	1.5	0	0	1	2	5
IMC	96128	27.3215	6.8	10.01	23.4	27.32	29.86	95.69
HbA1c_level	96128	5.5326	1.07	3.5	4.8	5.8	6.2	9
Glycémie à jeun	96128	138.218	40.91	80	100	140	159	300
Diabetes	96128	0.0882	0.3	0	0	0	0	1

D’après letraitement de la base de données DBI, le taux de diabète est de 8,82 %. Les valeurs moyennes observées sont de 41,80 ans pour l’âge, 138,22 mg/dL pour la glycémie, 5,53 % pour le taux d’HbA1c et 27,32 kg/m² pour l’indice de masse corporelle (IMC).

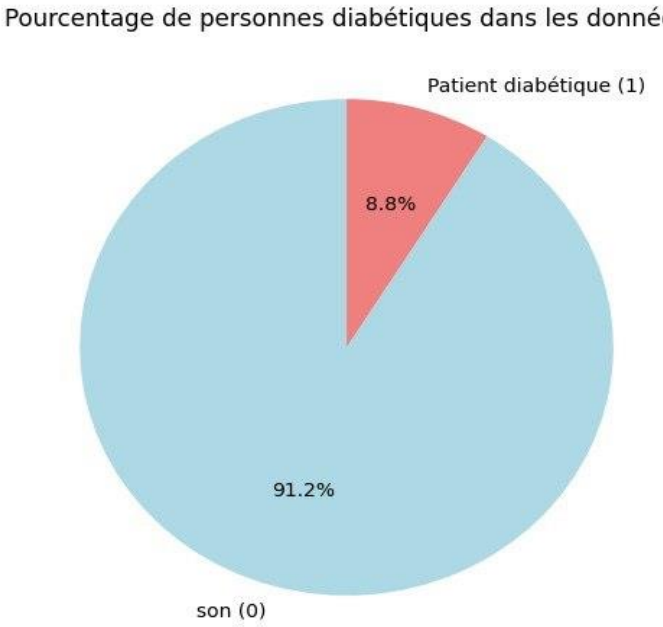


Figure III.1: Diagramme circulaire du pourcentage du nombre de personnes atteintes de diabète dans la base de données.

Le diagramme circulaire illustre la répartition du pourcentage de personnes diabétiques dans la base de données. On observe que le groupe témoin représente la plus grande part (91,2 %), tandis que le groupe diabétique représente la plus petite part (8,8 %). Cette répartition met en évidence une disparité entre les groupes, indiquant un déséquilibre dans l'ensemble de données.

2.2.4.Histogramme des indicateurs DBI traitée

Le relevé des histogrammes relatives à différents indicateurs peut être très utile pour l’analyse statistique. Les figures III.2àIII.5 présentent des histogrammes relatifs à DBI.

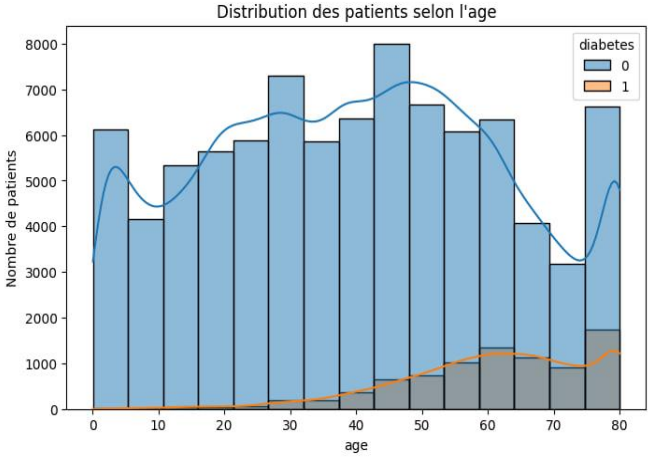


Figure III.2 : Distribution des patients selon l'age.

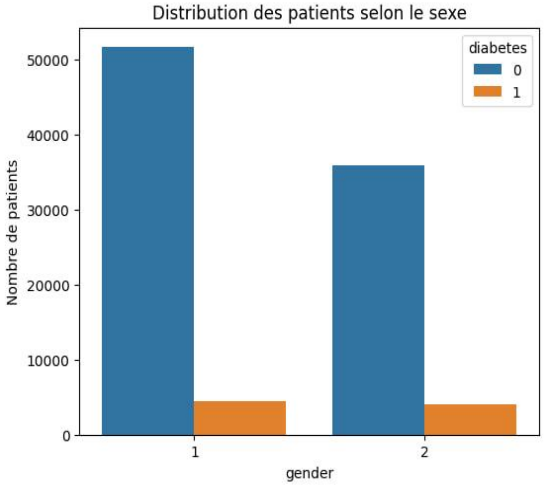


Figure III.3 : Distribution des patients suivant le genre.

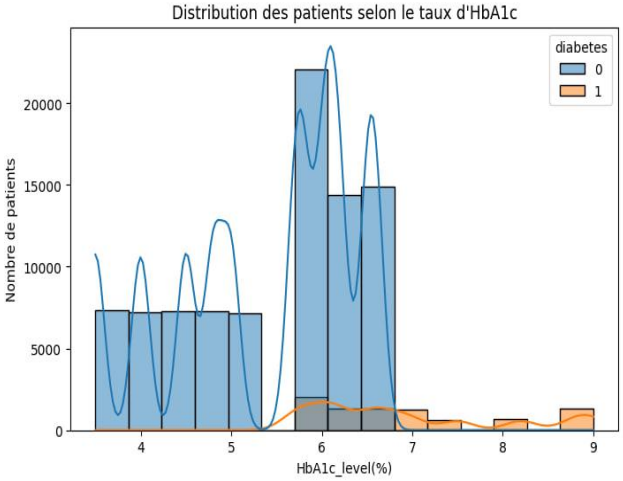


Figure III.4 : Distribution des patients selon le taux de HbAlc.

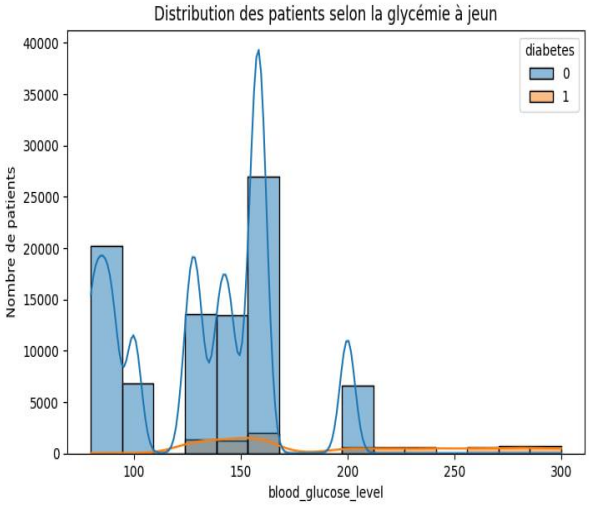


Figure III.5 : Distribution des patients selon la glycémie à jeun.

La Figure III.2 indique que la majorité des personnes saines ont entre 25 et 55 ans, alors que le risque de diabète augmente nettement après 40 ans, liant le vieillissement à la maladie.

La Figure III.3 indique que le nombre de malades (en orange) est presque identique chez les deux sexes malgré la dominance des hommes (genre 1) dans l'échantillon, montrant que le genre n'influe pas sur le diabète.

La figure III 4 représente la distribution de l'HbA1c dans le grand échantillon de référence, où la concentration du groupe des non-diabétiques (0) est observée dans les plages normales inférieures à 6% avec un pic net à 6%, tandis que la courbe des diabétiques (1) s'étend doucement et vers le haut vers des valeurs élevées comprises entre 6,5% et 9%.

La figure III 5 représente la répartition du taux de sucre dans le sang à jeun (Glycémie), où apparaît une concentration dense de personnes non infectées (0) dans les plages sûres et stables (entre 80 et 160 mg/dl) avec un pic dépassant 35 mille personnes, tandis que la courbe des personnes infectées (1) se distribue de manière faible et continue s'étendant vers des lectures élevées qui dépassent 200 mg/dl.

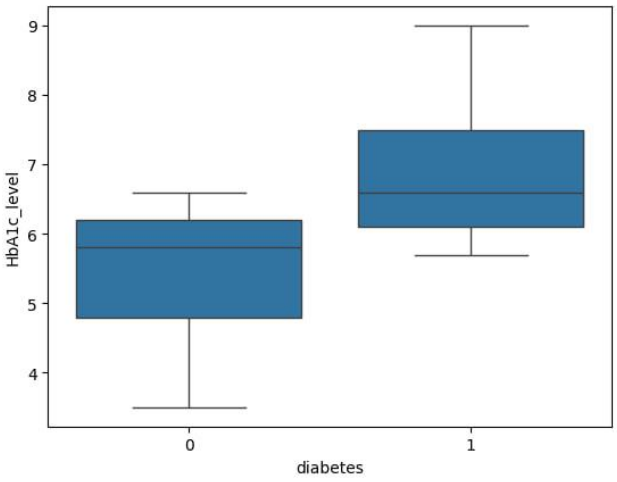


Figure III.6 : Distribution du taux d'HbA1c selon l'état de diabète.

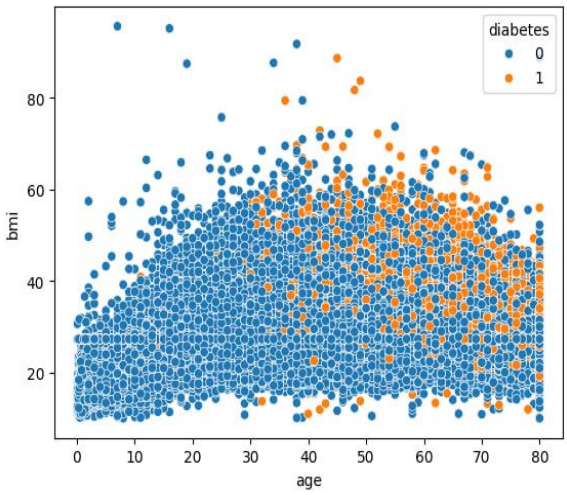


Figure III.7 : Nuage de points du BMI en fonction de l'âge selon l'état de diabète.

La figure III 6 représente un diagramme en rectangles pour la répartition de la glycémie cumulée (HbA1c) par diabète, car il montre une nette variation de la médiane statistique, qui se stabilise en dessous de 6% dans le groupe des personnes non diabétiques (0), tandis qu'elle augmente de manière significative dans le groupe des personnes diabétiques (1), étendant la plage de ses valeurs maximales à 9%.

La figure III 7 représente un nuage de points de la distribution de l'indice de masse corporelle (IMC) en fonction de l'âge et de l'infection, car elle montre une concentration intense de cas de diabète (1) dans la tranche d'âge avancée comprise entre 40 et 70 ans et à des seuils d'obésité qui dépassent 30 kg/m^2, par rapport à la catégorie des personnes non infectées (0) qui est distribuée uniformément dans les jeunes âges et les indicateurs de santé, avec des valeurs fausses nulles observées sur le bord inférieur qui nécessitent un prétraitement.

1.3 Application du modèle d'Apprentissage Profond à DBI traité

1.3.1 Normalisation des données

À l'aide de la classe ***MinMaxScaler*** de la bibliothèque Scikit-learn, les variables ont été transformées afin que leurs valeurs soient comprises dans un intervalle fixe,généralement entre 0 et 1. Le processus se déroule en plusieurs étapes :

- ✓ ***fit()*** : calcul de la valeur minimale et maximale de chaque variable.
- ✓ ***transform()*** : application de la normalisation à ces valeurs.
- ✓ ***fit_transform()*** : combinaison des deux opérations en une seule étape.

Ceci permet de transformer toutes les variablesa la même échelle (entre 0 et 1), d'éviter que les variables a grandes valeurs dominent le modèle, et d'améliorer la stabilité du processus d'apprentissage.

1.3.2 Division des données

Nous divisons les données pour allouer 80 % à l'entraînement et 20 % à l'évaluation du modèle. On sélectionne***Random_state=42*** pour garantir la même évaluation à chaque fois, rendant ainsi les résultats reproductibles.

1.3.3 Construction du modèle ANN (MLP)

La phase de sélection du modèle dépend de la nature des données disponibles et du type de problème. Dans cette étude, notre problème est une classification binaire visant à prédire la présence ou l'absence de diabète à partir d'un ensemble d'indicateurs médicaux. Par conséquent, le modèle de réseau de neurones artificiels (ANN), et plus précisément le sous-modèle MLP, a été utilisé, comme d'autres modèles (...), en raison de son efficacité dans les problèmes de classification binaire et de sa capacité à traiter des données médicales complexes et non linéaires présentant des relations intervariables, telles que celles utilisées dans cette étude. Après avoir justifié ce choix, nous procéderons à la construction de l'architecture du réseau de neurones artificiels.

C’est là qu’interviennent la bibliothèque Keras et son modèle séquentiel, qui permet de construire un modèle de réseau neuronal couche par couche comme suit:

Couche d’entrée
Couche cachée 1
Couche cachée 2
Couche de sortie

Notre problème étant simple et ne nécessitant pas de structures de réseau complexes, le modèle séquentiel est adaptéetil est facile à implémenter. Les raisons scientifiques justifiant le choix des différents composants du réseau sont expliquées ci-après.

- ■ Dense : Cela rend tous les neurones d'une couche entièrement connectés à tous les neurones des couches précédentes.
- ■ 16 neurones : ce nombre est suffisant pour extraire toutes les relations entre les variables. Il n’est ni trop faible, ce qui affaiblirait le modèle, ni trop élevé, ce qui le rendrait trop complexe et entraînerait un sur-apprentissage et un allongement du temps d’entraînement.
- ■ Neurones 12 : Cela suffit pour continuer à traiter les informations extraites de la première couche.

- La fonction ***ReLU*** a été choisie pour sa simplicité et sa rapidité, accélérant l'apprentissage davantage que la fonction Sigmoïde. Elle offre également de bons résultats en classification et en prédiction, et gère les bases de données volumineuses.
- Dans la couche de sortie, un seul neurone a été utilisé avec une fonction sigmoïde pour dériver une valeur unique entre 0 et 1, qui est facilement interprétable par la fonction susmentionnée pour une probabilité claire.
- Adam : Il a été choisi parce qu'il est facile à utiliser, ne nécessite pas de formation, gère l'apprentissage profond et les données massives, réduit les erreurs et offre une précision accrue.
- Loss (BCE) : convient à la classification binaire et aide Adam.
- Accurcy: Convient à l'évaluation.

1.3.4 Entraînement du modèle ANN

À ce stade, nous avons réglé les paramètres d'entraînement du réseau de neurones artificiels (ANN), notamment le nombre d'époques (epochs), la taille des lots (batch size) et le taux d'apprentissage (learning rate), afin d'améliorer les performances du modèle. Ensuite, nous avons entraîné le modèle à l'aide des données d'entraînement préparées.

1.3.5 Évaluation du modèle

Après l'entraînement, nous avons évalué les performances du modèle en utilisant les données de test. Nous avons utilisé plusieurs métriques telles que l'accuracy, la précision (Precision), le rappel (Recall) et le score F1. Nous avons également généré une matrice de confusion afin d'analyser les résultats des prédictions et de comparer les valeurs réelles avec les valeurs prédites. Ensuite, nous avons étudié et analysé les résultats afin d'évaluer l'efficacité du modèle.

1.3.6 Prédiction du diabète

Enfin, nous avons utilisé le modèle entraîné pour effectuer des prédictions sur de nouvelles données. Les résultats obtenus ont permis de classer les cas en diabétique ou non diabétique. Nous avons ensuite analysé ces prédictions afin d'évaluer la capacité du modèle à généraliser sur des données inconnues.

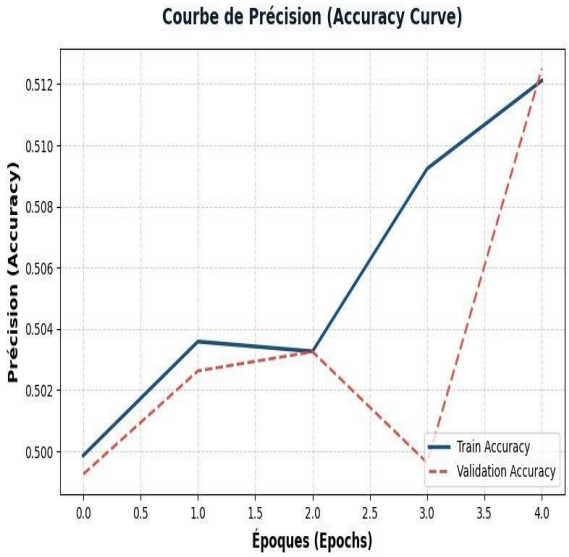


Figure III 8 : Courbe de précision du modèle (Train vs Validation) en fonction des époques.

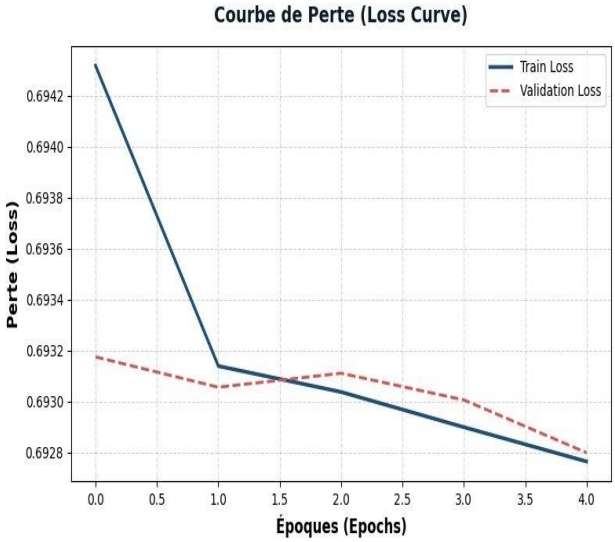


Figure III 9 :Courbe de perte du modèle (Train Loss vs Validation Loss) en fonction des époques.

La figure III.8 montre une amélioration progressive de la précision d'entraînement (Train Accuracy), tandis que la précision de validation (Validation Accuracy) subit une fluctuation marquée (baisse au cours 3) avant de remonter, ce qui indique une instabilité temporaire du modèle avant sa convergence finale.

La figure III.9montre une diminution continue et progressive de la perte d'entraînement (Train Loss) et de la perte de validation (Validation Loss) au fil des époques, ce qui indique que le modèle apprend correctement et converge efficacement sans signe de surapprentissage (overfitting).

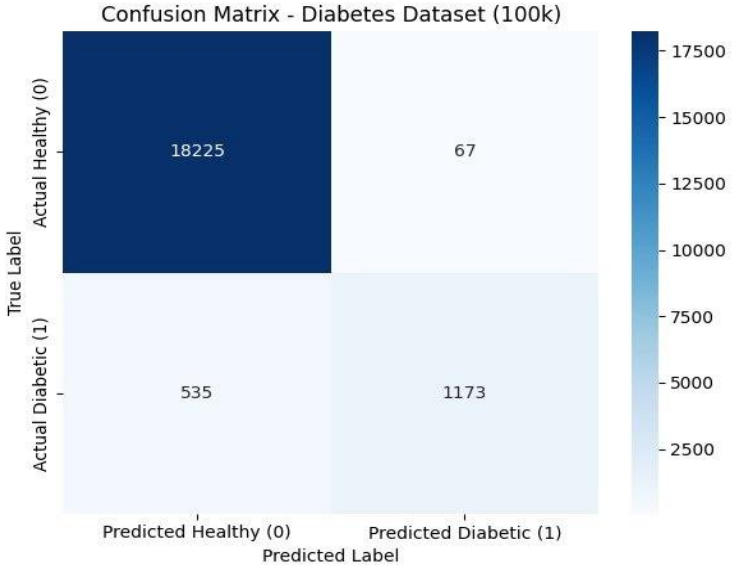


Figure III.10 :Matrice de confusion du modèle sur l'ensemble de données du diabète.

La Figure III.10est relative à la matrice de confusion. Elle montre une haute performance du modèle pour classer les individus sains avec 18225 prédictions correctes (TN), tout en identifiant 1173 cas diabétiques réels (TP), contre 535 faux négatifs (FN) et 67 faux positifs (FP).

$$Accuracy = \frac{TF+TN}{TP+TN+FP+FN} = \frac{(1173+18225)}{20000} = 96.99\ \%$$

$$Sensibilité = \frac{TP}{TP+FN} = \frac{1173}{(1173+535)} = 68.68\ \%$$

$$\text{Précision} = \frac{\text{TP}}{\text{TP+FP}} = \frac{1173}{(1173+67)} = 94.60\ \%$$

$$F1\text{-}Score = \frac{2*(Précision*Sensibilité)}{Précision+Sensibilité} = 79.58\ \%$$

2. Implémentation sur la Base de Données de la Maison du Diabète de Ouargla BDMD

2.1. Identification de la Base de Données BDMD

Étant donné que cette base de données, qui a été collectée auprès de Maison du diabète de Ouargla, est à double classification, nous avons appliqué les mêmes étapes précédentes dans la Première expérimentation avec quelques modifications qui sont cohérentes avec la base de données :

- Nombre de patients ou d'observations : 700 ;

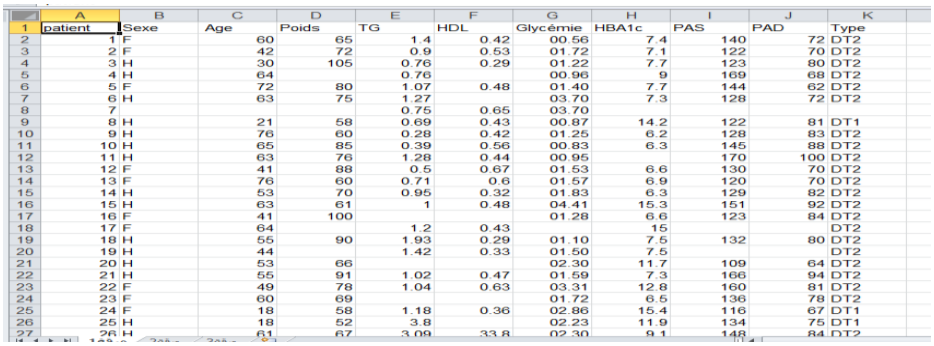
- Nombre de variables : 10 ;

- variable cible : Type de diabète ; Deux classes : DT1 / DT2 (Le type pré-diabète a été omis des calculs et analyse).

2.2 Importation et prétraitement de DBMD

2.2.1 Importation des données de BDMD

La base de données (Figure III.11) a été importée après avoir été placée dans un fichier Excel et nommée " df ".



	A	B	C	D	E	F	G	H	I	J	K
1	patient	Sexe	Age	Poids	TG	HDL	Glycémie	HBA1c	PAS	PAD	Type
2	1	F	60	65	1.4	0.42	00.56	7.4	140	72	DT2
3	2	F	42	72	0.9	0.53	01.72	7.1	122	70	DT2
4	3	H	30	105	0.76	0.29	01.22	7.7	123	80	DT2
5	4	H	64		0.76		00.96	9	169	68	DT2
6	5	F	72	80	1.07	0.48	01.40	7.7	144	62	DT2
7	6	H	63	75	1.27		03.70	7.3	128	72	DT2
8	7				0.75	0.65	03.70				
9	8	H	21	58	0.69	0.43	00.87	14.2	122	81	DT1
10	9	H	76	60	0.28	0.42	01.25	6.2	128	83	DT2
11	10	H	65	85	0.39	0.56	00.83	6.3	145	88	DT2
12	11	H	63	76	1.28	0.44	00.95		170	100	DT2
13	12	F	41	88	0.5	0.67	01.53	6.6	130	70	DT2
14	13	F	76	60	0.71	0.6	01.57	6.9	120	70	DT2
15	14	H	53	70	0.95	0.32	01.83	6.3	129	82	DT2
16	15	H	63	61	1	0.48	04.41	15.3	151	92	DT2
17	16	F	41	100			01.28	6.6	123	84	DT2
18	17	F	64		1.2	0.43		15			DT2
19	18	H	55	90	1.93	0.29	01.10	7.5	132	80	DT2
20	19	H	44		1.42	0.33	01.50	7.5			DT2
21	20	H	53	66			02.30	11.7	109	64	DT2
22	21	H	56	91	1.02	0.47	01.59	7.3	166	94	DT2
23	22	F	49	78	1.04	0.63	03.31	12.8	160	81	DT2
24	23	F	60	69			01.72	6.5	136	78	DT2
25	24	F	18	58	1.18	0.36	02.86	15.4	116	67	DT1
26	25	H	18	52	3.8		02.23	11.9	134	75	DT1
27	26	H	61	67	3.09	33.8	02.30	9.1	148	84	DT2

Figure III.11 : Base de données brute de BDMD

2.2.2 Exploration et prétraitement des donnéesde BDMD

Les types de variables et le nombre de valeurs manquantes et vides ont été affichés. Les résultats des statistiques descriptives initiales sont présentés dans le Tableau III.4.

Tableau III.4 : Statistiques de la base de données BDMD avant traitement.

	Age	Poids	Glycémie	PAD
Count	691	374	606	403
Mean	52.186686	78.358289	12.503587	77.858561
Std	14.030966	17.159764	183.092168	13.281766
Min	11	28	0.560000	0
25%	44	67	1.350000	70
50%	52	76	1.740000	80
75%	62	88	2.557500	85.500000
Max	88	150	4,025	130

Le tableau III.4 présente les statistiques descriptives avant traitement pour la BDMD : Age moyen : 52,18 ans ; sexe : 78,35 ; glycémie à jeun : 12,50 mg/dL ; pression artérielle diastolique moyenne : 77,85 mmHg. Les autres valeurs ont été omises car Python les a catégorisées comme du texte.

Pour le prétraitement, nous avons traité les données comme suit :

- Codage du genre : (H = 0, F = 1);
- Codage du type : (Prédiabète = 0, DT1 = 1, DT2 = 2);
- Convertir les colonnes contenant du texte en nombres;
- Compenser les valeurs manquantes avec la médiane ou la valeur moyenne ;
- Supprimez la ligne contenant le type prédiabétique (car c'est la seule ligne).
- Calcul des statistiques après le nettoyage de la base (***'BDMD traitée'***)

2.3 Calcul des tendances statistique sur BDMD traitée

Le Tableau III.5 présente les valeurs des tendances des indicateurs de BDMD.

Après le traitement de la base de données de la Maison du Diabète de Ouargla, l'échantillon comprend 699 patients. Les statistiques descriptives montrent que l'âge moyen des patients est de 52,17 ans, avec un poids moyen de 77,22 kg. La glycémie moyenne est de 11,07 mmol/L, tandis que le taux moyen d'HbA1c est de 14,86 %, ce qui reflète un mauvais contrôle glycémique chez une grande partie des patients. En outre, la moyenne de la variable « Type » est de 1,98, indiquant que la majorité des patients de l'échantillon sont atteints du diabète de type 2.

Variable	Count	Mean	Std	Min	25%	50%	75%	max
Sexe	699	0	0	0	0	1	1	1
Age	699	52.1731	13.9472	11	44	52	62	88
Poids	699	77.2175	12.5488	28	75	76	77	150
TG	699	3.4195	38.9456	0.0600	0.9250	1.2450	1.5950	1,010
HDL	699	1.0121	4.6188	0.1000	0.3700	0.4200	0.4700	53
Glycémie	699	11.0726	170.4981	0.5600	1.4050	1.7400	2.4000	4,025

Tableau III.5 : Statistiques préliminaires de la base de données DBMD.

HbAlc	699	14.8635	75.8642	0.3000	7.5000	9	11.2000	1,505
PAS	699	132.2074	16.6368	79	130	130	134	219
PAD	699	78.7625	10.1344	0	77	80	80	130
Type	699	1.9757	0.1542	1	2	2	2	2

2.4 Calcul des histogrammes sur BDMD traitée

Les figures ci-après présente des histogrammes de quelques indicateurs de BDMD.

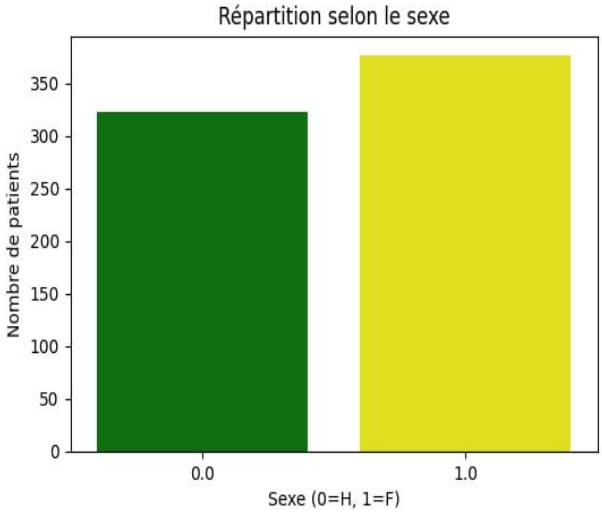


Figure III.12:Distribution du nombre de patients selon le genre

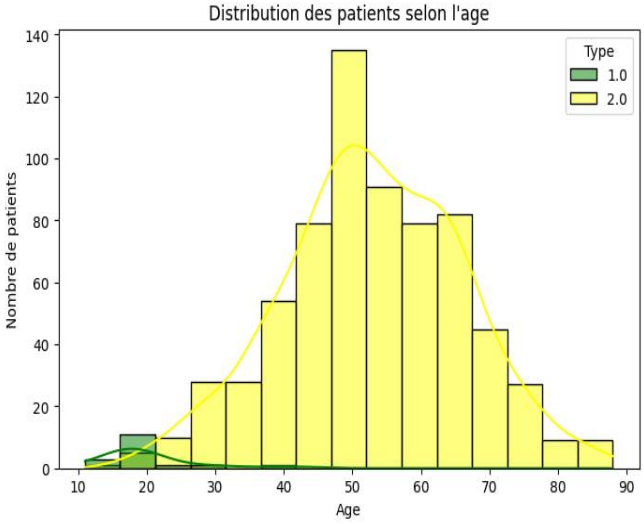


Figure III.13:Distribution du nombre de patients selon l'âge

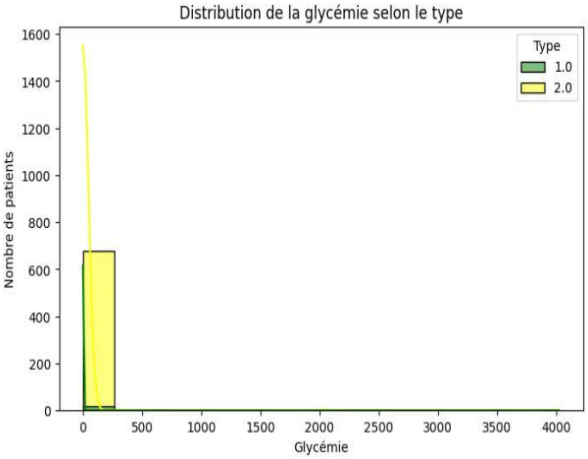


Figure III.14:Distribution du nombre de patients selon La glycémie

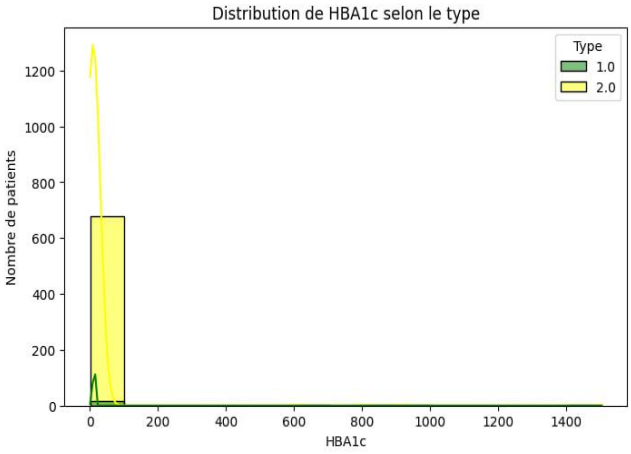


Figure III.15:Distribution du nombre de patients selon le HbA1c patients selon le HbA1c

La figure III.12représente la répartition des patients par genre dans l'échantillon de l'étude, où l'on observe une convergence équilibrée entre les sexes, avec une légère supériorité pour la catégorie des femmes (plus de 370 patients) par rapport à la catégorie des hommes (environ 320 patients).

La figure III.13 représente la répartition des patients selon l'âge dans l'échantillon de référence, car il est clair qu'il existe une concentration totale de patients diabétiques de type 2 dans les tranches d'âge avancées, avec un pic dépassant 130 patients à l'âge de cinquante ans, tandis que le diabète de type 1 est limité à un petit nombre de jeunes de moins de vingt ans.

La figure III.14 représente la répartition de la glycémie à jeun (glycémie) dans l'échantillon de référence, où une forte concentration de diabétiques de type 2 apparaît dans la première colonne du graphique, au milieu d'une absence presque totale de mesures de type 1.

La figure III.15 représente la distribution de la glycémie cumulée (HbA1c), où une concentration écrasante de patients diabétiques de type 2 est clairement visible dans la plage initiale du graphique, compensée par une courbe très basse et étroite pour les patients de type 1.

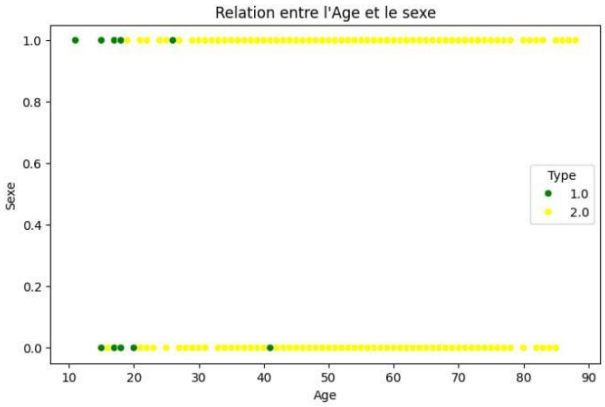


Figure III.16 : Relation entre l'âge et le genre Diagramme en boîte de l'HbA1c

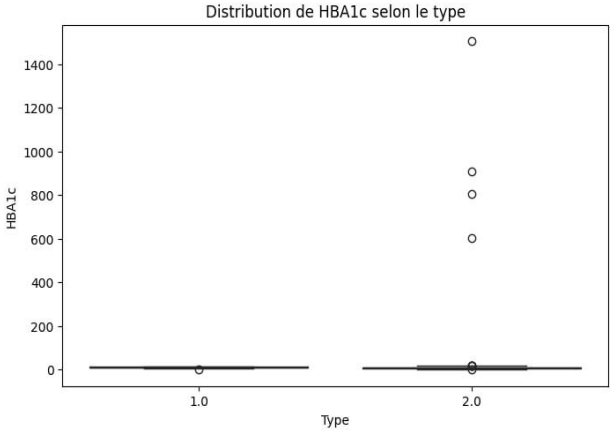


Figure III.17 : Diagramme en boîte de l'HbA1c

La figure III.16 représente un nuage de points de la relation entre l'âge et le sexe par type de diabète, où une prévalence étendue et intense de patients diabétiques de type 2 est observée à différents âges et chez les deux genres, tandis que les patients de type 1 sont principalement confinés aux groupes d'âge jeunes et à ceux de moins de 25 ans pour les hommes et les femmes.

La figure III.17 représente une boîte à moustaches pour la distribution de l'HbA1c par type, car elle montre la stabilité des lectures normales pour les patients proches de zéro, avec l'apparition de valeurs anormales qui se limitent uniquement aux patients de type 2.

2.5 Application du modèle Deep Learning à DBMD traitée

Nous avons opté pour la même procédure du modèle ANN.

- ***Normalisation des données;***

- ***Division des données;***

- ***Construction du modèle ANN (MLP).***

Le modèle a été évalué à l'aide de plusieurs indicateurs de performance, notamment la fonction de perte, la précision et la matrice de confusion, afin d'analyser la qualité d'apprentissage et l'efficacité du modèle de classification. Les résultats suivants ont été obtenus :

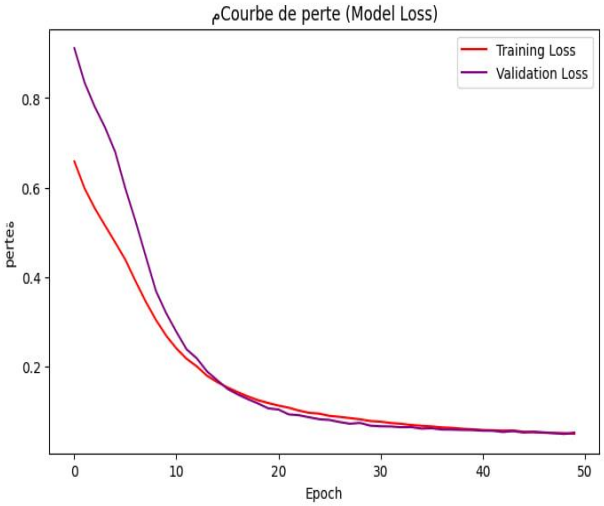


Figure III.18 : Courbe de la fonction de perte (LossFunction)

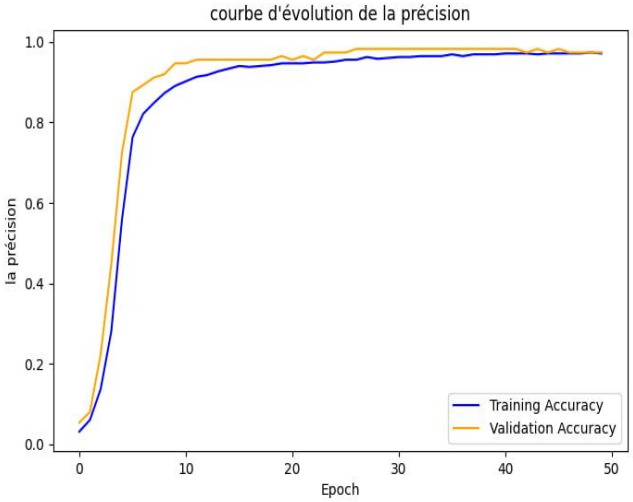


Figure III.19 : Courbe de précision (Accuracy)

La figure III.18 représente la courbe de perte du modèle en fonction du nombre de cycles d'entraînement (Epochs), où une diminution rapide et continue des valeurs de perte est observée lors des premiers cycles, avant que les pertes d'entraînement et de validation ne se stabilisent de manière identique et idéalement à leurs niveaux les plus bas (environ 0,08), ce qui confirme la stabilité du modèle et sa grande capacité de généralisation.

La figure 19 représente la courbe d'évolution de la précision du modèle en fonction du nombre de cycles d'entraînement (Epochs), où une montée rapide et stable de la précision est observée lors des premiers cycles avant que la précision d'entraînement et de validation ne se stabilise de manière équilibrée avec une excellente adéquation supérieure à 93%, ce qui confirme l'efficacité du modèle et l'absence de surapprentissage (surapprentissage).

Matrice de confusion représentant les performances du modèle de classification :

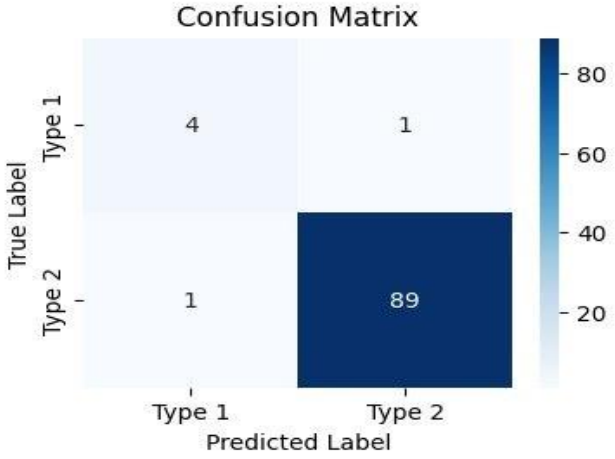


Figure III.20 : Matrice de confusion

La matrice de confusion obtenue (Figure III.20)permet d'évaluer les performances du modèle de classification. À partir des valeurs des vrais positifs (TP), des vrais négatifs (TN), des faux positifs (FP) et des faux négatifs (FN), les principales métriques d'évaluation, à savoir l'accuracy, la sensibilité, la précision et le F1-Score, seront calculées afin de mesurer la capacité du modèle à distinguer les patients atteints de diabète de type 1 et de type 2

$$Accuracy = \frac{TF+TN}{TP+TN+FP+FN} = \frac{(89+4)}{95} = 97.89\ \%$$

$$Sensibilité = \frac{TP}{TP+FN} = \frac{89}{(89+1)} = 98.89\ \%$$

$$Précision = \frac{TP}{TP+FP} = \frac{89}{(89+1)} = 98.89\ \%$$

$$F1\text{-}Score = .\frac{2*(Précision*Sensibilité)}{Précision+Sensibilité} = 98.89\ \%$$

Les résultats obtenus sur 95 cas de test montrent que le modèle a atteint une précision globale de 97.89 %. Il a correctement classé **89cas positifs (TP)** et **4 cas négatifs (TN)**, avec seulement **un faux positif (FP)** et **un faux négatif (FN)**. Les métriques de précision, de sensibilité et de F1-Score ont toutes atteint **98,89 %**, ce qui témoigne d'excellentes performances de classification.

Ainsi, ces performances confirment que le modèle proposé est fiable et suffisamment performant pour distinguer efficacement les patients atteints de diabète de type 1 et de type 2.

2.6 Décision finale basée sur l'évaluation des performances du modèle

Les prédictions obtenues montrent que le modèle est capable de classer correctement les patients comme atteints de diabète de type 1 ou de type 2.

2.7 Expérimentation avancée sur BDMD

La base de données BDMD initiale porte plusieurs lignes avec des informations manquantes. Dans les sections précédentes, nous avons opté pour remplacer les vaeurs manquantes par les valeurs médianes. Dans cette section, on se propose de supprimer les lignes contenant plusieurs valeurs manquantes. Les résultats sont présentés dans les deux sous-section suivantes 2.7.1 et 2.7.2.

2.7.1 Base de données réduite DS3 :

DS3_DataSet_Centre_Diabete_Ouargla_.xlsx
Description :DT1, DT2, avec des champs principaux non remplacés par la Médian

Le Tableau III.6 présente les valeurs des tendances des indicateurs de DS3. Le nombre de patients est 404. L'âge moyen est 52,61 années, la glycémie moyen est 17,68 et la valeur moyenne de l'HbAlc est 14,61.

Tableau III.6 : Relevés statistiques de DS3

	count	mean	Std	Min	25%	50%	75%	Max
Age	404.0	51.616337	13.921575	15.00	43.00	52.00	62.0000	88.0
Glycémie	404.0	17.685455	224.151857	0.56	1.35	1.73	2.4625	4025.0
HbAlc	404.0	14.615916	80.058535	0.30	7.40	8.95	11.1250	1505.0

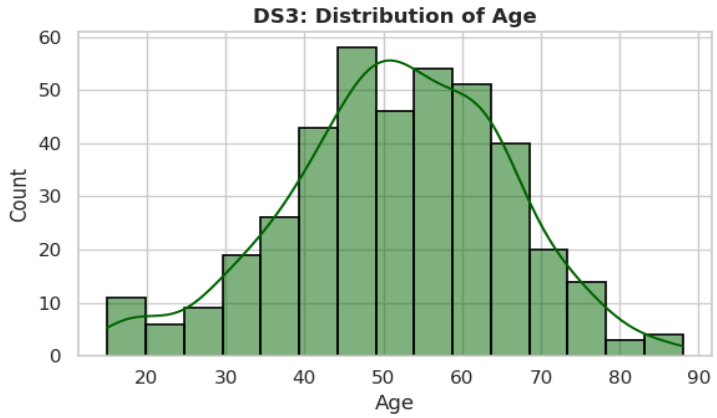


Figure III.21: Histogramme d'âge pour DS3

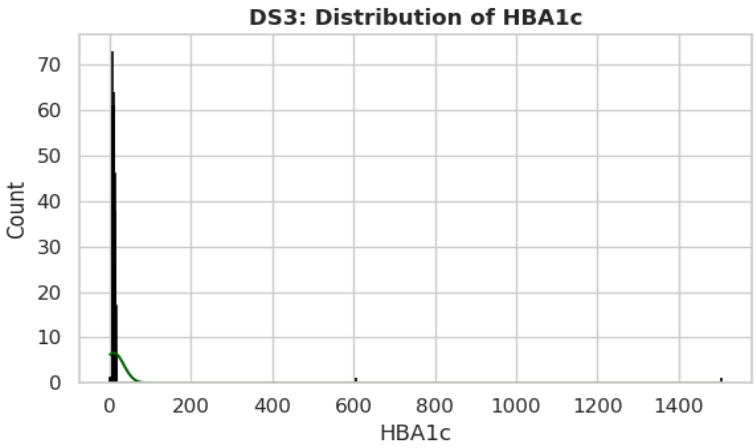


Figure III.22: Histogramme d'HbAlc pour DS3

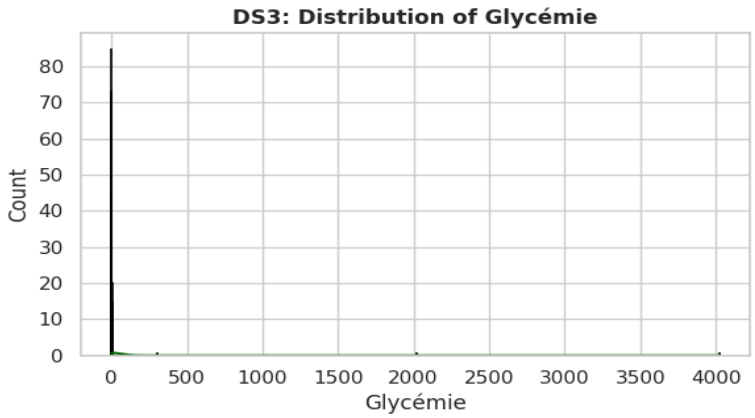


Figure III.23: Histogramme de Glycémie pour DS3

Le Tableau III.7 présente les résultats de l'application de DS3. Les deux classes sont les deux types DT1 et DT2.

Tableau III.7 : Application du modèle ANN à DS3

	Precision	**Recall**	**f1-score**	**support**
DT1	0.000000	0.000000	0.000000	2.000000
DT2	0.975309	1.000000	0.987500	79.000000
Accuracy	0.975309	0.975309	0.975309	0.975309
macro avg	0.487654	0.500000	0.493750	81.000000
weighted avg	0.951227	0.975309	0.963117	81.000000

2.7.2 Base de données réduite DS4_DT2 :

DS4_DataSet_Centre_Diabete_Ouargla_.xlsx
Description : DT2, avec des champs principaux non remplacés par la Médian
Etude Multi-classes après transformation des valeurs d'un indicateur en classes.

Tableau III.8 : Statistiques DS4_DT2

	Count	Mean	std	min	25%	50%	75%	max
Age	390.0	52.717949	12.803335	16.00	44.00	53.00	62.000	88.0
Glycémie	390.0	18.240779	228.130188	0.56	1.35	1.73	2.450	4025.0
HBA1c	390.0	14.742385	81.481667	0.30	7.40	8.90	11.075	1505.0

Le Tableau III.8 présente les valeurs des tendances des indicateurs de DS4_DT2. Le nombre de patients est 390. L'âge moyen est 52,71 années, la glycémie moyen est 18,24 et la valeur moyenne de l'HbAlc est 14,74. C'est valeurs sont légèrement supérieures à celle du Tableaux III.6 ; en effet le nombre de patients de type 2 est dominant.

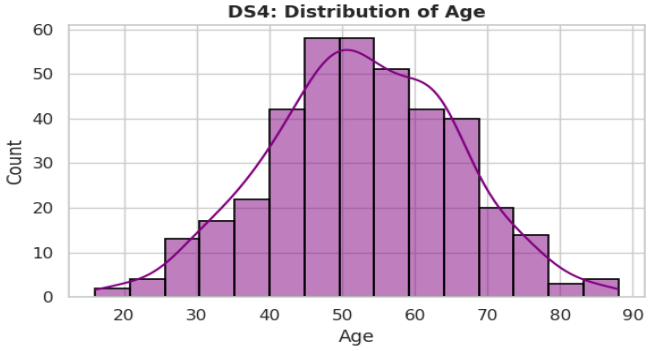


Figure III.24 : Histogramme d'âge pour DS3

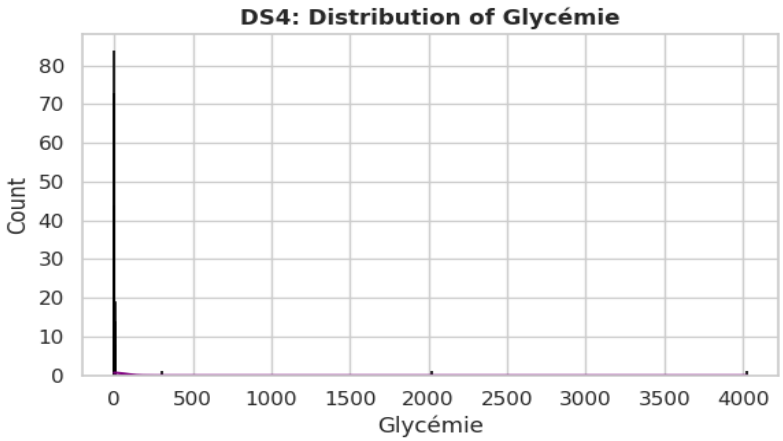


Figure III.25 : Histogramme de glycémie pour DS4_DT2

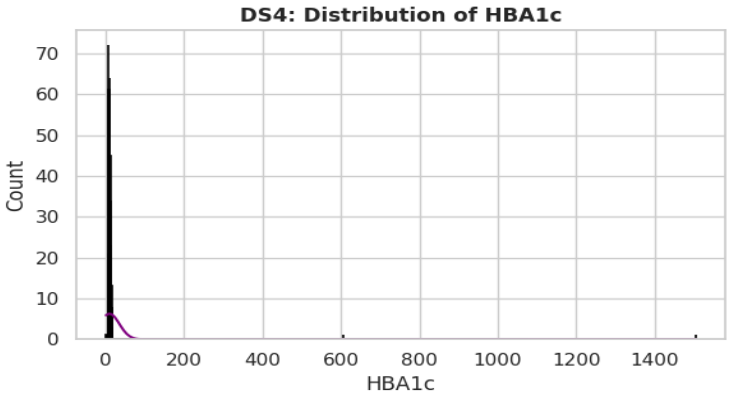


Figure III.26 : Histogramme d'HbAlc pour DS4_DT2

Conclusion

Ce chapitre montre l'importance de travaillé en langage Python et sous l'environnement Google Colab. Il montre aussi la possibilité de télécharger sur Kaggle des données de données pédagogiques.

Le travail sur la première base DBI de classification binaire (diabétique, non diabétique) était très bénéfique) en terme d'analyse de données et en terme d'application du modèle. Les tendances statistiques, les histogrammes et la métrique sont calculés et discutés. Le traitement de la base de données DBI, le taux de diabète est de 8,82 %. Les valeurs moyennes observées sont de 41,80 ans pour l'âge, 138,22 mg/dl pour la glycémie, 5,53 % pour le taux d'HbA1c et 27,32 kg/m² pour l'indice de masse corporelle (IMC).

La deuxième base de données est locale BDMD, du Centre de diabète de Ouargla. La base est à deux classes : Diabètes de Type 1 (DT1) et Diabètes de Type 2 (DT2). Les statistiques descriptives montrent que l'âge moyen des patients est de 52,17 ans, avec un poids moyen de 77,22 kg. La glycémie moyenne est de 11,07 mmol/l, tandis que le taux moyen d'HbA1c est de 14,86 %, ce qui reflète un mauvais contrôle glycémique chez une grande partie des patients. En outre, la moyenne de la variable « Type » est de 1,98, indiquant que la majorité des patients de l'échantillon sont atteints du diabète de type 2.

Le modèle développé a atteint une précision de classification supérieure à 93 %.

Conclusion générale

Conclusion générale

L'humanité donne une importance capitale à la santé et la médecine. Ces dernières décennies, l'expansion des maladies chroniques a connu une croissance exponentielle à l'échelle mondiale. Ces pathologies représentent un enjeu majeur pour les systèmes de santé, en raison de leur impact direct sur la santé des individus et des sociétés. Parmi ces maladie le diabète.

Les spécialistes du secteur de la santé et de la médecine s'intéressent aux différents indicateurs cliniques, biologiques et autres. Pour le suivi, le traitement, la prévention et prédiction, les spécialistes s'intéressent à plusieurs outils. Les calculs statistiques présentent les premiers outils. Le développement de l'informatique et des techniques d'intelligence artificielle a permis l'utilisation d'autres outils. L'intelligence artificielle est devenue un outil essentiel qui contribue à résoudre des problèmes complexes que les méthodes traditionnelles sont incapables de résoudre. Parmi les branches les plus développées et les plus influentes dans ce domaine, se distinguent les technologies d'apprentissage automatique et d'apprentissage profond.

Dans ce mémoire de Master on s'intéresse à l'application de l'Intelligence Artificielle dans l'étude de la maladie du diabète.Quels sont les fondements théoriques et méthodologiques ?Comment les données cliniques brutes collectées sur le terrain peuvent-elles être nettoyées et traitées pour les rendre adaptées aux modèles de calcul ? Comment développer et construire un modèle intégré de réseau neuronal artificiel (ANN) capable de classer avec précision les données médicales ?

Dans le premier chapitre nous avons présenté les intérêts et concepts fondamentaux du diabète : types DT1 et DT2, aspects cliniques et biologiques, diagnostics et indicateurs, traitements et prévention.

Dans le deuxième chapitre, nous nous sommes concentrés sur les fondements théoriques et méthodologiques, en expliquant l'environnement de travail du logiciel et les bibliothèques utilisées. Le langage utilisé est python, l'environnement est Google Colab.Pour le calcul statistique des grandeurs (telles que valeurs moyennes, écart-type, mode, médiane, quartiles), il est nécessaire de faire un prétraitement et nettoyage des bases de données. Pour les modèles et algorithmes pour le deep-Learning et les réseaux neuronaux, on présente aussi les étapes de traitement statistique et par ANN (Normalisation, Séparation des données X et y, Répartition des données entre données d'entrainement et données de test, choix du Modèle, fonctions d'activation, métrique d'évaluation).

Le chapitre II présente aussi nos activités du stage sur terrain et des étapes de collecte de données du Centre de diabète de Ouargla. La base de données locale portant 700 patients (notées anonymes pour l'étude pour confidentialité).

Le troisième présente les résultats de nos travaux.

Une première base de données internationale de 100000 individus, téléchargé sur Kaggle, nous était pédagogique. La base est de classification binaire (diabétique, non diabétique). Les tendances statistiques, les histogrammes et la métrique sont calculés et discutés.Le traitement de la base de données DBI, le taux de diabète est de 8,82 %. Les valeurs moyennes observées sont de 41,80 ans pour l'âge, 138,22 mg/dL pour la glycémie, 5,53 % pour le taux d'HbA1c et 27,32 kg/m² pour l'indice de masse corporelle (BMI).

La deuxième base de données est locale ; c'est celle collectée au Centre de diabète de Ouargla. Le nombre de patients de 700. Nous avons travaillé sur deux classes : Diabètes de Type 1 (DT1) et Diabètes de Type 2 (DT2).Les tendances statistiques, les histogrammes et la métrique sont calculés et discutés. Les statistiques descriptives montrent que l'âge moyen des patients est de 52,17 ans, avec un poids moyen de 77,22 kg. La glycémie moyenne est de 11,07 mmol/L, tandis que le taux moyen d'HbA1c est de 14,86 %, ce qui reflète un mauvais contrôle glycémique chez une grande partie des patients. En outre, la moyenne de la variable « Type » est de 1,98, indiquant que la majorité des patients de l'échantillon sont atteints du diabète de type 2.

Le modèle développé a atteint une précision de classification supérieure à 93 % avec une excellente stabilité dans les courbes d'apprentissage. La matrice de confusion a montré une grande capacité à séparer avec précision le type 1 (DT1) et le type 2 (DT2) avec la marge d'erreur la plus faible possible,

La base de données BDMD initiale porte plusieurs lignes avec des informations manquantes. Nous avons proposé de supprimer les lignes contenant plusieurs valeurs manquantes et de travailler sur deux nouvelles bases de données. La première à deux classes (DT1 et DT2). La deuxième est sur le type 2 seulement permettant un passage au calcul multi-classes (âge, ou Glycémie, ou HbAlc, …).

En perspectives, il serait envisageable de voir les points suivants :

- Calcul pour multi-classes ;
- Utiliser d'autres modèles de Deep Learning ;
- Elargir la collecte de données locales et proposition d'indicateurs d'étude ;
- Travail avec les institutions locales et régionales pour une numérisation des données relatives aux patients, aux médicaments et à la gestion des services.